

Robert L. Sielken, Jr., Ph.D. and Ciriaco Valdez-Flores, Ph.D., P.E.
Sielken & Associates Consulting Inc.
3833 Texas Avenue, Suite 230, Bryan, TX 77802
Tel: 979-846-5175; Fax: 979-846-2671;
Email: SielkenAssoc@aol.com

Comments on Appendix J.3.1. of EPA's Draft IRIS Assessment of Ethylene Oxide September 30, 2014

EPAs modeling approach for lymphoid and breast cancer remains incorrect. The methodological problems identified in Valdez-Flores and Sielken (2013) are relevant despite EPA's dismissal in Appendix J.3.1. Weaknesses include:

- Despite the 2007 SAB's recommendation for EPA to focus on individual data, EPA's modeling continues to focus on a few categorical rate ratios.
- EPA inappropriately rejects scientifically credible modeling options that are based on individual (not summary data).
- EPA's criterion for model selection (or equivalently model rejection) is fundamentally flawed. EPA fails to recognize that the categorical rate ratios are summary statistics tied to a specific assumed exposure-response model. Rate ratios are comparisons to a "control" group, and the data for the control group are random and not equal to some fixed underlying true value. In fact, the value for the "control" rate is an estimated value that depends on the fitted model. The rate ratios are ratios of the estimated "treatment" rate to the estimated "control" rate when all rates are estimated from the same model. Thus, it is not appropriate to plot a particular set of rate ratios based on an assumed model and then compare a second fitted model to these rate ratios. It is inappropriate because the relevant rate ratios for the second fitted model have a different denominator (different "control" rate) than the rate ratios based on the assumed model- hence different rate ratios. Comparing a fitted model to rate ratios based on a different model is inappropriate. That is, rejecting a second fitted model because it does not compare well to the wrong set of rate ratios (as EPA does) is inappropriate.

Valdez-Flores and Sielken (2013) does not advocate a specific model, but rather advocates using the individual data and model selection based on valid comparisons.

The following text is EPA's Appendix J.3.1 with Sielken & Associates Consulting, Inc.'s comments inserted in italics and numbered.

J.3.1. Valdez-Flores and Sielken (2013)

Valdez-Flores and Sielken (2013) criticize the approach employed by EPA in this and earlier drafts of the EtO carcinogenicity assessment of using a weighted linear regression of the RR estimates based on categorical exposure groups to derive exposure-response relationships for lymphoid cancer mortality and breast cancer mortality, stating that exposure-response modeling is best based on individual data. Valdez-Flores and Sielken (2013) express concern, for example, that because all the categorical RR estimates share the same denominator, if the rate in the reference group differs from the true rate, the RR estimates will be systematically biased.

Valdez-Flores and Sielken (2013) also contend that, because the data for the unexposed group are random, the intercept should be estimated for any exposure-response model fit to the categorical RR estimates rather than being fixed at 1. To illustrate their arguments, Valdez-Flores and Sielken (2013) fit several models to the Steenland et al. (2004) breast cancer mortality data: a categorical log-linear model (model 1); a linear model fit to the categorical RR estimates, with the intercept fixed at 1 (model 2); a log-linear model fit to the categorical RR estimates, with the intercept unrestricted (model 3); and a continuous log-linear model (fit to the continuous exposure data; model 4).

1. EPA uses the terms “continuous data” and “individual data” synonymously. On page iii, EPA says that continuous data are data across the full exposure range without first converting the exposure and individual occurrence data into categorical data for exposure-response modeling or the derivation of unit risk estimates.

With four exposure groups (not counting the unexposed reference group)—the number Steenland et al. (2004) used for their categorical modeling—the RR estimates and the slopes for models 2 and 3 are similar, but the slope for model 4, the continuous log-linear model, is much shallower. With more exposure groups (20 and 61, the latter being chosen so that there was exactly one breast cancer death in each exposure group), the results for models 2 and 3 differ more appreciably and the slope for model 3 (unrestricted intercept) is more similar to the slope for model 4, the continuous log-linear model.

2. In other words, model 2 (EPA’s model using categorical summary rate ratios and the intercept fixed to 1) is substantially different than the model fit to the individual data (model 4). Furthermore, as the number of categories increases, the model with unrestricted intercept (model 3) approaches the model fit to the individual data (model 4.) EPA does not mention that a similar behavior of the models was seen for the other two endpoints, lymphohematopoietic and lymphoid cancers, and that their results were shown in the supplemental material to Valdez-Flores and Sielken (2013).

Valdez-Flores and Sielken (2013) further suggest that it is inappropriate to use comparisons with categorical RR estimates when evaluating models of the continuous individual data. Valdez-Flores and Sielken (2013) disregard the apparent supralinear shape of the exposure-response relationship, suggesting it is the result of exposure measurement error, citing work by Crump (2005).

3. **None** of the categorical RRs in Figures 1, 2 and 3 for breast cancer mortality in Valdez-Flores and Sielken (2013) show a clear “supralinear” shape as argued by EPA. Furthermore, there is no piecewise – two-, three- etc. – supralinear model that fits the categorical RRs statistically significantly better than the linear model at the 5% significance level. In addition, EPA fails to acknowledge that Valdez-Flores and Sielken (2013) wrote “grouping of mortality deaths into only a **few** cumulative exposure intervals results in **oversimplification** of the underlying data and **masks the true exposure-response relationship.**” [Emphasis added]. The effect of very few groups can be seen in Table 2 in Valdez-Flores and Sielken (2013), where the slope of the correct model that estimates the intercept approaches the model fit to the individual data as the number of cumulative exposure intervals increases.

EPA notes that it did use continuous exposure models for the breast cancer incidence data.

4. However, breast cancer incidence is the only endpoint for which the final model (i.e., the model used for calculating risk estimates) was not based on a few grouped categorical RRs and a linear model restricted to being equal to one at the origin.

Extensive modeling of the individual data based on continuous exposure for the data sets for lymphoid cancer mortality and breast cancer mortality, however, yielded no preferred

5. Does EPA have an objective scientific definition of “preferred”?

models for the purposes of deriving unit risk estimates because of the strong supralinearity of the data (i.e., the response rate increases rapidly at low exposure levels and then attenuates or plateaus), which results in a very high slope in the low-exposure range (see Sections 4.1.1.2 and 4.1.2.2). Thus, EPA retains the approach of using the linear regression of the categorical results for those two cancer data sets (although EPA does not rely on the breast cancer mortality data set for the derivation of a unit risk estimate because the breast cancer incidence data are more suitable for the derivation of the desired incidence estimates). Valdez-Flores and Sielken (2013) state that “[u]nder no circumstances should an appropriate and relevant exposure-response model fit to individual epidemiological data be discarded in favor of an exposure-response model fit to summary data”; however, EPA did consider the continuous exposure log-linear model that Valdez-Flores and Sielken (2013) advocate and did not find it to be appropriate.

6. EPA’s criterion for model selection (or equivalently model rejection) is fundamentally flawed. EPA’s problem stems from their misinterpretation of the categorical rate ratios. EPA fails to recognize that the categorical rate ratios are summary statistics **tied to a specific assumed exposure-response model**. Rate ratios are comparisons to a “control” group, and the data for the control group are random and not equal to some fixed underlying true value. In fact, the value for the “control” rate is **an estimated value that depends on the fitted model**. The rate ratios are ratios of the estimated “treatment” rate to the estimated “control” rate when all rates are **estimated from the same model**. Thus, it is inappropriate for EPA to plot a particular set of rate ratios

based on an assumed model and then compare a second fitted model to these rate ratios. It is inappropriate because the relevant rate ratios for the second fitted model have a different denominator (different “control” rate) than the rate ratios based on the assumed model and hence different rate ratios. Comparing a fitted model to rate ratios based on a different model is inappropriate. That is, rejecting a second fitted model because it does not compare well to the wrong set of rate ratios (as EPA does) is inappropriate.

Valdez-Flores and Sielken (2013) does not advocate a specific model, but rather advocates using the individual data and model selection based on valid comparisons.

For example, the log-linear model does not provide a statistically significant fit to either the lymphoid cancer mortality or breast cancer mortality data sets.

7. EPA inappropriately rejects the fitted log-linear model for lymphoid cancer mortality and for breast cancer mortality because these fitted models do not visually compare well with the wrong categorical rate ratios (i.e., rate ratios based on a different estimated control rate and a different fitted model – as discussed in Comment 6 above).

EPA uses the phrase “does not provide a statistically significant fit” inappropriately. A likelihood ratio test does not test whether a fitted model provides “a statistically significant fit”. Instead, a likelihood ratio test (in this case) is testing whether the fitted model provides a better fit than the same model with the “slope” (or some other parameter) being fixed equal to the null hypothesis value (frequently, zero).

Valdez-Flores and Sielken (2013) ignore the fact that when EPA modeled the continuous exposure data, the supralinear models (e.g., log-linear models with log-transformed exposure and two-piece spline models) fit the data better than linear and sublinear (e.g., log-linear) models, consistent with the supralinear shape suggested by the categorical results.

*8. The likelihood of a composite model – e.g. piecewise linear model – is greater than or equal to the likelihood of a simpler (nested) model – e.g., a linear model. That is, the fact that the likelihood of the two-piece linear model is greater than or equal to the likelihood of the linear model cannot be challenged, but the real scientific question is whether the likelihood of two-piece linear model is **statistically significantly** greater than the likelihood of a linear model (i.e., is there sufficient statistical evidence to justify a more complicated model). It can be shown, that **none** of the piecewise linear models used by EPA fit the lymphoid and breast cancer mortality data statistically significantly better than the linear or log-linear models at the 5% significance level.*

The attenuation in the exposure-response relationship results from the influence of a small number of subjects with very high exposures. Such attenuation is commonly observed in occupational epidemiology studies of cancer (Stayner et al., 2003). Consideration of the categorical data can help avoid the influence of the subjects with very high exposures; in fact, EPA omits the highest exposure category in its linear

regressions of the categorical results. The supralinear log-transformed exposure model similarly mitigates the influence of the subjects with very high exposures.

9. *EPA still considers the supralinear log-transformed exposure model as reasonable model even though the major proponent of this model has rejected it because it is biologically implausible -- see Steenland, K., R. Seals, M. Klein, J. Jinot and H.D. Kahn. 2011, "Risk Estimation with Epidemiological Data When Response Attenuates at High-Exposure Levels", Environmental Health Perspectives 119(6): 831-837. Similarly, other authors have warned about supralinear models when fitting epidemiological data – see Ginevan, M. E. and D. K. Watkins, 2010, "Logarithmic Dose Transformation in Epidemiologic Dose-Response Analysis: Use with Caution," Regulatory Toxicology and Pharmacology 58: 336-340; Crump, K.S., 2005, "The effect of random error in exposure measurement upon the shape of the exposure response," Dose-Response 3: 456–464 (Formerly Nonlinearity in Biology, Toxicology, and Medicine); and Valdez-Flores, C., R. L. Sielken, Jr, and M. J. Teta, 2010, "Quantitative Cancer Risk Assessment Based on NIOSH and UCC epidemiological data for workers exposed to ethylene oxide," Regulatory Toxicology and Pharmacology 56: 312-320.*

Moreover, EPA sees no reason to suppose that there is a problem with the reference group that would result in notable systematic bias. Rothman (1986) (p. 345) notes that the linear regression approach he presents for estimating a continuous slope based on categorical data can be inefficient if the reference category contains small frequencies; such is not the case for the Steenland et al. (2004) data, however. For the breast cancer mortality data set, the reference group contains 40% of the breast cancer deaths, and for the lymphoid cancer data set, the reference group contains about 17% of the deaths.

10. *The 40% of breast cancer deaths in the reference group does not necessarily mean that the frequency of breast cancer deaths in the reference group is not a small frequency. That is, (the number of breast cancer deaths corresponding to 40% of the breast cancer deaths) / (total number of people in the reference group) can still be a small frequency, and Rothman's warning about the effect of small frequencies applies.*

*EPA ignores the other warnings that Rothman (1986) (p. 345) gives about the linear model they used; namely, "By forcing the regression line through the reference point, this regression model effectively places an extremely large weight on the location of that point in relation to the remaining array of points. ... If the reference category has small frequencies, then the **relation of the remaining points to one another will reflect the trend better than their relation to the arbitrarily fixed reference point.**" [emphasis added].*

Furthermore, the reference groups are highly comparable internal comparison groups made up of other workers in the same facilities with short follow-up times since the time of first exposure (e.g., with exposures all within 20 years of follow-up for the breast cancer mortality analyses) who were "lagged out" in the analyses (i.e., assigned zero exposure). There is no reason to expect that these workers differ from those with non-zero exposure estimates with respect to background cancer rates.

In addition, EPA disagrees with the contention that the intercept in the linear regression model should be estimated. In conducting its weighted linear regressions, EPA used the approach of Rothman (1986), in which the RR at the intercept is explicitly fixed at 1 (see Appendix F). The categorical rate for the unexposed reference group, although not defined explicitly, is the best estimate available for the rate in that group, and it is appropriate to fix the baseline rate for the RR estimates at that level.

*11. EPA ignores the other warnings that Rothman (1986) (p. 345) gives; namely “By forcing the regression line through the reference point, this regression model effectively places an extremely large weight on the location of that point in relation to the remaining array of points. ... If the reference category has small frequencies (see also Comment 10 above), then the **relation of the remaining points to one another will reflect the trend better than their relation to the arbitrarily fixed reference point.**” [emphasis added]. In addition, EPA forces the linear model to go through the rate ratio for the first group. What would be the result of the model if the line is forced to go through the RR of any other group? Why does the linear model have to assume that the reference group has to be perfectly predicted at the cost of missing the RRs at other groups?*

When the underlying exposure-response relationship is supralinear, as is the case for these two data sets (see Sections 4.1.1.2 and 4.1.2.2), using the data for all the categorical groups to “estimate” the rate in the unexposed group – as proposed by Valdez-Flores and Sielken (2013) – would result in the data with the high exposures in the plateau region receiving a lot of weight in the linear regression. This would “pull” the linear model down at the higher exposures and concomitantly “pull” the model up at the lower exposures, thus “overestimating” the implicit rate in the unexposed group. In other words, the results for the higher exposure levels would unduly influence the estimate for the unexposed group, yielding a flawed estimate of the baseline rate.

*12. Valdez-Flores and Sielken (2013) does not suggest “using the data for all the categorical groups to ‘estimate’ the rate in the unexposed group.” Valdez-Flores and Sielken (2013) suggest that, as Rothman suggests, “the **relation of the remaining points to one another will reflect the trend better than their relation to the arbitrarily fixed reference point.**” [emphasis added]*

Moreover, Valdez-Flores and Sielken (2013) appear to be using an unweighted linear regression model, which undervalues the fact that a lot of data are reflected in that group (40% of deaths for breast cancer mortality) relative to the exposed groups and exacerbates the over-influence of the high-exposure results. (How one would obtain a weight for the reference group in their approach is unclear because the variability in the RR estimates is built into the estimates for the exposed groups and there is no SE or CI for the RR = 1 value of the reference group.)

13. It seems that EPA misinterprets the weighted least squares fit. For the linear model that goes through the origin, the weight for the unexposed/reference group is infinite, regardless of the number or percentage of deaths in the group. Using an un-weighted procedure does not undervalue the reference group but rather assigns equal weight to all rate ratios. Furthermore, Valdez-Flores and Sielken (2013) state that “EPA used a

weighted least-squares procedure instead of a simpler un-weighted least squares procedure; however, the difference between these two procedures virtually disappears if the number of cancer deaths per non-zero exposure interval is approximately the same.” [emphasis in the original]. Valdez-Flores and Sielken defined exposure intervals with approximately equal number of cancer deaths.

This over-influence of the high-exposure results in the Valdez-Flores and Sielken (2013) approach is illustrated in their findings. In their Table 2, for example, the increasing intercept and decreasing slope values obtained for model 3 with the increasing number of exposure intervals reflect the very high exposure values getting more weight. This is visually depicted in their Figures 2 and 3, which show the high-exposure values “pulling” model 3 down to a greater extent than model 2.

14. EPA’s own reference, Rothman (1986) (p. 345), indicates the reasons for what EPA observed in Table 2 of Valdez-Flores and Sielken (2013) when Rothman states “... the relation of the remaining points to one another will reflect the trend **better** than their relation to an arbitrarily fixed reference point.” [emphasis added]. Model 4 in Valdez-Flores and Sielken (2013) is the model fit to the individual data and provides the best estimate of the trend. Model 3 is a model fitted to the categorical RRs, and the estimate of the trend based on this Model 3 approaches the trend of the best model (model 4) as the number of exposure intervals increases.

Valdez-Flores and Sielken (2013) find affirmation in the fact that the slope of model 3 (the unrestricted log-linear model fit to the categorical results) approaches the slope of the log-linear model fit to the continuous exposure data, but they are aspiring to a (sublinear) continuous exposure model that does not provide a good fit to the data.

15. The log-linear model fit to the individual data is the **best** log-linear model **fit to individual data** based on Cox proportional hazards regression and a maximum likelihood criterion. Although the fit to the individual data is most appropriate, if the modeling is done on the basis of the categorical RRs, then the **unrestricted** log-linear model fits better than EPA’s **restricted** linear model.

As discussed in Comment 6, EPA’s rejection of the best log-linear model fit to individual data is inappropriate.

In EPA’s analyses, the log cumulative exposure model provides a better fit to the continuous exposure data, reflecting the underlying supralinear pattern of the exposure-response relationship.

16. Although the likelihood of the log cumulative exposure model is greater than the likelihood of the log linear model, the main proponent of this log cumulative exposure model has dismissed this model because it is biological implausible -- see Steenland, K., R. Seals, M. Klein, J. Jinot and H.D. Kahn, 2011, “Risk Estimation with Epidemiological Data When Response Attenuates at High-Exposure Levels,” Environmental Health Perspectives 119(6): 831-837. Similarly, other authors have warned about supralinear models when fitting epidemiological data – see Ginevan, M.

E. and D. K. Watkins, 2010, "Logarithmic Dose Transformation in Epidemiologic Dose-Response Analysis: Use with Caution," Regulatory Toxicology and Pharmacology 58: 336-340; Crump, K.S., 2005, "The effect of random error in exposure measurement upon the shape of the exposure response," Dose-Response 3: 456-464 (Formerly Nonlinearity in Biology, Toxicology, and Medicine); and Valdez-Flores, C., R. L. Sielken, Jr, and M. J. Teta, 2010, "Quantitative Cancer Risk Assessment Based on NIOSH and UCC epidemiological data for workers exposed to ethylene oxide," Regulatory Toxicology and Pharmacology 56: 312-320.)

The Breslow and Day (1980) quote cited by Valdez-Flores and Sielken (2013) recognizes that although the scale of the relative risk estimates is arbitrary, the shapes of the curves have meaning; Valdez-Flores and Sielken (2013), however, largely ignore the underlying shape.

*17. Estimates of categorical rate ratios cannot (and should not) be interpreted as the raw data. Comparing fitted models to categorical RRs is not the same as comparing fitted models to the individual data. (See also Comment 6.) Categorical rate ratios are **estimates** of the ratio of hazard rates for different exposure intervals to the hazard rate for a reference group. The reference group is usually selected to be the unexposed group but any group can be used as a reference group. The estimated rate ratios give an indication of the relative hazard rate in one exposure group to the hazard rate in another exposure group.*

There could be a lot of information contained in the individual data that are "absorbed" into the rate ratio estimates (i.e., "lost") and not properly accounted for by models fit to a few rate ratios. For example, age, gender, race, calendar year, co-exposures, job, plant, etc. are variables that often influence the hazard ratios of different exposure groups. Models fit to the individual data, on the other hand, can appropriately account for all of this information and may result in different models than models fit to a few categorical rate ratios. The apparent supra-linearity that EPA observes in the four rate ratios of the lymphoid data looks more like random variation when the number of cumulative exposure intervals and rate ratios is increased—say to 20 or more categories as in the supplementary material for Valdez-Flores and Sielken (2013) which is publicly available on line at <http://dx.doi.org/10.1016/j.yrtph.2013.07.011> and attached for CAAC members' convenience.

EPA's analyses of the effects of excluding the top 5% of exposures confirm the impact of these high-exposure results in dampening the slopes of the log-linear models. For example, Figure D-2c

18. Presumably, EPA meant D-2b.

in Appendix D shows the considerable increase in the slope of the log-linear model when the top 5% of exposures are omitted from the breast cancer mortality data set. As discussed above, the unrestricted intercept approach used by Valdez-Flores and Sielken (2013) is similarly susceptible to the undue influence of a small number of subjects with very high exposures. For unit risk estimation, it is the lower exposure

region that is of greatest interest for low-exposure extrapolation, and EPA's approach of fixing the RR for the unexposed group at 1 and omitting the highest exposure group (due to the plateauing of the responses at high exposure levels) from the linear regression of the categorical results provides a better representation of the exposure-response relationship in the lower exposure range than the approach advocated by Valdez-Flores and Sielken (2013).

Furthermore, EPA disagrees that the categorical results are not useful depictions of the underlying exposure-response relationship. Valdez-Flores and Sielken (2013) themselves note that because there is a separate beta parameter estimate for each exposure group in the categorical model, there is considerable flexibility for the beta parameters to represent changes from the background hazard rate. For this reason, categorical models can be useful to show underlying exposure-response patterns (shapes) in the data. Such patterns can be obscured with single-parameter continuous models, such as the log-linear model, which presupposes a sublinear shape.

19. (As previously stated in Comment 17). *“Estimates of categorical rate ratios cannot, and should not be interpreted as the raw data. Categorical rate ratios are **estimates** of the ratio of hazard rates for different exposure intervals to the hazard rate for a reference group. The reference group is usually selected to be the unexposed group but any group can be used as a reference group. The estimated rate ratios give an indication of the relative hazard rate in one exposure group to the hazard rate in another exposure group. There could be a lot of information contained in the individual data that are “absorbed” into the rate ratio estimates (i.e., “lost”) and not properly accounted for by models fit to a few rate ratios. For example, age, gender, race, calendar year, co-exposures, job, plant, etc. are variables that often influence the hazard ratios of different exposure groups. Models fit to the individual data, on the other hand, can appropriately account for all of this information and may result in different models than models fit to a few categorical rate ratios.”*

EPA also disagrees with the discounting by Valdez-Flores and Sielken (2013) of the apparent supralinear shape of the exposure-response relationship as the result of exposure measurement error based on work by Crump (2005). Exposure misclassification error is a complicated issue, and Crump's (Crump, 2005) conclusions appear to rely on the simplifying assumptions that the measurement error is classical and multiplicative.

20. *Crump (2005) cautions that this may be a problem with any type of measurement error. Crump states “This formulation should be general enough to approximate a wide range of conditions involving random, independent exposure errors. A qualitatively very similar result was obtained assuming a uniform distribution of true exposures (Figure 3). Thus, the effect of random exposure error seems to be in the direction of making low exposures appear more dangerous than they actually are.” Furthermore, Crump (2005) warns of other reasons that the response may be misinterpreted to be supra-linear when he states “In addition to random exposure errors, there are other sources of distortion of the shape of the exposure response. If a study utilizes an inappropriate*

control group in which the response is low compared to that expected in the study population, the exposure response will tend toward an appearance of supra-linearity.”

However, in epidemiological studies, such as the NIOSH ethylene oxide study, in which job-exposure matrices are used to estimate individual worker exposures, exposure measurement error is generally considered to be largely of the Berkson type and is often treated as additive (Armstrong, 1990). Heid et al. (2004) have demonstrated that different assumptions about exposure measurement error can have different impacts on the observed exposure-response relationship. For a specified log-linear relationship, Heid et al. (2004) found that multiplicative classical error could make the observed exposure-response relationship appear supralinear, consistent with the findings of Crump (2005); however, additive classical error dampened the log-linear relationship, multiplicative Berkson error intensified the log-linear relationship, and additive Berkson error had no impact on the log-linear relationship, although the precision would be reduced. Moreover, NIOSH conducted an extensive exposure assessment that included the development of a regression model that had high validity when tested against independent measurement data (Section A.2.8 of Appendix A); thus, the existence of substantial exposure measurement error in the Steenland et al. (2004) data is speculative.

21. EPA failed to document any check on the nature of the exposure errors in the NIOSH data. Although not published as part of our 2013 paper, Sielken and Associates Consulting, Inc., used the techniques suggested by Heid et al. (2004) and the individual data, to show that the measurement errors in the NIOSH data appear to be multiplicative at least for plant and calendar year.

EPA’s statement above suggests that “For a specified log-linear relationship” the measurement errors in the NIOSH data “could make the observed relationship appear supralinear.” Given that the categorical rate ratios can be misleading and can be misinterpreted depending on the shape/form/behavior of measurement errors, the most sensible alternative is to fit a model to the individual data and determine scientifically whether a model fits the data appropriately for the sake of risk assessment.

There are other sources of the apparent supralinearity that EPA “sees” when EPA looks at only a few categorical RRs and mistakenly rejects fitted models by comparing the fitted models to inappropriate categorical RRs (as discussed in Comment 6).