Comments on the Draft Cancer Inhalation Unit Risk (IUR) Factor for Ethylene Oxide

Draft Released on May 14, 2026, by the Office of Environmental Health Hazard Assessment California Environmental Protection Agency (CalEPA OEHHA)

Contributors:

Abby Li Ph.D., James Bus Ph.D., DABT, Fellow ATS, Kenneth Bogen, Dr.P.H., DABT, and Rick Reiss Sc.D., Exponent Inc.

B. Bhaskar Gollapudi, Ph.D., Independent Consultant

Christopher R. Kirman, M.S., SciPinion

Ciriaco Valdez-Flores, Ph.D., Independent Consultant

M. Jane Teta, Dr.P.H., M.P.H., Independent Consultant

Date:

July 29, 2026

Contents

List of Figures

List of Table

Executive Summary

Determining a scientifically robust Inhalation Unit Risk (IUR) for ethylene oxide (EtO) requires a rigorous evaluation of available data and alignment with current biological evidence. Errors in OEHHA's rating of studies lead to incorrect evaluation of the epidemiological evidence that informs the selection of continuous statistical models to apply to the NIOSH study. OEHHA's rationale for biological plausibility of its spline model is based on speculative possibilities for chemicals that have different mechanisms than EtO and are inconsistent with the EtO-specific biological mode of action and toxicological data. Instead, the draft OEHHA IUR is based on incorrect visual "eyeballing" of continuous models with the shape of a few grouped categorical estimates, not the individual hazard rates that were modeled. OEHHA's primary rationale for rejecting the biologically plausible and statistically superior standard Cox proportional hazards model relies on a flawed comparison of relative-rate magnitudes, despite its own figures explicitly acknowledging that such visual comparisons are not an appropriate basis for interpretation. This approach can be applied to both breast and lymphoid mortality cancers

We highlight key errors in OEHHA's evaluation related to 1) exposure measurement error and 2) dismissal of the findings from the UCC study that contradicts their choice of a model. The unvalidated nature of the NIOSH exposure regression model introduces a high likelihood of exposure measurement error for >80% of the lymphoid and breast cancer cases exposed prior to 1978. Because the high-probability of exposure error impacts cases across the entire spectrum of cumulative exposures, any reliable IUR derivation must utilize models that incorporate the full spectrum of data rather than on low-dose segments that are driven by a small fraction of cases. Utilizing the entire dataset ensures greater statistical power and stabilizes risk estimates against the uncertainties inherent in the unvalidated NIOSH exposure model for the period prior to 1978, during which time no monitoring data were available.

Exposure Assessment Uncertainties and Double Standards

There are two major cohort studies informing lymphoid mortality: the NIOSH study of sterilization workers and the UCC study of manufacturing workers. Both studies should be considered when evaluating the plausibility of alternative dose-response models. Although both studies lack contemporaneous exposure measurements for certain historical periods, this limitation is substantially more severe in the NIOSH cohort. OEHHA's draft heavily discounts the UCC exposure assessment, even though it incorporated exposure data from UCC facilities (including the same and/or closely comparable plants) collected between

1957 and 1988, covering more than 75% of the cohort. For earlier years, plant medical records documenting worker overexposure incidents were used to validate EtO department exposure classifications. Nevertheless, OEHHA assigned the UCC exposure estimates its lowest rating of "critical." In contrast, OEHHA appears to place disproportionately greater weight on the unvalidated assumptions underpinning the NIOSH exposure reconstruction for the period from 1940 to 1978—a span of nearly four decades during which exposure data were either extremely limited (1977–1978) or entirely absent (1940–1976).

Although the NIOSH exposure model was well-validated for a brief 10-yr window (1976-1985) when exposure data were available; the model was subsequently altered without sound justification to estimate exposures for the preceding four decades. Peer-reviewed analyses (e.g., Bogen et al., 2019) indicate that the NIOSH model fails to capture major historical shifts in engineering and industrial hygiene practices. As a result, the NIOSH model produces counterintuitive estimates of lower exposures in early years, directly contradicting documented evidence of much higher historical exposures. In general, underestimating exposures results in overestimation of cancer risk. OEHHA's evaluation of study quality lacks objectivity by ignoring the substantial uncertainty and unreliability of the NIOSH reconstruction for the majority of exposure period.

It is indefensible to fault UCC for relying on surrogate exposure data from a comparable UCC facility to characterize exposures during earlier periods while simultaneously assigning greater weight to NIOSH's approach, which assumes that calendar year—the primary predictor of exposure in the post-1978 period—had no effect on exposures in earlier years and that other key predictors identified in the later post-1978 period were equally applicable to the pre-1978 period. Moreover, the NIOSH exposure regression model was developed using post-1978 exposure data from 36 sterilization facilities, only 14 of which were included in the study cohort, reflecting NIOSH's own reliance on data from comparable facilities to estimate historical exposures.

Biological Plausibility -A Required Criteria for Model Selection

Any risk assessment relying on human epidemiology must utilize a dose-response model consistent with established biological and mechanistic evidence (EPA SAB, 2016). In addition to EPA, the Texas Commission on Environmental Quality (TCEQ) also modeled the NIOSH lymphoid data and considered biological plausibility as a key criterion for model fit[1]. Consideration of biological plausibility is especially necessary when there is no difference in statistical p-value between TCEQ's vs OEHHA's model approaches for both

[1] TCEQ (2020) stated that "The TCEQ considers multiple factors when deciding on the dose-response model . . .First and foremost is the consideration of the chemical's MOA" (P. 35 Section 4.2).

lymphoid and breast cancer. Based on sound statistical principles alone, the TCEQ's model approach is superior because it better fulfills OEHHA's/EPA's criteria of parsimony (e.g., is the simplest model with fewer model parameters). However, we agree with EPA SAB's repeated advice that whatever model is selected must be biologically plausible.

Recent molecular dosimetry and genotoxicity studies (e.g., Gollapudi et al., 2026; Liu et al., 2026) clarify this relationship at low exposure levels. These studies demonstrate that key mechanistic endpoints—including systemic dosimetry measured as hemoglobin adducts, DNA adduct formation, mutations, and chromosomal damage—do not exhibit a steep low-dose response. Instead, the biological response demonstrates (at most) a shallow linear response at low exposures, which steepens only at higher exposures possibly due to saturation of glutathione-dependent EtO detoxification depletion as well as possible saturation of DNA and cellular repair processes. This pattern aligns with the original conclusions of the NIOSH epidemiological studies:

- **Lymphoid and breast cancer mortality:** "Positive exposure-response trends for lymphoid tumours were found for males for lymphoid mortality, and for females for breast cancer mortality. Male and female workers of each sex with the highest cumulative exposures and longest latency had statistically significant excesses for these two cancers, respectively" (Steenland et al. 2004).

- **Breast cancer incidence:** "Our data suggest that is associated with breast cancer, but a causal interpretation is weakened due to some inconsistencies in exposure-response trends and possible biases due to non-response and incomplete cancer ascertainment." (Steenland et al. 2003)

Collectively, these DNA adduct, genotoxicity, toxicology, and pharmacokinetic evidence contradict all two-piece spline models or truncated linear models (e.g., applied to a small subset of cases at lower exposures) featuring a biologically implausible steep low-dose slope based on a small subset of cases. Rather, the weight of scientific evidence, including data from the NIOSH studies, supports the opposite model shape, a shallow, single-slope linear dose-response model, such as a standard log-linear Cox proportional hazards (CPH) model. This was the model used by Steenland et al. 2003, 2004, as it is the most consistently used epidemiology approach to statistical analyses of continuous data.

The Texas Commission on Environmental Quality TCEQ (2020) successfully applied this standard CPH approach to lymphoid mortality. This same approach can also be applied to breast cancer mortality or incidence data. We provide values derived from TCEQ, Valdez-Flores et al. (2010) and EPA (2016) slope models based on the NIOSH lymphoid mortality,

breast cancer mortality and breast cancer incidence data. However, it should be noted that the Steenland et al. (2004) breast cancer incidence study is of low quality for quantitative risk assessment due to substantial under-ascertainment.

We urge OEHHA to reevaluate its study evaluations in a more balanced manner and align its model selection with current biological evidence. Finally, OEHHA should include an alternative IUR derivation based on animal cancer bioassay data for a more robust scientifically sound risk assessment.

OEHHA's Selection Criteria for Dose-Response Modeling is Flawed

OEHHA incorrectly treats categorical relative risks (RRs – odds ratios) as the data and the "gold standard" for evaluating fit of different forms of continuous Cox Proportional Hazards (CPH) models. This misinterpretation led OEHHA to visually judge continuous model fit based on how closely the curves conform to a few categorical RR estimates in shape and along the y-axis. Consequently, OEHHA prioritizes incorrect subjective visual fit over more objectively rigorous statistical measures of goodness of fit, which OEHHA completely ignores in the draft assessment.

The erroneous practice of using categorical RRs to visually judge continuous model fit has been rigorously addressed by TCEQ (2020) and Valdez-Flores and Sielken (2013). In fact, OEHHA's y-axis comparisons directly contradict explicit footnotes in OEHHA own Figures 4 (and EPA IRIS (2016) documentation) stating that such comparisons are invalid. Valdez-Flores and Sielken (2013) demonstrate how the exposure-response pattern varies with the number of exposure categories utilized. Thus, categorical estimates should not be a basis for fitting the individual hazard rates that are being modeled by different forms of CPH models.

To illustrate this systemic issue, the standard log-linear (TCEQ's approach), linear and 2-piece linear spline (OEHHA's approach) CPH models are statistically significant for breast cancer incidence ($p=0.02$, $p=0.01$, $p=0.04$). However, the linear and log-linear models visually appear lower than the categorical estimates ($p=0.11$). OEHHA improperly rejects the log-linear model simply because it appears out of alignment with categorical relative rate (RR) estimates, even though those RR estimates do not represent the individual hazard rates. OEHHA's framework for continuous model selection is fundamentally flawed and internally inconsistent.

Reality Check on IURs

Utilizing a variety of methods and distinct datasets outside of the NIOSH cohort, we evaluated the consistency of the IRIS value against independent benchmarks:

- **The UCC Manufacturing Cohort**: The independent Union Carbide Corporation (UCC) cohort study of EO manufacturing workers tracked an equivalent number of lymphoid cases to the NIOSH male cohort. Valdez-Flores et al. (2025) concluded that the UCC cohort data does not support the steep dose-response model for lymphoid mortality predicted by IRIS.
- **Epidemiological Evidence from Smokers**: Cigarette smoke contains measurable concentrations of EtO, resulting in demonstrable increases in systemically absorbed EtO above natural background levels. Despite this elevated exposure, the large and extensively studied risks of cancer among smokers do not exhibit an elevated risk for lymphoid cancers. As further examined in these comments, a new analysis of a larger U.S.-based study of cigarette smokers (Troy et al., 2010) shows that if the IRIS potency value were true, it should yield detectable increased risk of lymphoid cancers—outcomes that were entirely absent in smoker epidemiology studies.
- **Endogenous Exposure Disconnect**: EtO is formed endogenously in the human body via normal ethylene metabolism. Even non-smokers exhibit significant background levels of endogenous EtO, as measured by EtO-derived hemoglobin adducts. These daily, naturally produced exposures are substantially higher than the exposure levels contributed by EtO in ambient air. Crucially, the U.S. EPA's risk-specific concentration is thousands of times lower than these natural endogenous levels.

OEHHA has a range of options that it can utilize to select a scientifically grounded unit risk level instead of the scientifically implausible IRIS value**.**

- There are several rodent carcinogenicity studies with the highest risk value being 8.33×10^{-2} per ppm from a mouse study. USEPA used this value in the Office of Pesticide Programs (OPP) risk assessment in 2008, well after the NIOSH studies were published (U.S. USEPA, 2008)
- It could use NIOSH data but with more biologically plausible standard log-linear CPH dose-response models (TCEQ's modeling approach) applied to

both lymphoid and breast cancer mortality. The values range from 5.64×10^{-2} to 7.01×10^{-2}.

These recommended values are broadly consistent with one another, biologically plausible, and present EtO cancer risks substantially lower than the IRIS unit risk of 9.1 per ppm.

Conclusion

In conclusion, the weight of scientific evidence demonstrates that OEHHA's draft risk assessment relies on a biologically untenable dose-response framework. A scientifically robust IUR must reject the implausibly steep low-dose slopes driven by a small fraction of cases. Furthermore, independent benchmarks—including the UCC manufacturing cohort, modern molecular dosimetry, and empirical "reality checks" from heavily studied smoker cohorts—stand in direct contradiction to deriving an exceptionally high inhalation IUR.

We urge OEHHA to correct its study quality criteria and evaluations, abandon incorrect and subjective visual fit comparisons, and align its final risk assessment with the overwhelming biological evidence that supports applying the standard log-linear CPH model to the NIOSH data. Given the severe lack of historical exposure data and major exposure reconstruction uncertainties within the NIOSH study, OEHHA should also derive an alternative IUR based on animal cancer bioassay data, which do not suffer from the exposure measurement errors.

1. OEHHA Ranking of the Quality Characteristics of the NIOSH and UCC epidemiology studies shows distinct and erroneous bias favoring the NIOSH study

OEHHA utilizes a qualitative analysis to evaluate the validity and utility of three ethylene oxide (EtO) epidemiological studies (NIOSH, UCC, or Sweden) for exposure-response modeling and unit risk calculation. To identify the study most appropriate for quantitative risk assessment, OEHHA assigned one of four qualitative ratings (high, adequate, low, or critical) to 35 relevant factors, as detailed in OEHHA Figure 3, with the rationale for each rating documented in Attachments D and E.

Based on this evaluation, OEHHA concluded that the NIOSH lymphoid mortality and breast cancer incidence analyses are high-quality studies that provide the most sensitive and reliable basis for quantitative risk assessment. In contrast, the Union Carbide Corporation (UCC) study received an overall "low" rating, including seven "low" and six "critical" rankings, while the NIOSH studies received no comparable deficiencies.

We agree that, among the available epidemiological studies, the NIOSH study by Steenland et al. (2004) is the most appropriate candidate for quantitative risk assessment, provided that the selected exposure-response model is biologically plausible and consistent with the overall epidemiological evidence. However, the NIOSH study also contains important limitations that are largely overlooked in OEHHA's evaluation. These limitations introduce substantial uncertainty into the resulting risk estimates. Consequently, the NIOSH study should not be used in isolation for quantitative risk assessment without considering evidence from the UCC studies and comparisons with EPA's animal-based inhalation unit risk.

A broader concern is that OEHHA's rating framework penalizes (or rewards) studies multiple times for highly correlated factors. For example, while the NIOSH cohort benefits from the inclusion of both sexes, the UCC cohort is effectively penalized multiple times for being a male-only study. This approach creates the appearance of substantially more "critical" deficiencies in the UCC studies than in the NIOSH studies, as reflected in OEHHA's color-coded heat map. The resulting concentration of red cells exaggerates the apparent difference in study quality.

In addition, OEHHA combines studies addressing different outcomes from multiple studies into a single rating which reduces the usefulness of the rankings for quantitative risk

assessment purposes. By combining multiple studies and outcomes together, the ratings are not helpful to supporting OEHHA's derivation of the IUR. For example, the Kelly-Reif et al (2025) is an update of the NIOSH cohort focused exclusively on breast cancer mortality, an endpoint OEHHA did not use for quantitative risk assessment purposes. By combining all NIOSH endpoints into a single evaluation, OEHHA obscures important differences in study quality of greatest relevance for the draft OEHHA IUR. Most importantly, OEHHA fails to critically evaluate the quality of the NIOSH breast cancer incidence analysis. As discussed in Appendix 1, evidence of case under-ascertainment in the breast cancer incidence dataset raises concerns about its suitability for quantitative risk assessment.

Study quality should be evaluated separately for lymphoid cancer mortality and breast cancer incidence to provide the level of granularity required for quantitative risk assessment. Likewise, breast cancer incidence should be evaluated independently from breast cancer mortality because OEHHA selected incidence—not mortality—as the critical endpoint for derivation of the draft IUR. For a more balanced comparison of the NIOSH and UCC cohorts, OEHHA should also consider an apples-to-apples comparison of lymphoid cancer mortality in males, particularly given that the original investigators reported associations only among males (Steenland et al. 2004). This allows consideration of the UCC cohort as weight-of-evidence regarding what type of continuous model should be applied to the NIOSH lymphoid mortality data.

OEHHA should also consider the implications of the currently available NIOSH updates. Kelly-Reif et al. (2025) did not report updated lymphoid mortality analyses, even though Picciotto et al. (2025) updated the broader lymphohematopoietic (LH) mortality category, of which lymphoid mortality is a subset. While there may be legitimate reasons why these analyses have not yet been published, the omission is nevertheless noteworthy given the historical importance of lymphoid mortality in EtO risk assessment. Complete reporting of all major outcomes, including analyses using the standard log-linear Cox proportional hazards (CPH) model employed by Steenland et al. (2004)—is necessary before the NIOSH updates can be considered a complete and transparent evaluation of the cohort.

The most striking example of OEHHA's preferential treatment of the NIOSH studies is its evaluation of exposure assessment quality. The NIOSH exposure estimates warrant a rating of "critical" rather than "adequate," while the UCC exposure assessment should be upgraded from "critical" to "adequate." OEHHA's downgrading of the UCC studies is due in part to its failure to consider the quantitative analyses of Valdez-Flores et al. (2010, 2025), which include evaluations of lag periods, categorical exposure analyses, and other important sensitivity analyses. At the same time, OEHHA relies on Picciotto et al. (2025), a companion analysis to Kelly-Reif et al. (2025), to support the NIOSH findings. Applying one

standard to the NIOSH studies while disregarding comparable follow-up analyses of the UCC cohort represents a clear double standard.

OEHHA's treatment of the Union Carbide cohort updates reflects an apparent bias against industry-funded research rather than an objective application of the National Toxicology Program's Office of Health Assessment and Translation (OHAT) framework (NTP, 2019), which OEHHA cites as the basis for its review. OHAT does not consider financial conflicts of interest to be an automatic basis for a critical risk-of-bias rating. Instead, it calls for an evaluation of whether a potential conflict actually influenced study design, conduct, analysis, interpretation, or reporting. OEHHA identifies no evidence that industry funding affected the methods or findings of the Valdez-Flores analyses. As a result, OEHHA's "critical" rating appears to be based primarily on sponsorship rather than demonstrated methodological deficiencies, which is inconsistent with the principles of the OHAT systematic review framework. More generally, the relevant question is not whether a potential conflict exists, but whether there is evidence that it affected specific methodological decisions or reporting of study findings. OEHHA identifies no such evidence for the UCC cohort updates.

Correcting OEHHA's study-quality ratings would materially change the weight assigned to the UCC cohort. While the UCC study may not be the preferred basis for deriving the IUR, its lymphoid mortality findings remain highly relevant to evaluating the shape of the exposure-response relationship. The UCC data therefore provide an important independent check on whether the dose-response models selected for the NIOSH cohort are consistent with the broader epidemiological evidence. Accordingly, the UCC lymphoid mortality study should be considered when selecting the most appropriate dose-response model for the NIOSH data. The UCC findings provide independent epidemiological evidence regarding the shape of the exposure-response relationship and do not support the steep low-dose response implied by OEHHA's preferred model.

Appendix 1 provides detailed comments and corrections of OEHHA's rankings comparing NIOSH and UCC studies.

1.1. NIOSH epidemiology study exposure assessment should be ranked as "critical"

The NIOSH cohort estimated worker exposures to EtO over a 50 year period, from 1937-1988. The NIOSH exposure regression model's estimates based on monitoring data after 1970 was rigorously validated by comparisons with actual independent exposures available from 1976-1985. However, little or no historical exposure monitoring data existed

for the 40-year period from 1976 to 1977. As a result, the exposure estimates assigned to this earlier period were derived by extrapolating the regression model beyond the range of available data and were never independently validated.

Critically, the post-1978 validated model was fundamentally altered for application to these earlier decades by fixing the "calendar year" variable at 1978 levels. This improper constraint ignored the impact on exposures of processes unique to the 1940s-era, i.e., industrial hygiene, air wash frequencies, and equipment integrity (e.g., inferior seals and gaskets), assuming they were equivalent to 1978 standards, despite published literature to the contrary (Bogen et al., 2019). By also reducing the exposure regression model effectively to a function of "sterilizer volume," a main exposure predictor post 1978, it produced implausibly low exposure estimates for the 1940s–1960s that contradict the historical record and have an opposite trend than the evolution of the American Conference of Governmental Industrial Hygienists (ACGIH) threshold limit value-time weighted averages (Figure 1).

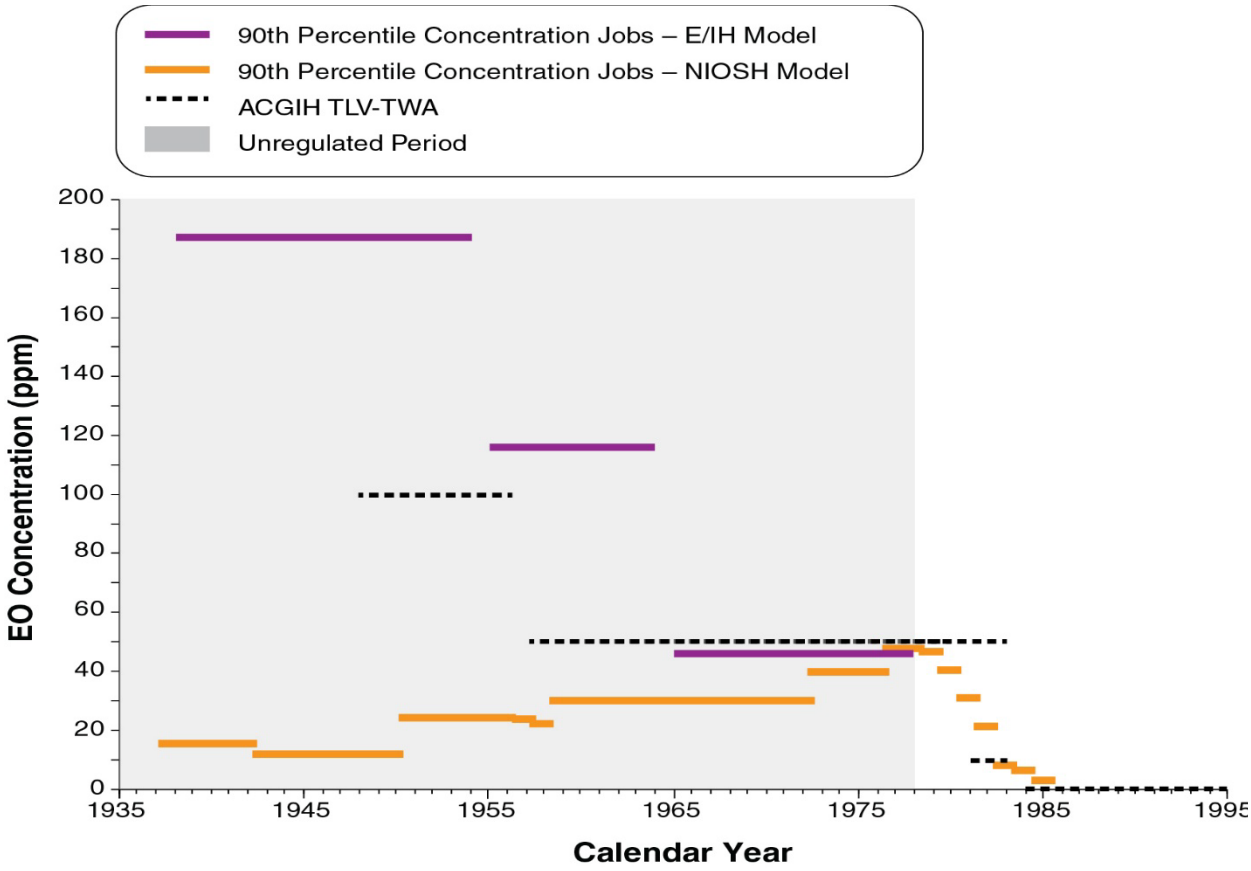


Figure 1. Comparison of Exposure model estimates for sterilizer operators by NIOSH (orange) and Bogen et al. (2019; purple) assuming sterilizer operators are 90th percentile exposures estimated by NIOSH for sterilizer workers in 1978.

Note: ACGIH TLV-TWA stand for the American Conference of Government Industrial Hygienists threshold limit value-time-weighted averages. Bogen et al. emphasized directional trends rather than absolute values based on published historical information and interviews with sterilizer operators who worked or had knowledge of working conditions before 1978.

These exposure estimation errors introduce broad uncertainty and a high potential for bias at the lower exposure range. Further extensive and fundamental uncertainty in the NIOSH assessment of EtO exposure in its EtO cohort is confirmed by a new analysis of NIOSH EtO-cohort occupational exposure to EtO discussed in Section I.B below. Specifically, this uncertainty implies that workers with high actual exposures in earlier decades are substantially likely to have been incorrectly modeled as having low exposures, making the NIOSH estimates an unreliable basis for quantitative risk assessment. These issues apply to all previous and recent publications from NIOSH.

1.1.1. OEHHA's evaluation of the NIOSH study exposure model is flawed and should be downgraded to "critical" due to lack of a validated exposure model for the 40 years prior to 1978.

OEHHA states incorrectly that "each participant's EtO exposure was estimated using a validated multiple regression exposure model that incorporated information on workplace air measurements, sterilization unit size, engineering controls, timing of sterilization, product type, calendar year, and historical process changes." This is incorrect for participants who worked in earlier periods prior to 1978. OEHHA made several critical errors in describing, objectively evaluating, and ranking the quality of this NIOSH exposure model.

Specifically, the model can only be described as validated for exposures between 1978-1985. OEHHA failed to note that the NIOSH multiple regression model was optimized to fit exposure data collected exclusively across this eight-year period The original model included "calendar year" as a necessary variable, which the NIOSH authors considered a surrogate for workplace conditions vital to predicting post-1978 exposures. The model variables consisted of:

- Exposure category (7 types, e.g., sterilizer operator)
- Product (5 types)
- Age (i.e. freshly sterilized =1)
- **Calendar Year** (considered an essential variable during the validation exercise, but then frozen to 1978 when estimating 40 years of exposures)
- Rear Exhaust – Yes/No
- Aeration – Yes/No
- **Volume -Cubic Feet**
- **Volume^2**

While the NIOSH exposure model was validated exclusively using post-1978 exposure data, its application to earlier periods was substantively altered. For the 40 years of exposures prior to 1978, the key "calendar year" variable was arbitrarily frozen to 1978 levels. Because it was never validated for the 40 years of exposure prior to 1978, this modified pre-1978 exposure model is unvalidated. Furthermore, it depends on the demonstrably false assumption that industrial hygiene practices, engineering controls, sterilizer equipment modifications (e.g. tighter gaskets and seals), and workplace conditions from the 1940's through the mid-1970s were identical to those in 1978.

With this arbitrary adjustment, sterilizer volume becomes the primary determinant of historical exposure levels. This unvalidated pre-1978 exposure model assumes that "increased sterilizer volume generally resulted in higher predicted average exposures until the late 1970's" as illustrated in EPA IRIS (2016) Table I-3 for Plant 5 below. Consequently, the model erroneously predicts that exposures were lower in earlier decades simply because historical sterilizer volumes were smaller.

Table I-3. Plant 5, sterilizer volume and predicted ETO exposure levels by year

Plant 5, Dept 1, Oper ZZ		
Year	Sterilizer volume (cubic ft)	Predicted ETO level (ppm)
1943–50	887	6
1951–61	1,679	15
1962–70	1,304	10
1971–72	1,964	18
1973–76	2,624	26
1977–78	3,284	32

I-27

The U.S. EPA then utilized the outputs generated by this altered, unvalidated pre-1978 exposure model to justify the validity of the model's own underlying structure. Using the predictions of an altered model to prove that the model itself is correct represents clear circular logic. Consequently, OEHHA should more objectively downgrade its scoring of the NIOSH exposure model to "critical" because it introduces a high likelihood of exposure measurement errors.

The resulting NIOSH pre-1978 exposure estimates by job category show a flawed pattern of lower exposures in earlier years (1940's) increasing to peak levels in 1978. In other words, the NIOSH model nonsensically predicts lower exposures for early-era workers who operated under the crudest control technology compared to workers in the 1970's using more sophisticated technology (Figure 2, see also Bogen et al. 2019).

The draft OEHHA challenges this pattern by selectively presenting the median and inter-quantile range of exposures for cases compared to controls (Figure 3A), without also presenting the 95 percentile which represents the most exposed workers such as sterilizer operators (Figure 3B, same as draft OEHHA Figure 4). These figures from EPA IRIS (2016) do not directly reflect what the NIOSH exposure regression model predicts. Figures 1 and 2

are a more direct representation of the NIOSH exposure regression model estimates because it focuses on exposures by job category (department/operation) rather than by individual worker (who can work different types of jobs). In contrast, Figure 3A and 3B reflect exposures for cases based on each workers work history.

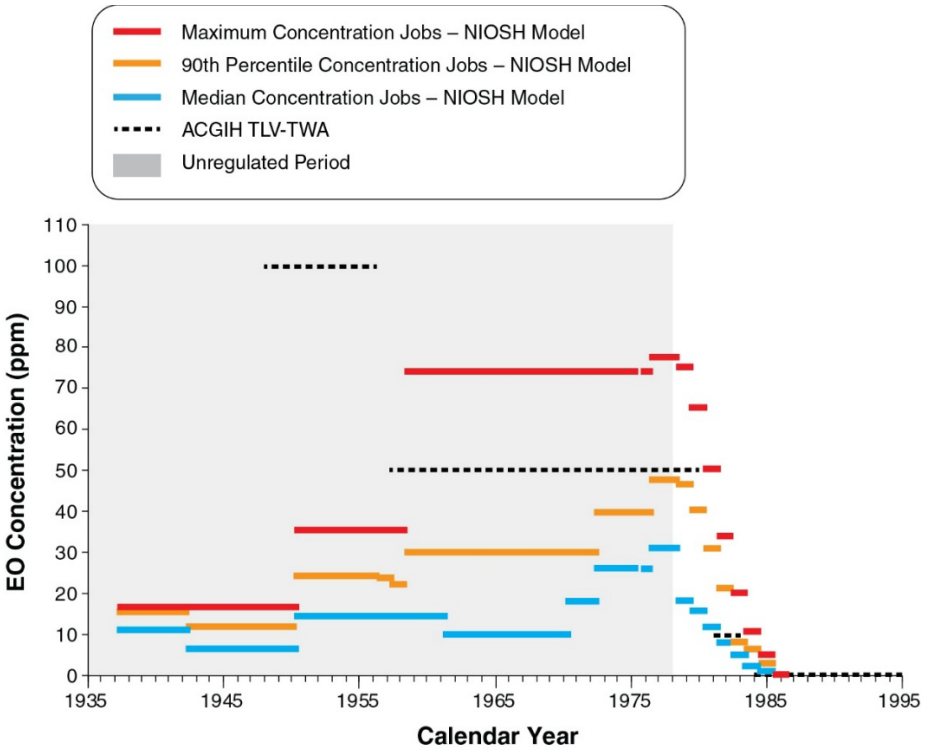


Figure 2. NIOSH pre-1978 exposure estimates used by IRIS based on NIOSH exposure model output file for different types of jobs (i.e. job/plant/department/operation/year combination)

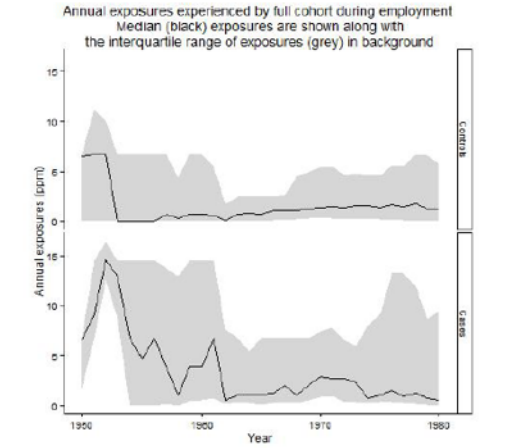


Figure D-22. Estimated annual exposures experienced by cases and controls in the full cohort while working[1]—medians and interquartile ranges[2]

[1]Annual exposure histories taken from NIOSH deidentified exposure records; include 382 workers ultimately removed from the analysis due to inconsistencies in the record.
[2]Prior to 1962, fewer than five cases were working in any given year. This resulted in a number of years where the 25th, 50th, and 75th percentiles of the exposure distribution were identical or nearly so.

Figure 3(A)

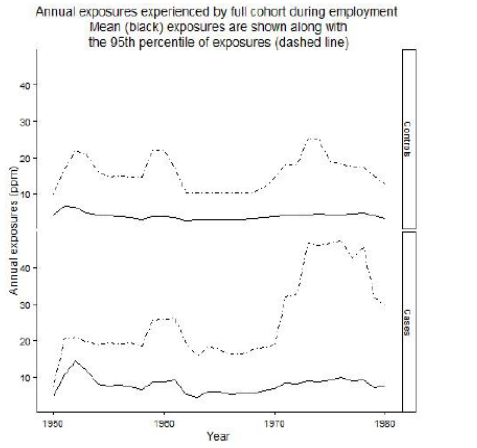


Figure D-23. Estimated annual exposures experienced by cases and controls in the full cohort while working[1]—means and 95th percentiles

[1]Annual exposure histories taken from NIOSH deidentified exposure records; include 382 workers ultimately removed from the analysis due to inconsistencies in the record.

Figure 3(B)

Figure 3. (A) EPA Figure D-22 and (B) EPA Figure D-23

Note: Figure 3A is the same as draft OEHHA Figure 4

The draft OEHHA Figure 4 (shown above as Figure 3A of these comments) is misleading because it does not reflects the exposure estimates of the NIOSH exposure regression model and defines "controls" in a manner that is unrelated to the dose-response model analyses. It is true that there were **53** cases, representing workers who died from lymphoid cancer. However, the "controls" shown in Figure 3 are the 17,070 workers who did not die from any lymphohematopoietic diseases. These workers do not represent the actual controls, or risk set of approximately 100 matched controls. Only these matched controls were included in the dose-response model analyses. In fact, a matched control could include a person who died of lymphohematopoietic disease after the case died. Consequently, draft OEHHA Figure 4 presents a characterization of "controls" that differs substantially from those actually used in the continuous dose-response analysis. At best, the figure is confusing; at worst, it obscures the true nature and definition of the control group and the historical exposure estimates.

OEHHA should remove its Figure 4 because it is misleading. If OEHHA elects to retain the figure, it should clearly explain that the individuals labeled as "controls" are not the controls (i.e. members of the risk sets) used in the dose-response analyses. Instead, the figure presents all workers who did not die from lymphohematopoietic cancer, whereas the dose-response analyses were based on matched risk sets constructed for each case.

OEHHA should also revise the text to clarify that the NIOSH exposure model was not validated for exposures prior to 1976 and that the model predicts lower ethylene oxide exposures for sterilizer operators in the 1940's and 50's than for sterilizer operators in the 1960's and 70's.

OEHHA should add to the explanation of risk sets on page 34 of the draft OEHHA IUR that EPA IRIS did not justify limiting each case's risk set to 100 matched controls, even though considerably larger numbers of eligible matched controls were available. In contrast, TCEQ and Valdez-Flores et al. (2010, 2025) did not impost a 100-control limit on the risk set for each case.

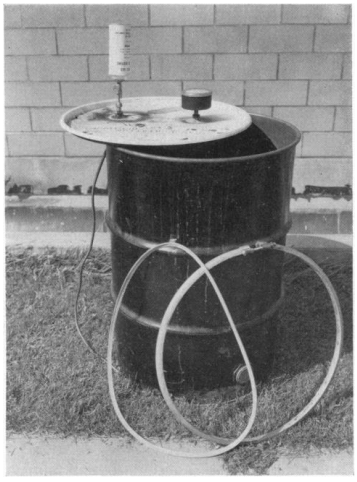
1.1.2. Ethylene oxide sterilization historical literature provides ample evidence that exposures were higher in earlier years.

Examples from two pioneers in the development of EtO as a sterilizing agent provide clear evidence that the NIOSH model is flawed in predicting lower exposures in earlier years when there were smaller sterilizer chambers:

The 1957 McDonald EtO Sterilizer Patent
(https://patents.google.com/patent/US3068064A/en)

The 1957 U.S. Patent for an EtO sterilizer developed by Robert McDonald that the sterilizer can “reduce to a minimum the objectionable odor of any gas remaining in the chamber”. This statement is informative because the odor detection threshold for EtO has been reported to range from approximately 260 to 700 ppm (National Research Council, 2010). The patent’s concern with minimizing residual odor suggests that in the 1950’s and 60’s workers may have been exposed to concentrations at or above the odor threshold, potentially exceeding 200 ppm.

The patent also notes that residual EtO could “cause blisters if the good being sterilized are goods such as rubber gloves”. The potential for blister formation indicates the presence of substantial residual EtO concentrations within sterilized materials and, consequently, very high airborne concentrations during unloading and handling operations. Such observations are difficult to reconcile with NIOSH’s assumption that exposures were lower during early sterilization operations.



Early Perceptions of EtO Toxicity

Additional historical context is provided by Phillips and Kaye (1949) who characterized EtO as “low toxicity to human beings.” Although the sterilizer shown in their publication was not widely commercialized, it illustrates the limited concern regarding occupational exposure and toxicity during the 1940s and 1950s. More importantly, it underscores the weakness of the pre-1978 NIOSH exposure model assumption that smaller volume sterilizers necessarily resulted in lower worker exposures. Early sterilization processes often involved minimal engineering controls, little awareness of occupational hazards, and practices that would likely have resulted in higher exposures to EtO in the 1940’s-60’s compared to 1978.

The draft IUR also fails to adequately consider the findings of Bogen et al. (2019), which directly address the substantial limitations of the historical exposure estimates used in the NIOSH cohort. Bogen et al. (2019) developed an engineering and industrial hygiene model specifically designed to reconstruct historical exposures among sterilizer operators. The model incorporated multiple sources of historical information, including:

- Ethylene oxide (EO) microbial kill concentrations
- EO residue levels in sterilized materials
- Post-wash EO chamber concentrations based on the number of wash cycles
- Historical facility design characteristics

To reconstruct workplace conditions during periods predating systematic exposure monitoring, Bogen et al. supplemented published information with interviews of sterilizer operators who possessed firsthand knowledge of historical operating conditions. Although U.S. EPA (2020[2]) largely dismissed the model because it relied in part on historical interviews, the Agency failed to recognize the importance of operational knowledge in interpreting and applying the historical information.

This data-driven, operational history approach is substantially more credible than the Hornung et al. (1994) pre-1978 exposure model used by NIOSH. Hornung et al. made no attempt to verify its core assumption that industrial hygiene practices, engineering controls, and workplace conditions from the 1940s through the mid-1970s were the same as those observed in 1978.

Although Bogen et al. (2019) emphasized directional trends rather than prediction of actual levels, U.S. EPA criticized Bogen et al. (2019) because the model projected sterilizer operator exposures in the early years (1930-1940s) to exceed the 100 ppm level established by ACGIH in 1948. However, this criticism misses the more important question. Rather than asking why Bogen's estimates exceeded historical exposure guidelines, EPA should have examined why the NIOSH model produces the opposite trend despite historical evidence suggesting potentially much higher exposures.

As discussed in the previous section, exposures at or above the EtO odor threshold—approximately 200 ppm and higher—are entirely plausible given the operating conditions described in historical patents, publications, and industrial practices. Consequently, the directional trends predicted by Bogen et al. appear more consistent with the historical record than the reverse trend of lower NIOSH exposures in earlier years that underlie EPA/OEHHA analyses.

1.1.3. The vast majority (>80%) of the exposure for lymphoid cases occurred in the pre-measurement period where the unvalidated pre-1978 NIOSH exposure model predicts lower exposures in earlier years

We have grouped the 53 lymphoid mortality cases by the following three cumulative exposure categories to show how much of their exposures occurred during the period pre-1978, when the invalidated NIOSH regression model was applied. For each case, the range

[2] EPA (2020) p. 87 states "Bogen et al. attempted to project EtO concentrations for time periods preceding the available measurement data; however, the methodology was not sufficiently document to evaluate how the model was derived. . .Finally, model projections of industry-wide exposures of sterilizer operators for much of the earlier time periods were in excess of then current ACGIH health criteria.

of years worked is represented by a horizontal line segment. Cases in Figure 4 are grouped into three exposure categories consistent with the USEPA IRIS (2016) two-slope spline model reflecting increasing cumulative exposure :

- **Lagged-out cases (n = 9):** assigned zero effective exposure under a 20-year lag structure;
- **Low-exposure group (n = 13):** cumulative exposure ≤ 1,600 ppm-days (below the spline knot); and
- **Higher-exposure group (n = 31):** cumulative exposure > 1,600 ppm-days (above the spline knot).

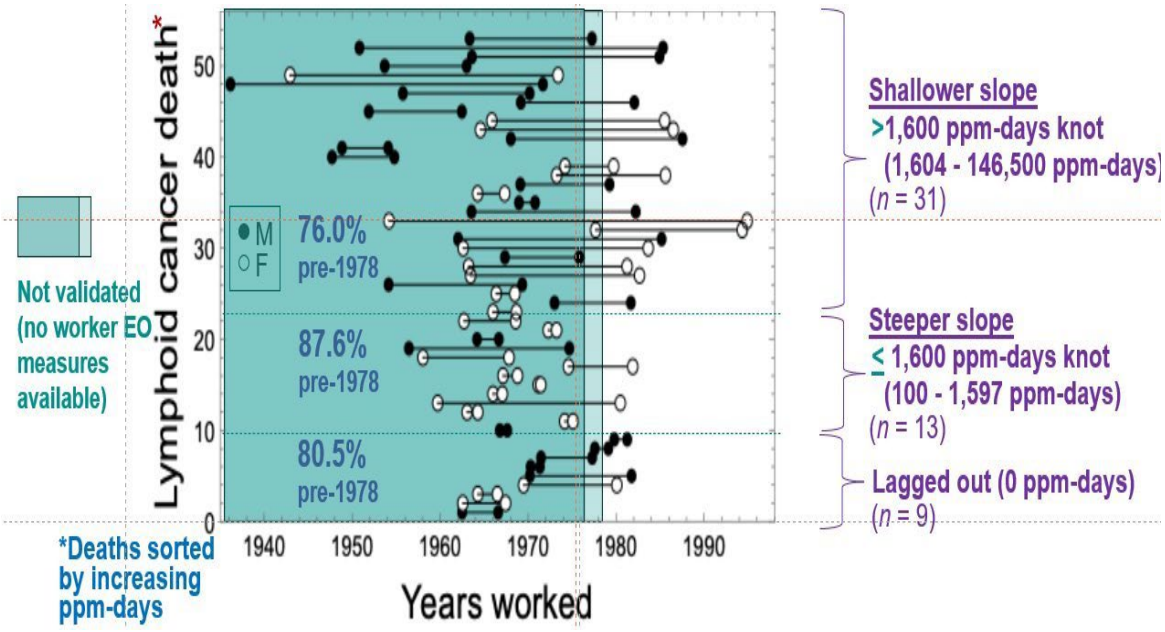


Figure 4. Range of employment years (x-axis) for 53 NIOSH cohort workers who died of lymphoid cancer (cases indexed on the y-axis), ordered by increasing cumulative exposure within three exposure categories: lagged-out (zero effective exposure), ≤1,600 ppm-days, and >1,600 ppm-days. Shaded regions indicate periods with no (dark shading) or limited (light shading) exposure measurement data. Male and female cases are denoted by closed and open symbols, respectively. Source: Analysis of summary tables supporting Valdez-Flores et al. (2010) provided by Dr. Ciriaco Valdez-Flores.

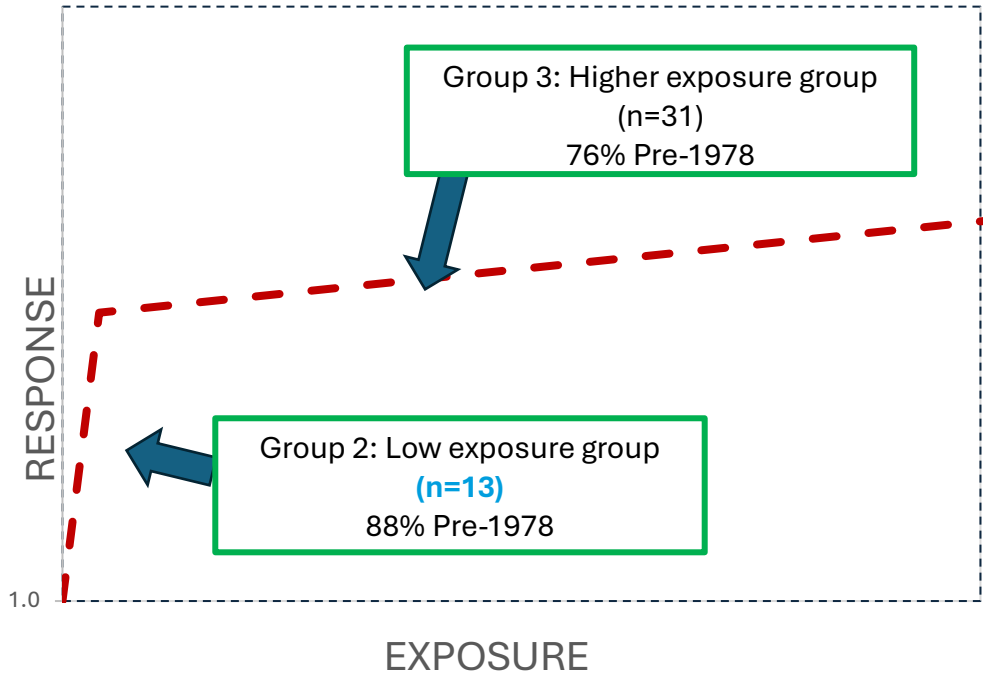


Figure 5. Graphical Representation of the USEPA IRIS (2016) 2-Piece Linear Spline CPH Model Cancer Risk Assessment for Ethylene Oxide.

Note: Group 1 (n=9 lagged out) not shown because all exposures in this group occurred within the 20-year lag period assumed for lymphoid cancer; these workers are thus assigned zero effective cumulative exposure in the risk model even though they were exposed to EtO.

The high exposure uncertainty also applies to breast cancer mortality (see Appendix B). More than 90% of the exposures contributing to cases that are lagged out, as well as those in both the low-exposure group below the knot and the high-exposure group, occurred before 1978. Consequently, there is substantial potential for exposure measurement error in the breast cancer mortality analysis. This limitation is also likely to affect the breast cancer incidence analysis.

1.2. Methodological double standard in OEHHA's quality ratings and why OEHHA's study quality rating for UCC should be higher

An objective evaluation demonstrates that the UCC study possesses a superior exposure assessment compared to the NIOSH study across their respective operational timelines:

- **Data-Rich Era (1957–1988):** Empirical exposure data was available for over 75% of the UCC cohort. This dataset included routine industrial hygiene monitoring, personal breathing-zone sampling, and a comprehensive plant-wide survey from a nearly identical UCC facility utilizing the exact same manufacturing processes.
- **Evidence-Based Historical Reconstruction (1940–1956):** For early operational periods lacking direct measurements, UCC exposure estimates were **based on** empirical data from a similar chemical operation in Sweden, **whose operations like UCCs were in enclosed buildings.** These estimates took into consideration internal UCC plant medical records documenting chemical injuries and acute respiratory irritation in high-exposure departments during that time period.
- **Proportional Historical Scaling (1925–1939):** UCC exposures during the earliest decades were systematically scaled upward. These adjustments logically accounted for less sophisticated manufacturing equipment and an early industry lack of operational experience. **A very small number of** study subjects **(98)** worked during this time period, **as the study was designed to follow workers active in 1940 or hired thereafter.**

Application of this methodology resulted in average worker cumulative exposure of 67 ppm years, compared to the estimated 27 ppm years for the NIOSH cohort. Unlike the NIOSH model, which predicts lower exposures during the earliest and least-controlled period, the UCC exposure assessment provides estimates that are more consistent with the expectation of higher exposures before the implementation of modern exposure controls and monitoring practices. Epidemiologic studies that include workers with higher cumulative exposures provide greater potential to detect exposure–response relationships, if such relationships exist, and therefore offer a stronger basis for identifying potential carcinogenic hazards.

OEHHA should upgrade the Union Carbide Corporation (UCC) cohort study rating from "critical" to "adequate," establishing it as a key dataset. The UCC study (Greenberg et al. (1990); Swaen et al. (2009)), provides superior exposure assessment through extensive monitoring data (1957–1988) and historical reconstruction (1940–1956, 1925–1939). This dataset, featuring detailed internal analyses, aligns with industrial hygiene principles to offer a more realistic basis for risk assessment than the NIOSH model.

2. The OEHHA approach to model selection relies on flawed visual fit comparisons with categorical estimates along the y-axis and an unjustified focus on modeling a fraction of cases at lower exposures

The Draft OEHHA IUR identifies six principles for selecting continuous exposure-response models, with particular emphasis on visual fit comparisons along the y-axis (i.e., magnitude of Relative Rate).

1. The use of individual level continuous exposure data over categorical data
2. Prioritizing the specific fit of the model at lower exposure levels rather than the overall fit of the model at high and low exposure levels combined. This principle is particularly important since some models appear to dramatically underestimate risks at lower exposures.
3. Parsimony
4. Not solely relying on Akaike Information Criteria (AIC) scores for model fit
5. Biological plausibility and statistical considerations.
6. Visual fit comparisons of the continuous data model results to the categorical odds ratio.

OEHHA's framework for continuous model selection is fundamentally flawed and internally inconsistent. The first principle emphasizes the use of individual-level continuous exposure data rather than categorical "data." The underlying "data" are the individual hazard rates that are modeled using continuous exposure-response functions. We agree with the importance of prioritizing individual-level continuous exposure-response models over categorical **estimates**. Accordingly OEHHA should revise all references to categorical "data" and replace them with categorical "estimates" because these categorical "estimates" are NOT the data modeled.

OEHHA departs from the first principle when selecting continuous models by effectively treating categorical estimates as though they are the data being modeled, comparing all the continuous models against these few categorical estimates. By relying on the pattern of a limited number of categorical estimates to determine which continuous model provides the "best fit" to the individual hazard rates, OEHHA abandons its stated preference for individual-level continuous models.

The inconsistency becomes even more pronounced when OEHHA relies on figures displaying **Relative Rate** (RR) values to compare continuous models against categorical estimates **along the y-axis**. OEHHA uses these comparisons to support its conclusion that

the TCEQ model underestimates the categorical estimates. However, OEHHA explicitly acknowledges that such comparisons are not valid:

> ". . .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." (OEHHA Figure 11 p. 59)

Thus, OEHHA's critique of the TCEQ model is based on a comparison that it expressly warns should not be used for model evaluation. Having recognized that RR values from different models are not directly comparable along the y-axis because of differing baseline risk estimates, OEHHA nevertheless relies on those same comparisons to conclude that the TCEQ model underpredicts the categorical estimates. This approach is both internally inconsistent and contrary to OEHHA's own stated principles for model selection.

OEHHA's criterion of prioritizing model fit at lower exposure levels over overall fit across the full exposure range places disproportionate weight on a relatively small subset of the available data. The more fundamental issue, however, is that OEHHA provides no scientific, biological, or statistical justification for allowing approximately 13 of the 44 unlagged cases (about 30% of the cases) to drive model selection while effectively discounting the remaining 70% of the observations. In contrast, the TCEQ approach evaluates model performance using all 44 unlagged cases and therefore incorporates the full body of evidence regarding the exposure-response relationship.

When exposure estimates are uncertain as they are in the NIOSH study, a cancer slope factor based on selection of a model that makes use of the entire dataset minimize uncertainty in any one portion of the exposure range. OEHHA's approach does the opposite by deriving an IUR based primarily on a small subset of the dose-response data This concern is particularly relevant because the NIOSH exposure regression model likely underestimated exposures in earlier years based on an incorrect assumption that industrial hygiene practices and awareness of EtO hazards in 1978 apply to earlier years. If earlier exposures were underestimated, then cumulative exposures assigned to some workers would also be underestimated, potentially affecting the very portion of the dataset that OEHHA has chosen to emphasize.

OEHHA unfairly characterizes TCEQ's approach as primarily focused on statistical criteria. In fact, TCEQ (2020) addressed statistical issues only because EPA IRIS (2016) incorrectly reported the p-values and then used those results to dismiss the more biologically plausible standard log-linear Cox proportional hazards (CPH) model. The corrected p-values demonstrate that the key models cannot be distinguished statistically (Table 1), except that the standard log-linear CPH model is statistically preferable under the principle

of parsimony. A more complex model should not be selected unless it provides a demonstrably better fit to the data, which neither US EPA nor OEHHA have demonstrated.

Table 1. Original and Corrected p-values for comparison of IRIS 2-slope linear spline and IRIS standard CPH model from the null hypothesis

	USEPA (**corrected**) 2-slope linear spline (draft OEHHA preferred model approach)	USEPA log-linear CPH (TCEQ's model approach)
Lymphoid Mortality	**p = 0.14** corrected from 0.07 (knot at 1600 ppm-days)	p = 0.22
Breast Cancer Incidence	**p = 0.04** corrected from 0.01 (knot at 5750 ppm-days)	p = 0.02
Breast Cancer Mortality	**p = 0.15** corrected from 0.07 (knot at 700 ppm-days)	p = 0.10

Source: Corrected and IRIS reported p-values are based on IRIS (2016; Tables 4-2, 4-4, 4-6, 4-12, 4-13, Appendix D) and TCEQ (2020). [3,4] Calculation based on USEPA IRIS Tables D-25 and D-24 values and Chi square calculation for linear model in excel =(Chisq.dist.rt(5.396,1) for 1 modeled parameter ; chi square calculation for linear spline model (Chisq.dist.rt (5.375,3) for 3 modeled parameters.

TCEQ DSD emphasis on biological plausibility is very clearly stated: "First and foremost is the consideration of the chemical's MOA . . . MOA information can help inform expectations about the shape of the curve at low doses . . ." . In contrast, EPA IRIS did not meaningfully consider biological plausibility in its model selection process, except when rejecting two extremely steep models that approached infinity. We provide new and existing published data that further support TCEQ's conclusion that the standard log-linear CPH model has much greater biological plausibility compared the OEHHA/EPA's approach (Section 4).

[3] USEPA, 2016. Evaluation of the Inhalation Carcinogenicity of Ethylene Oxide (Final Report). https://iris.epa.gov/document/&deid=329730

[4] TCEQ (2020)

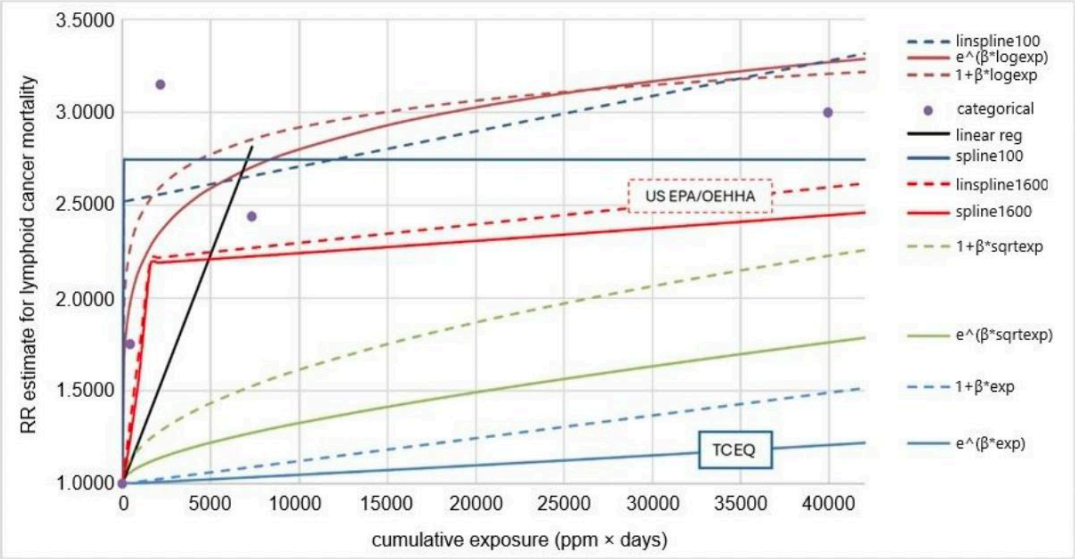
3. OEHHA's inferences on the shape of the dose-response relationship based on categorical estimates and steep log-cumulative models are inappropriate and misleading.

In addition to making incorrect comparisons between the graphs of continuous models with the categorical estimates along the y-axis, OEHHA also incorrectly relies on the shape of the categorical estimates and log cumulative models to support the selection of the 2-linear slope model. TCEQ (2020) and Valdez-Flores and Sielken (2013) demonstrate that a few categorical estimates do not represent the pattern of the individual hazard rates that the continuous models are fitting.

3.1. Categorical estimates should not be the basis for selecting continuous models

In Figure 11 of OEHHA (2026)(provided in **Figure 6A** below), OEHHA relies upon categorical estimates for lymphoid cancer mortalities in NIOSH workers to make inferences about the shape of the dose-response relationship and to support the selection of a nonlinear dose-response model (i.e., two-piece linear spline). Missing from this figure is important information regarding the variation (e.g., no x- or y-error bars are presented for the 4 data points) as well as potential sex differences in the underlying data. The absence of error bars misleads readers in the precision in the four categorical estimates depicted in this figure. The absence of presenting data by sex does not allow the reader to assess the appropriateness of combining data. To remedy this deficiency, OEHHA should rely upon a figure that contains more detailed information as provided below (**Figure 6B**).

6A


RR estimate for lymphoid cancer mortality
3.5000
3.0000
2.5000
2.0000
1.5000
1.0000
0
5000
10000
15000
20000
25000
30000
35000
40000
cumulative exposure (ppm × days)
US EPA/OEHHA
TCEQ
linspline100
e^(β*logexp)
1+β*logexp
categorical
linear reg
spline100
linspline1600
spline1600
1+β*sqrtexp
e^(β*sqrtexp)
1+β*exp
e^(β*exp)

6B

Figure 6. **(A)** Exposure-Response Data and Models for Lymphoid Cancer Mortality in the NIOSH Cohort (male and female workers combined) (from Figure 11 from OEHHA 2026). **(B)** Detailed Plot of Lymphoid Dose-Response Mortality in Male and Female Workers (1 case per data point, source: Dr. Ciriaco Valdez-Flores summary tables supporting Valdez-Flores et al. 2010).

Note: To provided finer granularity than quartiles on the underlying data, the worker mortality data was divided into 1 lymphoid case per categorical group, resulting in 26 female worker groups (blue circles) and 27 male worker groups (orange X's). The finer categorical groups were plotted as a function of cumulative exposure to examine the dose-response relationships for each sex.

From this figure two important insights can be gained:

- **First,** there is a large degree of scatter in the underlying response data (i.e., variation along y-axis), such that attempts to make inferences by USEPA/OEHHA about the shape of the dose-response relationship based on the four categorical estimates (which are not the individual data[5] modeled) depicted in **Figure 6A** without error bars is a fallacy. Uncertainty along the x-axis for cumulative exposure is also present in these data (i.e., high likelihood of exposure error for cases who worked prior to 1978 when NIOSH had zero exposure data and applied an unvalidated pre-1978 exposure model)
- **Second,** The underlying data for male and female workers exhibit very different dose-response behaviors. The male worker data (orange X's in Figure 6B) are consistent with a monotonic dose-response relationship (i.e., highest responses are associated with the highest cumulative EO exposures; depicted by orange dashed line in Figure 6B). In contrast, the female worker data (blue circles in Figure 6B) exhibit an increased response across a narrower exposure range, but without a monotonic dose-response relationship (depicted by blue dashed line in Figure 6B). This pattern is being driven largely by two data points (circled in yellow in Figure 6B), that warrant further examination (e.g., potential diagnostic or exposure measurement errors) The apparent plateau behavior that OEHHA is attempting to capture with the two-piece linear spline is an artifact of the inappropriate decision to combine dissimilar data sets, with the high end of the apparent plateau

[5] For continuous models the individual hazard rates are modeled. Hazard rates compare each lymphoid cancer case to it's matched set of controls. NIOSH only selected 100 controls matched on age, plant, race and sex and randomly selected from the pool of all those who had survived without ***lymphohematopoietic*** cancer to at least the age of the case in that set (Steenland et al. 2004; EPA IRIS Section D.5.2., p. D69). TCEQ's approach was more rigorous because it included the **entire pool** (not just 100) of all those who had survived without ***lymphoid*** cancer to at least the age of the case in that set (Valdez-Flores et al. 2010, 2025).

dependent on the male worker data set and the low end of the apparent plateau dependent on the two female worker data points.

Figure 6A illustrates how the U.S. EPA IRIS quantitative risk assessment EtO relies on data dredging without regard for biological plausibility. In contrast, the TCEQ DSD (2020) explicitly roots its selection of the standard log-linear Cox Proportional Hazards (CPH) model in biological plausibility and corrected statistical methods.

3.2. Log-Cumulative Exposure Models Should Not Be a Statistical or Biological basis to justify the EPA's Spline model.

OEHHA relies on EPA IRIS assessment evaluation of multiple models including the log-cumulative model. EPA (2016) focused on log cumulative exposure models as a major driving force but did not take into account that these models force supralinearity that could be simply linear (Valdez-Flores et al. 2010).

In addition, fitting log cumulative models to the pooled male and female data, results in a model that tries to accommodate the decreasing trend in females (with smaller cumulative exposures) and the increasing trend in males (with larger cumulative exposures), when there is clear epidemiological evidence that the data from both sexes should not be combined (Steenland et al. 2004). At the same time, EPA rejects the log cumulative model because it approaches infinity, and a slope derived from this model would be based on zero cases represented. Thus, it is inappropriate to use this model to justify applying a very steep model, while at the same time concluding that this model is biologically implausible.

Valdez-Flores et al. (2010), a paper that OEHHA overlooked, discussed the inappropriateness of the log cumulative exposure models. They enumerate six reasons: 1) the log cumulative exposure model fits linear or sublinear increasing dose-response relationship with a supra-linear model (supra-linear here means that slope at the origin is infinite and continuously decreases as cumulative exposure increases), 2) the shape of the model with log cumulative exposure has infinite slope at zero exposure, followed by very steep slope at small cumulative exposures and very shallow slopes at high cumulative exposures rendering a biological implausible model, 3) because the logarithm of zero cumulative exposure is minus infinity by definition, in doing numerical approximations in model fitting, an arbitrary value has to be used to replace the minus infinity resulting in model fits that depend on the arbitrary values used to replace minus infinity, 4) even if the data seems to behave in a supra-linear manner, it could be due to several reasons; e.g., incorrect dose metric, exposure measurement error, inappropriate adjustment for

covariates, etc. 5) the degree of supra-linearity of the model depends on the numerical definition of minus infinity, and 6) the implausibility of the shape of the exposure-response model with little change at moderate cumulative exposures. Figure 7 illustrates the behavior of the model with log-cumulative exposure model when the true relationship is linear.

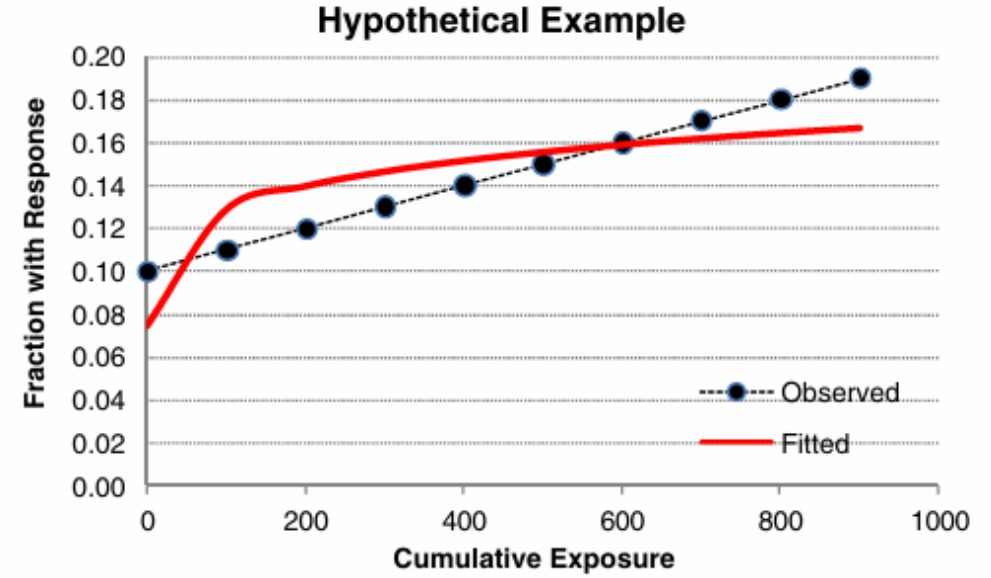


Fig. 1. Inappropriate supralinear fit when the observed data follow a linear relationship with cumulative exposure and there are equal numbers of observed person-years in each of 10 exposure intervals.

Figure 7. Log-cumulative Distortion of a Linear Function

4. Flaws in OEHHA's Assertion of Greater Biological Plausibility for the Two-Slope Spline Model

In its May 2026 public review draft (p. 64), OEHHA deemed the two-piece linear spline model biologically plausible for the following reasons:

1. EtO is a direct-acting genotoxic carcinogen, and dose-response functions that are linear at low doses are considered the most plausible.
2. A plateauing or attenuation of odds ratios at higher exposure levels has been observed in several other occupational carcinogen studies, with Stayner et al. (2003) cited as providing a plausible explanation for this pattern.

The following sections discuss why these reasons fail to establish the biological plausibility of the two-piece linear spline model for EtO risk assessment. Additionally, we present scientific evidence that supports the biological plausibility of a log-linear Cox proportional hazards (CPH) model.

4.1. OEHHA's Rationale #1: EtO is a direct-acting genotoxic carcinogen, and dose-response functions that are linear at low doses are considered the most plausible.

Both the EPA's and TCEQ's model approach are linear at lower exposures, and both EPA and TCEQ use the same linear extrapolation to derive the IUR. Thus, this rationale applies to both the EPA's and TCEQ's model approaches

While the default regulatory assumption for direct-acting genotoxic carcinogens is a linear response at low doses, the slope of that response is chemical-specific. For EtO, there is little evidence to support a hyper-steep, supra-linear response at low doses for the early key events involved in its genotoxic mode of action (MOA) for EtO-induced tumors. These early key events include systemic EtO concentrations measured as hemoglobin (HEV) adducts, DNA adduct formation, gene mutations, and chromosomal aberrations.

Liu et al. (2026) and Gollapudi et al. (2026) examined the dose-response relationship for these early events to evaluate whether any low-dose supra-linearity exists. Liu et al. measured DNA adduct formation across multiple tissues (lung, liver, bone marrow, and mammary gland), while Gollapudi et al. assessed apical genotoxicity endpoints, including gene mutations and cytogenetic effects. In these studies, male and female mice were exposed via whole-body inhalation for 28 consecutive days to seven different

concentrations of EtO, ranging from 0.05 to 200 ppm. This design successfully captured the shape of the dose-response curve from low parts-per-billion environmental levels up to high parts-per-million occupational levels.

4.1.2. EtO is a Relatively Weak In Vivo Genotoxicant

In tissue samples collected after the 28-day exposure, Liu et al. (2026a) found that N7-(2-hydroxyethyl)guanine (N7-HE-G) adducts increased linearly with exposure across all tested concentrations and tissues. Crucially, pro-mutagenic O^6-hydroxyethylguanine O^6-HE-dG) adducts appeared only at high exposures ≥50 ppm), despite the use of highly sensitive detection methods.

Dose-response slopes for both adduct types steepened only at higher concentrations, which occurs when metabolic detoxification and repair systems become saturated (Figure 8). This disproportionate increase in adduct load matches predictions from a validated physiologically based pharmacokinetic (PBPK) model (Fennell and Brown, 2001), which attributes elevated blood EtO levels to EtO-induced glutathione depletion at high concentrations. Importantly, this same 28-day study also directly demonstrated EtO systemic doses measured as HEV increase dose-disproportionately at high exposures (Liu et al., 2026a, 2026b).

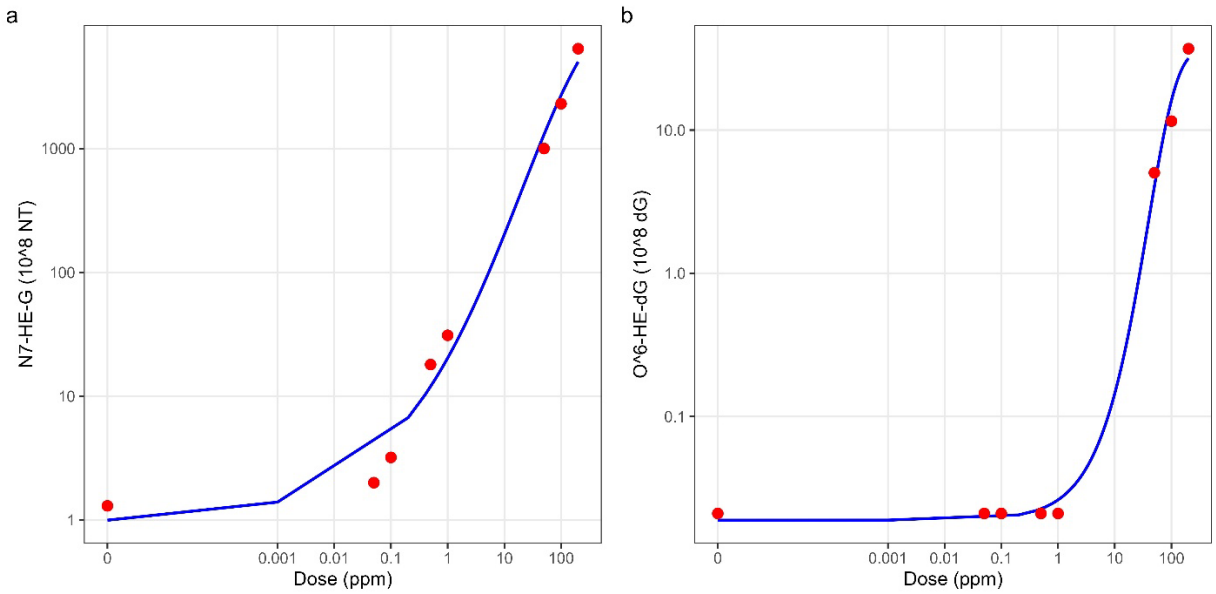


Figure 8. N7-HE-G and O6-HE-dG adduct levels in the lung tissue of female mice following 28 days of ethylene oxide exposure (Liu et al., 2026a). Data were fit using the PROAST (version 70.0; EFSA web-based program) dose-response model, which includes a meta-analysis of four model types, including an exponential, Hill, inverse exponential, and log-normal model. The meta-analysis fits are plotted and only the inverse exponential model weighted in the meta-analysis.

Gollapudi et al. (2026) assessed genotoxicity by measuring the frequency of *Pig-a* mutations in reticulocytes (RET) and mature red blood cells (RBC) on Day 28. Cytogenetic damage was evaluated using the erythrocyte micronucleus (MN) test on Days 5 and 28. EtO proved to be a relatively weak genotoxicant, with treatment-related increases in *Pig-a* and MN frequencies occurring primarily at the maximum dose of 200 ppm. Although the dose-response curve resembled a hockey-stick shape (Figure 9), this was not interpreted as a threshold response. Instead, it was conservatively treated as a linear response with a shallow slope, showing little evidence for supra-linearity at low exposures. This shallow slope represents the anticipated worst-case scenario for a DNA-reactive mutagen in normal cells due to active cellular defenses like detoxification and DNA repair.

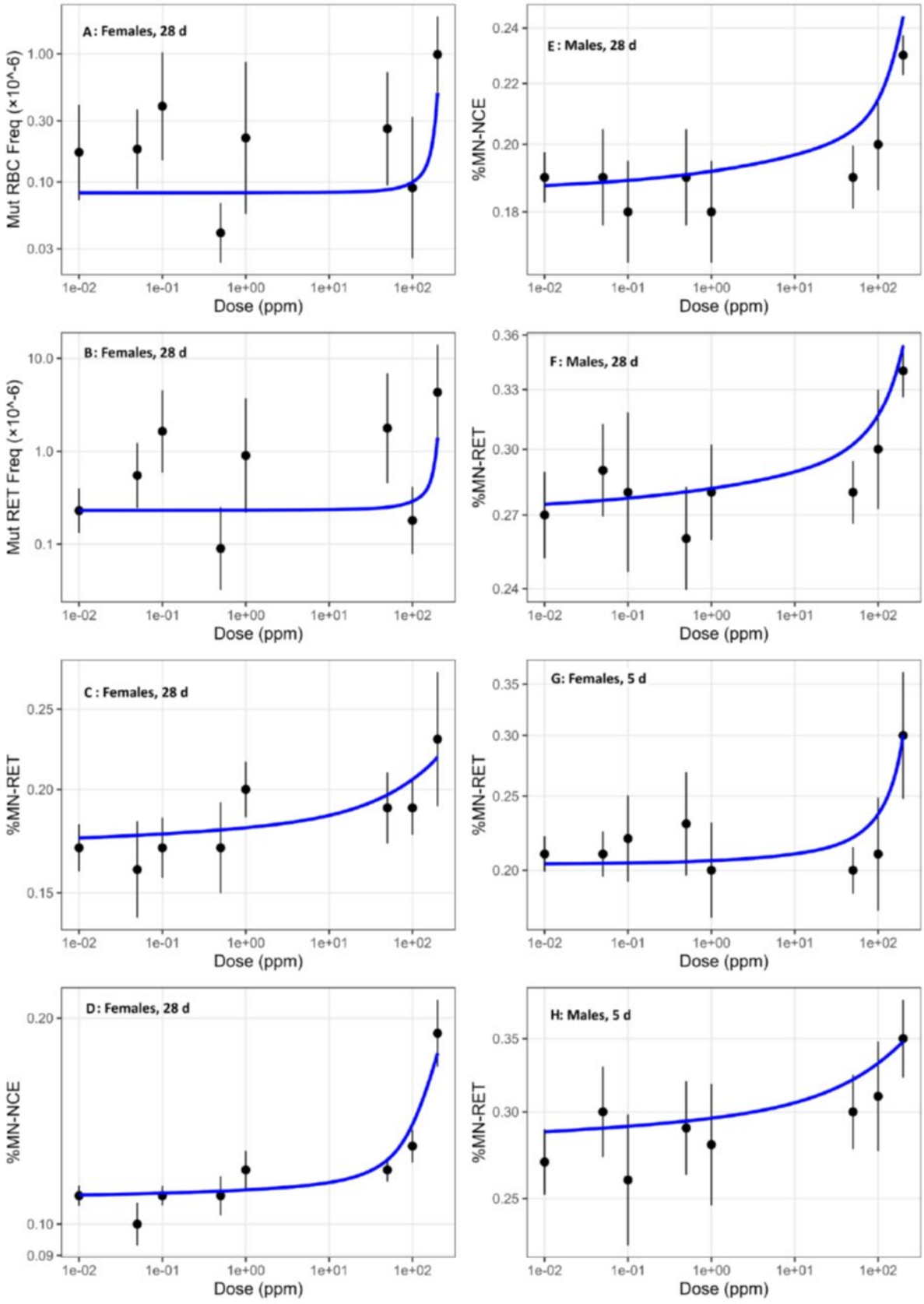

A: Females, 28 d
Mut RBC Freq (×10^-6)
1.00
0.30
0.10
0.03
1e-02
1e-01
1e+00
1e+01
1e+02
Dose (ppm)
E: Males, 28 d
%MN-NCE
0.24
0.22
0.20
0.18
Dose (ppm)
B: Females, 28 d
Mut RET Freq (×10^-6)
10.0
1.0
0.1
Dose (ppm)
F: Males, 28 d
%MN-RET
0.36
0.33
0.30
0.27
0.24
Dose (ppm)
C: Females, 28 d
%MN-RET
0.25
0.20
0.15
Dose (ppm)
G: Females, 5 d
%MN-RET
0.35
0.30
0.25
0.20
Dose (ppm)
D: Females, 28 d
%MN-NCE
0.20
0.10
0.09
Dose (ppm)
H: Males, 5 d
%MN-RET
0.35
0.30
0.25
Dose (ppm)

Figure 9. Dose–response modeling of genotoxicity endpoints following whole-body inhalation exposure of B6C3F1 mice to ethylene oxide. (A) Pig-a mutant RBC, females at 28 days; (B) Pig-a mutant reticulocytes (RET), females at 28 days; (C) micronucleated reticulocytes (MN-RET), females at 28 days; (D) MN-normochromatic erythrocytes (MN-NCE), females at 28 days; (E) MN-NCE, males at 28 days; (F) MN-RET, males at 28 days; (G) MN-RET, females at 5 days; and (H) MN-RET, males at 5 days. Dose-response modeling: PROAST version 70.0 (EFSA web-based program), applying a model-averaging approach across four continuous-response model families: exponential (EXP), Hill (HILL), inverse exponential (INVEXP), and log-normal (LOGN). From: Gollapudi et al. (2026).

These findings align with previous research in the bone marrow (Recio et al., 2004) and lung tissues (Manjanatha et al., 2017) of Big Blue® mice, where mutations increased only at high exposures (≥100 ppm and 200 ppm, respectively) and after prolonged durations (up to 48 weeks). Similarly, LeBaron et al. (2012) reported no MN induction in mouse erythrocytes exposed to 10–200 ppm EtO for 28 days. This weak genotoxic profile is further supported by a comprehensive review of 40 *in vivo* studies compiled by Gollapudi et al. (2020).

4.1.3. Relevance of 28-Day Mouse Studies to Long-Term Risk Assessment

The 28-day exposure regimen used by Gollapudi et al. (2026) and Liu et al. (2026) is directly relevant for assessing long-term human and animal risks. A 28-day exposure period is fully sufficient for DNA adducts to reach a steady state in mice, rats, and humans (Filser and Klein, 2018; Walker et al., 2000). Because steady-state adduct levels reflect a balance between formation, repair, and cell turnover; extending exposures beyond 28 days is not expected to substantially alter the shape of the dose-response pattern especially at lower exposure concentrations.

For the apical genotoxicity endpoints reported by Gollapudi et al. (20216), gene mutations and chromosomal damage were mainly observed at the 200-ppm exposure, yet a linear, no-threshold dose-response with a shallow slope was conservatively assumed. The obvious question is whether longer exposures would have resulted in a dose-response pattern markedly different from what was observed with the 4-week exposure. There are two reports in the literature examining the exposure duration on EtO genotoxicity. Recio et al (2004) and Donner et al. (2010) measured gene mutations and chromosomal aberrations, respectively, following EtO exposure (25, 50, 100 or 200 ppm) for 12, 24, or 48 weeks in mice. In the context of informing the dose-response for carcinogenicity, the data reported by these authors in somatic cells is of relevance. While exposure duration did have an effect on the ability to detect a statistically significant increase and/or the

magnitude of response, the shape of the dose-response can be best described as sublinear in these mice exposed for more than half of their lifespan (Figures 10 and 11). It is worth noting that for genotoxic carcinogens, mutation as the initiating event occurs early in the carcinogenic process, rather than after extended exposure periods (Moore et al., 2008). Given the default assumption that EtO is expected to be mutagenic at all exposure concentrations, the available data supports the conclusion that it is highly unlikely that the response at low doses would have a steep slope as predicted by the 2 piece-spline model.

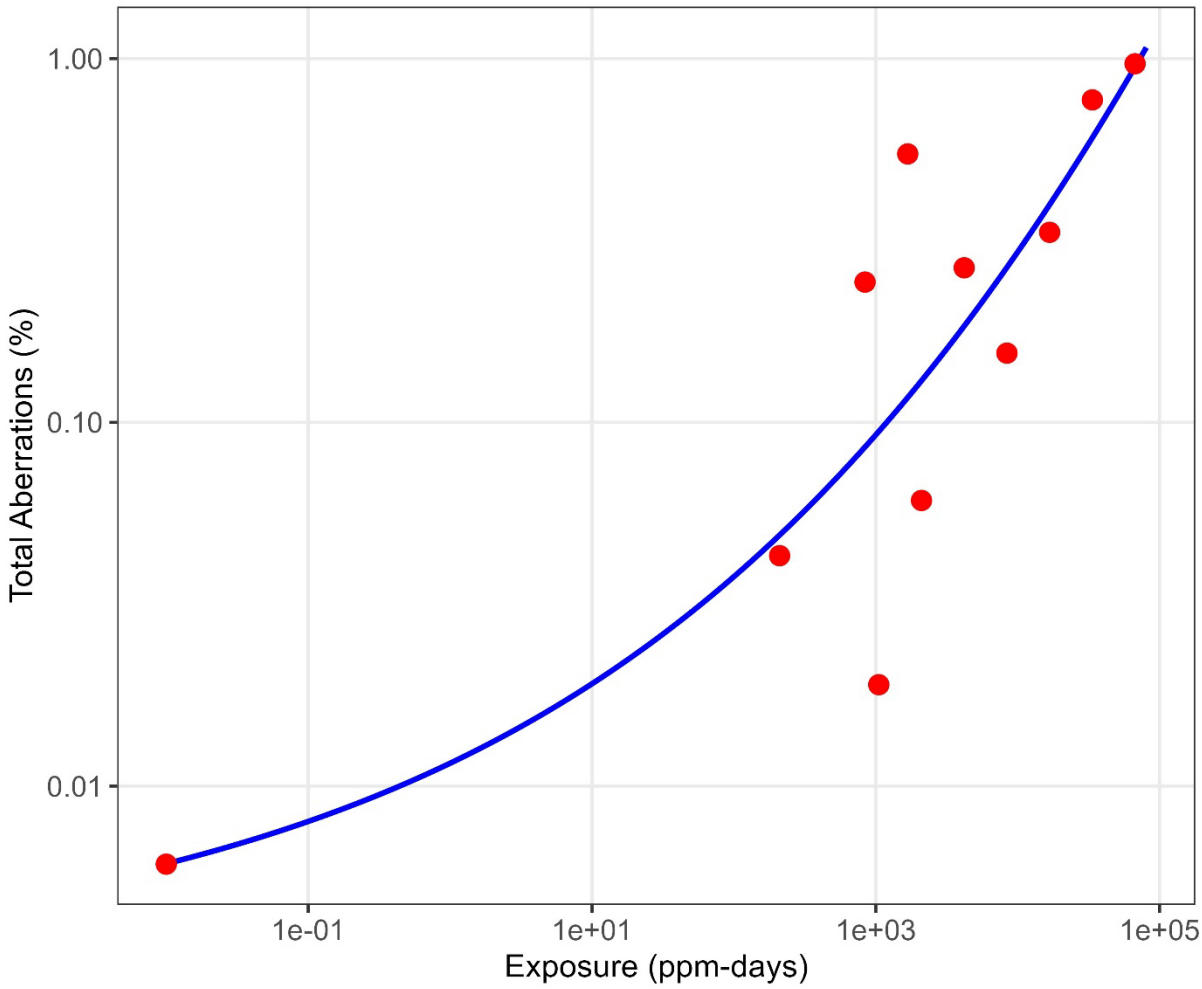


Figure 10. Frequency of total aberrations following 6, 12, 24, and 48 weeks of inhalation exposure to ethylene oxide (Donner et al., 2010). Data were fit using the PROAST (version 70.0; EFSA web-based program) dose-response model, which includes a meta-analysis of four model types, including an exponential, Hill, inverse exponential, and log-normal model. The meta-analysis fits are plotted.

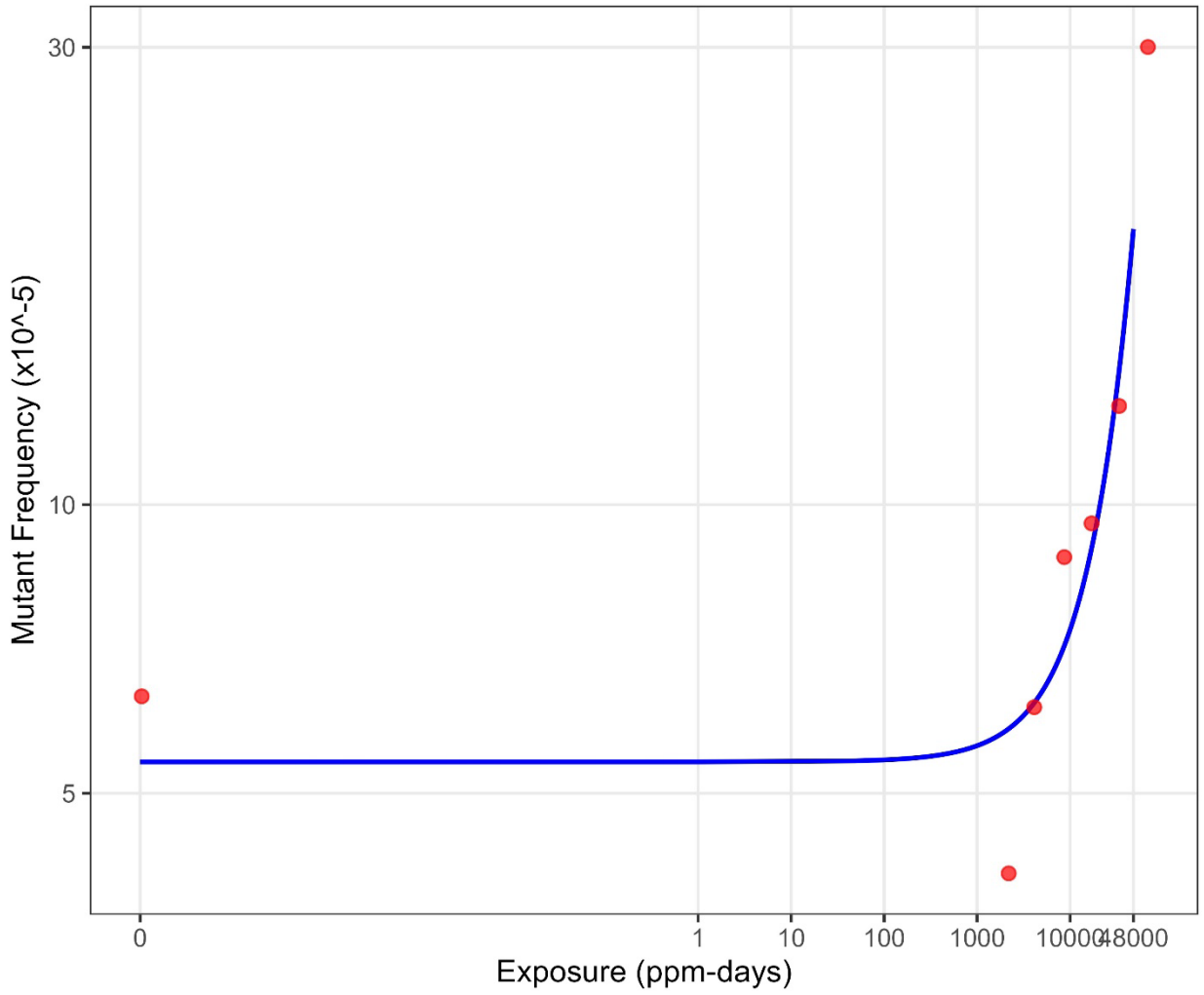


Figure 11. lacI Mutant Frequency in the bone marrow of B6C3F1 transgenic mice following 12, 24, or 48 weeks of inhalation exposure to ethylene oxide (Recio et al., 2004). Data were fit using the PROAST (version 70.0; EFSA web-based program) dose-response model, which includes a meta-analysis of four model types, including an exponential, Hill, inverse exponential, and log-normal model. The meta-analysis fits are plotted.

4.1.4. Relevance of Mouse Data for Human Cancer Risk Assessment

An essential consideration in using data from the mouse studies is their value to inform the selection of an appropriate model for human cancer risk assessment, particularly since the EtO cancer risk assessment relies on human epidemiological data. Evidence shows that human DNA repair capacity is comparable to, if not greater than, that observed in mice (MacRae et al., 2015). This similarity extends to critical DNA repair mechanisms, such as O^6-methylguanine-DNA methyltransferase (MGMT), a highly conserved enzyme involved in

the direct repair of the pro-mutagenic O6-HE-dG and other O^6-dG alkyl adducts (Roy et al., 1995; Klapacz et al., 2016). Furthermore, PBPK modeling reveals that systemic blood concentrations of EtO following exposures $\leq$ 100 ppm are highly similar between mice and humans, indicating that EtO detoxification kinetics are consistent across both species (Fennell and Brown, 2001).

This finding supports the use of mice as a conservative and valid surrogate model for predicting the genotoxic effects of EtO in humans, especially at exposure levels that do not overwhelm detoxification pathways essential for maintaining homeostasis. At higher exposure levels (>200 ppm), PBPK modeling indicates that humans possess a greater capacity to maintain EtO detoxication, for example, through glutathione conjugation, than mice. Consequently, the mouse response at these levels serves as a worst-case scenario, further reinforcing the conservative nature of using mouse data for human risk assessment (Fennell and Brown, 2001).

4.2 OEHHA's Rationale #2: A plateauing or attenuation of odds ratios at higher exposure levels has been observed in several other occupational carcinogen studies, with Stayner et al. (2003) cited as providing a plausible explanation for such a pattern with EtO.

OEHHA cites Stayner et al. (2003) three times to provide possible explanations of the pattern of categorical estimates (p. 35), and to justify selection of the two-piece spline model for lymphoid and breast cancers based on visual fit comparisons between continuous data models at categorical estimates (p.60, p.63). However, Stayner et al. (2003) is a theoretical discussion paper that explores possible explanations for attenuation of exposure-response relationships at higher exposures in occupational epidemiology studies. It does not contain a rigorous analysis of EtO and does not provide evidence that any of the proposed explanations apply to the NIOSH EtO cohort. OEHHA further claims that the "supra-linear" models are "the statistically best fitting models in terms of both AIC scores and p-values". As discussed in Section 2, this statement is incorrect. When the p-values are corrected, the standard log-linear CPH model cannot be statistically distinguished from OEHHA's preferred spline model. Under these circumstances, model selection should rely more heavily on biological plausibility. The standard CPH model has the advantage of being a simpler more parsimonious model that is more consistent with the biological evidence.

Even if one assumes that attenuation occurs at higher exposures, this does not provide biological support for a steeply increasing risk at lower exposures. The central issue in OEHHA's model selection is not whether risk attenuates at high exposures, but whether a steep low-dose exposure-response relationship is biologically plausible. Stayner et al. (2003) does not address this question.

Stayner et al. (2003) identifies six possible explanations for attenuation of exposure-response relationships at higher exposures, including healthy worker survivor effects (HWSE), depletion of susceptible individuals, high background disease rates, exposure misclassification, confounding by other risk factors, and saturation of biological processes involved in disease development. These explanations were presented as general hypotheses relevant to occupational epidemiology and were not demonstrated to apply to EtO or the NIOSH cohort. More importantly, even if one or more of these mechanisms contributed to attenuation at higher exposures, none provides a biological rationale for the steep low-dose slope implied by OEHHA's preferred two-piece spline model. Consequently, Stayner et al. does not provide direct support for OEHHA's selection of the spline model for derivation of the IUR for EtO.

Three of Stayner's hypothesis warrant brief consideration:

- **First, a healthy worker survivor effect (HWSE) could theoretically attenuate the exposure-response relationship if less healthy workers leave employment while healthier workers remain employed and accumulate higher exposures.** Picciotto et al. (2025) hypothesized that a HWSE may contribute to attenuation of the exposure-response relationship at higher exposures for breast cancer mortality and lymphohematopoietic cancer mortality. However, the authors also acknowledged that the presence of a HWSE does not identify the correct shape of the exposure-response curve and, at most, lends plausibility to a flattening of the curve at higher exposures. Therefore, even if a HWSE exists, it does not provide evidence regarding the steepness of the low-exposure portion of the curve that distinguishes OEHHA's preferred spline model from the standard log-linear CPH model. Consequently, the HWSE hypothesis cannot be used as justification for selecting a model with a steep low-dose slope.

- **Second, Exposure misclassification or measurement error is one of the few hypotheses identified by Stayner et al. (2003) that is directly relevant to the NIOSH EtO cohort.** The NIOSH exposure reconstruction was not validated for most historical exposures. Consequently, exposure measurement errors and

misclassification could contribute to the apparent shape of few categorical estimates, which are not the individual hazard rates that were modeled. As discussed in detail above, the NIOSH exposure regression model systematically underestimated historical exposures. In other words, individual cases plotted at lower exposures could actually have received higher exposures. This could potentially result in the appearance of a steeper response at lower estimated exposure levels.

- **Third, Saturation of biological processes is not a plausible explanation for the EtO exposure-response relationship.** The vinyl chloride example discussed by Stayner et al. (2003) depends on saturation of metabolic activation pathways that reduce formation of the carcinogenic metabolite at high exposures. In contrast, EtO is a direct-acting DNA-reactive carcinogen that does not require metabolic activation. If anything, this mechanism would be expected to produce an upward inflection in the dose-response relationship at higher exposures rather than the attenuation proposed by Stayner et al. for vinyl chloride. Thus, OEHHA's reliance on Stayner et al. (2003) as biological support for a two-piece spline model is inconsistent with the known toxicokinetics and mode of action of EtO.

4.3. Conclusion and Recommendations to Correct OEHHA Statements

In conclusion, OEHHA's assertion that the two-piece spline model is biologically plausible rests on two weak and largely speculative arguments. First, although EtO is a direct-acting genotoxic carcinogen, OEHHA provides no mechanistic evidence supporting the steep low-dose slope implied by the spline model. To the contrary, mode of action studies consistently demonstrate a shallow, near-linear, or sublinear response at low exposure concentrations, with steeper responses occurring only at high occupational exposures where detoxification pathways become saturated.

Second, OEHHA's reliance on Stayner et al. (2003) is misplaced. Stayner discusses possible explanations for attenuation of exposure-response relationships at higher exposures but provides no evidence regarding the shape of the exposure-response relationship at lower exposures. Even if attenuation occurs at high exposures, it does not follow that the low-dose slope must be steep.

In contrast, the biological evidence directly addresses the question at issue and consistently supports a shallow low-dose response for EtO. Because human detoxification

kinetics and DNA repair capacities are comparable to, or greater than, those observed in mice, these findings are directly relevant and conservative for human risk assessment. Moreover, once the statistical errors identified in EPA IRIS (2016) are corrected, the standard log-linear Cox proportional hazards (CPH) model cannot be distinguished statistically from OEHHA's preferred spline model.

Therefore, biological plausibility and parsimony favor the standard log-linear CPH model. OEHHA has not provided a scientific, statistical, or biological basis for preferring a two-piece spline model with a steep low-dose slope. Accordingly, the weight of evidence supports application of the standard log-linear CPH model to the NIOSH data rather than OEHHA's biologically unsupported spline model.

5. Uncertainty in EPA's Dose Response Modeling of NIOSH Data

5.1. Exposure estimates in NIOSH are too uncertain to support inferences on the shape of the dose-response relationship for EtO

As described in detail in Section 1, the NIOSH exposure estimates are unvalidated for over 80% and 90% of exposures to lymphoid and breast cancer mortalities. Bogen et al. (2019) developed an engineering/industrial-hygiene (E/IH) model to reconstruct historical EO exposures in the NIOSH sterilization worker cohort and found that the NIOSH statistical regression (NSR) model used by EPA to derive its 2016 IRIS inhalation unit risk (IUR) substantially underestimated pre-1978 exposures — by approximately 4-fold for the middle period (1955–1964) and 16-fold for the early period (1938–1954) (Figure 12) for sterilizer operators who had the highest exposures (90th-maximum exposures).

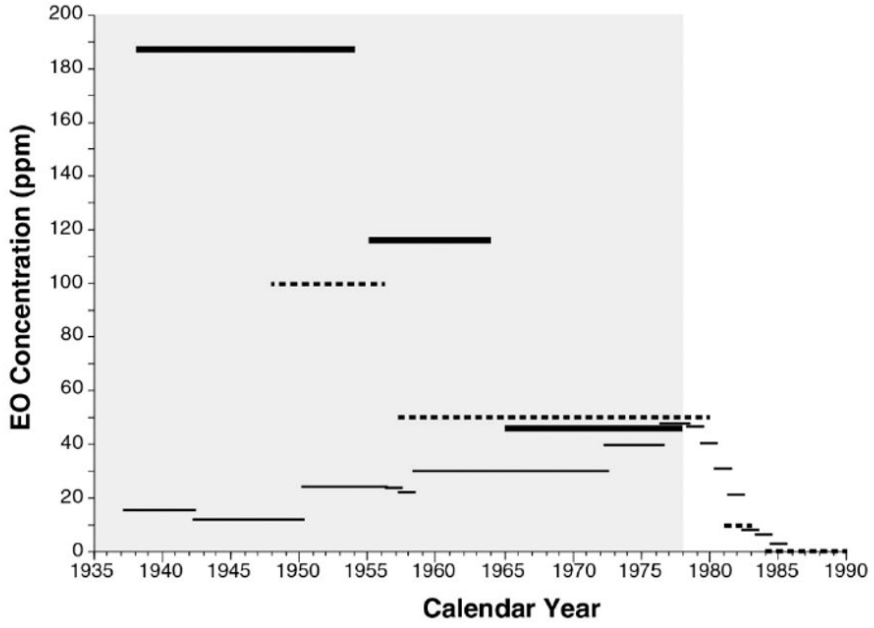


Figure 12. Comparison of E/IH (thick lines) and NSR (thin lines) exposure model estimates of occupational respiratory exposures to EO in facilities that sterilized medical/health products and prevailing ACGIH TLV limits for EO (dashed lines).

Note: Shaded area represents the period during which very limited or (pre-1976) no contemporaneous measurements were available to validate NSR model predictions and during

which no EO-specific regulations were in place to limit occupational EO exposures (From Bogen et al., 2019).

Ironically, EPA dismissed the Bogen model's projections of industry-wide sterilizer operator exposures because the historical levels exceeded the ACGIH maximum allowable concentration recommendation of 100 ppm in 1948 (EPA Response to Comments on the HON, 2022). EPA should instead question the NIOSH exposure model, which counterintuitively predicts *lower* exposures in the 1940s than in the 1960s. This runs entirely counter to historical ACGIH guidelines, which dropped from 100 ppm to 50 ppm between the 1950s and 1960s.

Bogen et al. (2019) emphasized evaluating the exposure trends in exposures. Supporting this trend, published 1940s and 1950s evidence detailed in Section 1 shows that sterilizer operators could smell EtO. Based on the odor thresholds for EtO, this strongly suggests that 90th-percentile workers regularly faced exposures between 200 and 700 ppm during those early decades. In contrast, the NIOSH exposure regression model assumes incorrectly that industrial hygiene and sterilizer equipment standards (e.g. leakiness of doors) were identical in the 1950s and 1970's. In general, if historical exposures are underestimated, the resulting dose-response analysis will tend to overestimate the apparent potency of the chemical

EPA's exclusive reliance on the NIOSH cohort, without accounting for the likely errors in historical exposures introduces substantial uncertainty into the dose-response assessment. Based upon this analysis, the error bars associated with cumulative exposure estimates (i.e., x-error bars) are expected to be very wide, such that inferences OEHHA makes about the shape of the dose-response relationship cannot be made based on OEHHA Figure 11 or 12 for lymphoid mortality and breast cancer incidence.

5.2. The dose-response modeling approach adopted by OEHHA Is inconsistent with USEPA guidelines.

USEPA's Guidelines for Carcinogen Risk Assessment (USEPA, 2005), advocates for the use of high-quality epidemiology data for dose-response assessment, but further cautions:

"Because epidemiologic data are usually limited and many models may fit the data (Samet et al.,1998), other factors may influence model choice. For epidemiologic studies, including those with grouped data, analysis by linear models in the range of observation is generally appropriate unless the fit is poor. The relatively small exposure range observed in many epidemiologic studies, for example, makes it difficult to discern the shape of the

exposure-or dose-response curve. Exposure misclassification and errors in exposure estimation also obscure the shape of the dose-response curve."

As noted above, due to the high degree of scatter in the dose-response data (Figure 6B), and the large uncertainty in the exposure estimates, the use of nonlinear models based upon inappropriate inferences made from data is ill-advised, and a simpler linear model serves as a more appropriate choice (e.g., standard log-linear Cox Proportional Hazards model is effectively linear for the range of observation for the NIOSH cohort).

In addition, there is an apparent difference in the pattern of the male and female lymphoid cancer exposure response when considering the individual hazard rate estimates (Figure 6B). It appears that EPA's justification for combining males and females is based on testing for interaction using the categorical grouped estimates which are not the data modeled (Appendix D). Steenland et al. (2004) concluded that positive exposure-response trends for lymphoid mortality were found for males only.

USEPA (2012) benchmark dose guidelines state:

"If it is plausible that the multiple datasets represent a homogeneous picture of the dose-response (for example, the responses at doses common to two or more datasets are essentially the same and statistically undifferentiable), then this [combining data sets] is a justifiable approach." (USEPA, 2012)

Because the dose-response data for lymphoid cancer mortality in male and female workers are clearly not homogenous (Figure 6B) based on conclusions by the authors of the NIOSH study, combining these data for dose-response modeling purposes is inconsistent with USEPA's benchmark dose guidelines (USEPA 2012).

USEPA's benchmark dose guidelines also caution against the use of dose-response models that predict implausibly steep slopes at low doses and plateau at higher exposures:

> "… a particular problem arises with datasets where all non-control doses have response levels above the selected BMR and response is flat or shallow. When this situation arises with quantal data, especially if the maximum response is less than 100%, using a model like the Weibull with no restrictions on the power parameter is tempting because such models reach a plateau of less than 100% and most modeling programs do not include other models for quantal data that have this property. Using such an unconstrained model, however, can result in very imprecise BMDs because the data do not constrain the dose-response curve in this lower dose range, where all the change in response is occurring. In theory, other models could be found that force the BMD to be anywhere between the lowest BMDL and the lowest administered dose. Thus, the BMD computed here depends solely on the

model selected, and goodness-of-fit provides no help in selecting among the possibilities. The unfortunate reality in such situations is that the data provide little useful information about the dose-response relationship…"

The apparent plateau and steep slope observed in the dose-response data for lymphoid mortality is an artifact of the decision to combining data for male workers, which provide useful dose-response information, and data for female workers, which do not provide useful dose-response information. As such, OEHHA should select the male worker data for assessing the dose-response relationship, consistent with the conclusions of the NIOSH peer-reviewed study (Steenland et al. 2004).

5.3. The steep slope calculated by OEHHA/USEPA is not supported by the weight of evidence provided by multiple evidence streams for EtO.

There are multiple evidence streams that refute the behavior of the 2-piece linear spline model (i.e., steep slope for low concentrations, shallow slope for high concentrations), including dose-response data from: (A) other epidemiology studies; (B) animal cancer bioassays for EtO; (C) animal cancer bioassay for ethylene; (D) U.S. smokers; and (E) data from mechanistic studies. Each of these evidence streams are summarized below.

5.3.1. A steep slope for EO exposure and lymphoid mortality Is not supported by other epidemiology data sets.

Valdez-Flores et al. (2025) assessed the exposure-response data for cancer mortality in an independent cohort (UCC male EO chemical workers), which can serve as a validation data set for slope predictions made based on the NIOSH cohort (Table 1 shows comparability of both studies). Despite the high average cumulative exposure, extensive follow up and detailed analyses of the UCC male cohort, neither external nor internal analyses found exposure related effects or any indication of a positive slope for male lymphoid cancers. These results are consistent across various UCC cohort updates spanning 40 years (Greenberg et al. 1990, Teta et al. 1993, Benson and Teta 1993, Swaen et al. 2009, Valdez-Flores et al. 2010, Valdez-Flores et al. 2025).

Table 1. Comparability of NIOSH and UCC data sets for **male** workers for quantitative exposure-response analysis (Valdez-Flores et al. 2025)

Characteristic	NIOSH (Steenland et al. 2004)	Updated UCC (Valdez-Flores et al. 2025)
Male Workers	7,634	2,053

% Deceased	19%	70%
EtO Exposure Period	1936-1986	1925-1988
Availability of Worker-EtO Exposure Measures	1976-1985 (10 yr)	1957-1988 (29 y)
EtO Only Exposures (ppm-yr)	27	67
Lymphoid mortalities	27	25

Based upon **analyses using external or internal comparisons, categorical or continuous approaches, cumulative or log cumulative exposure metrics, and various exposure lags, the results show no indication of increasing slopes at low or high cumulative exposures for lymphoid tumors. These findings raise serious questions about the validity of the** steep EPA/OEHHA spline model **which therefore** should not be used to estimate risks associated with low exposures (e.g., environmental).

5.3.2. A steep slope for EO is inconsistent with potency estimates based on animal data.

Multiple animal data sets are available to assess the dose-response relationship for EO. USEPA (2016) calculated inhalation unit risk (IUR) values for EO based on five rodent data sets:

- Mononuclear cell leukemia, testicular tumors, peritoneal mesothelioma, and brain tumors male rats (Lynch et al., 1984);
- Mononuclear cell leukemia, testicular tumors, peritoneal mesothelioma, and brain tumors male rats (Snellings et al., 1984);
- Mononuclear cell leukemia and brain tumors female rats (Snellings et al., 1984);
- Lung tumors in male mice (NTP, 1987); and
- Lung tumors, lymphoma, uterine tumors, and mammary gland tumors in female mice (NTP, 1987)

The distributions of IUR values supported by the animal data sets are compared to those based upon epidemiology data in Figure 13.

From this figure, several conclusions can be made:

- The IUR distribution derived using a 2-piece linear spline model by USEPA and selected by OEHHA is very broad and includes a substantial fraction of negative slope values (Figure 7A), which reflects the large uncertainty in the model parameters used in the two-piece spline analysis. This distribution

would be even wider if the uncertainty in the exposure estimates (see comment above) were quantified.

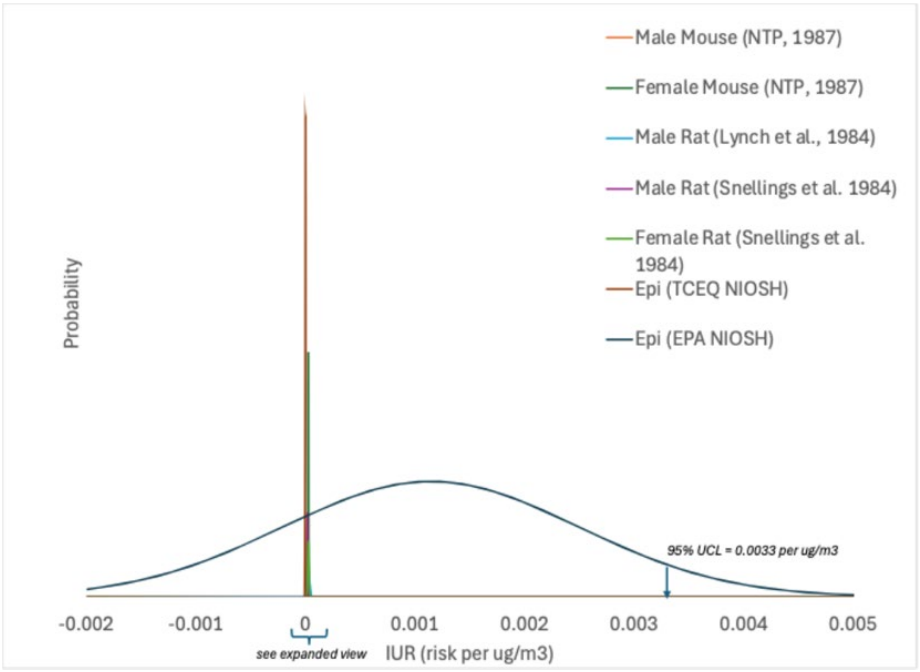
- The IUR distributions from rodent cancer bioassays (Figure 9), which reflect the combined contribution across multiple tumor types, are remarkably consistent (MLE values are all within a factor of 2 of one another; UCL values are also within a factor of 2 of one another), and define much narrower distributions for the IUR (i.e., less model uncertainty) (Figure 13)The UCL for the IUR calculated by USEPA (0.0033 per ug/m3 for adult only exposures) is orders of magnitude higher than the corresponding values calculated from animal data sets. The USEPA's IUR distribution based on the NIOSH data is inclusive of the animal IUR UCL values (i.e., the epidemiology data do not rule out the possibility of animal-based IUR values). In contrast, animal IURs are exclusive of USEPA's IUR UCL value (i.e., the animal data indicate USEPA's IUR UCL value is highly unlikely). There is no mechanistic basis to support the presence of such a large species difference in sensitivity (humans>>rodents) for a direct acting genotoxic agent such as EO (see previous section). In contrast, the distribution IUR calculated by TCEQ (based on lymphoid cancers alone) is lower than the corresponding distributions in animals.

With respect to relying upon epidemiology data, OEHHA (2026) notes that,

"The evaluation of animal and epidemiological cancer studies for EtO indicates that human epidemiological studies are more relevant and sensitive than animal studies for deriving an IUR for EtO. The use of human data also avoids the potential uncertainties involved in extrapolating risks from laboratory animal studies."

However, this consideration also needs to balance the large uncertainties associated with model parameters and underlying exposure estimates, as described above (Figures 8 and 9).

A



B

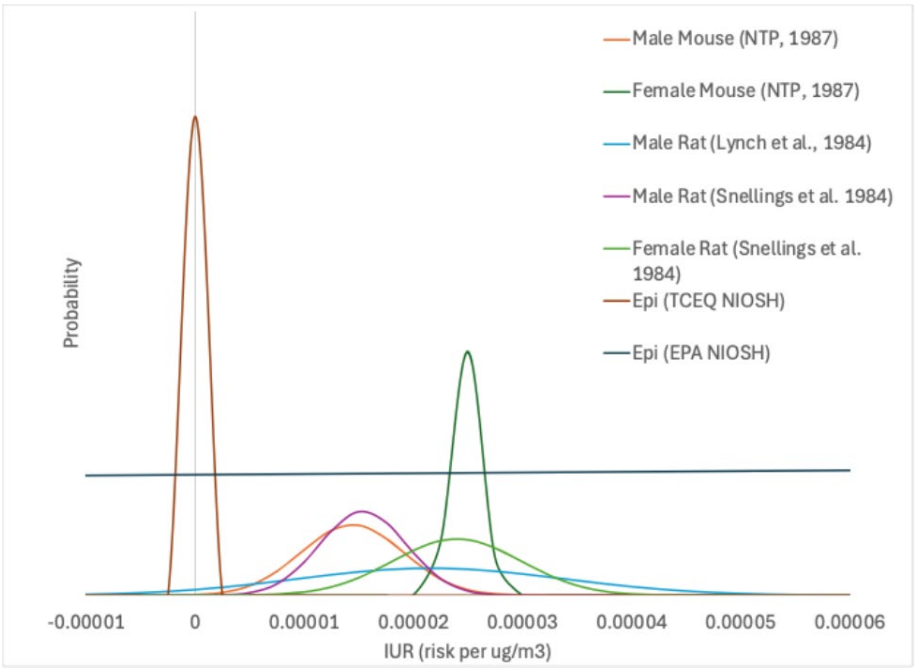


Figure 13. (A) IUR distributions (prior to ADAF application; normality assumed) based on difference data sets for EtO (human, rat, mouse) and assessors (EPA, TCEQ); (B) expanded view. Distributions for inhalation unit risk (IUR) values were derived from animal data sets are based upon MLE and UCL values reported in USEPA (2016). Similarly IUR values derived from epidemiology data sets are based upon MLE and UCL values reported by USEPA (2016) and TCEQ (2015). All distributions reflect an assumption of normality.

5.3.3. A steep slope for EtO is inconsistent with animal bioassay data for ethylene.

A cancer bioassay was conducted for ethylene in rats (Hamm et al., 1984; not noted by OEHHA as an EtO-related animal cancer bioassay). Because ethylene is metabolically converted to EtO, internal dose levels (HEV [6]adducts or PBPK modeling) can be used to combine this study with other rat EtO bioassays to characterize the dose-response relationship across a broader exposure range. Crucially, the negative results in rats following moderate ethylene-derived EtO exposures—approximately 50–160 times higher than background—provide a critical "reality check" for USEPA IRIS dose-response assessment. These findings indicate that potency estimates based solely on high-exposure EtO-only studies significantly overestimate the response slope at low-to-moderate concentrations (see Kirman et al., 2021, Figure 7). In this study, groups of 120 rats of each were exposed to 0, 300, 1000, or 3000 ppm ethylene (6 hours/day, 5 days/week) for up to 24 months. Gross examination of rats dying during the study, or of those that were sacrificed as scheduled, did not reveal any lesions attributable to ethylene exposure. Histologically, a variety of proliferative, degenerative, and inflammatory lesions were observed in both the control and 3000-ppm groups. These lesions were typical of those seen in this strain of animal and were considered unrelated to ethylene exposure. The authors concluded there was no evidence that ethylene at these concentrations causes chronic toxicity or carcinogenicity rats. Because rats convert ethylene to EO (e.g., resulting in equivalent concentrations of approximately 0, 2.4, 5.3, and 5.5 ppm EO; Filser and Klein, 2018), this bioassay provides information on the shape of the exposure-response relationship for EO and tumors at moderate concentrations. The lack of a tumor response in this study indicates that the slope is not steep (Figure 14).

[6] **N-(2-hydroxyethyl)valine (HEV) adducts** are stable biomarkers formed when ethylene oxide reacts with hemoglobin; they serve as a precise metric for cumulative internal dose

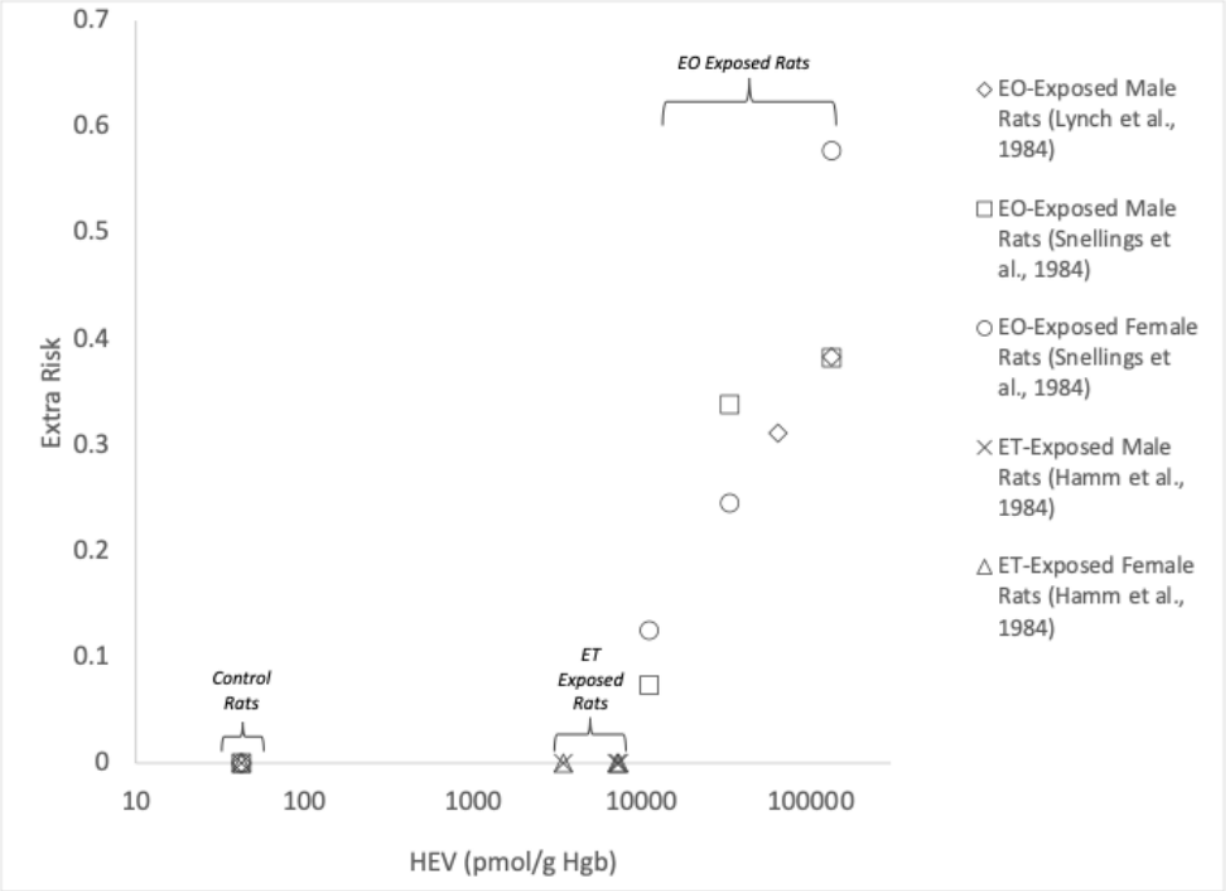


Figure 14. Total extra risk vs. total ethylene oxide internal dose (HEV, pmol/g Hb) in rats exposed to ethylene or EtO. X-axis displayed on log-scale to clearly depict the broad range of internal doses, and not to accentuate a nonlinear dose-response relationship (from Kirman et al., 2021). Extra risks for total cancer risk were calculated for each dose group from available rat data sets (Snellings et al., 1984; Lynch et al., 1984) by summing across tumor sites for which USEPA (2016) identified as increased over control animals for EtO. For ET, extra risk was set equal to zero based upon the reporting of no increased tumors (Hamm et al., 1984). HEV levels reflect values reported by Filser and Klein (2018) for similarly exposed rats or estimated by interpolation.

In contrast, if the steep slope calculated for EPA IUR is assumed to be operable in rats (i.e., that the plateau behavior was missed by the high test concentrations used in the EO bioassays), the predicted tumor response in the ethylene bioassay would be near 90% at the lowest concentration, and 100% in the mid and high concentration groups. As such, the lack of a significantly increased tumor response in the ethylene bioassay excludes the possibility of a steep slope for EO.

5.4. A Steep Slope for EO is Inconsistent with Data from U.S. Smokers.

USEPA's gold standard science policy includes a mandate to be "skeptical" "of findings and assumptions" and "structured for falsifiability of hypotheses" (EPA, 2025). These principles refer to conducting science with an aim towards independently testing findings. This should be particularly true of the risk estimates from the NIOSH study, given that the OEHHA/USEPA IUR predicts cancer risk at extraordinarily low levels of EtO, less than background or endogenously produced EtO (see below).

One way to assess the plausibility of the IRIS cancer risk estimate for lymphoid cancer is using the reality check of studies examining smoking and lymphoid cancer. Cigarette smoke is recognized as a significant source of EO exposure in the U.S. population (Kirman et al., 2021, 2025; Lin et al., 2026). As demonstrated in greater detail below, smokers have significant analytically quantified EtO exposures measured as HEV that can directly correlate with demonstrated concentrations of EtO in cigarette smoke. Smoking-related exposures measured as daily ppb to EtO can be reliably estimated using pack-year smoking data available in epidemiology studies (Kirman et al., 2025). The Prostate, Lung, Colorectal, and Ovarian (PLCO) Cancer Screening Trial, which also evaluated the risk of non-Hodgkin Lymphoma (NHL), is the largest U.S.-based cohort study that evaluated smoking and lymphoid cancer outcomes (Troy et al., 2010). NHL is the primary cancer type in the lymphoid grouping for EtO. In terms of number of cases, NHL contributed the most to lymphoid mortality cancers, and used as a surrogate for the lymphoid group.

Based on the pack-year smoking data, known concentrations of EtO present in cigarette smoke and varying levels of smoking intensity, EtO exposures were estimated for the subjects in the study and the IRIS value applied to estimate lymphoid cases. The estimate of lymphoid cases, assuming IRIS as a correct predictor of risk, was large enough that they would have been statistically detectable based on the large sample size of the study (>1 million person-years of follow-up). Yet, there was no statistically elevated level of lymphoid cancer in the study[7]. In fact, non-smokers had slightly higher (non-statistically significant) incidence of lymphoid cancer. This key smoker reality check is yet another analysis that shows the implausibility of the IRIS value.

USEPA's response to lack of a lymphoid cancer signal for smokers (USEPA, 2024) implausibly posits, without presenting any evidence, that the absence of a lymphoid cancer signal in smokers is due to possible interactions between EtO and other

[7] Based on more than 1,200 observed NHL cases hazard ratios of 1.0, 0.91, 0.99, 0.94, and 0.95 were reported for nonsmokers (referent) and each pack-year quartile for smokers, respectively

carcinogenic agents in cigarette smoke that attenuate possible EtO lymphoid cancers. Importantly, and in contrast, USEPA (2022) itself recommended that if EtO concentrations could be quantified in cigarette smoke, such data would be a valuable "forwards" validation opportunity of the EtO IRIS low-exposure cancer risk predictions, applicable for testing against a smoker cancer reality check as well as to informing the reality check context of contribution of external air exposures to endogenous production of EO (see below).

The plausibility of the above smoker reality check can be further graphically visualized when the smoking exposures are converted to lifetime equivalent EtO exposures (based upon observed relationships between hemoglobin adducts and smoking intensity, hemoglobin adducts and air concentrations; data from CDC NHANES; Kirman et al., 2025), and then used to support a linear regression (solid blue line in Figure 15), the overall slope defined by these data (black circles in Figure 11) is near zero, but slightly negative. However, the 95% confidence interval defined by these data (light blue shading bounded by orange dashed lines in Figure 11) includes the possibility of a shallow positive slope at the high end. The slopes predicted by USEPA's IUR are very steep (solid red and green lines for MLE and UCL, respectively, in Figure 11) and fall well above the 95% confidence interval. Therefore the USEPA's IUR is inconsistent with the data from Troy et al. (2010). In contrast, the UCL slope predicted by TCEQ's IUR, which defines a shallower slope (purple line in Figure 11), falls within the 95% confidence interval, and therefore is considered consistent with evidence provided by smoking data.

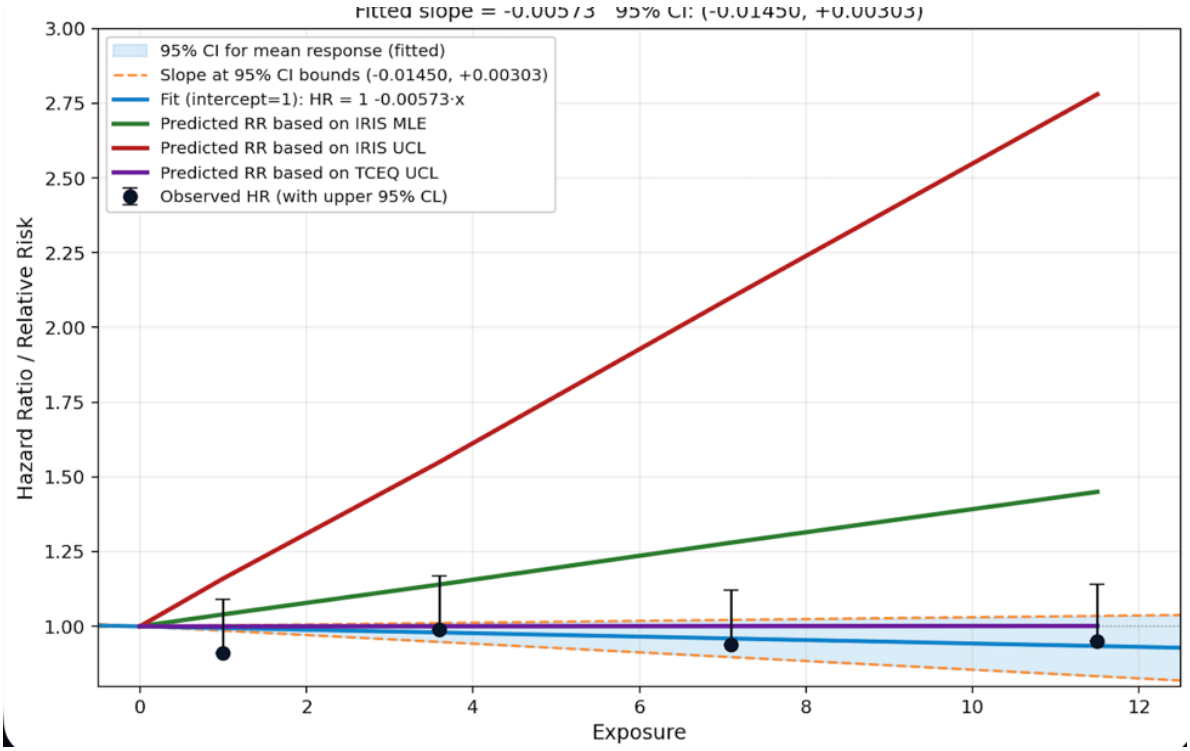


Figure 15. Comparison of epidemiology-based IUR predictions (USEPA, TCEQ) to data for U.S. smokers.

Rate ratios for NHL in smokers stratified by quartiles for pack-years of smoking (Troy et al., 2010) were plotted as a function of equivalent lifetime exposures to EtO. Equivalent EtO exposures were determined using published linear relationships between smoking intensity (cigarettes per day) and HEV levels, and between HEV levels and workplace air. Linear regression analysis was performed to estimate the slope (MLE and 95% confidence interval). Predicted risks for NHL from USEPA and TCEQ inhalation unit risks were adjusted for the relative contribution of NHL to total lymphoid cancer responses (~60%).

OEHHA (2026) challenges this type of analysis, by stating:

“It is not appropriate to attribute specific cancer outcomes in smokers, of which there are many (IARC, 2012), to EtO or any other single carcinogenic chemical component of tobacco smoke.”

We agree that attributing any observed cancer response solely to EtO is not a reasonable assumption, as cigarette smoke contains numerous constituents that may contribute to cancer risk. However, this consideration is only relevant where there is an observable and statistically significant increase in the cancer endpoint of interest. In the case of NHL – the primary contributor to the lymphoid tumor category in the NIOSH study – the smoking data provide a useful bounding exercise. Specifically, by conservatively assuming that the entire observed NHL response is attributable to EtO, the analysis estimates the maximum response that could plausibly be associated with EtO exposure.

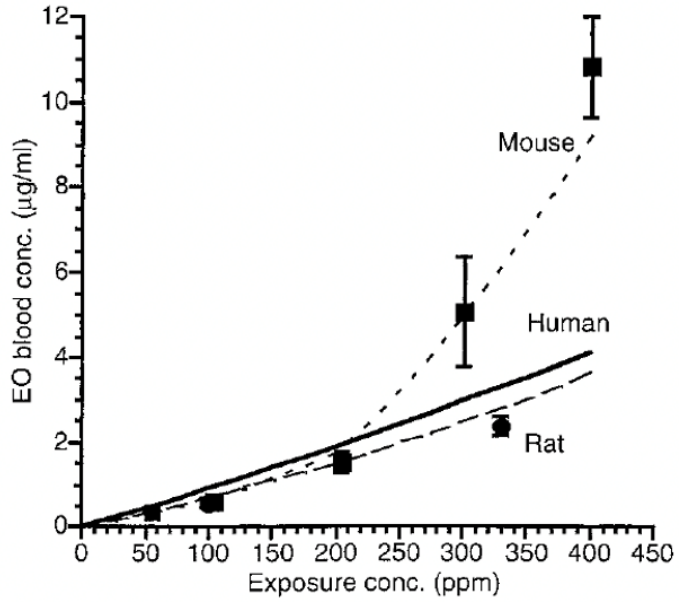
Accordingly, the assessment based on U.S. smokers should be viewed as conservative. As shown in Figure 15, if the EPA IUR is assumed to be correct, the implied EtO attributable NHL signal among smokers would exceed the observed response. This would require the implausible conclusion that the potential EtO-induced NHL signal in smokers is attenuated by other components in cigarette smoke. There is no convincing evidence to support this assumption.

Thus, the smoker data provides an independent reality check on the validity of EtO inhalation unit risk estimates. EPA's IUR predicts an NHL response that exceeds that observed among smokers, indicating that EPA's IUR is overly conservative and not supported by epidemiological evidence in smokers. In contrast, the response predicted using the TCEQ IUR falls within the confidence interval of the observed smoking data.

5.5. A Steep Slope for EO is Inconsistent with Mechanistic Data.

Mechanistic data for EtO do not support the behavior of the 2-piece linear spline model (i.e., steep slope at low concentrations, shallower slope at high concentrations), and in fact support the opposite behavior (i.e., shallow slope at low concentrations, steeper slope at high concentrations). For example, the toxicokinetics for EtO in mice, rats, and humans at concentrations below 200 ppm are linear and closely overlap each other, but exhibit an inflection point (i.e., transition to steeper slope) due to glutathione depletion in mice (Figure 16 B). Similarly, the relationship between EtO exposure from cigarette smoke is linearly proportionate to measurements for HEV hemoglobin adducts in U.S. smokers and nonsmokers (Kirman et al., 2025; Figure 16B).

(16A)


EO blood conc. (µg/ml)
Exposure conc. (ppm)
Mouse
Human
Rat
0
2
4
6
8
10
12
0
50
100
150
200
250
300
350
400
450

(16B)

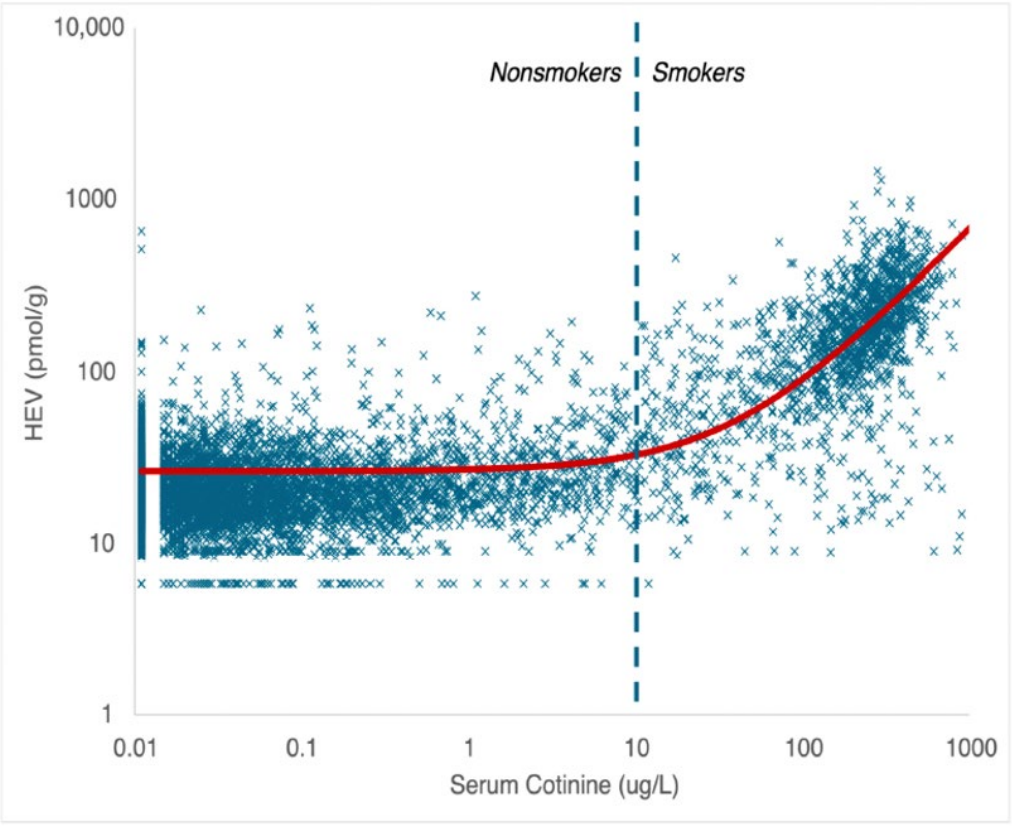

HEV (pmol/g)
Serum Cotinine (ug/L)
Nonsmokers
Smokers
10,000
1000
100
10
1
0.01
0.1
1
10
100
1000

Figure 16. EO toxicokinetics exhibit linear behavior at low concentrations: (A) in mice, rats, and humans (Fennell et al., 2001); (B) in U.S. smokers and nonsmokers (Kirman et al., 2025)

Note: Red line = linear regression of NHANES data. Data from NHANES (2013-2020) for individual smokers and nonsmokers with reported values for both HEV adducts and serum cotinine were plotted to depict a consistent linear relationship between the two variables from background levels in nonsmokers to levels approximately an order of magnitude higher than background levels in smokers.

With respect to toxicodynamic considerations, recent data collected in mice indicate that DNA adduct formation at low concentrations is consistent with sub-linear/shallow linear dose-response relationships (Liu et al., 2026a; Section 4 above). In this study, male and female B6C3F1 mice were exposed to EtO (0, 0.05, 0.1, 0.5, 1, 50, 100, and 200 ppm) 6 hours/day for 28 consecutive days. Immediately following the last exposure, DNA was extracted from lungs, liver, bone marrow, and mammary gland, and further utilized to measure DNA adduct levels using the most sensitive analytical platforms published to date. Formation of the non-mutagenic N7-HE-G was detected in all tissues and exposure groups, showing linear dose-response relationships in the low-dose range (≤1 ppm) and increased sharply and exposure-disproportionately in the high-dose range (≥50 ppm) (Figure 17A & B), mirroring the toxicokinetic behavior in Figure 16A for mice. Despite the very low limit of detection, the putative promutagenic O^6-HE-dG adduct was not detected at exposures <50 ppm in any tissue consistent with an apparent threshold response (Figure 17C). At higher exposures (≥50 ppm), O^6-HE-dG exhibited a dose-response pattern of N7-HE-G.

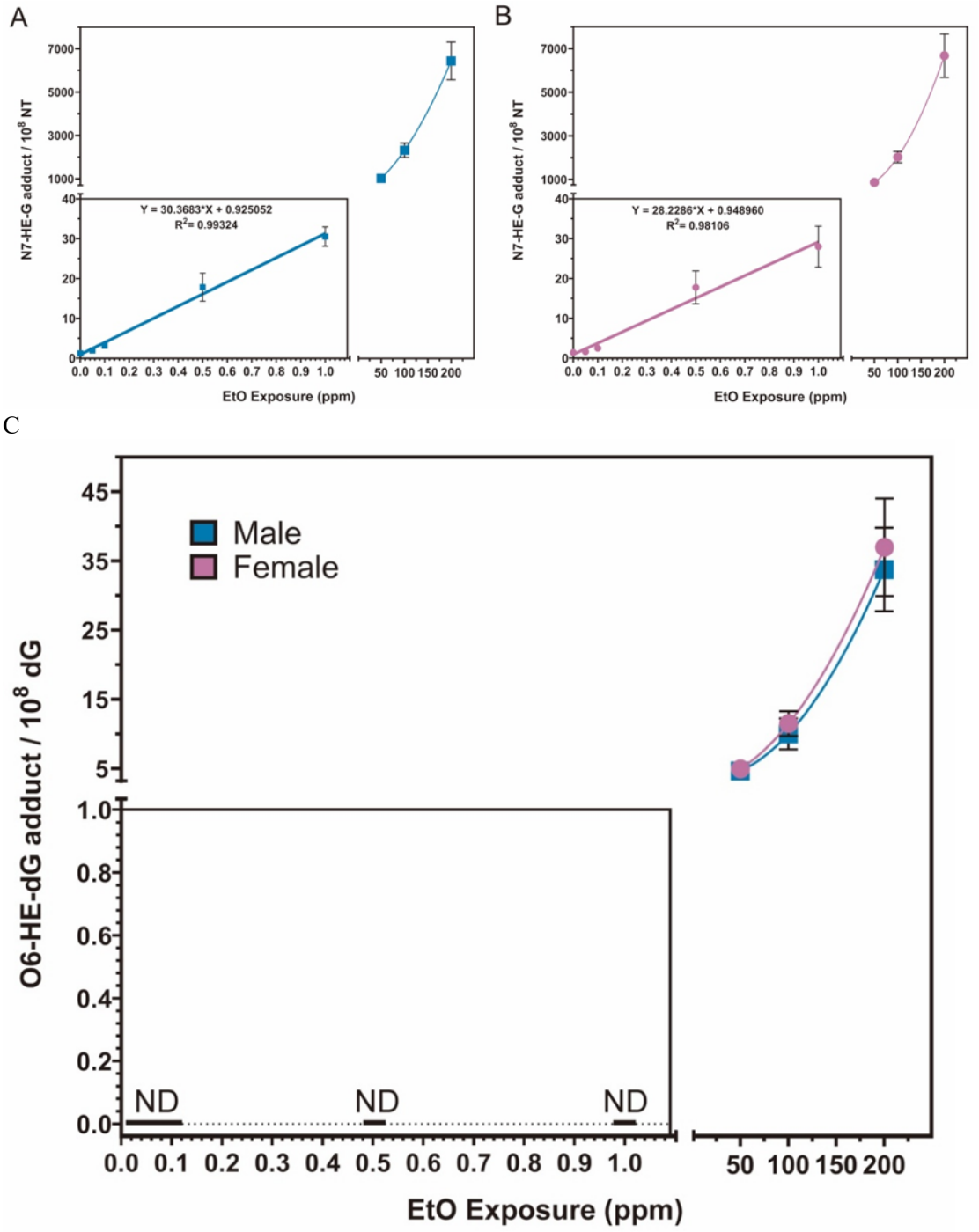


Figure 17. Dose-response data for DNA adducts in lungs of mice exposed to a broad range of EO concentrations exhibit: (male: A & female: B) linear relationship for N7HEG at low concentrations; (C) apparent threshold relationship for O^6HEdG at low concentrations.

As described in detail in Section 4 of these comments, the same animals in the Liu et al. (2026) study were also characterized for apical *Pig-a* mutations and cytogenetic damage

(micronucleus formation; Gollapudi et al., 2026). The dose-response patterns for these endpoints (Figure 9) were consistent with similar dose-response patterns observed for DNA adducts and systemic EtO HEV, and importantly, all conservatively only reflect at most a single linear shallow dose-response shape.

6. Recommended IURs

Due to the high probability of exposure measurement error, we recommend that OEHHA provide IUR values based on (1) application of the biologically plausible standard log-linear CPH model to the Lymphoid and Breast Cancer Data, and (2) EPA IRIS (2016) model of the Animal Data

6.1. Derivation of IURs based on TCEQ's Approach for both Breast Cancer Mortality and Lymphoid Mortality

If the NIOSH study remains the critical study for cancer risk assessment, the primary cancer endpoints considered by USEPA IRIS and TCEQ for quantitative risk assessment include lymphohematopoietic (LH) cancer mortality, lymphoid cancer mortality (a subset of LH), breast cancer mortality, and breast cancer incidence from the NIOSH cohort study that was updated through 1998 (USEPA, 2016; TCEQ, 2020; Steenland et al., 2003, 2004). The overall weight of epidemiologic evidence is stronger for LH cancers than for breast cancer (Marsh et al., 2019). USEPA's IRIS (2016) cancer risk value is based on lymphoid mortality and breast cancer incidence, with lymphoid mortality contributing approximately 85% of the USEPA's risk estimate. In contrast, TCEQ (2020) relies solely on lymphoid mortality from the same cohort. We note that TCEQ's decision to not address breast cancer does not undermine its analysis with respect to lymphoid cancer. Breast cancer can be addressed using TCEQ's model approach as discussed below and as reflected in Table 2.

We agree with TCEQ (2020) that the breast cancer incidence data from Steenland et al. (2003) is not appropriate for quantitative risk assessment. However, our primary reason for excluding this data is due to the substantial under-ascertainment (missing cases) and the data are not publicly available for independent analysis. The authors of the NIOSH study indicated that workers with shorter employment duration and, hence, lower cumulative exposures were more difficult to locate. This, together with the exposure errors in earlier years, indicates that this study should NOT be used for quantitative risk assessment purposes. This recommendation is consistent with USEPA Office of Pesticide Programs (OPP) Guidance[8] for use of open literature toxicity studies to support human health risk assessment. Although written by USEPA OPP, the principles are universally applicable including emphasis on adequate assessment of exposures and reasonably valid and reliable outcome ascertainment, which the Steenland et al. (2003) breast cancer

[8] USEPA https://www.epa.gov/pesticide-science-and-assessing-pesticide-risks/guidance-considering-and-using-open-literature

incidence data lacks. In contrast, the mortality data for breast cancer and lymphoid cancers from Steenland et al. (2004) are fully ascertained, but still suffers from exposure errors in earlier years. As described in greater detail below, the epidemiological, biological, and toxicological evidence support application of the standard log-linear approach to both lymphoid and breast cancer mortality data from the NIOSH cohort.

Unit risk values based on the standard log-linear CPH model were applied to the NIOSH lymphoid and breast cancer mortality (Table 2). In the majority of cases, these values could be found in TCEQ (2020) or USEPA IRIS (2016), or be derived from model slopes published by Valdez-Flores et al. (2010) or USEPA IRIS (2016). As explained in greater detail above, the standard log-linear CPH model is essentially a linear model that has much greater biological plausibility based on existing and new mode of action data. These values are derived using USEPA IRIS approach to modeling with the following exceptions:

- The point-of-departure for cancer risk assessment should be at the lower range of the NIOSH data. The POD of 1E-05 was correctly demonstrated by TCEQ (2020) to be at the lower range of the NIOSH data. In most cases, the POD of 1/100 used by USEPA is at the higher range or extrapolates above the experimental exposure range.
- Excess risks were calculated at age 70 instead of 85 years. The estimated exposure-response models are based on **worker** exposures. Calculating extra risks through age 85 years involves extrapolation of the fitted models beyond the range of the data upon which they are based. Calculating excess risks through age 85 years makes the life table analysis unstable because <1% of the cohort lived past 85 involving extrapolations of the fitted models beyond the range of the data upon which they are based (see Valdez-Flores et al. 2010 for detailed discussion).
- This analysis followed the US EPA IRIS (2016) approach for life-table calculations, modeling excess risk for lymphoid and breast cancer incidence based on the lymphoid and breast cancer mortality data from Steenland et al. (2004). There is a high potential for substantial error if life-table calculations for incidence are made using life-table formulas designed for mortality (Sielken and Valdez-Flores, 2009). Nevertheless, we used CDC WONDER (Wide-ranging Online Data for Epidemiologic Research) Database 1999-2022 rather than the NCI SEER (Surveillance, Epidemiology and End Results) Database. The CDC WONDER is more representative of the entire U.S. Population because it provides cancer incidence data for all 50 states, the District of Columbia, and Puerto Rico. In contrast, the SEER program

alone covers approximately 48 % of the U.S. population through selected cancer registries.

To increase transparency regarding these limitations and uncertainties, OEHHA could provide IURs for lymphoid and breast cancer mortality by first applying the SEER-type mortality database to estimate extra risk and then applying the SEER-type incidence database. OEHHA could then explicitly state that the incidence-based risk estimates assume the same dose-response relationship as the mortality-based risk estimates. The difference between the mortality- and incidence-based extra-risk estimates could then be treated as an adjustment factor. Specifically, when SEER-type incidence data are used to estimate extra risk (as shown in Table 2 below), it is assumed that the dose-response pattern for incidence mirrors that for mortality. This assumption may not be correct as detailed by Sielken and Valdez-Flores et al. (2009). This approach would allow calculation of both incidence-based and mortality-based extra risks, thereby improving the transparency of the risk assessment. By clearly documenting these assumptions, OEHHA would enable future revisions to the risk assessment as additional data or better approaches become available.

Table 2. Summary of biologically plausible EtO cancer incidence unit risk values based on rat and mouse cancer and human cancer risk values. Values are reported without ADAF to be comparable to the 2019 USEPA Memo

Cancer/Model Type Note: All estimates for lymphoid and breast cancer are based on NIOSH study by Steenland et al. 2004)	Model slope per ppm-day MLE (SE)	Incidence Unit risk per ppm MLE (95% UCL)[7]
Rat Cancer Risk (USEPA IRIS, 2016, Table 4-20, 4-26)	-	**4.01 – 6.70 E-02 (95% UCL)**
Mouse Cancer Risk (USEPA IRIS, 2016, Table 4-26)	-	**5.51 – 8.33 E-02 (95% UCL)**
Lymphoid (15 yr lag) + Breast Cancer (20 yr lag) from NIOSH Steenland et al. (2004) study converted to incidence unit risk (USEPA's method)	See below for each cancer type	
log-linear CPH (TCEQ, 2020 and Valdez-Flores 2010)		**2.79E-02 (5.64E-02)**
log-linear CPH (USEPA IRIS, 2016)		**3.89E-02 (7.01E-02)**
linear CPH (USEPA IRIS)		8.66E-02 (1.93E-01)
Lymphoid Cancer Mortality (15 yr-lag) – these values were used to develop the combined URE		
log-linear CPH (TCEQ, 2020)	**3.12E-06 (2.61E-06)[1]**	**1.93E-03 (4.59E-03)**
log-linear CPH (USEPA IRIS, 2016)	**4.74E-06 (3.35E-06)[2]**	**2.93E-03 (6.34E-03)**
linear CPH (USEPA IRIS, 2016)	1.23E-05 (2.12E-05)[3]	7.59E-03 (2.91E-02)
Breast Cancer Mortality (20 yr-lag)		
log-linear CPH (Valdez-Flores et al. 2010; Supplementary tables with lagged-out slopes)	**9.42E-06 (5.48E-06)[4]**	**2.78E-02 (5.43E-02)**
-log-linear CPH (USEPA IRIS)	**1.22-E05 (6.41E-06)[5]**	**3.60E-02 (6.70E-02)**
-linear CPH (USEPA IRIS)	2.68E-05 (2.15E-05)[6]	7.90E-02 (1.83E-01)

[1]Table 9 in TCEQ (2020) – mortality – males only (higher cancer risk than for males and females combined)
[2]Table 4-2 in USEPA (2016)
[3]Table D-36 in USEPA Appendices (2016). Calculated SE=(4.71E-05-1.227E-05)/1.645
[4]Table S-10 in Supplemental of Valdez-Flores (2010)
[5]Table 4-9 in USEPA (2016)
[6]Table D-25 in USEPA Appendices (2016).
[7]POD = 1E-05 (Sielken and Valdez-Flores, 2009), risk at 70 yrs age estimated for lymphoid mortality based on combined male and female background incidence rates, and for breast cancer mortality based on female background incidence rates. CDC Wonder Database 1999-2022 (https://wonder.cdc.gov/cancer-v2022.html)

was used to estimate extra risk. It is a much more comprehensive database covering all 50 states, Puerto Rico and Washington D.C.

6.2. Additional Model Slope Comparisons Based On NIOSH Breast Cancer Incidence, Updated NIOSH Breast Cancer Mortality, UCC + NIOSH Combined Cancer Mortalities

For comparison purposes only, the following model slope values for breast cancer incidence (USEPA IRIS) and breast cancer mortality (Kelly-Reif et al. 202**5**) were also estimated (Table 3). Breast cancer incidence is not appropriate for quantitative risk assessment purposes because there are a substantial number of missing cases. Based on analysis of Steenland et al. (2003) if all women were interviewed, the analysis would have included 134 or more breast cancer cases in addition to 233, which became the data incorporated in the USEPA model[9]. The missing cases were due mostly to location problems of short term workers, i.e., those with lower cumulative exposures. This can result in serious concern about selection bias where proportionally more cases are found and interviewed among long-term workers, and fewer cases interviewed in shorter term workers.

For the updated breast cancer mortality study, the "unadjusted" linear CPH slope reported by Kelly-Reif et al. (202**5**) is nearly identical (and numerically smaller) than the USEPA IRIS (2016) linear CPH slope based on Steenland et al. (2004), indicating consistency in linear model results across datasets. These results do NOT indicate that the results of this update support USEPA's steep spline model. The "unadjusted" linear CPH model is identical to the USEPA IRIS (2016) application of the linear CPH model. Notably, neither USEPA IRIS nor Kelly-Reif et al. (202**5**) evaluated the log-linear CPH model, despite its application to the breast cancer mortality data in Steenland et al. (2004).

[9] Steenland et al. (2003) identified 319 cases. The SMR = 0.87. 233 (73%) of cases interviewed. SMR = O/E; 0.87 = 319/E, solving for E: E=367 so 319 were found (observed) but expected is 367. Therefore 367-319 =48 cases under-ascertained. Assuming no association with EtO, i.e., SMR = 1. Observed would have to be 367, same as expected to get SMR of 1.0. If EtO caused cases then more than 48 would be missing to get an SMR>1.0. Of 319 cases, only 233 (73%) were interviewed, so 319-233 =86 were lost at this phase of study. So at least 48 + 86 =134 more cases should have been included in the interview portion of the study.

Table 3. Model slope values derived from UCC and/or NIOSH cohorts based on linear and log-linear models for comparison purposes only

Cancer/Model Type (lag included unless otherwise noted)	Model slope MLE (SE) per ppm days
Breast Cancer Incidence	
log-linear CPH (USEPA IRIS, 2016)	5.4E-06 (3.5E-06)[1] all 9.5E-06 (4.1E-06)[1] interview
linear CPH (USEPA IRIS, 2016)	2.30E-05 (1.44E-05)[2] interview
Breast Cancer Mortality	
linear CPH (Kelly-Reif et al. 2025;Supplemental)	1.92E-05 (1.83E-05)[3]
Linear CPH (USEPA IRIS analysis of Steenland et al. 2004)	2.68E-05 (2.15E-05)[4]
UCC + NIOSH studies for Male and Female combined	
Log-linear CPH (no lag) (Valdez-Flores et al. 2010)	1.92 E-06 (2.08 E-06)

[1]Table 4-12 in USEPA 2016

[2]Table 4-13 in USEPA (2016). Calculated SE=(4.67E-05-2.30E-05) /1.645

[3]MLE and SE for the slope were calculated from the unrestricted linear model in Table S3 of Kelly-Reif et al. (2025). MLE = (1.07-1)/3650 = 1.92E-05, 95%UCL = (1.18-1)/3650=4.93E-05, SE=(95%UCL-MLE)/1.645.

[4]USEPA IRIS, 2016 Table D-25

6.3. Endogenous exposure estimates are a useful reality check for risk-based concentration levels.

OEHHA generally appropriately summarizes the clear presence of endogenous EO production. A few places where substantive improvements could be made are provided below.

OEHHA makes several statements regarding uncertainty in endogenous EO estimates:

> "The usefulness of HEV levels in predicting endogenous production of ethylene or EtO in humans is uncertain at present as there are no validated toxicokinetic models or data to support it."
>
> "OEHHA recognizes the uncertainty regarding the relative contribution of endogenous sources of ethylene and EtO to the background levels of HEV in non-smokers (i.e., 8.27–656 pmol/g Hb)."
>
> is based on extra risk above endogenous and ambient background exposures to EtO." "However, this uncertainty regarding the relative contribution of endogenous

> EtO sources to background HEV levels does not impact the derivation of the IUR, which

While some uncertainty in endogenous estimates (Kirman et al., 2021, 2025) is recognized, because the estimates are based upon high-quality data sets (NHANES hemoglobin biomarkers in non-smokers and smokers; USEPA air monitoring of background ambient and near-facility), and sound scientific principles (e.g., linear toxicokinetics at low concentrations), the uncertainties are relatively modest when compared to the much larger uncertainties associated with the IUR value for EO as noted above.

OEHHA further states,

> "However, there are no tenable toxicokinetic models or data to support using the HEV levels to back-calculate endogenous equivalent air concentrations of EtO."
>
> "In this regard, ATSDR (2022b) concluded that data are not available to demonstrate that background HEV levels in non-smokers are direct indicators of internal, endogenous EtO exposures."

HEV are biomarkers of total exposure, from which endogenous can be estimated indirectly by difference when contributions from other contributing pathways (exogenous exposure by air) pathway are quantified. Although having a toxicokinetic model would help improve confidence in endogenous estimates, it is not necessary. A continuously linear relationship between hemoglobin adducts and exogenous EO has been factual established for ppm-level exposed workers down to low-ppb levels in U.S. smokers, and smoking-contributed EO exposures were demonstrated as merging within the variability of total HEV in non-smokers (Fig., 5; Kirman et al., 2025).

These data offer an important reality check opportunity to examine whether EPA/OEHHA IUR 10^{-6} to 10^{-4} risk exposures (0.0001 – 0.011 ppb) contribute a biologically plausible extra burden above that of total background HEV in nonsmokers. The smoker EO-HEV dose-response indicates exposure to the IUR risk concentrations results in extra contributions to total background HEV that fall well below the first percentile of total background HEV in non-smokers, i.e., they cannot be meaningfully differentiated from total and endogenously dominated background HEV (Kirman et al., 2025).

While we agree that “endogenous EtO sources to background HEV levels does not impact the derivation of the IUR”, total background HEV indeed offers an informative and independent human dataset for gauging performance of the low-exposure dose-response model underpinning the IUR against real-world data. This approach is akin to how PBPK model performance is routinely reality checked (validated) and modified accordingly against appropriate human/animal toxicokinetic datasets. The observation that the

performance of the IUR dose-response model fails to predict true extra risk beyond substantial endogenously generated background exposure thus highly challenges the biological plausibility of the model predictions.

Although Lin et al. (2025) concludes that endogenous EtO production explains less than 20% of total HEV in nonsmokers and therefore the source of the remaining 80% remains uncertain, this conclusion rests on questionable inputs. For example, the authors relied upon a small subset of available data without consideration that breath ethylene measurements vary widely depending on analytical method (laser photoacoustic vs. gas chromatography), or that previous attempts to incorporate data from the laser photoacoustic method into a PBPK model were problematic (Berkelmans et al., 2003). Secondly, the ambient air exposures assumed in Lin et al.'s assessment overestimate real-world levels for ethylene. Lastly, the underlying PBPK model (Filser and Klein, 2018), which was modified by Lin et al. to include the endogenous production pathway of Csanady et al., 2000), has not been validated against HEV data for unexposed humans and animals. The underprediction HEV in unexposed across mice, rats, and humans was discussed by Csanady et al. (2000), but not by Lin et al. (2025). Part of this underprediction may reflect that the endogenous ET production pathway is included in the model as a single arterial-compartment pathway rather than the three expected biological sources (gut bacterial production, liver metabolism, and oxidative stress).

Lastly, OEHHA states,

> "No studies have characterized the exposure response relationship between endogenous EtO and cancer, and there are no data to indicate whether the nature of the dose-response relationship between endogenous EtO and cancer is identical to the exogenous EtO versus cancer relationship derived from the NIOSH study."

This statement is somewhat unusual, and problematic. Endogenous exposure estimates (Kirman et al., 2021, 2025) rely upon a linear relationship for the toxicokinetics of EO at low exposure levels (i.e., the relationship between exposure and hemoglobin adduct formation is linear). OEHHA appears to be questioning the validity of this assumption, despite the strong evidence supporting it. Cancer risk assessment performed by OEHHA and other agencies also rely upon a linear relationship for the toxicokinetics of EO at low exposure levels and extends this assumption further by also including the toxicodynamics of EO in the assumption (i.e., the relationship between exposure and cancer risk is assumed to be linear). If OEHHA claims that the former assumption of linearity is too uncertain, OEHHA then needs to explain how the latter assumption of linearity would not also be too uncertain. That is, if linearity is not too uncertain for dose response assessment, then How does OEHHA conclude that it is too uncertain for the toxicokinetics of initiating responses?

Endogenous exposure estimates for EO provide context for interpreting the significance of potential exogenous exposures greater than the ambient background and provide value as a reality check on the utility of agency cancer risk assessment risk-specific concentrations in managing and communicating the risk of excess exogenous EO exposures potentially experienced by populations residing near emitting industrial facilities. Attempts to use IUR values based on the 2-piece linear spline model result in risk-based concentrations for EO that are not biologically meaningful within the context of background exposures, i.e., the Kirman et al. (2025) smoking analyses. As such, OEHHA should consider other options for selection of dose-response models exhibiting performance consistency with considerations of endogenous EO production (e.g., rely on TCEQ assessment, animal data) for determining the IUR value for EO.

References

Agency for Toxic Substances and Disease Registry (ATSDR). 2022. Toxicological Profile for Ethylene Oxide.

Benson LO, Teta MJ. Mortality due to pancreatic and lymphopoietic cancers in chlorohydrin production workers. British Journal of Industrial Medicine. 1993;50(8):710–716.

Berkelmans HWA, Moeskops BWM, Bominaar J, Scheepers PTJ, Harren FJM. Pharmacokinetics of ethylene in man by on-line laser photoacoustic detection. Toxicol Appl Pharm. 2003;190:206–13.

Bogen KT, Sheehan PJ, Valdez-Flores C, Li AA. Reevaluation of Historical Exposures to Ethylene Oxide Among U.S. Sterilization Workers in the National Institute of Occupational Safety and Health (NIOSH) Study Cohort. Int J Environ Res Public Health. 2019 May 16;16(10):1-16.

Centers for Disease Control and Prevention. (n.d.). National report on human exposure to environmental chemicals: Ethylene oxide hemoglobin adducts. Available online: https://www.cdc.gov/exposurereport/index.html.

Csanady GA, Denk B, Putz C, Kreuzer PE, Kessler W, Baur C, Gargas ML, Filser JG. A physiological Toxicokinetic model for exogenous and endogenous ethylene and ethylene oxide in rat, mouse, and human: Formation of 2-hydroxyethyl adducts with hemoglobin and DNA. 2000. Toxicol Appl Pharm165, 1–26.

Donner EM, Wong BA, James RA, Preston RJ. Reciprocal translocations in somatic and germ cells of mice chronically exposed by inhalation to ethylene oxide: implications for risk assessment. Mutagenesis. 2010 Jan;25(1):49-55. doi: 10.1093/mutage/gep042. Epub 2009 Nov 1. PMID: 19884119; PMCID: PMC2796186.

Dosemeci M, Wacholder S, Lubin J. Does nondifferential misclassification of exposure always bias a true effect toward the null value? Am J Epidemiol 1990;132:746–8.

Fennell TR, Brown CD. A physiologically based pharmacokinetic model for ethylene oxide in mouse, rat, and human. Toxicol Appl Pharmacol. 2001 Jun 15;173(3):161-75. doi: 10.1006/taap.2001.9184. PMID: 11437638.

Filser JG, Klein D. A physiologically based toxicokinetic model for inhaled ethylene and ethylene oxide in mouse, rat, and human. Toxicol Lett. 2018 Apr;286:54-79. doi: 10.1016/j.toxlet.2017.07.896. Epub 2017 Jul 31. PMID: 28774830.

Gollapudi BB, Bus JE, Cassidy P, Weinberg JT, Bemis JC, Torus, DT, Dertinger SD, Lu K, Reiss, R, and Li AP. Investigation of ethylene oxide genotoxicity dose-response to inform cancer risk assessment. Environ Mol Mutagen. 2026 (In Press).

Gollapudi BB, Su S, Li AA, Johnson GE, Reiss R, Albertini RJ. Genotoxicity as a toxicologically relevant endpoint to inform risk assessment: A case study with ethylene oxide. Environ Mol Mutagen. 2020 Nov;61(9):852-871. doi: 10.1002/em.22408. Epub 2020 Sep 28. PMID: 32926486; PMCID: PMC7756744.

Greenberg HL, Ott MG, Shore RE. Men assigned to ethylene oxide production or other ethylene oxide-related chemical manufacturing: a mortality study. British Journal of Industrial Medicine. 1990;47(4):221–230.

Hamm, T. E., Jr, Guest, D., & Dent, J. G. (1984). Chronic toxicity and oncogenicity bioassay of inhaled ethylene in Fischer-344 rats. Fundamental and applied toxicology : official journal of the Society of Toxicology, *4*(3 Pt 1), 473–478. https://doi.org/10.1016/0272-0590(84)90205-7

Hornung, R. W., Greife, A. L., Stayner, L. T., Steenland, N. K., Herrick, R. F., Elliott, L. J., Ringenburg, V. L., & Morawetz, J. (1994). Statistical model for prediction of retrospective exposure to ethylene oxide in an occupational mortality study. *American journal of industrial medicine*, *25*(6), 825–836. https://doi.org/10.1002/ajim.4700250607

Joyner RE. Chronic Toxicity of Ethylene Oxide: A Study of Human Responses to Long-Term Low-Level Exposures. Archives of Environmental Health. 1964;8:700–710

Kelly-Reif, K., Bertke, S. J., Stayner, L., & Steenland, K. (2025). Exposure to Ethylene Oxide and Relative Rates of Female Breast Cancer Mortality: 62 Years of Follow-Up in a Large US Occupational Cohort. *Environmental health perspectives*, *133*(5), 57013. https://doi.org/10.1289/EHP15566

Kirman, C. R., Li, A. A., Sheehan, P. J., Bus, J. S., Lewis, R. C., & Hays, S. M. (2021). Ethylene oxide review: characterization of total exposure via endogenous and exogenous pathways and their implications to risk assessment and risk management. *Journal of toxicology and environmental health. Part B, Critical reviews*, *24*(1), 1–29. https://doi.org/10.1080/10937404.2020.1852988

Kirman, C. R., Sheehan, P. J., Li, A. A., Bus, J. S., Su, S. H., Dopart, P. J., Watson, H. N., Moynihan, E. E., & Reiss, R. (2025). Characterization of Background Exposures to Ethylene Oxide in the United States: A Reality Check on Theoretical Health Risks for Potentially Exposed Populations near Industrial Sources. International journal of environmental research and public health, 22(4), 597. https://doi.org/10.3390/ijerph22040597 7

Klapacz J, Pottenger LH, Engelward BP, Heinen CD, Johnson GE, Clewell RA, Carmichael PL, Adeleye Y, Andersen ME. Contributions of DNA repair and damage response pathways to the non-linear genotoxic responses of alkylating agents. Mutat Res Rev Mutat Res. 2016 Jan-Mar;767:77-91. doi: 10.1016/j.mrrev.2015.11.001. Epub 2015 Dec 2. PMID: 27036068; PMCID: PMC4818947.

LeBaron, M. J., Schisler, M. R., Zhang, F., Jeong, Y. C., Bartels, M. J., Sura, R., Hotchkiss, J., Geter, D. R., Gollapudi, B. B., and Moore, N. P. (2012). Bioanalytical, molecular, and genetic alterations in ethylene oxide-exposed male B6C3F1 mice. Environ. Mol. Mutagen. 53, S65.

Lin, Y. S., Thayer, K. A., White, P., Morozov, V., & Persad, A. S. (2025). Uncovering the connection: ethylene exposure and endogenous ethylene oxide levels in humans. Journal of exposure science & environmental epidemiology, 36(2), 361–374. https://doi.org/10.1038/s41370-025-00826-7

Liu CW, Peng J, Feng J, Zhao H, Wang X, Gollapudi BB, Li AA, Bus JS, Lu K. Molecular dosimetry of DNA adducts in mice exposed to ethylene oxide. Toxicol Sci. 2026 Jul 1;209(7):kfag063. doi: 10.1093/toxsci/kfag063. PMID: 42391620.

Liu, CW, Feng J, Peng J, Zhao H, Wang X, Gollapudi BB, Li AA, Bus JS Lun K. Molecular dosimetry of hemoglobin adducts in mice exposed to ethylene oxide. Arch Toxicol (2026a). https://doi.org/10.1007/s00204-026-04419-7.

Liu, CW, Feng, J, Peng, J, Zhao H, Wang X, Gollapudi B, Li A, Bus JS, Kirman C, Lu K. Sensitive Quantification of the Ethylene Oxide Hemoglobin AdductN-(2-Hydroxyethyl)-L-valine from Minimal Blood Volumes. Chem Res Toxicol 2026b. 39, 1084−1087; https://doi.org/10.1021/acs.chemrestox.6c00200.

Lynch, D. W., Lewis, T. R., Moorman, W. J., Burg, J. R., Groth, D. H., Khan, A., Ackerman, L. J., & Cockrell, B. Y. (1984). Carcinogenic and toxicologic effects of inhaled ethylene oxide and propylene oxide in F344 rats. Toxicology and applied pharmacology, 76(1), 69–84. https://doi.org/10.1016/0041-008x(84)90030-9

MacRae SL, Croken MM, Calder RB, Aliper A, Milholland B, White RR, Zhavoronkov A, Gladyshev VN, Seluanov A, Gorbunova V, Zhang ZD, Vijg J. DNA repair in species with extreme lifespan differences. Aging (Albany NY). 2015 Dec;7(12):1171-84. doi: 10.18632/aging.100866. PMID: 26729707; PMCID: PMC4712340.

Manjanatha MG, Shelton SD, Chen Y, Parsons BL, Myers MB, McKim KL, Gollapudi BB, Moore NP, Haber LT, Allen B, Moore MM. Dose and temporal evaluation of ethylene oxide-induced mutagenicity in the lungs of male big blue mice following inhalation exposure to

carcinogenic concentrations. Environ Mol Mutagen. 2017 Apr;58(3):122-134. doi: 10.1002/em.22080. Epub 2017 Mar 22. PMID: 28326610.

Marsh, G. M., Keeton, K. A., Riordan, A. S., Best, E. A., & Benson, S. M. (2019). Ethylene oxide and risk of lympho-hematopoietic cancer and breast cancer: a systematic literature review and meta-analysis. International archives of occupational and environmental health, 92(7), 919–939. https://doi.org/10.1007/s00420-019-01438-z

Moore MM, Heflich RH, Haber LT, Allen BC, Shipp AM, Kodell RL. Analysis of in vivo mutation data can inform cancer risk assessment. Regul Toxicol Pharmacol. 2008 Jul;51(2):151-61. doi: 10.1016/j.yrtph.2008.01.015. Epub 2008 Feb 1. PMID: 18321622.

National Research Council (NRC). Ethylene Oxide Acute Exposure Guideline Levels. In: Acute Exposure Guideline Levels for Selected Airborne Chemicals, Volume 9. Washington, DC: National Academies Press; 2010. Section 2.2.1 (Odor Threshold), pp. 51–52.

National Toxicology Program (1987). NTP Toxicology and Carcinogenesis Studies of Ethylene Oxide (CAS No. 75-21-8) in B6C3F1 Mice (Inhalation Studies). National Toxicology Program technical report series, 326, 1–114.

Phillips, C.R., and Kaye, S. (1949). The sterilizing action of gaseous ethylene oxide; a review. American journal of hygiene, 50(3), 270–279. https://doi.org/10.1093/oxfordjournals.aje.a119359

Picciotto S, Kelly-Reif K, Eisen EA, Stayner LT, Costello S. How to identify the healthy worker survivor effect empirically and how to interpret results from published studies: the NIOSH ethylene oxide cohort as a case study. Am J Epidemiol. 2026 Mar 9:kwag052. doi: 10.1093/aje/kwag052. Epub ahead of print. PMID: 41800785.

Recio L, Donner M, Abernethy D, Pluta L, Steen AM, Wong BA, James A, Preston RJ. In vivo mutagenicity and mutation spectrum in the bone marrow and testes of B6C3F1 lacI transgenic mice following inhalation exposure to ethylene oxide. Mutagenesis. 2004 May;19(3):215-22. doi: 10.1093/mutage/geh017. PMID: 15123787.

Roy R, Shiota S, Kennel SJ, Raha R, von Wronski M, Brent TP, Mitra S. A comparative study of the biochemical properties of human and mouse recombinant O6-methylguanine-DNA methyltransferases. Carcinogenesis. 1995 Feb;16(2):405-11. doi: 10.1093/carcin/16.2.405. PMID: 7532116.

Sexton RJ, Henson EV. Experimental Ethylene Oxide Human Skin Injuries. A.M.A. Archives of Industrial Hygiene and Occupational Medicine. 1950;2(5):549–564. PMID: 14777190.

Sielken RL, Valdez Flores C. (2009). Life-table calculations of excess risk for incidence versus mortality: ethylene oxide case study. Regul Toxicol Pharmacol 55(1):82-89.

Snellings, W. M., Weil, C. S., & Maronpot, R. R. (1984). A two-year inhalation study of the carcinogenic potential of ethylene oxide in Fischer 344 rats. *Toxicology and applied pharmacology*, *75*(1), 105–117. https://doi.org/10.1016/0041-008x(84)90081-4

Stayner L, Steenland K, Dosemeci M, Hertz-Picciotto I. Attenuation of exposure-response curves in occupational cohort studies at high exposure levels. Scand J Work Environ Health. 2003 Aug;29(4):317-24. doi: 10.5271/sjweh.737. PMID: 12934726.

Steenland, K., Stayner, L., & Deddens, J. (2004). Mortality analyses in a cohort of 18 235 ethylene oxide exposed workers: follow up extended from 1987 to 1998. *Occupational and environmental medicine*, *61*(1), 2–7.

Steenland, K., Whelan, E., Deddens, J., Stayner, L., & Ward, E. (2003). Ethylene oxide and breast cancer incidence in a cohort study of 7576 women (United States). *Cancer causes & control : CCC*, *14*(6), 531–539. https://doi.org/10.1023/a:1024891529592

Swaen GMH, Burns C, Teta JM, Bodner K, Keenan D, Bodnar CM. Mortality study update of ethylene oxide workers in chemical manufacturing: a 15 year update. Journal of Occupational and Environmental Medicine. 2009;51(6):714–723.

Teta MJ, Benson LO, Vitale JN. Mortality study of ethylene oxide workers in chemical manufacturing: a 10 year update. British Journal of Industrial Medicine. 1993;50(8):704–709.

Texas Commission on Environmental Quality (TCEQ). (2020). Ethylene oxide carcinogenic dose-response assessment; CAS Registry Number: 75-21-8; In Development Support Document Final, May 15, 2020. Texas Commission on Environmental Quality: Austin, TX, USA.

Troy, J. D., Hartge, P., Weissfeld, J. L., Oken, M. M., Colditz, G. A., Mechanic, L. E., & Morton, L. M. (2010). Associations between anthropometry, cigarette smoking, alcohol consumption, and non-Hodgkin lymphoma in the Prostate, Lung, Colorectal, and Ovarian Cancer Screening Trial. American journal of epidemiology, 171(12), 1270–1281. https://doi.org/10.1093/aje/kwq085

U. S. Environmental Protection Agency. Guidelines for carcinogen risk assessment. 2005. Available at: https://www.epa.gov/sites/default/files/2013-09/documents/cancer_guidelines_final_3-25-05.pdf. Accessed 3 March 2026.

U.S. Environmental Protection Agency (US EPA). (2022). Summary of public comments and responses for the reconsideration of the 2020 National Emission Standards for Hazardous Air Pollutants: Miscellaneous Organic Chemical Manufacturing Residual Risk and Technology Review. Office of Air Quality Planning and Standards Sector Policies and Programs Division (EE-143-01), Research Triangle Park, NC 27711.

U.S. Environmental Protection Agency (US EPA). (2024). Summary of Public Comments and Responses for New Source Performance Standards for the Synthetic Organic Chemical Manufacturing Industry and National Emission Standards for Hazardous Air Pollutants for the Synthetic Organic Chemical Manufacturing Industry and Group I & II Polymers and Resins Industry. Office of Air Quality Planning and Standards, Research Triangle Park, NC. Docket ID No. USEPA-HQ-OAR-2024-0303, posted by US EPA May 16, 2024. https://www.regulations.gov/document/USEPA-HQ-OAR-2022-0730-2764

U.S. Environmental Protection Agency (USEPA SAB). (2015). Science Advisory Board Review of the USEPA's Evaluation of the Inhalation Carcinogenicity of Ethylene Oxide (Revised External Review Draft—August 2014); USEPA-SAB-15-012 (2015 SAB Review). Washington, DC.

U.S. Environmental Protection Agency Implementation of Gold Standard Science (available at https://www.epa.gov/scientific-leadership/gold-standard-science

U.S. Environmental Protection Agency, Integrated Risk Information System (US EPA IRIS). (2016). Evaluation of the Inhalation Carcinogenicity of Ethylene Oxide (CASRN 75-21-8) In Support of Summary Information on the Integrated Risk Information System (IRIS). U.S. Environmental Protection Agency, Office of Research and Development, National Center for Environmental Assessment: Washington, DC.

U.S. Environmental Protection Agency. (2008). U.S. Environmental Protection Agency. Reregistration Eligibility Decision for Ethylene Oxide. USEPA 738-R-08-003. U.S. Environmental Protection Agency, Office of Pesticide Programs. Washington, DC. March.

Valdez-Flores C, Sielken RL, Teta MJ. (2010). Quantitative cancer risk assessment based on NIOSH and UCC epidemiological data for workers exposed to ethylene oxide. Regul Toxicol Pharmacol 56: 312-320.

Valdez-Flores, C., & Sielken, R. L., Jr (2013). Misinterpretation of categorical rate ratios and inappropriate exposure-response model fitting can lead to biased estimates of risk:

ethylene oxide case study. Regulatory toxicology and pharmacology: RTP, 67(2), 206–214. https://doi.org/10.1016/j.yrtph.2013.07.011

Valdez-Flores, C., Li, A. A., Bender, T. J., & Teta, M. J. (2025). Use of updated mortality study of ethylene oxide manufacturing workers to inform cancer risk assessment. Risk analysis : an official publication of the Society for Risk Analysis, 45(9), 2822–2837. https://doi.org/10.1111/risa.70057

Walker, V. E., Wu, K. Y., Upton, P. B., Ranasinghe, A., Scheller, N., Cho, M. H., Vergnes, J. S., Skopek, T. R., & Swenberg, J. A. (2000). Biomarkers of exposure and effect as indicators of potential carcinogenic risk arising from in vivo metabolism of ethylene to ethylene oxide. Carcinogenesis, 21(9), 1661–1669. https://doi.org/10.1093/carcin/21.9.1661

Appendix A

Comments on OEHHA Ranking of the Quality Characteristics of the NIOSH and UCC epidemiological studies.

Overview of NIOSH vs UCC study rating considerations.

OEHHA presents a biased study quality evaluation favoring the NIOSH study such that the UCC study is not considered in the EPAs weight of evidence as a reality check on the results from the exposure-response analysis based on the NIOSH study.

The UCC cohort of 2,053 chemical workers producing or using EtO have been followed for mortality most recently from 1940-2013 (Valdez-Flores 2025). These workers and their work history data were derived from a NIOSH cohort study of 29,139 male UCC workers in the Kanawha Valley (KV) of West Virginia (Risky et al., 1988). NIOSH collected the work history records and job categories, departments and chemicals used in the UCC plants.

Although mortality follow-up began in 1940, some of the workers had begun employment as early as 1925, the year EtO production first began at UCC facilities. Publications by UCC medical directors in the early years of EtO production and use documented the acute effects of EtO due to frequent accidental over exposures (Sexton and Henson, 1950; Joyner, 1964). The UCC cohort has been updated several times over more than three decades (Greenberg, 1990; Teta et al. 1993; Benson and Teta, 1993, Swaen, 2009; Valdez-Flores, 2025). These publications provide extensive discussion of cohort assembly, exposure assessment methods, potential confounding exposures to other known carcinogens such as chlorhydrin.

OEHHA's evaluation of the UCC cohort appears to have relied on a limited subset of the available UCC literature while giving relatively little consideration to the full series of cohort updates and supporting analyses published by Greenberg et al. (1990), Teta et al. (1993), Benson and Teta (1993), Swaen et al. (2009), and Valdez-Flores et al. (2025). These publications provide important information regarding cohort definition, exposure assessment, confounding by chlorohydrin-related processes and other workplace chemicals, and the interpretation of cancer findings across successive follow-up periods.

By not comprehensively incorporating these UCC-related publications into its review, OEHHA may have undervalued the strengths of the UCC cohort and given disproportionate weight to perceived limitations of the study, resulting in a lower rating scores than warranted by the totality of the UCC literature. Furthermore, OEHHA did not appear to have applied the same level of critical scrutiny to limitations of the NIOSH exposure assessment or of recently published updates. As described in detail above, direct exposure monitoring data were unavailable for a substantial portion of the period before 1978, requiring retrospective reconstruction of exposures and limiting opportunities for empirical validation of exposure estimates. This applies to all NIOSH studies including recent

updates. This unequal treatment contributed to a disproportionate downgrading of the UCC cohort's rating scores compared to the NIOSH sterilizer-worker cohort.

Published results of the mortality follow up periods (1978, 1988, 2003, 2013) provide no empirical support for supra-linearity. The workers have now been followed an average of 41 years and 70% of them are deceased, compared to only 19% deceased in the NIOSH cohort, as of the follow-up from Steenland et al. 2004. OEHHA proposes to use the data from this follow-up period, similar to what was used by EPA IRIS 2016. The OEHHA report provides sparse background information on the UCC study found primarily in tabular form and quality rankings (Figure 3 and Attachment D).

OEHHA has leaned on EPA IRIS 2016 arguments to ignore the nonpositive findings from the extensive analyses over time of this historically important cohort of EtO workers, originally developed by NIOSH themselves in the 1980s as part of their cohort study of 29,139 UCC chemical workers (Rinsky et al.1988). Figure #3 in the OEHHA report ranks the UCC study poorly and the NIOSH sterilant study favorably along several important characteristics to justify their failure to utilize this important epidemiological evidence as a check on the validity of the OEHHA risk assessment. This is described in detail below referring to Figure 3 and Table D1 in the OEHHA report.

OEHHA gave high scores to the NIOSH study on their characteristic,“relevant outcome #2”, totally ignoring the under-ascertainment of breast cancer cases clearly recognized by Steenland et al. 2003. The SIR of 0.87 (a significant deficit) based on 319 identified cases means the expected number of cases is 367 (observed/expected = SIR). Of the 319 cases, 233 were interviewed (73%). The overall participation rate for cases and controls was 68%, below the generally accepted rate of at least 70% and which Steenland notes is “of concern”. OEHHA correctly notes that short-term workers “who may have had lower cumulative exposures were more difficult to locate and therefore less likely to be interviewed.” Losses of cases from the total population (at least 48 cases [367-319] and from the interview portion of the study (86 cases, 319-233) and losses of controls (32%) from the interview sub-study raises unanswerable questions about the impact on the exposure-response relationship. It is this data that form the basis for IRIS and are proposed to also be used by OEHHA.

OEHHA makes a biased comparison of UCC males to the combined NIOSH cohort of males and females and downgrades the UCC study along the lines of power and follow up (Figure 3). OEHHA argues that the overall UCC EtO sample size is much smaller than sterilant workers that included over 17,000 workers. The power of a study is not determined

by total sample size, as a large young population may be less informative than a smaller but older population followed over a longer time period. OEHHA ignores the importance of the more extensive UCC follow up that resulted in 25 male cases of lymphoid cancers, comparable to 27 identified among male sterilant workers (Steenland et al. 2004). OEHHA states 25 deaths are too small for analyses, even though Steenland et al. 2004 fully analyzed the NIOSH 27 male cases, including lagged analyses, and did not combine them with the female cases considered by these authors to not be at risk. In addition, the UCC study identified more deaths due to all LH cancers (41 versus 37 among NIOSH males).

The UCC study of males is unable to address breast cancer but is informative with respect to the shape of the dose-response model based on the strongest evidence (male lymphoid tumors) reported by Steenland et al. 2004 mortality study of sterilant workers and which comprise 85% of the IRIS and OEHHA unit risk estimates. UCC study results, therefore, should not be dismissed or treated as critically weak (Figure 3) because there are no females in the cohort.

OEHHA criticizes Valdez-Flores 2025 stating “few individual models assessed”. The authors, however, conducted continuous analyses with both cumulative and log cumulative exposure metrics and numerous lagged analyses, using the CPH model. They also conducted categorical analyses to complete a comparison with Steenland et al 2004. Additional models or analyses were not indicated as the continuous slopes were all negative and since lagged analyses made no difference in the results.

It is unclear what demographic comparisons are missing as the UCC cohort is all male. Although race data was incomplete in the KV study as described Rinsky et al., these authors noted the high percentage of whites (96%) among those whose race was known. A critically poor classification for UCC EtO workers along this criterion or for not adjusting for race is not justified.

OEHHA speculates it is “unclear if lagged analyses could be impacted by outlying values” for the UCC study but states “no concerns regarding… outlying values” for the NIOSH study. Classifying the UCC study as critically poor in this characteristic is not justified.

Potential confounding by other chemicals in the workplace provides another OEHHA justification to dismiss the results from the UCC study. Greenberg et al. in fact identified an increased risk of leukemia and pancreatic cancer among workers producing ethylene chlorohydrin, although EtO was used only sporadically in this department. Teta et al. (1993) described this in detail and removed the workers in this department from the cohort to eliminate confounding. OEHHA states, however, that other “carcinogenic exposures in the low EtO exposure group could bias EtO RRs towards the null”. This is implausible to

suggest given that increased risks for pancreatic cancer and leukemia were not seen in any other of the numerous high or low EtO exposure departments (Valdez-Flores et al 2025).

Uncertainty related to exposure assessment is a third area OEHHA ranks the UCC study with poor quality ratings, while giving higher ratings to the NIOSH exposure assessment. OEHHA fails to recognize the serious uncertainties in the NIOSH exposure assessment covering 6 decades that renders its use questionable for risk assessment: 1) the inappropriate application of the NIOSH regression model, based on post 1978 predictors of exposure, to estimate exposures in the time period prior to 1978, when processes and workplace conditions were vastly different; 2) holding the effect of calendar year constant at 1978, when applying the model to the earlier years, despite the fact that the 1940s to 1970s represent a period of significant improvement in industrial hygiene workplace standards; 3) ability to validate the regression model only for the more recent time period due to the absence of any actual exposure data prior to 1976 and very little prior to 1978) implausible prediction of lower estimated exposures in earlier decades (1940s to 1960s) compared with later periods (1960s to 1980s) (Bogan et al. 2019) and 5) the fact that the vast majority of the exposures for the cancer cases of interest occurred prior to the period when EtO measurements were available on which to validate any NIOSH estimates of occupational exposures. OEHHA claims without evidence that any exposure error in the NIOSH study would be non-differential and therefore the true risk would be greater. Stayner, Steenland and colleagues 2003 point out that the bias in the slope may be in either direction. Furthermore, they state, “Regarding exposure mismeasurement in analyses using continuous exposure variables, even if the error is random, either the attenuation or enhancement of an exposure-response may occur.”

OEHHA’s evaluation of the UCC exposure assessment (referred incorrectly as “modeled exposure” in table D1) states there was limited air monitoring data, even though 1) a large-scale monitoring program was in place since 1976 at the plants in the study and 2) data from an extensive industrial hygiene monitoring survey at a comparable UCC plant representative of the period 1957-1973 was available. Use of the comparable plant has been criticized by OEHHA but NIOSH’s use of exposure data from 36 sterilant plants for their regression model, when only 14 0f these plants were included in the study, is nowhere mentioned. Furthermore, OEHHA dismisses the medical record review by Greenberg et al. 1990 that validated the exposure classifications for 40 EtO using or producing departments. From the start of follow up 1940 to 1956, reasonable estimates were made in the absence of actual data for this 16-year period, in contrast to the nearly 40-year period of uncertainty (1940-1978) in the NIOSH study. The uncertainties associated with the UCC exposure assessment do not justify dismissing the findings of the UCC study, as an independent reality check on the NIOSH exposure-response assessment.

A steep exposure-response curve at low exposures as presented by EPA IRIS 2016 and the OEHHA draft indicates a potent carcinogen. Neither the UCC study nor the large database of other cohort studies that have examined EtO carcinogenicity suggest in any way a potent carcinogen (Marsh**, 2019**). IARC and EPA agree that the human evidence of EtO carcinogenicity is limited, a conclusion based on the results of numerous epidemiology studies over an extensive time period (IARC, EPA). Potent carcinogens are readily recognized in occupational epidemiology studies (e.g., vinyl chloride, benzene), even without detailed individual exposure estimates, provided study subjects have been exposed and many during time periods of heavy exposures. This is the case for the UCC EtO cohort followed repeatedly over more than 7 decades, starting at the infancy of the production and use of EtO in the chemical industry. Even without extensive exposure-response modeling, as has been conducted by Valdez-Flores et al. 2025, the nonpositive mortality results for LH and lymphoid cancers, using both external and internal analyses should raise considerable uncertainty about the use of a supra-linear model for quantitative risk assessment (QRA).

The OEHHA ranking of the quality characteristics of the NIOSH and UCC epidemiology studies shows distinct and erroneous bias towards favoring the NIOSH study as the most suitable study for supporting EO risk assessment.

The NIOSH study clearly has the advantage of including both males and females, and having the largest total cohort size compared to the UCC study. Since the UCC study only included males, it only can inform quantitative risk assessments for male lymphoid cancers. However, this is important because lymphoid cancer from the NIOSH study has the greatest impact on the EPA/OEHHA IUR (85% of the IUR is based on NIOSH lymphoid cancers). In addition, the original conclusion of the published NIOSH study is that there was an increase in males and not female lymphoid cancers (Steenland et al. 2004). Thus, the UCC study provides critical weight-of-evidence (WOE) informing the selection of dose-response model that OEHHA should apply to the NIOSH lymphoid cancers.

Two of the most egregious examples of OEHHA's bias in ratings merit discussion because they have major implications on the reliability of the NIOSH dataset for quantitative risk assessment purposes.

Exposure rating: OEHHA's rating of "critical" (lowest ranking criteria) for exposures in the UCC study vs "high" for the NIOSH study is incorrect. If OEHHA keeps the rating as "high" for the NIOSH study, then it should promote the ranking for UCC to "high". Alternatively, if one considers the years for which there was actual data informing the exposure estimates, then the UCC study should be ranked as Adequate-High, while the NIOSH study should be downgraded to "critical".

The NIOSH study relies on an exposure regression model which was unvalidated for the vast majority of exposure period. While the NIOSH exposure model was very well validated for the short 10 year-period after 1976, the exposure model was completely unvalidated for the earlier 40 years of exposure. NIOSH had virtually no exposure data during the majority of exposure periods for both lymphoid and breast cancer. In contrast, the UCC study has superior exposure data because it has contemporary exposure measures from the 1950's to 1980's in the same or identical plants.

NIOSH then altered the well-validated model based on an incorrect assumption that there were no worker industrial hygiene or equipment changes that might impact exposures for 40 year period when NIOSH had zero exposure data. The available published literature provides ample evidence that this assumption is wrong. The NIOSH exposure model has very few variables, with calendar year and volume of sterilizer driving the exposure model outcome. Thus, setting the key variable calendar year to 1978 levels results in the volume of sterilizer chambers driving the exposure estimates for the 40 year period before 1978. This is a fact that EPA IRIS (2016) agreed to in their response to EPA SAB query on this unusual pattern, where EPA stated, "increased sterilizer volume generally resulted in higher predicted average exposures until the late 1970's" (EPA IRIS p.I-26, 27).

Inclusion Rates: The NIOSH study should be downgraded to low because one of two key endpoints used for quantitative risk assessment is breast cancer incidence which was under-ascertained. The rationale that it was fully ascertained because the **interviewed** subset of those found were fully ascertained does not take into account that there were many missing cases unaccounted for to get to the subset of cases that OEHHA claims is fully ascertained. If OEHHA continues to rely on breast cancer incidence data, instead of the fully ascertained breast cancer mortality data, then the rating must be reflected in the rating scale.

Conflict of Interest: OEHHA's treatment of the Union Carbide cohort updates reflects an apparent bias against industry-funded research rather than an objective application of the National Toxicology Program's Office of Health Assessment and Translation (OHAT) framework (NTP, 2019), which OEHHA cites as the basis for its review. OHAT does not consider financial conflicts of interest to be an automatic basis for a critical risk-of-bias rating. Instead, it calls for an evaluation of whether a potential conflict influenced study design, conduct, analysis, interpretation, or reporting. OEHHA identifies no evidence that industry funding affected the methods or findings of the Valdez-Flores analyses. As a result, OEHHA's "critical" rating appears to be based primarily on sponsorship rather than demonstrated methodological deficiencies, which is inconsistent with the principles of the OHAT systematic review framework. More generally, the relevant question is not whether a

potential conflict exists, but whether there is evidence that it affected specific methodological decisions or reporting of study findings. OEHHA identifies no such evidence for the UCC cohort updates.

Sweden Study (Mikoczy study): Our detailed comments focus on comparison of the NIOSH and UCC studies because they are the most informative for quantitative risk assessment purposes. However, OEHHA should take into consideration that the Mikoczy (2011) study's referent group exhibited an anomalous breast cancer rate at just 50% of the general population—a deficit far too extreme to be explained by the Healthy Worker Effect—limiting it's use to inform quantitative risk assessments. Specifically, Mikoczy et al. cannot be used in support of EPA IRIS selection of exposure response models with steep initial slope.

Detailed Comments on OEHHA Study Selection

OEHHA's Figure 3 Rating

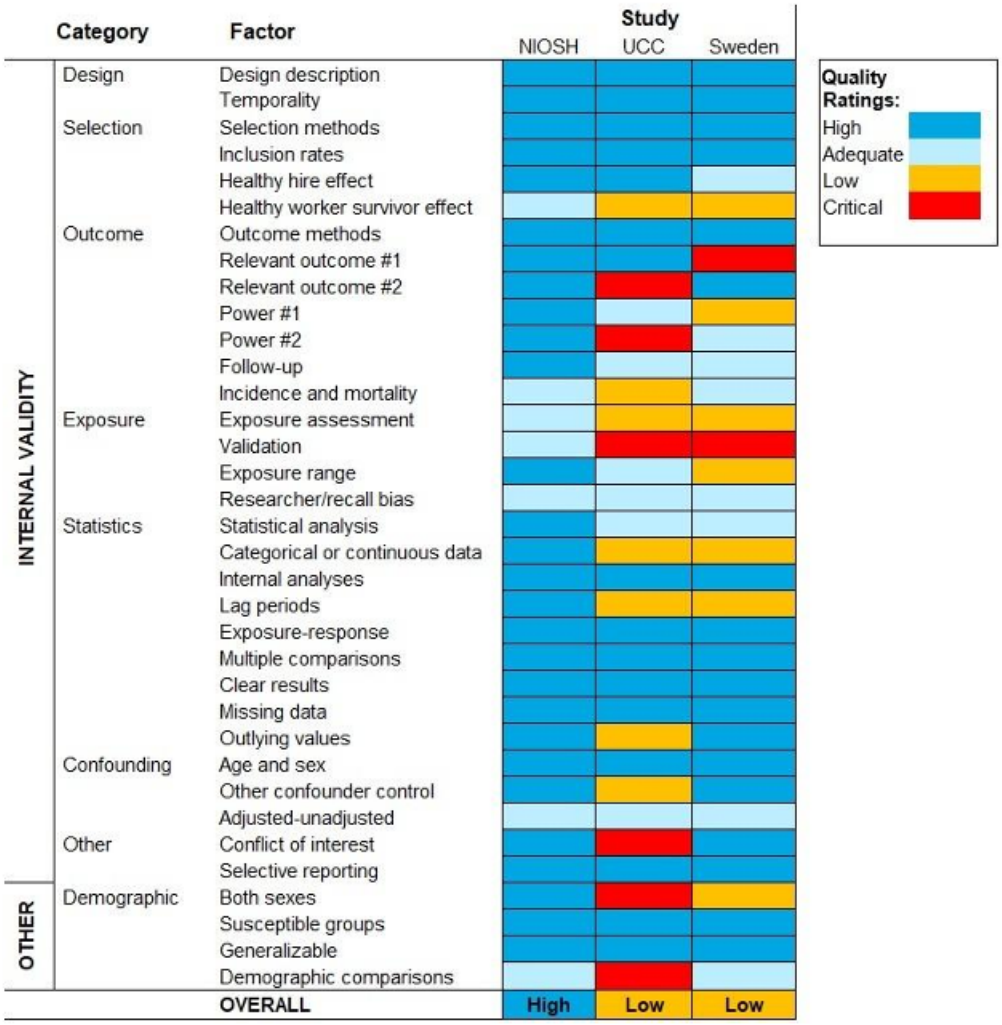
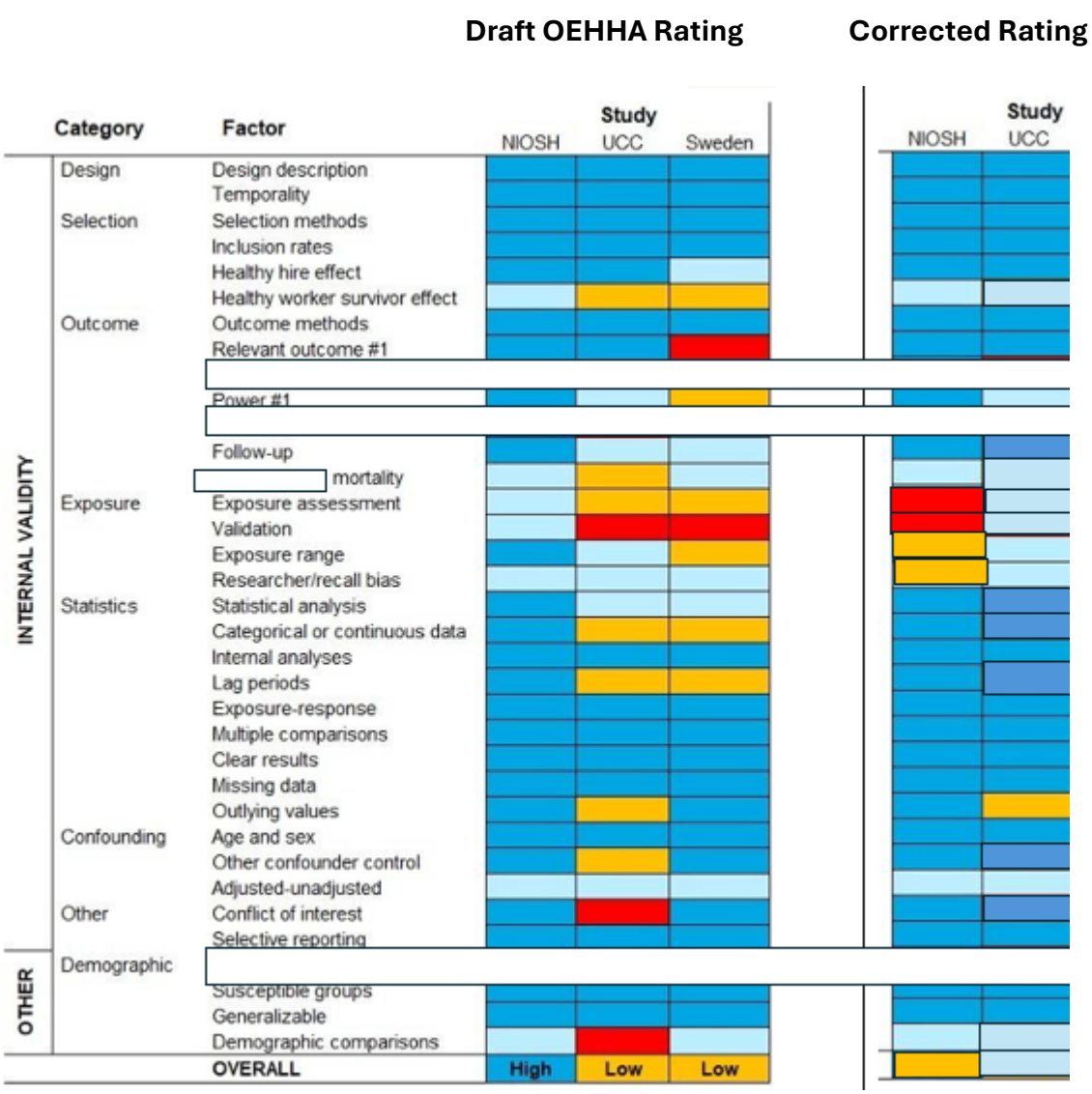


Figure 3. Detailed evaluations of the epidemiologic studies considered for exposure-response and IUR calculations.

CORRECTION OF RATINGS FOR MALE LYMPHOID MORTALITY FOR PURPOSES OF INFORMING THE QUANTITATIVE RISK ASSESSMENT

Steenland et al. (2004) only reported lymphoid mortality findings in males only. Thus, a comparison of quality of study for quantitative risk assessment purposes for male lymphoid mortality is especially relevant for determining what dose-response model to apply to the NIOSH lymphoid mortality data.

See text for detailed justification for corrected ratings for NIOSH and UCC studies as they pertain to male lymphoid mortality. **The low overall rating for NIOSH is due to the severe exposure uncertainty for more than 80% of the lymphoid mortality cases described in detail in our comments above.**



OEHHA uses a qualitative analysis to determine the most "useful and valid" of three studies (NIOSH, UCC, or Sweden) for exposure-response modeling and unit risk calculation of EtO. OEHHA assigned one of four qualitative ratings (high, adequate, low, or critical) to each of 35 factors they considered relevant. Figure 3 of the OEHHA document (replicated above) lists the results of such analysis and Attachments D and E discuss the rationale for every rating assigned.

OEHHA concludes that lymphoid mortality and breast cancer incidence studies from the NIOSH cohort were of high quality and "the most sensitive and valid epidemiological studies." In contrast the UCC study is assigned an overall "low" quality based on 7 "low" and 6 "critical" quality assessment rankings while no low or critical ranks were identified for the NIOSH study.

OEHHA presents a lengthy exposition of the NIOSH study, very similar to IRIS 2016 discussion. Attachments D and E have extensive narratives justifying the ratings assigned for each factor evaluated in the comparative ranking of each of the studies. However, discussions of the comparative quality factor considerations of the NIOSH and particularly the UCC study require clarification in that the described rationales otherwise appear preferentially bias to quality rankings favoring the NIOSH study. Comments on the OEHHA quality rationales are as follows:

1. **The rationale for the HWSE ratings was based on the following (Attachment E).**

Criteria	Factors to consider	High	Adequate	Low	Critical
			where the HHE is likely to be similar		
Healthy worker survivor effect (HWSE)	• Was the appropriate evidence provided to evaluate this potential bias? • Did this potential bias have an impact on study findings? • This criterion applies to occupational studies that use the general population as the "unexposed" comparison group	• A full assessment was done, and it was found that the HWSE is unlikely to have an impact on study conclusions • A full assessment was not done but there is some evidence that the HWSE did not impact study results	• A full assessment was done, and it was found that the HWSE likely had only a minor impact on study results • Evidence suggests that there was no discernable HWSE, but the assessment had some minor methodological weaknesses	• A full assessment was done, and it was found that the HWSE likely had a moderate impact on study results • An incomplete or no assessment but some other evidence (e.g., results from similar cohorts) that the HWSE had a moderate impact on study results • There were some moderate methodological weaknesses in the assessment	• Incomplete or no assessment but some other evidence that the HWSE had an important impact on study conclusions

Ratings based on this table for the HWSE for the NIOSH study ("Adequate") should be lower than the UCC study ("Low") or at least be the same for all studies.

The NIOSH cohort is the only study that has been suspected of having HWSE and has not been adjusted for that bias (Kelly-Reif at al. 2025 and Picciotto et al. 2026), and therefore falls into the "incomplete or no assessment" category for critical. If OEHHA relies on the recent EPA Kelly-Reif et al. and Picciotto et al. results, then the rating of the HWSE for the NIOSH study should be "Critical."

The UCC study, on the other hand, did not indicate any bias due to the HWSE. Valdez-Flores et al. 2025 analyses indicated that "no decreased risks with additional extensive follow-up of the UCC cohort were seen, as might be expected with a Healthy Worker Survivor Effect bias." The lack of support for the presence of a HWSE should warrant at least an "Adequate" rating for the UCC study.

2. **The criteria for "Relevant Outcome #1" (lymphoid cancer), "Relevant Outcome #2" (breast cancer), "Power #1" (lymphoid cancer), "Power #2" (breast cancer), and "Both sexes" are correlated and the OEHHA ratings assigned study similar ratings for these related factors.**

Criteria	Factors to consider	High	Adequate	Low	Critical
Relevant outcome #1	• Did the study include the primary outcome of interest?	• Includes the primary outcome of interest	• Does not include the primary outcome of interest but includes an outcome that is strongly related to or correlated with the primary outcome of interest • Not the primary outcome but the outcome assessed is in the direct causal chain	• Does not include the primary outcome of interest but may include a metric with some moderate correlation with the primary outcome of interest	• Does not include the primary outcome of interest or a related outcome • Not in the causal chain
Relevant outcome #2	• Same as above • Only used if there is more than one primary outcome of interest	• Same as above	• Same as above	• Same as above	• Same as above

For example, when a study did not include male and female workers OEHHA ratings were automatically "Critical" for "Both sexes", "Relevant Outcome #2" (breast cancer), and "Power #2" (breast cancer). That is, OEHHA penalized (or rewarded) the study three times for the same reason.

Criteria	Factors to consider	High	Adequate	Low	Critical
Power #1	• Did the study have adequate statistical power for the primary analyses of interest? • Review the precision of the main effects and any important subgroup or stratified analyses	• The study had good statistical power (generally ≥80%) for the primary analyses of interest (e.g., lag periods, sex specific analyses)	• The study had good statistical power for the primary analyses of interest but with some minor weaknesses (e.g., the findings were borderline precise) • Good statistical power but with some minor weaknesses in evaluating power	• The study had only moderate statistical power (e.g., <60-70%) • The study had adequate statistical power for some related analyses but not for the primary analysis of interest • The study had good statistical power for the study population overall but not for important subgroups or other key analyses • Limited information leads to moderate difficulties in evaluating power	• The study had inadequate statistical power for all important analyses • No or very limited information is available for evaluating statistical power
Power #2	• Same as above • Only used if there is more than one primary outcome of interest	• Same as above	• Same as above	• Same as above	• Same as above

Criteria	Factors to consider	High	Adequate	Low	Critical
Both sexes	• Are both sexes included in the analyses? • Are separate analyses done by sex?	• Both sexes were included and full results by sex are provided	• Separate analyses not provided by sex but there is evidence (empiric, mechanistic, or strong theoretical) that sex differences are unlikely to have an important impact	• Separate analyses not provided by sex and there is some evidence or theoretical concern that sex differences may have had a moderate impact on	• There is some evidence that failure to include or fully evaluate both sexes likely had an important effect on study findings, relevancy, or interpretation
	• This criterion does not apply to sex specific outcomes		on findings or relevancy • Analyses by sex are provided although with some minor limitations	study findings or relevancy • Analyses by sex are provided although with some moderate limitations	(Steenland et al., 2004)

Thus, because the NIOSH cohort includes male and female workers and breast cancer was analyzed the study received a rating of "High" for the factors "Both sexes", "Relevant outcome #2" (breast cancer), and "Power #2" (breast cancer). In contrast, the UCC cohort, which includes only male workers, was not analyzed for breast cancer (for obvious reasons) and the study was penalized with a rating of "Critical" for the factors "Both sexes", "Relevant outcome #2" (breast cancer), and "Power #2" (breast cancer). The multiplicity of

ratings to highly correlated factors give the impression of a bigger advantage of the NIOSH study over the UCC study.

The “Adequate” rating for “Power #1” (lymphoid cancer) for the UCC study seems **due to** models fit to the UCC cohort **not adjusted** for sex. Models were not adjusted for sex because there were all male workers. Models fit to the UCC study were also fit to lagged exposures. Valdez-Flores et al. 2025 state “Exposure-response models were also fit assuming non-zero lags and lags of 5–30 years in increments of 5 years with cumulative exposure to EO. Since there were no statistically significant differences in results based on lag, zero lag was used in the final main analyses (see Table S5 for model fits using exposures with non-zero lags).” Given this evidence, the UCC study should receive a rating of “High” for “Power #1” (lymphoid cancer).

3. **The subjective ratings for “Follow-up” are “High” for the NIOSH study and “Adequate” for the UCC study.**

Criteria	Factors to consider	High	Adequate	Low	Critical
Follow-up	• Was the period between exposure assessment and outcome measurement adequate? • For cancer studies this is generally several years or more • This criterion rates whether the underlying study design included an appropriate follow-up period. Bias due to low or differential follow-up rates is rated in the “Participation” criterion. Failure to evaluate appropriate lag periods in the statistical analysis is rated in the “Lag period” criterion	• There is good evidence or reason to believe that the follow-up period was appropriate	• There is evidence or reason to believe that the follow-up period was not too long or too short, but the weaknesses or concerns are minor • Some minor gaps in the description of the follow-up period • The follow-up period likely covers most of the appropriate follow-up period	• There is evidence or reason to believe that the follow-up period was not too long or too short and this led to moderate weaknesses or concerns • Some moderate gaps in the description of the follow-up period • The follow-up period likely covers some but not most of the appropriate follow-up period	• There is evidence that the follow-up period was likely too long or too short and this led to important bias • Critical elements are missing from the description of the follow-up period

OEHHA apparently based the rating of “Adequate” to the UCC study on the quote from Valdez-Flores et al. 2025 “longer **post-exposure** follow-up could potentially make it more difficult to distinguish background disease from exposure-related disease". The publication showed the results of all previous UCC updates, and the quoted text was only a cautionary statement to give more weight to internal comparisons as opposed to comparisons to general population background rates. OEHHA chose to ignore the fact that the length of follow-up of the UCC study (>40 years) is much longer than the length of follow-up of the NIOSH study (26.8 years) used in exposure-response modeling. In summary, the rating for

the “Follow-up” period for the UCC study should be higher than the rating for the NIOSH study.

4. **For the category exposure assessment, OEHHA provided ratings for four factors: “Exposure Assessment”, “Validation”, “Exposure Range”, and “Researcher/Recall Bias”.**

STUDY	NIOSH	UCC	Sweden
Exposure Range	Cumulative exposure: median = 2044 ppm-days, mean = 9818 ppm-days, 75th percentile = 8486 ppm-days; average exposure intensity = 3–5 ppm; overall: there appears to be a good range in cumulative exposure and exposure intensity	Cumulative exposure: median = 7480 ppm-days, mean = 24,455 ppm-days, 75th percentile = 28,224 ppm-days; information on average exposure intensity not provided; overall: there appears to be a good range in cumulative exposure although details on exposure intensity are lacking	Cumulative exposure: median = 0.13 ppm-years (47.5 ppm-days), mean = 2.92 ppm-years, 75th percentile = 0.21 ppm-years; highest: 40–75 ppm although these may have been rare and short-term; average exposure intensity: no information; overall: exposures appear to be relatively low compared to other studies
Exposure Assessment	Adequate	Low	Low
Validation	Adequate	Critical	Critical
Exposure Range	High	Adequate	Low
Researcher/Recall Bias	Adequate	Adequate	Adequate

The ratings for these four factors are summarized by OEHHA in Attachment D as follows:

OEHHA details the bases for the rankings. First, there is the following list of characteristics considered in the exposure assessment.

EXPOSURE:	
Exposure: description	• Describe how the exposure was assessed • Describe the timing, frequency, number of measurements, detection rates, and duration of the exposure assessment • Is the exposure likely to be reasonably consistent or is it highly variable over time (e.g., long or short half-life)? • Was exposure assessed similarly in people with and without the outcome of interest and in people with high and low exposure levels? • Were the researchers blinded, or was exposure measured in a way that researcher, interviewer, or subject recall bias was unlikely? • Was exposure assessed in the most relevant period? • Was any exposure misclassification likely to be differential or non-differential? • What was the possibility of significant exposure outside the time periods that exposure was assessed?
Validation	• Was the exposure assessment process validated? • If so, describe the validation process and its results
Exposure range	• Provide information on exposure means, medians, and percentiles • Describe how the study exposures differ from general population exposures

OEHHA then defined the ratings for each factor and each study as follows.

a) Exposure Assessment. The NIOSH study received an “Adequate” rating while the UCC study was assigned a “Low” rating. The UCC study should be assigned a rating of “Adequate”, and the NIOSH study rating should be ranked as “critical”.

Criteria	Factors to consider	High	Adequate	Low	Critical
Exposure assessment	• Was there a complete and clear description of how the exposure was assessed? • Is this a well-established, widely accepted or well validated exposure metric? • Does the exposure metric reliably distinguish between people with higher and lower exposure levels considering intensity, frequency and duration of exposure (consider the sensitivity and specify of the exposure metric)? • Was the exposure measured independently of the outcome? • Was exposure assessed in the most relevant period (consider life stage and half-life)? • Does the degree of exposure misclassification vary by the level of the outcome? • Weaknesses specifically related to blinding, researcher, interviewer, or recall bias are rated under “Researcher/Recall bias”	• A direct, well-described, well-established, widely accepted and common method was used • Modeled exposure with appropriate validation data showing good correlations • Exposure measured using the same methods in all participants (i.e., independent of the outcome) • Exposure misclassification is likely to be minimal or not discernable if present • Exposure was assessed in the most relevant period	• Modeled exposure without validation data but using direct, clear, thorough, and well justified and sound methods • Some gaps in the description of the exposure assessment but these are minor • Exposure misclassification is likely to be relatively minor • Minor differences in the validity or accuracy of exposure assessment across outcome groups • Somewhat less direct exposure measurements used (e.g., wide area measurements rather than personal sampling or biologic matrices) • The exposure covers a significant part but not all of the relevant exposure period • Occupational studies with some minor missing information on work histories	• Evidence suggests the method used to measure exposure is only moderately related to the relevant exposure • Exposure assessment methods are somewhat unclear or incomplete • Exposure misclassification likely to lead to moderate bias • Moderate differences in the validity or accuracy of exposure assessment across outcome groups • Some moderate concerns about whether exposure was assessed in the most relevant period • The exposure covers only a portion of the relevant exposure period • Some concern that exposure measurement was not fully independent of outcome status • Occupational studies with some moderate gaps in work histories	• Evidence shows that the method used to measure exposure is not related to the exposure of interest • Exposure misclassification is likely to lead to important bias • Large differences in the validity or accuracy of exposure assessment across outcome groups • Modeled exposure with unclear or poorly rationalized methods and no or poor validation data • Exposure was assessed in a period that is likely not related to exposure in the most relevant period • Critical information about exposure assessment is not provided • Exposure measurement was not independent of outcome status • Occupational studies with important gaps in work histories

It is not clear how OEHHA decided such ratings, with a current appearance of bias in favor of the NIOSH exposure estimates. The NIOSH exposure estimates were based on a model fit to some measurements after 1976, with calendar year being the main predictor of exposure concentration. The value of calendar year to estimate exposure concentrations before 1976 was fixed to 197**8**, resulting in severe underestimation of exposures for most of the **earlier** follow-up period (Bogen et al. 2019). NIOSH’s exposure model implausibly predicts lower estimated exposures in earlier decades (1940s–1960s) compared with later periods (1960s–1980s) (Bogen et al. 2019). The error in the NIOSH exposure model is that NIOSH assumed that industrial hygiene and workplace standards for equipment in 1978 apply to all years previously. Yet, the 1940s to 1970s represent a period of significant

improvement in industrial hygiene workplace standards with associated significant reductions in chemical exposures during this period. A new analysis of previously published NIOSH data presented earlier in this document shows the substantial impact of USEPA's exposure error. The vast majority of the exposures for lymphoid and breast cancer mortality cases occurred prior to the period when EtO measures were available on which to validate any NIOSH estimates of occupational EtO exposures. Critically, for the 13 cases accounting for low exposure group below the "knot" low exposure group, 90% of the cumulative exposures were modeled rather than based on any measured values of EtO exposure

In contrast, the UCC study used monitoring measures of EtO concentrations. The UCC historical concentrations were in line with regulatory exposure limits while NIOSH exposure estimates before 197**8** were well below regulatory occupational exposures (Bogen et al.2019). Including these facts in the validation rating should rate the UCC study exposure validation higher than the NIOSH study. Furthermore, the NIOSH exposure concentrations that were only available after 197**8** were used to model exposures for the whole exposure period, do not exist so that the NIOSH model-generated exposures cannot be validated. Swaen et al. 2009, on the other hand, thoroughly describes and substantiates EtO exposure concentrations in the UCC study.

b) Validation. The NIOSH study was rated "Adequate" while the UCC was rated "Critical".

Criteria	Factors to consider	High	Adequate	Low	Critical
Validation	• What is the quality and thoroughness of the validation procedures? • This criterion focuses on the quality of the validation procedures. Ratings related to actual validation results are considered in the "Exposure assessment method" criteria • This criterion is only used when exposure models or similar procedures are used	• Clear and complete description of validation procedures and results • Sound validation procedures • Thorough and clear validation results • Validation took place in a random subgroup of the study population • Validation took place in a population that is directly comparable to the study population	• Minor weaknesses in the validation procedures • Missing some validation results but these are minor • Minor lack of clarity in the description of the validation procedures or results • Minor differences between the study population and the validation study population	• Moderate weaknesses in the validation procedures • Missing some but not all relatively important validation results • Moderate lack of clarity in the description of the validation procedures or results • Moderate differences between the study population and the validation study population	• Unusual exposure assessment method with no validation results or clear underlying rationale • Important weaknesses in the validation procedures • Key elements are missing from the description of the validation procedures • The study population and the validation study population are not comparable on some critical elements

As described above, the NIOSH validation was limited only to exposures measured after 1976 and mostly after 1978, and 80% of the total cumulative exposures of the 13 cases driving the low-exposure dose-response characterization had no available exposure data. Including these facts in the validation rating should rate the UCC study exposure validation higher than the NIOSH study which is critically deficient in its exposure validation of EO low

exposure cases.. Furthermore, the NIOSH exposure concentrations that were only available after 1976 were used to model exposures for the whole exposure period, do not exist so that the NIOSH model-generated exposures cannot be validated. Swaen et al. 2009, on the other hand, thoroughly describes and substantiates EtO exposure concentrations in the UCC study. Thus, the NIOSH exposure validation should be ranked low or critical with respect to its quality for reliability estimating EO exposures associated with primary cancer cases driving the low-exposure response characterization.

c) Exposure Range. The NIOSH study received "High" rating for exposure range, while the UCC study was assigned an "Adequate" rating. Table 1 in Valdez-Flores et al. 2010 (reproduced here for convenience) lists the distribution of cumulative exposure in ppm-days (the exposure metric used for exposure-response models). Clearly, the range of the cumulative exposure experienced by the UCC workers is not less than the range of the cumulative exposure experienced by the NIOSH workers. According to the OEHHA criteria listed, both studies should receive the same "high" ratings for the "Exposure Range" factor. However, the NIOSH study is downgraded due to unreliability of exposures

Table 1
Distribution of the cumulative EO exposures (ppm-days) at the end of follow up among the workers.

Percentiles		NIOSH Females	NIOSH Males	UCC Males
100.0	Maximum	293,538	642,925	422,933
99.5		107,249	239,777	250,489
97.5		45,363	102,167	135,224
90.0		19,420	38,687	72,643
75.0	Third quartile	7178	12,160	28,119
50.0	Median	1767	2807	7364
25.0	First quartile	520	668	1876
10.0		166	134	391
2.5		45	31	68
0.5		17	14	14
0.0	Minimum	5.1	4.9	2.1

Criteria	Factors to consider	High	Adequate	Low	Critical
Exposure range	• Did the study have a wide range of exposure and include some people with high exposures? • Considerations of statistical power here should be primarily based on the exposure levels taking into account sample size, confounding control, and the expected effect size at these exposures	• The study involves a very broad range or good contrast of exposure • Examples might include median levels >5-fold higher than NHANES, an occupational study, or an area with known high exposures • There is good evidence that the potency of the agent is such or the study is large enough that detectable effects are expected despite a limited exposure range	• Moderate exposure range although statistical power to detect an effect is likely to be adequate • Some details on the exposure range are lacking but these are minor • There is some less strong evidence that the potency of the agent is such or the study is large enough that detectable effects are expected despite a limited exposure range	• Exposure range is only somewhat higher than general population levels and this is likely to lead to a moderate deficiency in study power • The potency of the agent is such or the study is large enough that borderline detectable effects may be seen despite a limited exposure range • Only a very small proportion of the participants are highly exposed • A few but not all key details on the exposure range are lacking	• Study exposures were likely too low to detect an effect even given a large sample size • Critical information on exposure levels is lacking

5. **Under Statistics the factors evaluated are: "Statistical Analysis", "Categorical or Continuous Data", "Internal Analyses", "Lag Periods", "Demographic Comparisons", "Exposure-response", "Multiple Comparisons", "Clear Results", 'missing Data", and "Outlying Values". The ratings of the factors in boldface are discussed here in detail.**

In Attachment E, OEHHA details the bases for their ratings. First, there is the following list of characteristics considered in the statistics.

STATISTICS:	
Statistics: description	• Describe the major aspects of the statistical analysis with a major focus on the analyses involving the most relevant results • Were these clearly presented? • Were demographic comparisons made across exposure and outcome groups? • How were exposure-response, confounders, effect modifiers, mediators, outliers, multiple comparisons, and missing data addressed? • Was there over-matching on exposure? • How were the exposure categories appropriately defined? • Were data distributions considered appropriately and are all statistical assumptions reasonable? • Were outlying values, missing data, and multiple comparisons issues evaluated and controlled for?
Lag periods	• Describe any known information about the possible latency period • Was this latency addressed appropriately and thoroughly in the statistical analyses and results?

a) Statistical Analysis. The NIOSH study received a "High" rating for statistical analysis that more appropriately should be "critical" while the UCC received just an "Adequate" rating.

STUDY	NIOSH	UCC	Sweden
Statistics			
Statistics: Description	Multiple models assessed; categorical and continuous results presented; internal analyses presented; some demographic comparisons provided for breast cancer (Steenland, 2003; Table 2), limited demographic data for lymphoid cancer (US EPA, 2016; Tables D-55 to 60); exposure-response fully assessed; no concerns regarding missing data or outlying values	Few individual models assessed; categorical results not provided for lagged analyses; internal analyses presented; demographic comparisons by exposure or case status not provided in Valdez. (2025); no concerns regarding missing data; unclear if lagged analyses could be impacted by outlying values	Few models assessed; no continuous data analyses but impacts likely only minor or moderate; internal analyses performed; limited demographic comparisons by exposure levels provided in their Table 2 (age, gender, employment history); no concerns regarding missing data or outlying values
Lag Periods	Multiple lag periods assessed for relevant analyses; lag periods applied to categorical and continuous data analyses	Multiple lag periods assessed but results only presented as maximum likelihood estimates and associated statistics	15 year lag assessed but only applied to external comparison results; unclear if there is sufficient statistical power for lagged exposure-response analyses
Statistical Analysis	High	Adequate	Adequate
Categorical or Continuous Data	High	Low	Low
Internal Analyses	High	High	High
Lag Periods	High	Low	Low
Demographic Comparisons	Adequate	Critical	Adequate
Exposure-response	High	High	High
Multiple Comparisons	High	High	High
Clear Results	High	High	High
Missing Data	High	High	High
Outlying Values	High	Low	High

Judging from the list of references, it seems that OEHHA is not aware of the Valdez-Flores et al. 2010 publication that analyzed both the NIOSH and UCC studies separately and

combined. **The** Valdez-Flores et al. 2010 analyses are in no way inferior. For example, Steenland et al. choose a lag-period for exposure metric without evaluating whether statistics support the inclusion of that variable into the model. Valdez-Flores et al., on the contrary, evaluate different lag periods and concludes that there is no significant statistical evidence for the inclusion of a non-zero lag period in the modeling. Valdez-Flores at al. 2025 used similar proven statistical analyses to evaluate the statistical relevance of the lag period.

Importantly, and similar to Steenland et al. 2004, there is substantial evidence that IRIS 2016 did not apply correct statistics for model selection, resulting in its selection of the IRIS 2-piece spline model as compared to the statistically correct and more parsimonious Cox proportionate hazard (CPH) modeling approach applied by TCEQ to the same Steenland data (see section x). In contrast, the UCC study employed statistically rigorous analyses should not receive a lower rating than the baseless choices for the NIOSH statistical analyses in IRIS 2016 that should otherwise result in a "critical" quality ranking. If

Criteria	Factors to consider	High	Adequate	Low	Critical
Statistical analysis	• Were the methods used in the statistical analyses clearly explained, appropriately applied, thorough, and unlikely to introduce bias? • Were data distributions appropriately addressed (e.g., log transformations)? • Appropriate and justifiable category cut-off points • Consider the possibility of over-adjustment or co-variance	• Appropriate and thorough statistical analyses without significant weaknesses	• Appropriate statistical analyses but with some minor missing elements, minor lack of clarity, or minor weaknesses unlikely to have a moderate or important impact on study conclusions • Minor weaknesses or limitations in the consideration of data distributions	• Moderately important missing or unclear elements or weaknesses that are unlikely to impact the direction of the effect but may moderately impact the effect size • Moderate weaknesses in the consideration of data distributions	• Major weaknesses or missing elements that are likely to cause important bias or to severely limit the interpretability of study findings • Incorrect analyses in the consideration of data distributions likely led to important bias

NIOSH is kept at high, so should UCC.

b) Categorical or Continuous Data.

First, it is not "Categorical or Continuous **Data**" but rather "Categorical or Continuous **Models**" because both are estimates based on individual longitudinal observations.

Criteria	Factors to consider	High	Adequate	Low	Critical
Categorical or continuous data	• Were results presented using both categorical and continuous data analyses?	• Thorough categorical and continuous analyses were presented for all important results • One type of analysis is not appropriate for the type of data or associations being assessed	• Both categorical and continuous analyses were presented although with some minor weaknesses or minor missing elements (e.g., only two exposure categories were used; continuous analyses were somewhat unclear or limited; both were not presented for a few relevant results) • Only one type was reported but this led to only minor weaknesses regarding study interpretation, validity, or completeness	• Both categorical and continuous analyses were presented although with a moderate weakness in one type (e.g., the continuous data model may have been a poor fit to the data; incomplete results in one type of analysis; category cut-offs poorly rationalized) • Only one type was reported and this led to moderate weaknesses regarding study interpretation, validity, or completeness	• Only one type of analysis was presented and this had important impacts on study interpretation, validity, or completeness

OEHHA assigns a "High" rating that is more appropriately "Low" or "Critical" for the NIOSH study, and a "Low" rating for the UCC study that should be elevated to "High" or possibly "Adequate". If NIOSH is kept at a high rating, then UCC should be elevated to "High" as well

These ratings are not surprising given the misunderstanding of EPA and OEHHA regarding categorical rate ratios (RR). Both agencies incorrectly believe that the categorical RRs are the data and the gold standard to compare to continuous models. This incorrect interpretation of categorical RRs led OEHHA to visually judge the models fit to individual data by how well the continuous models conform to a few categorical estimates of the RRs. In addition, OEHHA (as well as EPA) give more weight to visual fit than to statistical measures of goodness of fit. The erroneous interpretation and use of categorical RRs to visually judge model fit has been pointed out by IRIS 2016, TCEQ 2020, and Valdez-Flores and Sielken 2013, and detailed in our comments above. Strangely, although IRIS 2016 recognizes the issue in a footnote to their comparison table, EPA still uses visual comparisons to judge goodness of model fit. Here, again, OEHHA fails to include the results in Table S.9 of the supplementary material of Valdez-Flores et al. 2010 that reports categorical (as well as continuous models) for all combinations of cancer endpoint (12), study (NIOSH and UCC), and sex (male, female, and combined). Given all the analyses reported by Valdez-Flores et al. 2010 for both NIOSH and UCC studies, the ratings for "Categorical or Continuous Data" should be the same for both studies.

c) Lag Periods. The NIOSH study received a "High" rating for lag periods, while the UCC received a "Low" rating that should be elevated to a similar "High" ranking as the NIOSH study.

Criteria	Factors to consider	High	Adequate	Low	Critical
Lag periods	• Were the likely latency and lag periods appropriately addressed in the statistical analyses? • This criterion rates the incorporation of latency in the statistical analyses. Weaknesses related to the inclusion of an appropriate follow-up period in the underlying study design is rated under the "Follow-up period" criterion	• Lag periods were considered in the statistical analyses, multiple lag periods were evaluated as needed, and the results for all relevant outcomes included appropriate lag periods	• Lag periods were considered in the statistical analyses although with some relatively minor weaknesses (e.g., only a limited number of potential lag periods assessed)	• Lag periods were considered in the statistical analyses although with some moderate weaknesses	• There is good evidence that an important lag period exists, but this was not addressed in the statistical analysis

Steenland et al. 2004 and IRIS 2016 chose a lag-period for exposure metric without evaluating whether statistics support the inclusion of that variable into the model (see Valdez-Flores et al. 2010). Valdez-Flores et al. 2010, on the contrary, evaluate different lag periods and concluded that there was no significant statistical evidence for the inclusion of a non-zero lag period in the modeling of either the NIOSH or the UCC study. Valdez-Flores at al. 2025 used similar proven statistical analyses to evaluate the statistical relevance of the lag period. That is, the UCC study was also evaluated for different lag periods, but their inclusion did not improve model fit significantly and were not considered in the final models. Given that lag periods were indeed evaluated for the UCC study but not applied based on data, it should not be penalized with a "Low" rating.

d) Demographic Comparisons. Although Table 2 in Steenland et al. 2003 do list some demographic data for female workers of the NIOSH study, there is no demographic details for male workers. Thus, the NIOSH study rank as "Adequate" should be comparably ranked as "Critical" to the UCC rating based on a lack of female demographics.

Criteria	Factors to consider	High	Adequate	Low	Critical
Demographic comparisons	• Does the study provide information comparing groups with different levels of exposure or outcome? • Are these comparisons complete enough to evaluate whether demographic differences are likely to cause bias? • Study design and statistical analyses (e.g., adjustments) should be considered in these ratings	• Full demographic comparisons are made based on exposure or outcome • The study design was such that these demographic comparisons are not needed	• Demographic comparisons are somewhat limited although complete enough to evaluate whether moderate or important bias is possible	• Demographic comparisons are not provided or are limited to an extent that moderate bias is possible	• Demographic comparisons are not provided and there is a reasonable likelihood that demographic differences across exposure or outcome could lead to important bias

The UCC study did not present demographic details as there were only male workers and received a “Critical” rating. It seems that the UCC study is being penalized with a “Critical” rating for not including female workers while the NIOSH study is not similar**ly** penalized for lack of male demographic data.

Furthermore, UCC adjusted for age in the analysis and considered an adjustment for race in Valdez-Flores et al. (2025). The CPH model was not adjusted for any other covariates because none of the covariates significantly improved the likelihood of the model. In the UCC dataset, workers’ race was classified as white or Caucasian, black, non-white, and unknown. Consistent with previous analyses of the UCC data, the primary analyses were not adjusted for race because of the substantial number of subjects of unknown race (Rinsky et al., 1988). The present update of the UCC study was fit to the six cancer endpoints and compared with the results from the previous UCC update reported by Valdez-Flores et al. (2010), who published statistical analyses of the UCC data updated through 2003. In addition, a post hoc analysis adjusting the standard CPH model for race and year of birth did not change the results, so no further analysis was performed with this adjustment (Valdez-Flores et al. 2025)

e) Outlying Values. The “High” rating of the NIOSH study relative the “Low” UCC rating is unclear

OEHHA states, without a reference, that there are “no concerns regarding missing data or outlying values.” In a similar fashion, without citing any references, OEHHA states that it is “unclear if lagged analyses could be impacted by outlying values” in the UCC study. Thus, the UCC study is rated “Low”. More information regarding how OEHHA had “no concerns regarding missing data or outlying values” for the NIOSH study but for the UCC study it was “unclear if lagged analyses could be impacted by outlying values.” Without a reasonable

explanation of these two statements, OEHHA seems to have arbitrarily ranked the studies to place their preferred study ahead.

Criteria	Factors to consider	High	Adequate	Low	Critical
Outlying values	• Did the researchers evaluate whether outlying values may have impacted their findings? • Were appropriate methods used to control for outlying values?	• Thorough and accurate methods were used to evaluate and control for the presence of outlying values • Categorical data analyses are clear and complete and indicate that bias from outlying values is unlikely • Individual data are provided and show no outlying values	• Categorical analyses with some minor weaknesses • Some minor concerns about the methods or procedures used to identify and control for outlying values • Some indirect data suggesting that outlying values had relatively little impact on study results (e.g., comparable results from other more complete studies) • Individual data are provided and show no strongly influential outlying values	• Categorical analyses with some moderate weaknesses (e.g., limited number of categories) • Moderate concerns about the methods or procedures used to identify and control for outlying values • Individual data are provided and suggests that outlying values may cause moderate bias	• No analyses were done to control for or evaluate this bias and there is some likelihood that a few outlying values could invalidate study conclusions • Individual data show strongly influential outlying values and no methods were used to control for this

6. Under confounding, OEHHA evaluated the factors "Age and Sex", "Other Confounder Control", and "Adjusted-Unadjusted". In Attachment D, OEHHA summarizes the ratings for these factors.

STUDY	NIOSH	UCC	Sweden
Confounding: Other	A major advantage of the study design is that this study took place in facilities without high exposures to other known carcinogens; limited unadjusted results but impact of adjustments likely minor	278 workers in the chlorohydrin unit who were not involved in EtO production were excluded; a variety of other chemical exposures in EtO areas were noted by Greenberg 1990; carcinogenic exposures in the low EtO exposure group could bias EtO RRs towards the null; not adjusted by race since race was unknown in a "substantial number" of participants; limited unadjusted results but impact of adjustments likely minor	A major advantage of the study design is that this study took place in facilities without high exposures to other known carcinogens; limited unadjusted results but impact of adjustments likely minor
Age and Sex	High	High	High
Other Confounder Control	High	Low	High
Adjusted-Unadjusted	Adequate	Adequate	Adequate

The rating of “Low” for “Other Confounder Control” in the UCC cohort is mainly due to the possibility of confounding caused by exposures to other chemicals in the UCC plants. Although workers included in the UCC study were exposed to other chemicals, Valdez-Flores et al. 2025 state “Other chemicals cannot explain the absence of non-positive SMRs, that is, mask a positive finding unless they are protective of disease, which is highly unlikely. Confounding can possibly mask a positive trend in exposure-response analyses, although there is no known chemical cause of lymphoid tumors in the occupational setting.” This statement should have been considered when assigning a rating for “Other Confounding Control” in the UCC study. Thus, the rating should be changed from “low” to “adequate”, because Valdez-Flores et al. (2025) directly addressed this issue.

7. **Other: Conflicts of Interest: The NIOSH study is rated “High” while UCC study is penalized with “Critical” due to author acknowledgments of industry associations and/or funding.**

STUDY	NIOSH	UCC	Sweden
Other			
Conflict of Interest	Primarily government funded	Multiple potential conflicts of interest cited	Appears to be university funded (Metalund project at Lund University)
Selective reporting	No obvious missing results	No obvious missing results	No obvious missing results
Conflict of Interest	High	Critical	High
Selective Reporting Score	High	High	High

Conflict of interest	• Did a study author, funder, or other important entity have a financial interest in the study results?	• Study funding and conflict of interest information is clear and no study author, funder, or other important entity had	• Study authors appear to have some minor financial or other related interest in the exposure or outcome	• Study authors appear to have some moderate financial or other related interest	• A study funder, author or other related entity had a strong financial interest in the study results

OEHHA's treatment of the Union Carbide cohort updates reflects an apparent bias against industry-funded research rather than an objective application of the National Toxicology Program's Office of Health Assessment and Translation (OHAT) framework (NTP, 2019), which OEHHA cites as the basis for its review. OHAT does not consider financial conflicts of interest to be an automatic basis for a critical risk-of-bias rating. Instead, it calls for an evaluation of whether a potential conflict actually influenced study design, conduct, analysis, interpretation, or reporting. OEHHA identifies no evidence that industry funding affected the methods or findings of the Valdez-Flores analyses. As a result, OEHHA's

"critical" rating appears to be based primarily on sponsorship rather than demonstrated methodological deficiencies, which is inconsistent with the principles of the OHAT systematic review framework. More generally, the relevant question is not whether a potential conflict exists, but whether there is evidence that it affected specific methodological decisions or reporting of study findings. OEHHA identifies no such evidence for the UCC cohort updates.

Another factor that should be considered in this group is "Data Availability". The NIOSH study would likely receive a "Critical" rating since the data are not available to any scientists outside NIOSH. The UCC study would likely receive an "Adequate" or "High" rating because the DOW company would be willing to provide the UCC data to scientists interested in analyzing the effects of EtO.

8. **Under "Overall" rating, OEHHA assigns a "High" rating to the NIOSH study and a "Low" rating to the UCC study. In Attachment D, OEHHA summarizes the ratings for this factor.**

STUDY	NIOSH	UCC	Sweden
Overall rating: description	A few weaknesses were identified (modeled exposure, limited demographic comparisons, no lymphoid cancer incidence data) but these were all minor and not expected to have more than minimal impacts on hazard identification or exposure-response assessments	Several moderate and important weaknesses identified including questionable exposure assessment and validation, limited statistical power for lagged analyses, and lack of information on breast cancer	Several moderate and important weaknesses identified including non-specific exposure assessment, limited exposure validation, small sample sizes and low statistical power, and limited range of exposure
Overall rating	**High**	**Low**	**Low**

The "Low" rating for the UCC study is mainly related to low ratings assigned to several correlated factors. These low ratings have been discussed previously. The overall rating for the UCC should not be lower than "Adequate".

Figure 3 is labeled as evaluations of epidemiology studies for exposure-response and IUR calculations. OEHHA clearly does not consider evaluation of the alternate studies as weight of the evidence and as a reality check on the NIOSH based exposure-response model. Uncertainty analysis is the final key stage in QRA. At the very least, OEHHA should address the uncertainty of their exposure assessment and the use of the steep spline model in the context of a strong non-positive UCC study.

Appendix B

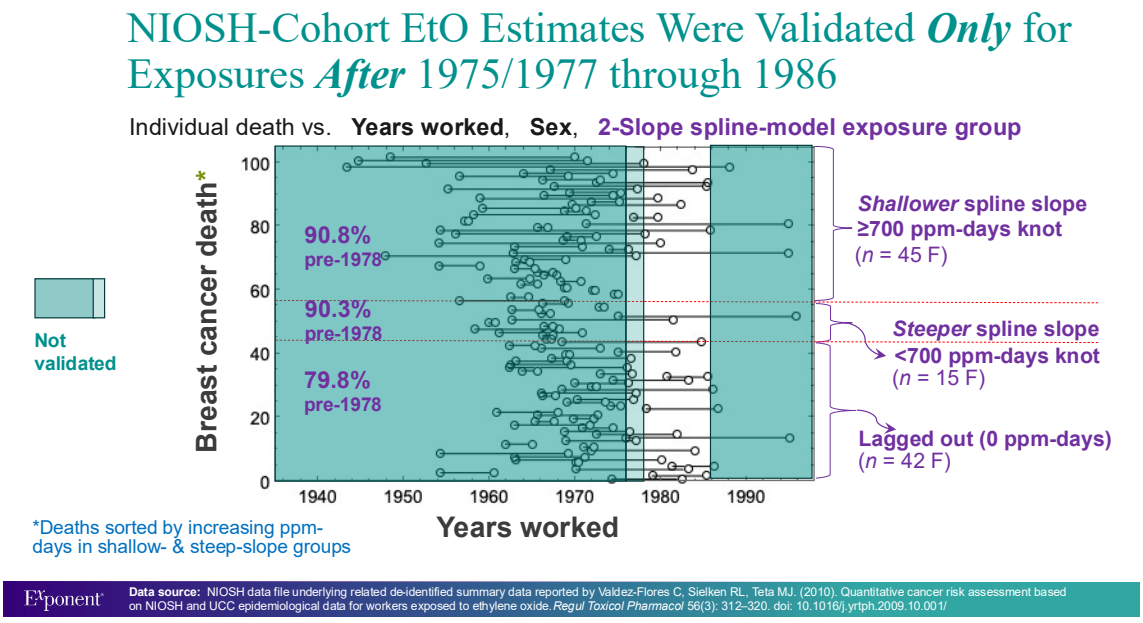
Graphs of NIOSH Breast Cancer Mortality Summary Data

NIOSH-cohort ethylene oxide estimates were validated only for exposures after 1976 through 1985.

All 102 Female cases are plotted by the range of years they each worked, by increasing exposure within each of 3 exposure ranges: the lagged out, the below-knot, and the above-knot exposure ranges presumed by the EPA-preferred 2-slope spline model of breast cancer mortality dose-response data.

Breast cancer incidence data is not available due to confidentiality concerns. However, breast cancer mortality was provided based on the NIOSH summary data file underlying related de-identified summary data reported by

- Valdez-Flores C, Sielken RL, Teta MJ. (2010) Quantitative cancer risk assessment based on NIOSH and UCC epidemiological data for workers exposed to ethylene oxide. *Regul Toxicol Pharmacol* 56(3): 312–320.
- Valdez-Flores C. and Sielken RL (2013). Misinterpretation of categorical rate ratios and inappropriate exposure-response model fitting can lead to biased estimates of risk: Ethylene oxide case study. *Reg Toxicol Pharmacol* 67:206-214.



Nearly All Pre-1978 Exposures to 103 Workers Who Died With Breast Cancer Lack *Any* Validation