UNITED STATES ENVIRONMENTAL PROTECTION AGENCY
Center for Public Health and Environmental Effects
109 T.W. Alexander Drive, Research Triangle Park, NC 27711

OFFICE OF
RESEARCH AND DEVELOPMENT

18 October, 2019

MEMORANDUM

SUBJECT: IRIS EtO Assessment - Modeling Comparisons and Assessment of Uncertainty

FROM: Kristina A. Thayer Kristina Thayer
Director, Chemical & Pollutant Assessment Division (CPAD)
ORD Center for Public Health and Environmental Assessment (CPHEA)

TO: Peter Tsirigotis
Director, Office of Air Quality Planning and Standards (OAQPS)
Office of Air & Radiation (OAR)

In response to the inquiry from OAR to ORD about the dose-response selection for the IRIS ethylene oxide inhalation unit risk (IUR), CPHEA's Chemical and Pollutant Assessment Division has reviewed the information available in the 2016 ethylene oxide IRIS assessment.

The attached analysis, developed by Paul White (ORD/CPHEA/CPAD), synthesizes the information on the range of model forms evaluated in the IRIS assessment, and considering statistical and biological factors, identifies additional models examined that can reasonably contribute to quantitatively characterizing model and statistical uncertainty in the risks of cancer associated with environmental exposures to EtO.

The alternative dose-response model forms tabulated here can aid assessors in understanding the uncertainties in the estimated risks from EtO exposures. It is important to note that this analysis relies entirely on results and equations presented in the final EtO IRIS assessment.

UNITED STATES ENVIRONMENTAL PROTECTION AGENCY

OFFICE OF RESEARCH AND DEVELOPMENT

18 October 2019

MEMORANDUM

SUBJECT: Sensitivity of ethylene oxide risk estimates to dose-response model selection

FROM: Paul White
Senior Advisor, Chemical & Pollutant Assessment Division (CPAD)
ORD Center for Public Health and Environmental Assessment (CPHEA)

Paul White

TO: Kristina A. Thayer
Director, Chemical & Pollutant Assessment Division (CPAD)
ORD Center for Public Health and Environmental Assessment (CPHEA)

In developing the cancer dose-response modeling and inhalation unit risk estimates for ethylene oxide (EtO), EPA had the advantage of utilizing human data from a large, high-quality, occupational epidemiology study performed by the National Institute of Occupational Safety and Health (NIOSH)[1] (Steenland 2003, Steenland 2004). Under the guidance received from two SAB reviews, EPA conducted extensive statistical modeling to examine multiple approaches to represent the risk information in this cohort. EPA initially developed risk estimates based on linear regression of published categorical rate estimates (cancer rates broken out by EtO exposure intervals) for lymphoid and female breast cancers. Following advice from the SAB, EPA worked with the principal investigator for the NIOSH study, Dr. Kyle Steenland, and developed models fit to the original individual data in the cohort. Again, following SAB advice, EPA focused on models that were compatible with the observed "plateauing"[2] shape of dose-response for these cancers where the risk rises rapidly at lower exposure levels and then rises more gradually for higher exposures. This plateauing was observed in published categorical results[3] and

---

[1] Steenland K, Stayner L, Deddens J. Mortality analyses in a cohort of 18 235 ethylene oxide exposed workers: follow up extended from 1987 to 1998. Occup Environ Med. 2004 Jan;61(1):2-7. PubMed PMID: 14691266; PubMed Central PMCID:PMC1757803.
Steenland K, Whelan E, Deddens J, Stayner L, Ward E. Ethylene oxide and breast cancer incidence in a cohort study of 7576 women (United States). Cancer Causes Control. 2003 Aug;14(6):531-9. PubMed PMID: 12948284.

[2] As characterized by the SAB, the study data indicate a pattern where "the risk rises rapidly with a small amount of exposure and then rises much more gradually for even higher exposures.".

[3] Plotting predictions from models in comparison with categorical breakouts of the data is a useful tool in epidemiology as it allows examination of the behavior of the continuous parametric models versus unstructured information on the levels of response (disease) within ranges of exposure. Plots of model fit to of individual dichotomous data points are difficult construct and comparisons with categorical data fill that gap. Note that the categorical breakouts and continuous models compared in the IRIS assessment are developed from the same set of individual response data and utilize the same definition of the referent group – individuals who do have no

available statistically significant continuous model fits. The two-piece spline modeling approach that the IRIS assessment selected to develop recommended cancer risk estimates were supported by the SAB and followed the EPA cancer guidelines for a direct acting mutagenic carcinogen. It represents a pattern that is linear at low-dose and provides and empirical fit of the shape of the response in the range of the observed data.

As the IRIS assessment examined a large number of potentially applicable models, the SAB (2015) offered advice on several aspects of model choice that the EPA considered in completing the assessment. Specifically, the SAB:

- Recommended prioritizing functional forms of the exposure that allow regression models with more local fits in the low exposure range (e.g., spline models)
- Preferred the use of continuous individual-level exposure data over the use of categorical results.
- Advised 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.

These considerations led the IRIS assessment to select the 2-piece linear spline models for dose-response assessment for both lymphoid and female breast cancers. Spline models were the only models identified that were fully consistent with SAB's advice (bullets above).

Models for lymphoid cancer

Lymphoid cancer presented the more statistically challenging data for modeling. The analysis in this memo utilizes the three SAB criteria above to review the multiple modeling results presented in the IRIS assessment in order to identify reasonable candidate modeling alternatives in addition to the Assessment's selected linear two-piece spline models[4]. A variety of other model forms that were evaluated in the assessment and found not suitable for quantitative cancer risk assessment are also briefly reviewed to provide a perspective on the modeling results.

Models fit to the individual level data

(1) In addition to the selected linear two-piece spline model (knot at 1600 ppm-days), the related log-linear two-piece spline model with the same knot value also provided a comparable fit to the data (considering AIC- a measure of model fit that EPA relies on in many dose-response applications, visual fit, and p-value). This model provided a similar maximum likelihood risk estimate, but showed narrower confidence limits, resulting in an upper bound risk estimate a factor of three below the selected linear two-piece spline model. While considerations of plausibility of model shape led IRIS to prefer the linear spline model, the log-linear spline model presents a reasonable alternative model.

---

calculable exposure after taking into account the lag period in the modeling. As noted in the IRIS assessment, these comparisons address model shape and do not mean that the different approaches would imply identical background rates.

[4] Following the preferences from the IRIS assessment and SAB review, models are for combined risks for men and women. For consistency in comparisons the best fitting lag overall (15 years) was selected for analyzing the lymphoid cancer mortality data.

(2) Models with steeper low-dose slopes (which would imply risk levels an order of magnitude greater than the selected spline model) also provided appropriate global fits the data. However, these models were judged to be limited in terms of interpretation and plausibility. The log-linear (Cox regression) model with log-scale dose was presented in Steenland (2004) and was useful in supporting EtO hazard conclusions in that it provided a simple representation of a plateauing dose-response pattern and indicated a statistically significant effect of EtO on increasing rates of lymphoid cancer. However, the slope of this model becomes increasingly steep at low-dose and thus unit risk estimates depend on the choice of the point of departure creating an additional modeling uncertainty. Similar observations apply for the linear model with a log-scale exposure variable. Additionally, spline models with knots at lower exposure levels (100 ppm × days) provided statistically significant global fits to the study data. However, in these models there are no lymphoid cancer cases with exposures below the "knot". Thus, the low-dose slopes of these model forms are not locally supported by study data. These models are judged not to be reasonable alternatives to inform risk assessment.

(3) Other models fit to the individual level data indicated lower, and sometimes markedly lower, risk estimates but did not provide an appropriate fit to the dose-response pattern in the study data. Among these the log-linear cumulative dose (standard Cox) model and a fully linear model were judged to fit poorly to the data, showing higher AIC values (lower is better), lack of significant fit, and a very inconsistent visual fit to categorical tumor rates (implying minimal increase in risk over the range where the categorical data and other better fitting models indicated substantial risks). Additionally, further evaluation indicates that while the cumulative dose log-linear model showed a shallow linear increase over most of the dose range, model predictions, particularly for the upper bound slope estimate, curve sharply upwards at the highest observed doses. This concave-up behavior is not supported by the observed data. Models incorporating a square root transformation of dose indicated somewhat higher risks than the Cox cumulative dose models, and also provided marginally better statistical fits, however they were also judged to fit poorly in comparison to the categorical rates. These several models are not considered reasonable alternative choices to inform EtO risk assessment.

Models fit to categorical tumor rates

Linear regression models were fit to the categorical rate estimates (with the highest categorical point excluded to improve local fit in the lower dose range) and provided an additional approach for the modeling of the lymphoid cancer data. This approach is limited in that it does not meet the SAB preference for models fit to the individual data. However, regression modeling of categorical rates has provided a useful tool in other epidemiological analyses. And the approach of excluding the highest dose data to improve fit in the low dose range is commonly used in EPA benchmark dose modeling and addresses a primary modeling goal for EtO. Linear regression results for the categorical rates did not reach statistical significance, an unsurprising result for an approach that has reduced statistical power. The linear regression fit to categorical rates is included here as a reasonable modeling alternative.

<u>Models for female breast cancer</u>

As in the case of lymphoid cancer, the EtO IRIS assessment contains extensive supplemental and sensitivity analyses that support an understanding of risk estimates for female breast cancer. Steenland

et al. (2003) identified 319 incident cases of breast cancer in the cohort, with interview/questionnaire data available for 73% (233 cases) to support further modeling of individual risk factors. As with lymphoid cancer IRIS selected a two-piece spline model as providing the most appropriate representation of the data for risk assessment. Two-piece spline models fit the data well providing both a statistically significant fit and importantly, a good visual representation of the categorical data for risks of breast cancer incidence over the full study dose range. In agreement with the categorical data, the two-piece spline model shows a plateauing effect with the slope of the dose response being steeper at lower dose than higher dose. However, in the case of breast cancer, the modeled change-point for slope was above the median dose, thus the model does not show a tendency towards high slope at the low end of the observed range. In short, for female breast cancer the IRIS/SAB modeling goals were robustly met by the fitted two-piece spline model.

Alternative modeling results for female breast cancer incidence[5] from the IRIS assessment are summarized here, relying on the three SAB goals for modeling as stated above.

Models fit to individual level data.

(1) Models using a square root of dose transformation fit the data without need for a spline modeling approach, achieved the best (lowest) AIC scores, and provided appropriate visual fit to the categorical data over the full dose range. The two square root of dose models implemented in the IRIS assessment would lead to unit risk estimates for EtO inhalation roughly 3-10 times higher than the selected two-piece spline model. However, the IRIS assessment did not prefer these models, noting that the slopes for square-root of dose models become increasingly steep at low-dose and thus unit risk estimates are dependent on the choice of the point of departure leading to an additional modeling uncertainty. The square root of dose models are supralinear in the low dose region and thus contrast with the two-piece spline models that are linear over the lower dose range of the data. Accordingly, the square root models are not suggested as desirable alternative models. The additional models fit using a log transform of dose didn't fit as well as the square root models and showed a more marked pattern of low-dose supra-linearity and are also not deemed useful as candidate alternative models.

(2) A linear model of risk using cumulative EtO dose was examined and provided a statistically significant global fit to the data and a roughly appropriate fit to the categorical data (IRIS, Figure 4-7), however the agreement with the categorical data is poorer in the low-dose region, indicating that the model does not fully meet the SAB goal of providing a local fit to the lower dose data. For the present analysis the linear model is retained as a potentially useful, but marginally supported, alternative model.

---

[5] Risk estimates based on cancer incidence data are generally preferred, when available, to estimates based on cancer mortality, especially for cancer types with relatively high survival rates, such as breast cancer. Following the IRIS assessment, for consistency in comparisons, the overall best fitting lag (15 years) was selected for analyzing the breast cancer incidence data. Also following the IRIS assessment and SAB advice models from the subgroup of women having interview/questionnaire data on other risk factors are used here.

(3) The log-linear (standard Cox) cumulative dose regression model, also provides a statistically significant fit to the global data set but shows notably worse agreement with the plateauing shape of the categorical rates. IRIS also provided a sensitivity analysis of behavior of the log-linear model where the data for women having the highest 5% of EtO doses are removed from the fit (IRIS EtO Appendix D, Figure D-4). The predicted breast cancer risks increase strongly when these high dose data points are removed. Additionally, further data plots for this review indicated that while the log linear model increased roughly linearly over most of the dose range, model predictions, particularly using the upper bound slope estimate, curve sharply upwards at the highest doses – a behavior not indicated by the observed data Accordingly this model (which would provide a unit risk estimate 13-fold lower than the recommended two-piece spline model) is not recommended as a reasonable alternative model.

Models fit to categorical rates

IRIS also fit a linear regression models to estimated categorical rates for breast cancer incidence with the highest categorical point excluded to improve local fit in the lower dose range. This model provides an appropriate visual fit to the categorical rates; being an inherently lower-power approach, it does not show statistical significance. However, regression modeling of categorical rates has provided a useful tool in other epidemiological analyses. And the approach of excluding the highest dose data to improve fit in the low-dose range is commonly used in EPA BMD modeling and addresses a primary modeling goal for EtO. The linear regression fit to categorical rates is included here as a reasonable modeling alternative

<u>Synthesis -- Quantitative comparison of modeling alternatives</u>

Table 1 below shows Inhalation Unit Risk (IUR) estimates for models fit to the NIOSH data for both female breast and lymphoid cancers, and total risk estimates for combined risks for these two cancers. Central estimates (maximum likelihood estimates) of risk are also shown for comparison. Risks estimates using the two-piece linear spline models chosen in the IRIS assessment are compared with predictions from the models discussed above as reasonable alternative models. The results from Table 1 are shown graphically in Figure 1. Additional graphs for lymphoid cancer (see Figure 2) and breast cancer (see Figure 3) show the shapes of the models included here as reasonable alternative choices.

For both cancer sites, considered individually, the upper bound unit risk estimates for the identified alternative models fall in a range of about 2 – 5 times lower than the selected linear spline models. The central estimates of risk for these models are generally about 2 – 4 times lower than corresponding upper bound unit risks. The lymphoid cancer dose-response estimates indicate somewhat greater variability in the estimated low-dose response; this is concordant with the considerably smaller numbers of lymphoid cancers than breast tumors observed in the NIOSH study.

Looking at estimates of total cancer risk, if the dose-response model for lymphoid tumors is varied, but the recommended breast cancer model (judged to have the more robust selected model) is held fixed, the alternative total cancer unit risk estimates would be approximately 2 - 3 times lower than the recommended IRIS value. If all combinations of models for both cancers are considered, then total cancer unit risk estimates range from essentially the same to 5 times lower than the IRIS recommended value.

When only the lymphoid cancer dose-response model is varied, the central estimates for total risk range from approximately equal to 2 times lower than the central estimate from the two-piece spline model. If all combinations of models for both cancers are considered, central estimates of total risk range from approximately equal to 3 times lower than the central estimate from the selected spline models.

Note, however, the IRIS unit risk, should not be considered a worst-case analysis. Higher estimates of risk were obtained using some other models providing statistically appropriate fits to the data. While there were limitations with these models, and we have not used them in this analysis, it is likely that a comprehensive analysis of alternative models (for example considering other spline models with knots somewhat lower than the selected values) would likely include some risk estimates higher than the IRIS unit risk.

As indicated in Table 1, the risk estimates presented here are not adjusted for ADAF factors for early life sensitivity to mutagenic carcinogens. As discussed in the EtO IRIS assessment, these factors should be applied in estimating cancer risks involving early life exposure to EtO. For the total risk estimates based on the linear spline models used in IRIS the ADAF adjusted full-life risk estimates are 1.5 (9.1/6.1) times higher than the unadjusted values. The IRIS assessment does not tabulate ADAF adjusted full-life risk values for the alternative models. The effects of ADAFs on estimated full-life risk estimates from the alternative models should be similar to, but not exactly equal to, the 1.5-fold factor seen with the selected linear-spline models. The process for applying ADAF values as presented in the EtO assessment involves several steps and if needed ORD can provide support for this application.

The alternative dose-response model forms tabulated here can aid assessors in understanding the uncertainties in the estimated risks from EtO exposures. It is important to note that this analysis relies entirely on results and equations presented in the final EtO IRIS assessment.

Table 1. EtO cancer unit risk UCL estimates including alternative reasonable models for the NIOSH worker cohort - risks per ppm continuous exposure. Values not ADAF adjusted.

Cancer/model type	**Unit risk female breast cancer** (upper bound / central estimates.)	**Unit risk lymphoid cancer** (upper bound / central estimates.)	**Unit risk combined female breast and lymphoid cancers** (upper bound / central estimates.)
Lymphoid cancer, linear spline, individual data; Breast cancer, linear spline, individual data ***IRIS recommended unit risk***	**1.5** / 0.72	**5.3** /1.3	**6.1** / 2.1
Lymphoid cancer alternatives - log-linear spline, individual data - linear regression, categorical data; Breast cancer, linear spline (IRIS recommended unit risk)	 **1.5** / 0.72	 1.9 / 1.0 0.97 / 0.44	 2.9 / 1.7 2.1 / 1.2
Lymphoid cancer, linear spline (IRIS recommended unit risk) Breast cancer alternatives - Linear regression, categorical data - Linear model, individual data (marginal choice)	 0.91 / 0.42 0.38 / 0.19	**5.3** / 1.3	 5.7 / 1.8 5.4 / 1.5
Alternatives for both cancers, - Lymphoid, log-linear spline, individual data; breast, linear regression, categorical data - Lymphoid, linear regression categorical data; breast, linear regression, categorical data - Lymphoid, log-linear spline, individual data; breast, linear model, individual data (marginal choice) - Lymphoid, linear regression, categorical data; breast, linear model, individual data (marginal choice)	 0.91 / 0.42 0.91 / 0.42 0.38 / 0.19 0.38 / 0.19	 1.9 / 1.0 0.97 / 0.44 1.9 / 1.0 0.97 / 0.44	 2.4 / 1.4 1.6 / 0.85 2.1 / 1.2 1.2 / 0.62

Table notes: Unit risk and EC01 values for linear regression of categorical results for lymphoid cancer are from IRIS Table 4-7 cancer incidence calculations. Corresponding values for female breast cancer incidence are from Table 4-15. The IRIS assessment's recommended unit risk values for lymphoid and breast cancer incidence are in Table 4-17; the notes to this table explain the Wald-type confidence interval approach taken to estimate the total unit risk value for these two cancers combined. The same formulas are applied to estimate total cancer risk for the model combinations shown here.

Figure 1

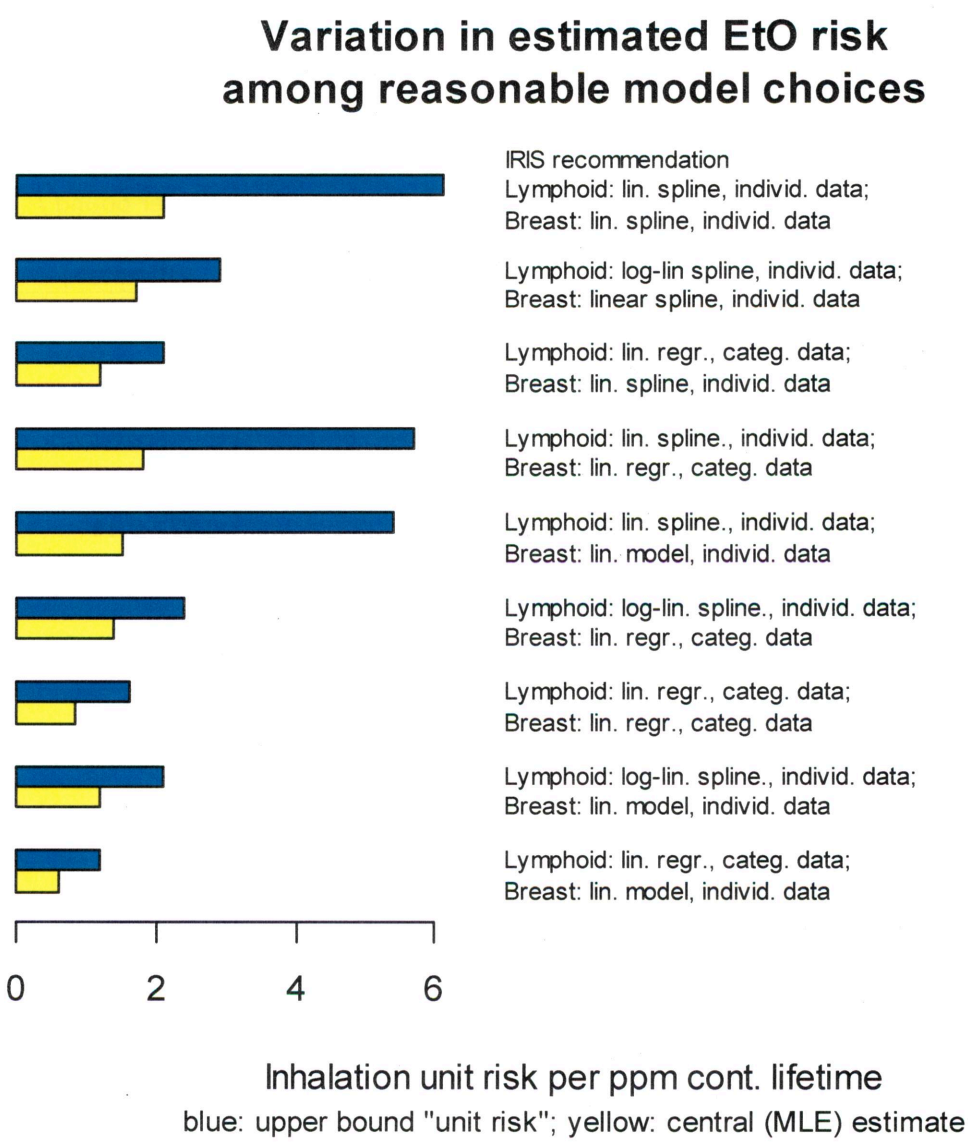

Figure notes: Results from Table 1 shown graphically

Figure 2

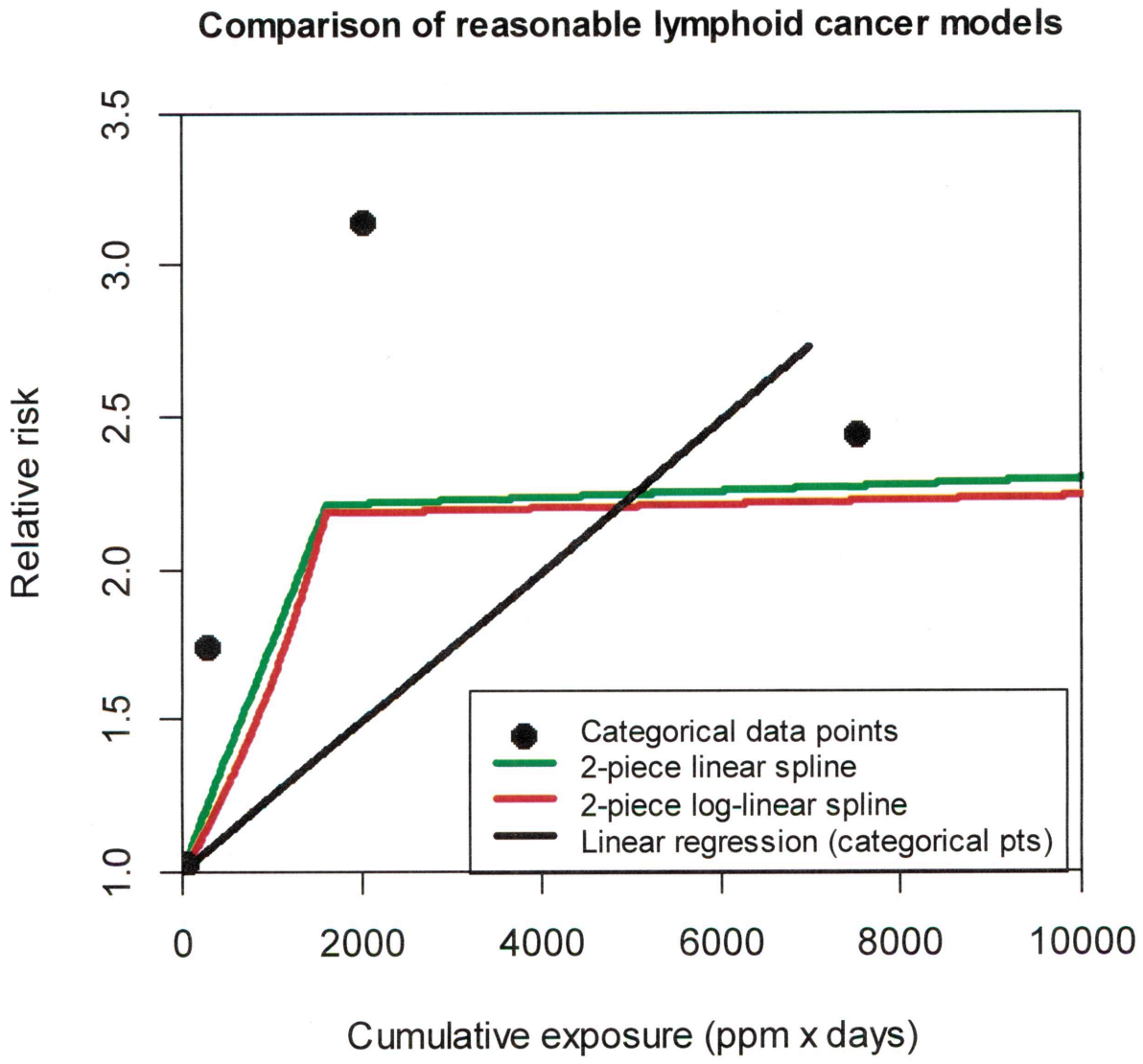

Figure notes:

Categorical data points: Relative risk averaged by exposure range to allow visualization.

2-piece linear spline: IRIS recommended approach - represents low-dose linear response and allows for plateau shape.

2-piece log-linear spline: Alternate spline approach affected by some curvature necessitated by model form - represents low-dose linear response and allows for plateau shape

Linear regression of categorical data (quintiles, without using highest group): Alternate approach to limit effect of highest dose data.

Figure 3

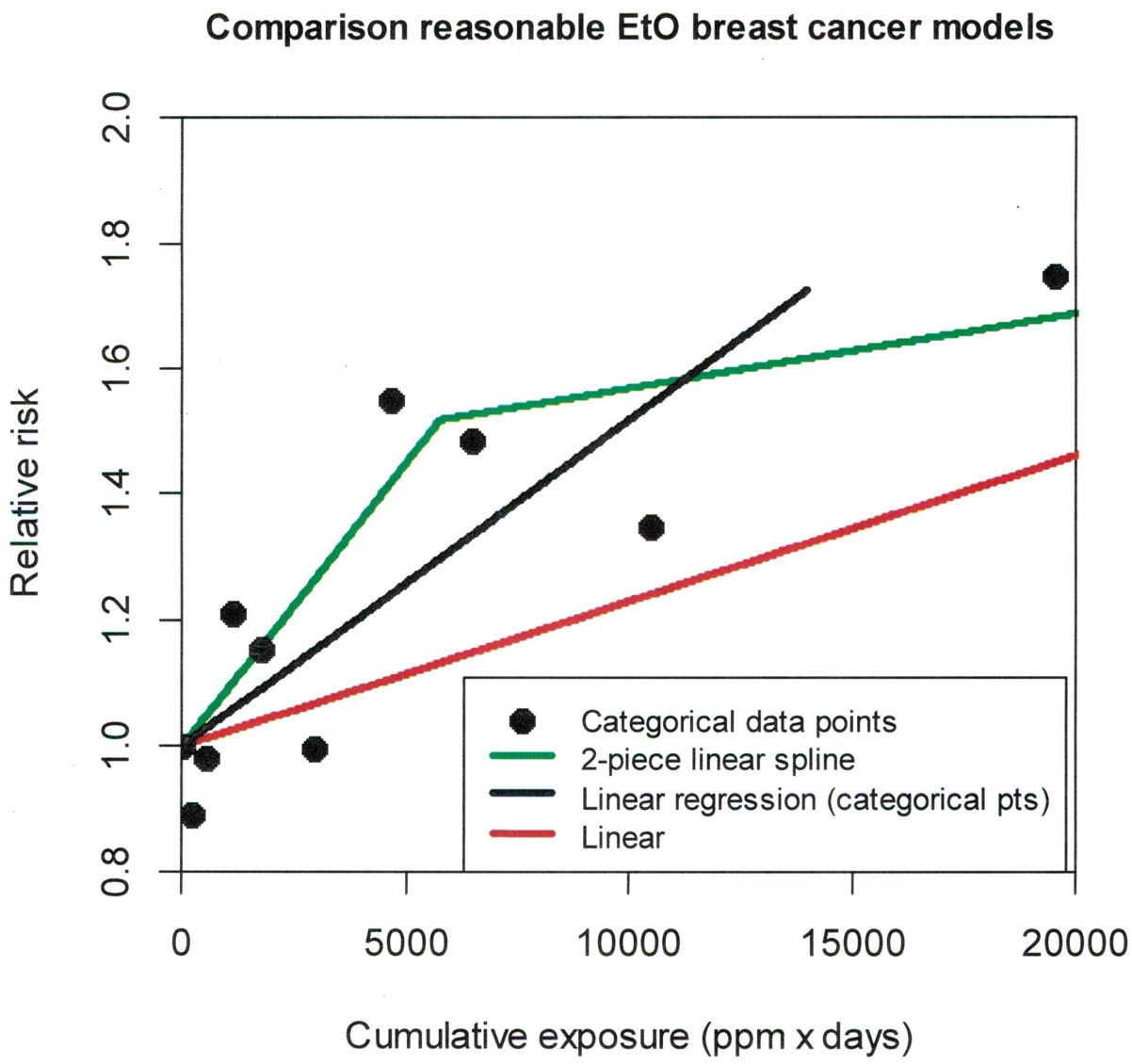

Figure notes:

Categorical data points: Relative risk averaged by exposure range to allow visualization.

2-piece linear spline: IRIS recommended approach - represents low-dose linear response and allows for plateau shape.

Linear regression of categorical data (quintiles without using highest group): Alternate approach to limit effect of highest dose data.

Linear model fit to full exposure range, individual data: Simple approach with marginal agreement with categorical data points.