COMMENTS TO THE CALIFORNIA AIR RESOURCES BOARD SCIENTIFIC REVIEW PANEL (SRP) ON TOXIC AIR CONTAMINANTS ON THE DRAFT OEHHA INHALATION UNIT RISK[1]

Jane Teta[2], Bhaskar Gollapudi[3], Kenneth T. Bogen[4], Steave Su[2], James Bus[5], Abby Li[6]

We appreciate the opportunity to provide comments to the California Air Resources Board Scientific Review Panel (SRP) regarding OEHHA's proposed Cancer Inhalation Unit Risk (IUR) Factor for ethylene oxide (EtO) and its accompanying Draft Technical Support Document (Draft IUR, 2023). These comments highlight key points related to the dose-response modeling and breast cancer that are discussed in greater detail in the attached American Chemistry Council (ACC) detailed comments on the Draft IUR that were submitted to OEHHA.

As reviewed by the Draft IUR, there are currently two government agency risk assessments for EtO, one conducted by USEPA and the other by TCEQ. Both TCEQ's and EPA's risk assessments are based on analyses of the same occupational cohort exposed to EtO. OEHHA's Draft IUR is based on the USEPA risk assessment. The principal endpoint of interest in both risk assessments is lymphoid cancer mortalities. TCEQ focuses solely on lymphoid cancers. EPA includes both breast and lymphoid cancers, but lymphoid cancer is the major contributor to the EPA IUR. Both use the Cox proportional hazards (CPH) modeling approach as the main statistical tool and cumulative exposure (ppm-days) to EtO as the metric of exposure. For the TCEQ model, the mathematical form of the CPH model is a single slope log-linear model, which approaches linearity at the lower exposures. In contrast, OEHHA and USEPA use two linear CPH models joined by a knot (inflection point) such that the first slope is much steeper than the second.

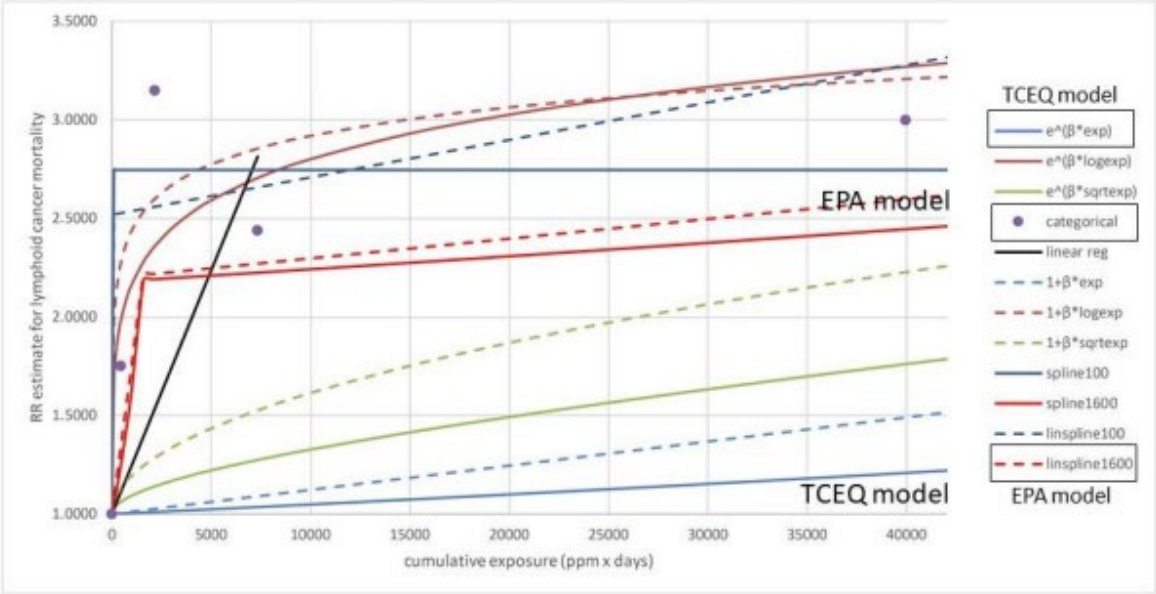
It is very clear that the OEHHA team put in a tremendous amount of effort to evaluate and clearly summarize the EtO epidemiology and animal toxicity studies. However, for the dose-response assessment, OEHHA relied heavily on the IRIS assessment due to the lack of access to the NIOSH data. The Draft IUR modeled "the categorical data", but these are 5 grouped estimates of the actual individual data including 53 cases of lymphoid mortalities. In fact, the EPA Science Advisory Board (SAB, 2015) recommended that EPA should re-do its modeling based on individual results instead of modeling the 5 categorical estimates (quintiles) as was proposed in the draft IRIS assessment. Yet, in the final IRIS assessment EPA graphically presents these quintiles as point estimates to visually determine which model fits the best. In essence, the IRIS assessment uses the very estimates EPA SAB considered to be not representative of the individual data to determine which continuous model of the individual data

[1] This white paper was funded by the American Chemistry Council
[2] Exponent, Inc. New York, NY.
[3] Toxicology Consultant, Midland, MI
[4] Versar Global Solutions through Exponent, Inc Oakland, CA
[5] Exponent, Inc. Alexandria, VA
[6] Exponent, Inc. Oakland, CA

to select from. While categorical models with a small number of odds ratios can be useful for identifying possible associations, they do not identify the shape of the dose-response curve based on continuous data modeling of as shown in detail by Valdez-Flores and Sielken (2013).

The EPA IRIS figures that the Draft IUR relies on to evaluate model-fit are not fit for this purpose. First, the figures do not convey the actual data that were fit. Second, they give the false impression that TCEQ's model "underestimates" the categorical RR estimates. This problem is best illustrated by public comments to TCEQ, which presented the following graph and incorrectly stated that the CPH model "underestimates" the categorical relative rate (RR) estimates (TCEQ response to public comments p. 41) based on visual fit comparisons. The reproduction of this graph in this public comment failed to include the easily missed footnote included in all EPA figures of RR correctly warning that the different models have different implicitly estimate baseline risks; thus, they are not strictly comparable to each other along the y-axis.

Figure 1. Figure from Public Comments to TCEQ (2020)



Reproduction of Figure 4-3 from EPA assessment, with black rectangles and text added to highlight TCEQ's model, the categorical RR estimates, and EPA's selected model.[54]

The Draft IUR (p.36) makes the same visual comparison error by stating "Other models, including the log-linear models (e.g., Cox regression) and the models using categorical data or exposure transformations, generally resulted in sloped that appeared to dramatically over- or under-predict the actual study results, especially in the lower-exposure ranges". In contrast, the ground-truthing approach used by TCEQ comparing expected and observed numbers of lymphoid cancers in the NIOSH and UCC cohorts comports with traditional well-accepted statistical practice (discussed in further detail below and in ACC detailed comments).

While the Draft IUR admirably attempts to conduct an "independent" evaluation of bias in EPA's model and the NIOSH epidemiological data, these efforts are based on assumptions in the absence of access to the actual data and repeats EPA's focus on categorical grouped estimates that are not the data modeled. ACC had access to the NIOSH exposure and mortality data prior to 2016, and our comments are based on our analysis of these individual data.

Recently, in response to public comments on new proposed EtO regulations, EPA has increasingly relied on the aforementioned misleading visual fit comparisons with grouped estimates as the basis for justifying a 2-piece spline model. EPA acknowledges that two statistical expert peer-reviewers for the TCEQ assessment agree that the EPA IRIS incorrectly calculated the p-values for likelihood-ratio tests of model-fit adequacy. With TCEQ's correction, neither the TCEQ model nor the EPA's model exhibit a statistically adequate fit by EPA's own stated criteria ($p<0.10$, EPA IRIS p.4-20).

It is relevant to consider here that the original NIOSH lymphoid study applied a large number of curve-fitting models (>50 for each cancer) in an exploratory statistical modeling exercise that included 5 different exposure metrics, 5 different lags for each exposure metric, males alone, females alone (Steenland et al. 2004). The authors reported only 3 continuous models with specific lags to be statistically significant, all of which are log cumulative CPH models, which EPA considers to be biologically implausible. Importantly, the NIOSH study authors concluded that male and female workers in the **highest** cumulative exposures and longest latency had statistically significant excesses for lymphoid cancer mortality (males) and the fully ascertained breast cancer mortality (females), respectively (Steenland et al. 2004).

While statistics provide an important objective method for assessing model-fit adequacy, the ACC comments emphasize the holistic integration of the epidemiological and biological evidence in the selection of a dose-response model. EPA has argued that if the statistics for the model fits are comparable, then why not select the most protective one? We disagree, and so does EPA SAB. EPA SAB (2015) stated that "Any model that is to be considered reasonable for risk assessment must have a dose-response form that is both biologically plausible and consistent with the observed data." While EPA may try to invoke the more protective option, that approach is valid only if the models are statistically significant **and biologically plausible.** In the case of lymphoid cancers, neither the TCEQ nor the EPA models indicates a significant relationship between EtO exposure and lymphoid cancers. More importantly, the TCEQ model is more consistent with the observed epidemiologic data and the genotoxicity data most relevant for lymphoid cancers.

The OEHHA Draft IUR states that US EPA concluded that the EPA 2-piece spline model provides the best biologically plausible fit to underlying NIOSH study data, especially in the lower exposure region. The Draft IUR also quotes EPA as stating that model selection weighs statistical and biological consideration. However, careful examination of the IRIS assessment indicates that EPA only considered biological plausibility in the two following respects:

1. when rejecting the statistically significant but implausibly steep log cumulative model
2. when justifying moving the knot of the EPA's favored model.

In both cases, EPA did not provide any biological evidence other than to state that it is biologically implausible for the dose-response to be so steep, and its IRIS assessment provides no explanation except merely to declare that the log cumulative and the initially selected knot these two cases are biologically implausible. We agree with EPA that the log cumulative models and the 2-piece spline model with the statistical best-fitting (100-ppm-days) knot are biologically implausible based on the NIOSH study alone, the epidemiological weight of evidence, the most relevant genotoxicity data, the carcinogenicity data and the toxicokinetic properties of EtO (see detailed ACC comments). But these same reasons support the TCEQ's standard (log-linear) Cox Proportional Hazards model over the EPA IRIS 2-piece spline model with knot at 1600 ppm-years. After rejecting the log-cumulative model, EPA then resurrects the biologically implausible log cumulative model to provide a statistical rationale for selecting the EPA IRIS model based on a flawed visual fit comparison with categorical estimates, which are not the data modeled.

Consistency with the Observed Data

The IRIS assessment for EtO is based on the 2-piece spline model with a steep increasing risk at low exposures, implying that EtO is a potent carcinogen. The NIOSH study of over 17,000 sterilant workers, the UCC study of over 2000 chemical workers from the 1930s, and the large number of other occupational cohort studies over a 40 yr. period do not support this implication.

There is no significant overall excess of lymphoid cancer mortality or of breast cancer incidence or mortality in the NIOSH studies, when compared to the general population. These findings cannot be attributed to the Healthy Worker Effect (HWE) as this study population has had extensive follow up known to diminish the HWE. The NIOSH authors also clearly dismiss the existence of a HWE for this study group. Their conclusions are consistent with that of the International Agency for Research on Cancer textbook on Cancer Epidemiology: Principles and Methods (IARC, 1999), which specifically notes that HWE "is known to vary with type of disease, being smaller for cancer than for other major diseases, and it tends to disappear with time since recruitment into the workforce." In worker-to-worker comparisons, increased lymphoid cancer mortality was observed in the NIOSH study, for males only, in only the highest exposure group with a 15-year lag. A statistically significant excess was observed for breast cancer mortality, again only in the highest exposure category with a 20-year lag. Neither corresponding overall trend exhibits both fit-adequacy and better fit than a corresponding CPH (i.e., approximately linear) fit.

The conclusions of the NIOSH authors with respect to the findings of their mortality study were very tempered.

> "Positive-exposure response trends were found for males only. Reasons for the sex specificity of this effect are not known. There was also some evidence of a positive exposure-response for breast cancer mortality."

The positive breast cancer incidence findings from worker-to-worker comparisons were also cautiously interpreted by the NIOSH authors who stated:

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

In any study of this type, in which multiple disease endpoints are considered and a multiplicity of models examined, there are the very real problems of multiple comparisons and data dredging, particularly when there is no consistency of findings within subgroups in the study. Although the authors argue that their study is exempt from these problems, the Steenland et al. (2004) study is no exception as described above. There appears to be no biological explanation for the discrepant findings for lymphoid cancer between males and females nor for the discrepant results between the NIOSH and UCC cohorts. A more defensible conclusion would be that this study generated hypotheses to be tested in future studies. Therefore, a more nuanced conclusion, based on the NIOSH study, is that evidence for a causal association between exposure EtO and any cancer in humans is weak and does not support a steeper exposure response model at lower exposures compared to higher exposures. The standard CPH dose-response model is more consistent with the observed epidemiology data.

Biological Plausibility

There are no in vitro or in vivo data specific to EtO that support the EPA spline-model assumption that risk of EtO-induced genotoxicity, mutations, or cancer increases sharply in relation to cumulative EtO exposure and then markedly saturates at relatively low levels of cumulative EtO exposure. To the contrary, mechanistic data indicate that DNA-damage-related EtO endpoints are expected to exhibit a linear, monotonic, no-threshold dose-response without any marked inflection point(s) in relation to exposure at relatively low exposure levels. Particularly in view of lack of any statistical evidence of improved data fit compared to traditional linear risk extrapolation, spline-model risk extrapolation is neither biologically nor statistically warranted. For the purpose of these overarching comments, we focus on the mutagenicity evidence, which was reviewed in the Draft IUR, but not considered with respect to the biological plausibility of model selection to apply to the NIOSH study.

There is little disagreement that mutagenicity is the presumed mode of action (MoA) through which EtO contributes to its carcinogenicity. Currently, there is no other alternate cancer MoA to exclude mutagenicity. EtO is a direct alkylating agent, meaning that it does not require any

metabolic conversion to become reactive with DNA. On the other hand, metabolic processes such as conjugation with GSH or hydrolysis through epoxide hydrolase are involved in the efficient detoxification of EtO at lower exposure levels. Accordingly, it is expected that overwhelming detoxification plays an important role in EtO induced DNA reactivity.

Reaction of EtO with DNA leads to the formation of various types of DNA adducts and some types of adducts contribute to the formation of mutations. In this context, it is important to mention that there are cellular repair processes that can effectively deal with the DNA damage at low doses, and it is only at doses that saturate these repair process that one should expect the unrepaired adducts leading to mutagenicity.

There is an extensive body of literature investigating the mutagenicity of EtO. EtO has been shown to be mutagenic both in vitro and in vivo test systems. From this perspective, one may conclude that the evidence for EtO's mutagenicity is strong. This does not mean that the mutagenic potency of EtO is strong. To the contrary, EtO is a weak mutagen, meaning that it requires relatively high doses and long exposure durations to elicit its mutagenic activity. The cellular detoxification and DNA repair processes coupled with its inability to bioaccumulate play a critical role in the weak mutagenicity of EtO. For a direct acting mutagen like EtO, the worst-case dose-response scenario for mutagenicity is linear with a single slope, no thresholds, and no inflection points at relatively low levels of cumulative exposure. For EtO, however, a more plausible dose-response is to have a shallow slope at lower doses and a steeper slope at higher doses due to dose-disproportionate increase in internal dose because of the saturation of detoxification processes and/or overwhelming of DNA repair processes. Multiple lines of experimental evidence support a shallower initial slope for the EtO-induced dose-response. For DNA adducts, the molecular initiating event for EtO-induced mutagenicity/carcinogenicity, Marsden et al. (2009) using one of the most sensitive analytical tools demonstrated little increase in DNA adduct formation in the livers of rats intraperitoneally injected with lower dose levels of EtO. Thus, even if we don't believe there are thresholds, these data at best conservatively demonstrate that the DNA adduct formation has a linear response with a single slope over the entire range of historical occupational exposures examined.

Further experimental evidence for the weak mutagenicity and shallower dose-response for EtO comes from the bone marrow (Recio et al., 2004) and lung (Manjanatha et al., 2017) tissues of Big Blue transgenic mice exposed to EtO. These two tissues are highly relevant to tumor findings in EtO-exposed mice. In both cases, the shape of the dose response is at best conservatively characterized as having a single linear slope. In fact, significant increases in mutations were observed only at the higher doses of EtO (i.e., 100 ppm and higher in the bone marrow and at 200 ppm in the lung) and that too after extended durations of exposure (i.e., after 48 weeks but not at 12 or 24 weeks in the bone marrow and at 8 but not 4 weeks in the lung). These data demonstrate that EtO is a relatively weak mutagen and the shape of the dose-response curve is conservatively linear with a single slope, but more likely with a shallower slope at relatively lower dose levels.

In the most recent EPA (2022) response to public comments regarding the lack of consideration of the biological evidence in the dose-response assessment, EPA conducted a highly subjective visual inspection of genotoxicity data to support their claim that the biological evidence cannot be used to inform biological plausibility. The EPA (2022) evaluation involved (a) plotting the data as point estimates without error bars, (b) drawing a straight line between the response levels for the lowest and highest dose levels, and (c) declaring the dose-response to be supralinear or sublinear depending on whether the responses for the mid-dose levels visually appeared to be above or below the line. This visual inspection did not involve any consideration of validly assessed statistical significance or evaluation of which data set and dose regimen is most relevant and useful to inform epidemiology data based on cumulative exposures. In their response to public comments, EPA opined that dose-response information obtained from animal genotoxicity studies would not allow selection of dose response models for human risk assessment for EtO. This position seems to be at odds with the agency's 2005 Cancer Guidelines stating that "[i]f dose-response analysis of nontumor key events is more informative about the carcinogenic process for an agent, it can be used in lieu of, or in conjunction with, tumor incidence analysis for the overall dose-response assessment." Thus, properly modeled genotoxicity dose-response information from animal studies should be considered in the selection of the model for EtO risk assessment.

Figure 2: EtO-induced Dose-Response for DNA Adduct Formation (Marsden et al., 2009).

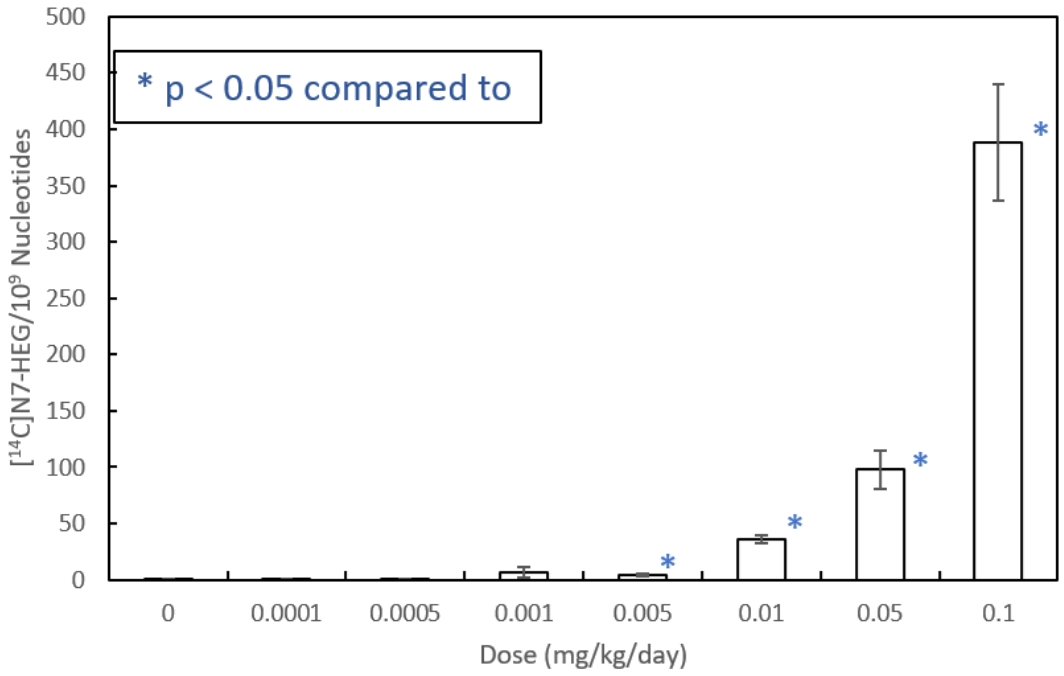


Figure 3: Dose-Response for the Induction of *lac I* Mutations in Big Blue Mouse Bone Marrow (Recio et al., 2004).

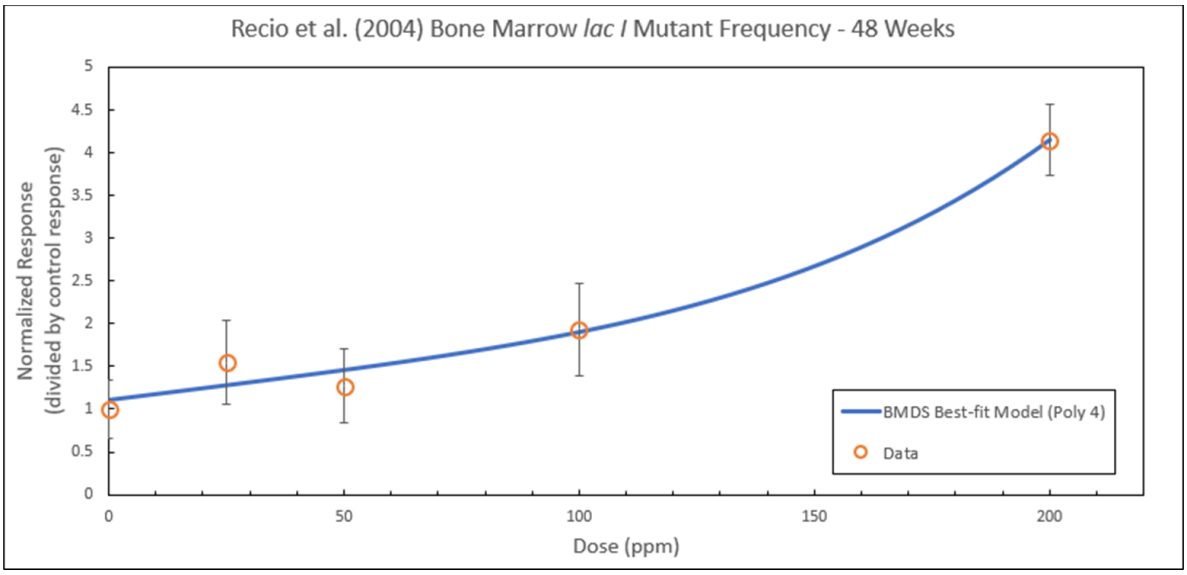
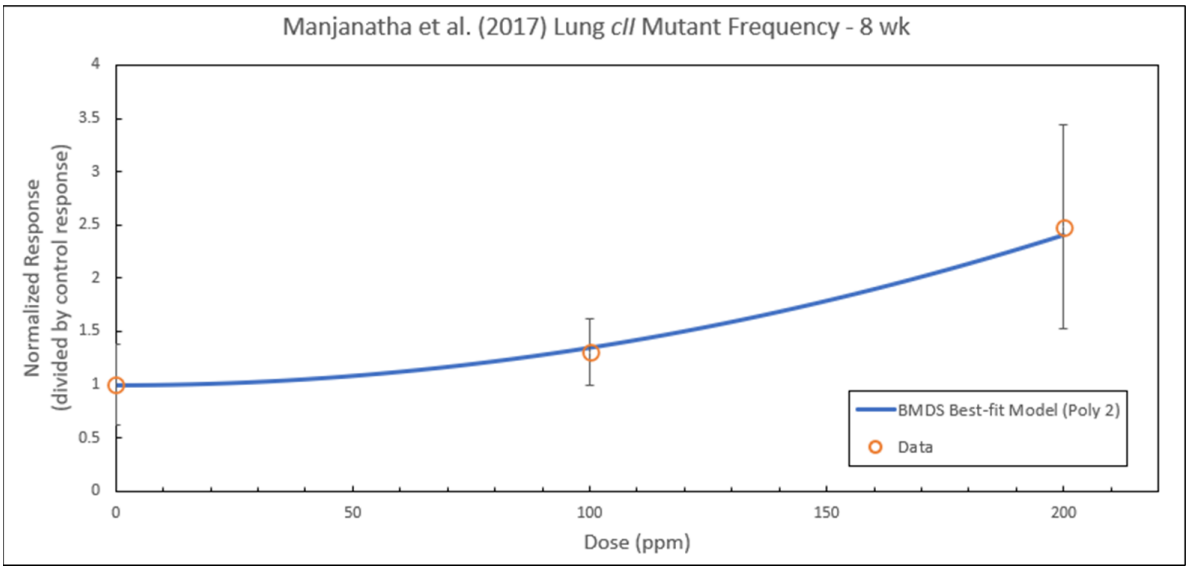


Figure 4: Dose-Response for the induction of *cII* Mutations in Big Blue Lung Tissue (Manjanatha et al., 2017).



Breast Cancer

In the ACC comments, we explain in detail why breast cancer is an appropriate endpoint to include as part of the epidemiological weight of evidence for carcinogenicity but is not appropriate to use for quantitative risk assessment.

There was a substantial number of missing cases in the NIOSH breast cancer incidence interview study, recognized by the study authors (Steenland et al. 2003). The bias concern is that these cases are more likely to be among short term, low cumulative exposure workers, who are more

difficult to locate, and therefore would have deficits of cancer in the low exposure category, suggesting a positive trend. The magnitude of missing data and potential bias argues against using these data in quantitative risk assessment. Breast cancer mortality results, also examined by NIOSH, do not suffer from underascertainment. These data are more consistent with the standard CPH model as was applied by Valdez-Flores et al. (2010). The standard CPH model, on the other hand, as employed by TCEQ for lymphoid cancers and generally applied to occupational epidemiology studies, more appropriately models a gradual increase in risk with increasing exposure.

The submitted ACC comments (at p. 20, Table 3) further note that in the case of breast cancer, the spline model fit by EPA, the corresponding corrected LR-test p-value of 0.04 indicates that fit is only marginally adequate and is in fact worse than that ($p = 0.02$) associated with the standard CPH model fit to the breast cancer data. Thus, if the California SRP agrees with OEHHA that breast cancer should be included in quantitative assessment, then we urge the SRP to recommend OEHHA to use the standard CPH model fit to the fully ascertained breast cancer mortality that is available in the EPA IRIS (2016).

Ground-truthing model selection

TCEQ's ground-truthing exercise is a more objective method than IRIS's visual fit comparisons to address how well the models predict the actual number of cancer mortalities. TCEQ demonstrated that the CPH model prediction for the full NIOSH cohort is more accurate than for the IRIS 2-slope model. This is also true when a healthy worker effect of 15% is included to represent differences between the NIOSH and general population even though the authors of the NIOSH study conclude there was no HWE. In addition, this conclusion remains true even when different methods for calculation of confidence intervals are used. Importantly, the CPH model prediction is also more accurate for each exposure quintile including the lowest exposure category that EPA considers most relevant for the general population.

The OEHHA IUR should be corrected to indicate that the TCEQ model has excellent overall and local fit based on the TCEQ's prediction analysis, which considers a possible HWE effect as a reasonable surrogate for differences that might exist between the general US population and the NIOSH worker cohort.

Summary

We appreciate OEHHA's efforts to independently evaluate the IRIS assessment. As part of this effort, we urge OEHHA to seriously consider our comments together with the more detailed ACC comments that address OEHHA's new analysis and evaluations. The attached ACC comments provide additional references and analysis supporting the following key points that summarize why dose-response analysis using a single log-linear CPH model has greater biological plausibility and is more consistent with the observed epidemiology data:

- The USEPA's 2-slope model is comprised of a steep slope in the low-dose region with high-dose plateau. This appears to be an artifact of embedded decisions made in the modeling, in particular:
    - Combining men's & women's data exhibiting dramatically different exposure-response behaviors
    - Incorrect statistics, misleading visual fit comparisons, over-reliance on biologically implausible log-cumulative models
- A Steep slope in low dose region is inconsistent with the epidemiology data.
    - Signals for LH, lymphoid and breast cancer are weak and inconsistent across available studies.
- A steep slope in the low-dose-region is inconsistent with the biological evidence
    - Genotoxicity for EtO do not exhibit this behavior.
    - EtO toxicokinetics do not exhibit the behavior of the EPA's steep initial slope.
    - The carcinogenicity data for ethylene and EtO do not exhibit this behavior.
- EPA's 2-slope model overpredicts risk
    - Overestimating cases in the range of observation.
    - The resulting IUR predicts unacceptable excess risk in ambient air, exhaled air and fruits
    - As such, the use of USEPA's IUR to assess, manage, and communicate risks from EtO is not recommended.
- The log-linear CPH model as performed by TCEQ is preferred
    - Accurately predicts the number of cancer cases in range of observation for the NIOSH cohort.
    - Is approximately linear in the low-dose range without exhibiting a plateau that is inconsistent with relevant mechanistic data.
    - Is a standard model used for epidemiology and is more representative of the epidemiological weight of evidence.
    - The behavior is consistent with the underlying genotoxicity, carcinogenicity and toxicokinetics of EtO.

All references cited are listed in the attached detailed ACC comments on the Draft IUR.