Regulatory Toxicology and Pharmacology 67 (2013) 206–214

Contents lists available at ScienceDirect

Regulatory Toxicology and Pharmacology

journal homepage: www.elsevier.com/locate/yrtph

Misinterpretation of categorical rate ratios and inappropriate exposure–response model fitting can lead to biased estimates of risk: Ethylene oxide case study ☆

Ciriaco Valdez-Flores *, Robert L. Sielken Jr.

Sielken & Associates Consulting Inc., 3833 Texas Avenue, Suite 230, Bryan, TX 77802, USA

ARTICLE INFO

Article history:
Received 20 March 2013
Available online 31 July 2013

Keywords:
Human epidemiological data
Categorical rate ratios
Individual data
Cancer exposure–response model
Goodness-of-fit evaluation
Estimated intercept
Ethylene oxide
Breast cancer
Lymphohematopoietic cancers and lymphoid neoplasms

ABSTRACT

There are pitfalls associated with exposure–response modeling of human epidemiological data based on rate ratios (RRs). Exposure–response modeling is best based on individual data, when available, rather than being based on summary results of that data such as categorical RRs. Because the data for the controls (or the lowest exposure interval if there are not enough controls) are random and not known with certainty a priori, any exposure–response model fit to RRs should estimate the intercept rather than fixing it equal to one. Evaluation of a model's goodness-of-fit to the individual data should not be based on the assumption that summary RRs describe the true underlying exposure–response relationship. These pitfalls are illustrated by Monte Carlo simulation examples with known underlying models. That these pitfalls are a practical concern is illustrated by the need for U.S. EPA to reconsider its most recent evaluation of ethylene oxide. If they had avoided these pitfalls, their exposure–response modeling would have been in better agreement with the log-linear model fit to the individual data.

© 2013 The Authors. Published by Elsevier Inc. All rights reserved.

1. Introduction

Exposure–response modeling is a necessary step in the development of quantitative risk assessments. Exposure–response modeling has been discussed and widely used in the analysis of animal studies (e.g. BMDS 2012, Crump et al. 1976). Researchers and regulatory agencies have used animal-based exposure–response models to fit experimental data. Recently, the availability of epidemiological data and the preference for basing human risk assessments on human data has sparked the development of exposure–response models for epidemiological data. Originally, Poisson regression modeling was widely used, but more recently Cox proportional hazards models have been embraced by researchers and regulatory agencies.

The use of exposure–response models based on epidemiological data results in better characterizations of human risks as the uncertainty in extrapolating from experimental species to humans is eliminated. In addition, the uncertainties due to the extrapolation from high animal exposures to lower human exposure levels and the uncertainties due to the extrapolation from animal exposure protocols to human exposures are reduced. These reductions in uncertainties increase the precision of risk characterizations unless other uncertainties are introduced. Misinterpreting epidemiological data along with a misunderstanding of exposure–response model fitting to epidemiological data can introduce uncertainties and biases to a risk assessment. These uncertainties and biases can result in risk characterizations that may be worse than risk characterizations based on animal bioassay data.

When individual exposure histories, age, calendar year, cause of death, follow up status, and other variables of interest are available, these individual epidemiological data should be used to fit exposure–response models. Analyses of epidemiological data using Poisson regression modeling require cause-specific deaths and person-years of individuals to be grouped into different age and other covariate strata and across different intervals of exposure measures. Each cell defined by the combination of a group and an exposure interval in the Poisson regression modeling defines a hazard rate within that cell. These hazard rates are assumed to be constant throughout the exposure interval and follow a Poisson distribution. Exposure–response models estimating the hazard rates can then be fit to the cell-specific hazard rates using maximum likelihood estimators (e.g., Breslow and Day 1980, 1987).

☆ This is an open-access article distributed under the terms of the Creative Commons Attribution-NonCommercial-ShareAlike License, which permits non-commercial use, distribution, and reproduction in any medium, provided the original author and source are credited.

* Corresponding author. Address: 3833 South Texas Avenue, Suite 230, Bryan, TX 77802, USA. Fax: +1 979 846 2671.

E-mail addresses: SielkenAssoc@aol.com, ciriaco@SielkenAssociates.com (C. Valdez-Flores).

0273-2300/$ - see front matter © 2013 The Authors. Published by Elsevier Inc. All rights reserved.
http://dx.doi.org/10.1016/j.yrtph.2013.07.011

More recently, the Cox proportional hazards models have become more widely preferred. The preference for Cox proportional hazards models over Poisson regressions model when individual epidemiological data are available is due to several reasons: (1) Cox models control optimally for the effect of age on the hazard when the age of each individual is used as the time variable and because, then, there is no need for defining age groups; (2) Checkoway et al. (1989) notes that "the Poisson regression model converges to the proportional hazards model as the age strata are made infinitely small"; (3) Cox models do not require that the data be grouped into exposure intervals; (4) Cox models use the exposure estimates for each individual in the epidemiological data as opposed to averages or other summary exposure estimates for exposure intervals; (5) the assumption of a Poisson distribution within a cell is not necessary as the Cox models do not require any groupings; (6) Cox models do not make any assumptions about the distribution of the survival probability; (7) confounders can be treated as either categorical or continuous variables without having to group the data into different categories and assigning a single value to all observations within the category, and (8) as with Poisson regression models, software systems have implemented modules that can handle a wide variety of Cox proportional hazards models (e.g., Allison 2010 and Cleves et al. 2004 present analyses of epidemiological data using Cox proportional hazards models based on SAS and STATA, respectively).

Even though Cox proportional hazards models have been well documented and widely used, there is still some lack of understanding and misinterpretation of their results. Just recently, the United States Environmental Protection Agency (EPA) (2006, 2011, 2013) discarded Cox proportional hazards models fit to individual epidemiological data and instead used summary data and a weighted least squares fit to the summary data (Rothman, 1986 and Van Wijngaarden and Hertz-Picciotto, 2004) to develop a risk assessment for ethylene oxide (EO). Some of the problems associated with using summary data in the context of meta-analyses are also discussed elsewhere (e.g., Easton et al., 1991; Greenland and Longnecker, 1992; Berlin et al., 1993, and Hartemink, et al. 2006).

1.1. EO cohort study data

In 1991 the National Institute for Occupational Safety and Health (NIOSH) published the results of a cohort **mortality** study with follow up through 1987 of male and female sterilant workers with potential for exposure to EO (Steenland et al., 1991). NIOSH developed job exposure matrices with concentrations for each combination of job and calendar year using measurements of EO concentrations from 1976 to 1985. Exposure concentrations of EO before 1976 were estimated using regression modeling with temporal and spatial variables correlated with EO concentration (Greife et al. 1988 and Hornung et al. 1994). The NIOSH cohort includes workers from 14 different plants. Thirteen of these plants had enough data to develop reliable exposure estimates and their workers were included in the exposure–response analyses reported in Steenland et al. (1991). Steenland et al. (2004) reported exposure–response modeling results based on a NIOSH data set that was more recently updated to include follow up through the end of 1998, for an average follow up of 26 years. The same job exposure matrix used for the original cohort was used for the updated cohort. The job exposure matrix for the updated NIOSH data set assumed that workers active at the end of 1987 (the end of the original follow up period) stayed working on the same job until their end of employment. This assumption did not substantially affect the cumulative exposures to EO in the updated NIOSH data set over the previous data set because concentrations after the mid 1980's, following reduction in the OSHA PEL and ACGIH TLV-TWA, were much lower than the concentrations in earlier years. A total of 17,493 workers in the updated NIOSH data set were followed up and had sufficient exposure information for exposure–response modeling. The NIOSH data used for the analyses presented herein includes 9,859 female, 7,634 male, 13,761 white and 3,732 non-white workers.

Although there were no workers with zero cumulative exposures to EO in the NIOSH study, the cumulative exposure was lagged 20 years for breast cancer and lagged 15 years for other endpoints resulting in several individuals with zero cumulative exposure once the lags were incorporated.

In another paper, Steenland et al. (2003) published summary results for breast cancer incidence among women in the NIOSH study. The EPA used breast cancer **incidence** as a response in some of its analyses. However, the NIOSH individual incidence data were not available to the authors; so that, the results published by Steenland and colleagues in 2003 could not be reproduced by us. Consequently, analyses of the breast cancer incidence in the NIOSH study are not discussed herein.

The Union Carbide Corporation (UCC) and Dow Chemical Corporation (Dow) have also developed an epidemiological data set with EO exposure information and length of follow up sufficient for exposure–response modeling (Greenberg and Ott, 1990, Teta et al. 1993, and Swaen et al. 2009). This UCC/Dow epidemiological data set has not been used by the EPA for the development of its risk assessment of EO. In order to focus on the analyses performed by EPA, the UCC/Dow epidemiological data are not included in the analyses presented herein.

2. Methods

Apparently, EPA's primary reason for their approach to modeling the NIOSH data in 2006, 2011, and 2013 was due to a lack of understanding of the consequences of their modeling approach and the underlying background hazard rates. Although these background rates are not explicitly estimated by the Cox proportional hazards models (Allison 2010), the background hazard rates do affect the estimation of the categorical rate ratios (RRs) and the RRs based on other models (as discussed in more detail in Allison (2010) and our sections on "Methods: Categorical Model", "Methods: Continuous Log-Linear Model", and "Results"). EPA seems to have fallen victim to the misconception that the epidemiological data ***are*** the categorical RRs. Then, EPA's model fit to these categorical RRs with the intercept fixed equal to one seemed to EPA to be a good fit to the *underlying epidemiological data* because their model with the intercept fixed equal to one provided a "good visual fit" to the categorical RRs whereas the model fit to the individual data was a "poor visual fit" and "not even close" to the categorical RRs.

In order to determine the impact of misinterpretation of exposure–response relationships, the NIOSH individual epidemiological data have been used to perform several analyses that mimic those used in EPA's risk assessment published in August 2006 and in draft form in July 2011 and in a Revised External Peer Review draft in 2013.

First, the individual NIOSH data on breast cancer mortality among the female workers in the NIOSH study were fit using a log-linear exposure–response model. (The models fit to the individual NIOSH data for breast cancer mortality are adjusted for race and assume a 20-year lag period in the cumulative exposure.) Then, we compare these results with the risk assessments that EPA published in August 2006, 2011, and 2013. We show that the exposure–response models fit to summary data (here categorical RRs) can be very different than the exposure–response models fit to the individual epidemiological data. These differences, however, can be reduced if the summary data are correctly interpreted and exposure–response models are appropriately fitted.

We begin by defining the exposure–response models and their associated background hazard rates.

2.1. Methods: categorical model

A categorical log-linear Cox proportional hazards model was fit to the NIOSH individual data. (A categorical "log-linear" model is identical to the "categorical model" used by Steenland (2010) to generate categorical RRs for EPA. However, we have added the term "log-linear" to emphasize the fact that both the continuous exposure and categorical exposure models have the same basic mathematical structure; i.e., the assumed background hazards rate is additive on the logarithmic scale and multiplicative on the linear scale.) To fit a categorical exposure model, cumulative exposure intervals were defined so that the number of breast cancer mortalities was approximately equally divided among the non-zero exposure intervals. This categorical log-linear model can be written as;

$$\mathrm{Log}[\mathrm{HR_i}(t)] = \alpha_{\mathrm{cat}}(t) + \beta_{\mathrm{i}} \tag{1}$$

where $\mathrm{HR_i}(t)$ is the hazard rate in the i-th exposure interval, $\alpha_{\mathrm{cat}}(t)$ is the logarithm of the unspecified background hazard rate, β_{i} is the change in the logarithm of the hazard rate from $\alpha_{\mathrm{cat}}(t)$ in the i-th exposure group and is estimated from the data. This expression can be rewritten as;

$$\mathrm{HR_i}(t) = \lambda_{\mathrm{cat}}(t) \times e^{\beta_{\mathrm{i}}} \tag{2}$$

where $\lambda_{\mathrm{cat}}(t) = \exp[\alpha_{\mathrm{cat}}(t)]$ is the assumed unspecified background hazard rate for the controls. The Cox proportional hazards model does not explicitly estimate the background hazard rate $\lambda_{\mathrm{cat}}(t)$ and does not make any specific assumptions about the functional form of $\lambda_{\mathrm{cat}}(t)$. That is, the Cox proportional hazards model only explicitly estimates the β_{i} that correspond to the changes in the hazard rate and define the rate ratio; that is,

$$\mathrm{RR_i} = \mathrm{HR_i}/\lambda_{\mathrm{cat}}(t) = e^{\beta_{\mathrm{i}}} \tag{3}$$

Because there is a separate β_{i} for each exposure group, there is considerable flexibility for the β's to represent the changes from the background hazard rate $\lambda_{\mathrm{cat}}(t)$. This minimizes the dependence of the β's on the $\lambda_{\mathrm{cat}}(t)$. As a result, the background hazard rate $\lambda_{\mathrm{cat}}(t)$ in the model, although not explicitly estimated in the Cox proportional hazards model, depends mainly on the hazard rate of the workers outside of the exposure groups corresponding to the β's, that is, the control or minimally exposed workers.

2.2. Methods: continuous log-linear model

Similarly, the continuous log-linear exposure–response Cox proportional hazards model fit to the same individual epidemiological data has an additive background hazard rate on the logarithmic scale but a multiplicative background hazard rate on the linear scale; that is,

$$\mathrm{Log}[\mathrm{HR}(t)] = \alpha_0(t) + \beta \times \mathrm{CumExp}(t) \tag{4}$$

where $\mathrm{HR}(t)$ is the hazard rate at age t, $\alpha_0(t)$ is the logarithm of the underlying unspecified background hazard rate at age t, β is the change in the logarithm of the hazard rate with a unit increase in the cumulative exposure and is estimated from the data, and $\mathrm{CumExp}(t)$ is the age-dependent cumulative exposure at age t and is calculated for each event time (i.e., for each risk set generated in the evaluation of the partial likelihood of the Cox proportional hazards model). This expression can be rewritten as;

$$\mathrm{HR}(t) = \lambda_0(t) \times e^{\beta \times \mathrm{CumExp}(t)} \tag{5}$$

where $\lambda_0(t) = \exp[\alpha_0(t)]$ is the assumed unspecified background hazard rate at age t. The Cox proportional hazards model does not explicitly estimate the background hazard rate $\lambda_0(t)$ and does not make any specific assumptions about the functional form of $\lambda_0(t)$. That is, the Cox proportional hazards model explicitly estimates the slope β that defines the change in the rate ratio with cumulative exposure; that is,

$$\mathrm{RR}(t) = \mathrm{HR}(t)/\lambda_0(t) = e^{\beta \times \mathrm{CumExp}(t)} \tag{6}$$

Because one value of β has to characterize the changes over the entire exposure range, there is considerably less flexibility (compared to the multiple β_{i}s in the categorical model) for this one β to represent the changes from the background hazard rate $\lambda_0(t)$. As a result, the assumed background hazards rate $\lambda_0(t)$, although not explicitly estimated in the Cox proportional hazards models, depends heavily on the functional form of the model fit to the data. That is, the background hazard rate $\lambda_0(t)$ depends not only on the hazard rate of the control or minimally exposed workers but also depends on the hazard rate of all of the exposed workers and the functional form of the representation of that hazard rate.*The major point here is that the implicit value of the background hazard rate in the categorical model$\lambda_{cat}(t)$ can be very different than the implicit value of the background hazard rate in the continuous log-linear model$\lambda_0(t)$.*

Other authors have explained the differences in these implicit background rates in different ways. For example, Easton et al. (1991) states that "A fundamental statistical principle, that a satisfactory summary of the data should include the sufficient statistics, underlies the inadequacy of the common practice of describing data on s categories by s-1 relative risks and their standard errors, a total of 2s-2 parameters." Thus, here, characterizing the individual data by only the RRs is inadequate.

Another description of the inadequacy of the RRs is that, when modeling exposure–response relationships from published categorical RRs, it is important to realize that all the ratios from a single study share the same denominator, namely the rate in the reference category. This means that when the number of cases in the reference category is higher than expected (based on the true incidence in this category), all rate ratios will be too low. There is systematic bias in that the RRs will all be too low with the same fraction. Similarly, if the number of cases in the reference category is lower than expected, all rate ratios will be too high. For example, when the number of cases in the reference group is by accident only half of the expected number, all RRs are expected to be twice as high as the true ratio.

3. Results

The continuous log-linear model (Model 4 as defined by Eq. 4) was fit to the individual NIOSH breast cancer mortality data. A categorical model (Model 1 as defined by Eq. 1) was also fit to the same individual data. The RRs estimated using this categorical model (Model 1) were fit using a least-squares procedure and a **restricted** continuous linear model (Model 2: EPA) similar to the models EPA used in their 2006, 2011, and 2013 risk assessments of EO. Model 2 is a linear model that EPA forced to go through an RR of one at zero cumulative exposure. In addition, an unrestricted continuous log-linear model (Model 3) (similar to the model fit to the individual data) was fit to the categorical RRs using least squares but not restricted to go through an RR of one at zero cumulative exposure.

Model 2 was a restricted linear model, rather than a restricted log-linear model, in order to parallel the analyses preformed by EPA. Similar results would have been obtained if a restricted log-linear model had been used. Similarly, 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.

Table 1
Rate ratios (RRs) estimated using four exposure intervals.

Cumulative exposure interval (ppm-days)	Breast cancer deaths	RRs from model 1[a]	RRs from model 2: EPA[b]	RRs from model 3[c]	RRs from model 4[d]
Controls	41	1.000	1.000	1.000	1.000
(0, 703.5]	16	1.232	1.041	1.021	1.003
(703.5, 2634.6]	15	1.421	1.195	1.104	1.016
(2634.6, 10900.0]	15	1.562	1.790	1.494	1.066
>10900.0	15	2.827	2.754	2.439	1.152

[a] Categorical model fit to individual data.
[b] Continuous linear model used by EPA with restricted intercept fit to categorical RRs from model 1.
[c] Log-linear model with unrestricted intercept fit to RRs from model 1.
[d] Continuous log-linear model fit to individual data.

Table 1 lists the estimated RRs from the four models of the NIOSH breast cancer mortality data. The four non-zero cumulative exposure intervals used in Table 1 were chosen so that there were approximately the same numbers of breast cancer mortalities in each non-zero cumulative exposure interval. Results were also obtained with the number of cumulative exposure intervals increased from 4 to 20 and then further increased to 61, respectively (corresponding tables are in the Supplementary Material). The 61 exposure intervals were chosen so that there was exactly one breast cancer mortality in each non-zero cumulative exposure group. Table 2 shows the parameter estimates for two different models (Models 2 and 3) fit to the categorical RRs as well as the continuous log-linear model (Model 4) fit to the individual NIOSH data.

The results in Table 1 show that the restricted linear model (Model 2: EPA) estimates of RR as a function of the cumulative exposure are close to the categorical RRs from the categorical model fit to the individual data (Model 1). Similarly, the unrestricted log-linear model (Model 3) estimates are close to the categorical RRs. However, the RRs calculated from the unrestricted log-linear model (Model 3) estimates are less than both the categorical RRs (from Model 1) and the restricted linear model (Model 2: EPA) estimates. It is not surprising that the RRs from the unrestricted log-linear model are less than the RRs from the restricted linear model because the two models do not have the same estimates of the background; that is, the two models have different intercept values (Table 2). Similarly, the estimates from the log-linear model fit to the individual NIOSH data (Model 4) are entirely below the categorical RRs for the very same reason; that is, as discussed in the Methods Section, *Models 1 and 4 have different unspecified background hazard rates*. Although these background hazard rates are not estimated by the Cox proportional hazards models, *the differences in the background hazard rates and the corresponding RRs and the impact of these differences on the fitted exposure–response models contributed to EPA':s difficulties in correctly modeling the NIOSH data.*

Characteristics similar to those observed in Table 1 are observed when the number of cumulative exposure intervals goes from four to 20 and 61. However, as the cumulative exposure intervals are refined to include a larger number of exposure groups, the RRs calculated from the unrestricted log-linear model (Model 3) fit to the categorical RRs (from Model 1) approach the RRs calculated from the continuous log-linear model (Model 4) fit to the individual NIOSH data using Cox proportional hazards methods. This is also seen in Table 2 where the slope of the unrestricted log-linear model (Model 3) fit using least squares to the categorical RRs (from Model 1) approaches the slope of the continuous log-linear model fit using Cox proportional hazards to the individual NIOSH data (Model 4) as the number of cumulative exposure intervals increases. However, this convergence does not happen when the intercept is restricted to being one for zero exposure (Model 2: EPA).

Figs. 1–3 show the same results for the number of cumulative exposure intervals being 4, 20, and 61, respectively. In these figures, the categorical RRs from Model 1 are plotted as well as the estimates of RR as a function of cumulative exposure from the restricted linear model (Model 2: EPA) fit to the categorical RRs, the unrestricted log-linear model (Model 3) fit to the categorical RRs, and the continuous log-linear exposure–response model (Model 4) fit to the individual NIOSH data. Because the intercept in Model 2 is one, the RRs calculated from Model 2 are the same as the model's estimates of RR as a function of cumulative exposure. However, because the intercept in Model 3 is greater than one, the RRs calculated from Model 3 are the less than the model's estimates of RR as a function of cumulative exposure and closer to the RR's calculated from Model 4 which was fit to the individual data rather than the categorical RRs. Furthermore, as the number of exposure intervals increases from four, to 20, to 61 in Figs. 1–3, respectively, the RRs calculated from Model 3 become almost identical to the RR's calculated from Model 4. In other words, as the number of exposure intervals increases, the intercept in Model 3 estimates the ratio of the Model 4 background hazard rate to the Model 1 background hazard rate. Thus, Model 4 which was fit to the individual data is implying the correct relationship between RR and cumulative exposure. Model 4 only "appears" to be incorrect because it is being compared to the categorical RRs (Model 1) and a model (i.e., Model 2: EPA with an intercept restricted to be one) with a different background hazard rate.

4. Discussion

Although we discuss the misinterpretation by EPA of the assessment of exposure–response model fitting using visual inspection specifically for the case of the risk assessment of EO, this problem has been anticipated by some other authors in similar situations. For example, Breslow and Day (1980) recognize the problem of comparing categorical or qualitative relative risks with the relative risks estimated using a continuous linear model when they observe in reference to a figure that "The estimated relative risks from the qualitative analysis lie entirely above those based on the log transform, which in turn lie above those derived from the linear model. The explanation for this apparently bizarre phenomenon is not hard to find. It is due to the arbitrary selection of 0 as a baseline value for tobacco, which constrains all three curves to pass through

Table 2
Parameter estimates for the Cox proportional hazards log-linear model fit to the breast cancer mortality in NIOSH epidemiological data and the models fit to the categorical RRs.

Model	Number of cumulative exposure intervals	Intercept	Slope
Cox proportional log-linear hazards model fit to NIOSH individual cohort data (Model 4)		Unspecified (not estimated)	9.42E-06
Unrestricted log-linear model fit to categorical RRs using least squares (Model 3)	4	1.13	5.93E-05
	20	1.62	1.08E-05
	61	1.96	9.68E-06
Restricted linear model fit to categorical RRs using least squares (Model 2: EPA)	4	Fixed at 1.00	1.17E-04
	20	Fixed at 1.00	4.77E-05
	61	Fixed at 1.00	6.98E-05

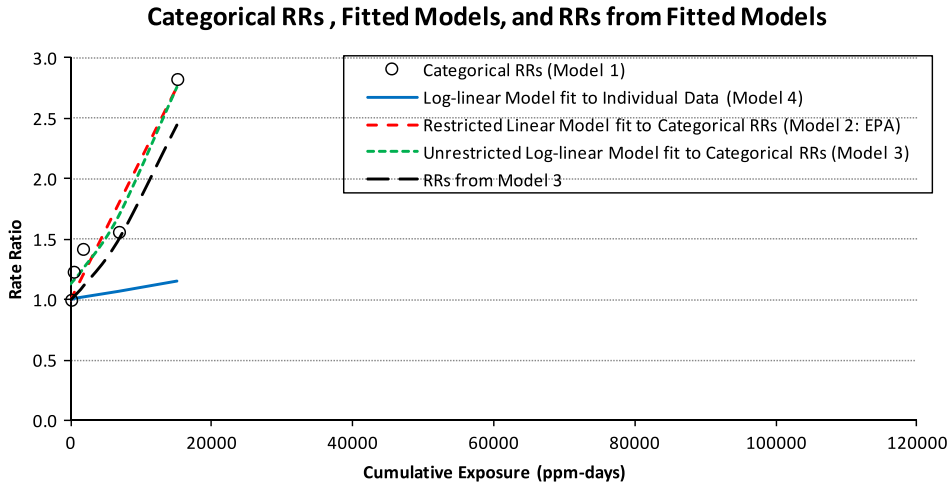

Fig. 1. Categorical RRs (Model 1) and continuous log-linear RRs (Model 4) estimated from the individual NIOSH epidemiological data versus the restricted linear model (Model 2: EPA) and the unrestricted log-linear model (Model 3) fitted to the categorical RRs when the data were split into four categories.

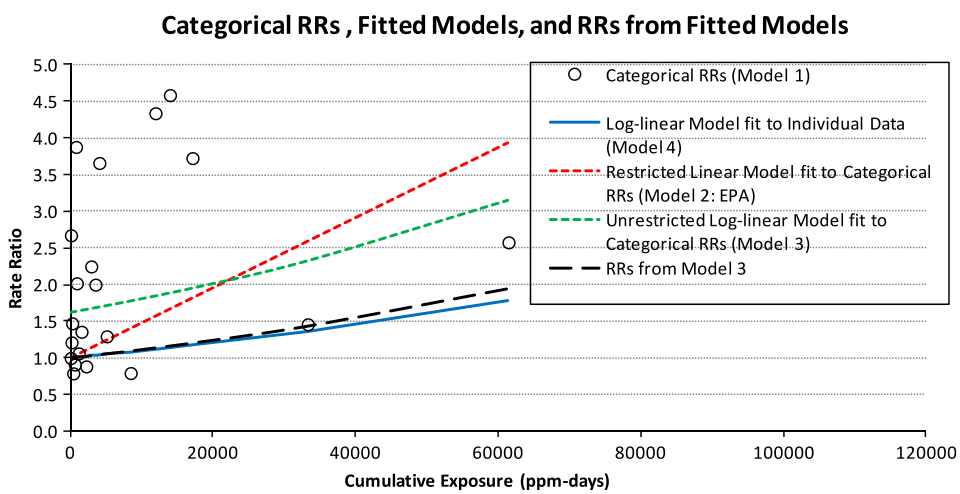

Fig. 2. Categorical RRs and continuous log-linear RRs estimated from the individual NIOSH epidemiological data versus the restricted linear model and the unrestricted log-linear model fitted to the categorical RRs when the data were split into 20 categories.

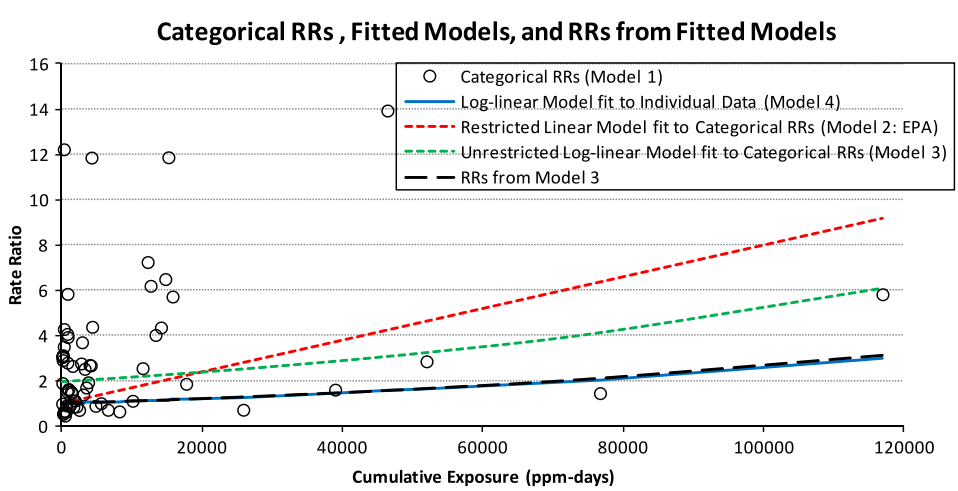

Fig. 3. Categorical RRs and continuous log-linear RRs estimated from the individual NIOSH epidemiological data versus the restricted linear model and the unrestricted log-linear model fitted to the categorical RRs when the data were split into 61 categories.

the origin of the graph. Any other value for tobacco could just as well have been chosen as baseline and assigned a 0 log relative risk, in which case the curves would all be displaced so as to pass through 0 at that point. In other words *the origin of the scale of log-relative risk is completely arbitrary and it is only the shapes of the curves which have any meaning*" [emphasis in the original].

The qualitative analysis cited by Breslow and Day (1980) refers to the categorical relative risks, rate ratios, or odds ratios that are usually published in the literature.

Although there are methods that can be appropriately used to fit exposure–response models when only summary data are available, it is important to emphasize that exposure–response modeling based on summary data can never be superior to exposure–response modeling based on individual epidemiological data. The methods that can be used to fit summary epidemiological data depend on the type of summary data. The TCEQ (2012) guidelines for developing toxicity factors present alternative methods for exposure–response modeling based on published epidemiological results when only limited information is available. The best method uses all the available information, so that exposure–response models based on more information are preferable. For example, Haney et al. (2012) used different methods, depending on the epidemiological data available, to fit exposure–response-models that were used to develop unit risk factors for nickel.

Ideally, individual epidemiological data should be used to fit exposure–response models using state-of-the-art modeling techniques like Cox proportional hazards models. If only summary epidemiological data are available, then the best available data should be used. Under 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 extracted from the individual data. A model based on summary data is dependent on the idiosyncrasies associated with the data summarization assumptions; e.g., exposure-interval categorizations, number of exposure intervals, characterizations of exposure in these intervals, stratifications used, definition of control group, etc.

The apparent discrepancies seen between the categorical RRs (from Model 1) and the continuous log-linear exposure–response model's (Model 4) fit to the individual NIOSH data are mainly due to the different non-estimated underlying background hazard rates assumed by the two different models. In addition, the apparent discrepancy seems more dramatic with the smaller number of exposure intervals selected. Figs. 1–3 show that the groupings of mortality deaths into only a few cumulative exposure intervals results in an oversimplification of the underlying data and masks the true exposure–response relationship.

In addition, any apparent supra-linear behavior of the categorical RRs is not surprising and actually is expected in epidemiological studies that usually include exposure errors. Crump (2005) has investigated this behavior and concluded that "Because of these potential distortions of the exposure–response shape, one should be cautious in drawing conclusions about the shape of the exposure response from epidemiological data. Since even random, unbiased errors in exposure measurement will convert a linear exposure response, and can a convert sub-linear response, into a seemingly supralinear shape, one should be particular[ly] cautious about concluding an exposure–response is truly supra-linear. In particular, it could be inadvisable to extrapolate an observed supra-linear exposure response to low exposures to predict human risk."

In its risk assessments of EO, EPA used visual comparisons of plots of categorical RRs and the RRs from alternative exposure–response models fit to the individual NIOSH data to determine the appropriateness of these alternative exposure–response models. EPA stated "In any case, models with relatively sparse data may not achieve conventional statistical significance (at the 0.05 level) but still provide a good fit to the data, judged by conformity with categorical and cubic spline analysis, and may still be useful for risk assessment." This visual inspection of the RRs presumes that the categorical RRs and/or cubic spline analysis describe the true exposure–response relationship in the NIOSH data. This presumption is incorrect because the categorical RRs and the cubic spline analysis are themselves models fit to the individual data and are not the true exposure–response relationship; so that, the apparent exposure–response shape of categorical RRs can be misleading (Crump 2005). In addition, EPA's visual comparison of the categorical RRs from Model 1 versus the RRs from the log-linear model (Model 4) assumes that the underlying background hazard rates are identical in the two models. The RRs estimated from continuous exposure–response models are not comparable to categorical RRs because the categorical RRs reflect their own implicit estimate of the underlying background hazard rate that can, and usually is, different than the underlying background hazard rate of other exposure–response models (e.g., linear, log-linear, etc.) (Breslow and Day 1980).

Results parallel to those obtained for breast cancer mortality were obtained for lymphohematopoietic cancers and lymphoid neoplasms and are shown in the Supplementary Material. Analyses for breast cancer **incidence** could not be performed because the individual NIOSH incidence data were not available to the authors. We, however, anticipate similar results for the analyses of incidence data because the categorical RRs and exposure–response models for breast cancer incidence presented by EPA have similar behavior to those for breast cancer mortality, lymphohematopoietic cancer mortality and lymphoid cancer mortality.

4.1. Analogous results from simulation studies

In order to further illustrate some of the problems associated with exposure–response modeling based on categorical rate ratios (RRs), several examples were generated using an Excel-based Monte Carlo simulator.

Example 1 refers to RRs where the comparison group is the lowest exposure group and this group contains some non-zero exposures. In this example, 10,000 observations were generated from a continuous linear regression model

$$Y = \alpha \times (1 + \beta \times \text{exposure}) + \varepsilon \tag{7}$$

where y is a continuous variable (not an RR), α and β were specified constants, and ε was a normally distributed error with mean zero.

Table 3
Simulated data and observed rate ratios (RRs) corresponding to 10,000 observations from a continuous linear regression model: example 1.

Exposure range for RR				Count in range	Average exposure in range	Average response in range	Rate ratio (RR)[d]
LB[b]	UB[c]	LB	UB				
0-th Perc[a]	20-th Perc	0	4.47	2000	2.15	5.28	1.00
20-th Perc	40-th Perc	4.47	10.14	2000	7.22	15.44	2.93
40-th Perc	60-th Perc	10.14	18.35	2000	13.96	28.93	5.48
60-th Perc	80-th Perc	18.35	32.03	2000	24.30	49.61	9.40
80-th Perc	100-th Perc	32.03	191.97	2000	52.87	106.74	20.22

[a] Percentiles (Perc) in the 10,000 simulated observations.
[b] Lower Bound.
[c] Upper Bound.
[d] RR = average response in the exposure interval/average response in the first (reference) exposure interval.

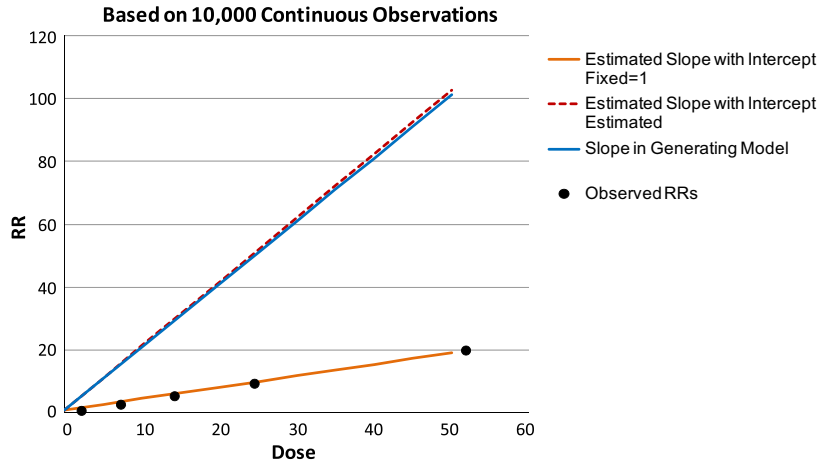

Fig. 4. Example 1 of the difference between fixing the intercept equal to one and estimating the intercept.

The simulated data and corresponding RRs for the first example are shown in Table 3. The results of fitting a continuous linear regression model to these data both with and without a fixed intercept are shown in Fig. 4. The average exposure in each exposure range in Table 3 is taken as the representative exposure value ("*x*-values") in the regression. In this example, the predicted RRs based on the slope in the model with the intercept estimated are very similar to the RRs in the generating model (almost super-imposed in Fig. 4) but very different than the RRs based on the slope in the model with the intercept fixed equal to one. There are three reasons for the differences. First, the lowest exposure group does not contain only zero exposures, hence the observed RRs do not reflect a comparison of zero exposure to a positive exposure. Second, the average response in the lowest exposure group is variable and not fixed; hence, the RRs are not a ratio based on a denominator that is known with certainty. Third, when the continuous linear regression model is estimated with the intercept forced to be equal to one, the first RR (which has been arbitrarily set equal to one) is given infinite weight compared to the other RRs even though all of the RRs are based on observed responses which are random variables. Even though the slope in the estimated continuous linear regression model with the intercept forced equal to one "fits" the observed RRs very well (in this example), the slope in the fitted model does not match the slope in the generating model. This is because the model is not fitting the underlying response data themselves but rather is fitting one summary characterization (the RRs) of that data. Fig. 4 also provides a good example of a common misinterpretation; namely, that, if the generating model were unknown and not plotted, then it would appear that the fitted model with the intercept fixed equal to one provides a "good fit" to the data whereas the fitted model with the intercept being estimated is a "poor fit" and "not even close."

Example 2 is analogous to example 1, except that in example 2 the comparison group contains only zero exposures. The simulated data and observed RRs for the second example are shown in Table 4. The results of fitting the continuous linear regression model to these data both with and without a fixed intercept are shown in Fig. 5. Fig. 5 provides another good example of the common misinterpretation that the fitted model with the intercept fixed equal to one provides a "good fit" to the data whereas the fitted model with the intercept being estimated provides a "poor fit." As in the first example, the predicted RRs based on the slope in the model with the intercept estimated are very similar to the RRs in the generating model but very different than the RRs based on the slope in the model with the intercept fixed equal to one. Unlike the first example, the predicted RRs based on the slope in the model with the intercept estimated are less than (rather than greater than) the predicted RRs based on the slope in the model with the intercept fixed equal to one. Similarly, the slope in the fitted model with the intercept estimated is less than the slope in the fitted model with the intercept fixed equal to one. Thus, considering both example 1 and example 2, it is clear that estimating the intercept does not necessarily result in RRs either greater than or less than the RRs based on fixing the intercept equal to one. Furthermore, in both examples 1 and 2, the slope in the fitted model with estimated intercept does a better job of estimating the slope in the generating model than the slope in the fitted model with the intercept fixed equal to one does. Examples 1 and 2 together indicate that, regardless of whether or not the comparison group contains some non-zero exposures, the problem of estimating the slope in

Table 4
Simulated data and observed rate ratios (RRs) corresponding to 10,000 observations from a continuous linear regression model with controls: example 2.

Exposure range for RR				Count in range	Average exposure in range	Average response in range	Rate ratio (RR)[d]
LB[b]	UB[c]	LB	UB				
0-th Perc[a]	20-th Perc	0	0	3020	0.00	0.67	1.00
20-th Perc	40-th Perc	0	3.13	980	1.52	4.58	6.80
40-th Perc	60-th Perc	3.13	11.30	2000	6.96	14.58	21.62
60-th Perc	80-th Perc	11.30	25.13	2000	17.43	35.74	53.01
80-th Perc	100-th Perc	25.13	201.11	2000	45.22	91.56	135.82

[a] Percentiles (Perc) in the 10,000 simulated observations.
[b] Lower Bound.
[c] Upper Bound.
[d] RR = average response in the exposure interval/average response in the first (reference) exposure interval.

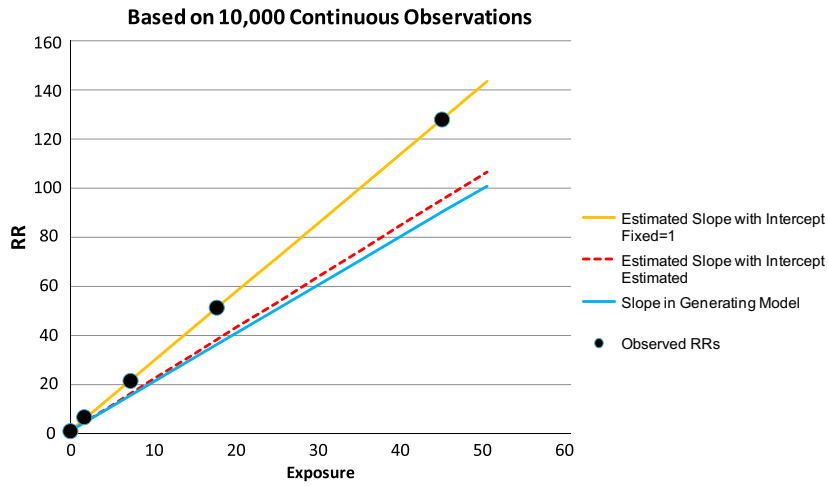

Fig. 5. Example 2 of the difference between fixing the intercept equal to one and estimating the intercept.

the generating model occurs when the exposure–response model is fitted with the intercept fixed equal to one.

When the underlying exposure–response relationship is more non-linear than in examples 1 and 2, fixing the RR = 1 for the lowest exposure group, or fixing the RR = 1 for the second lowest exposure group, etc. substantially changes the fitted continuous linear regression model "with fixed intercept" (i.e., with the fitted model forced to pass through one at the first exposure, or one at the second exposure, etc.) even when the fitted models are rescaled back to the original scale. This does not happen when the intercept is estimated rather than fixed. Thus, the somewhat arbitrary choice of which group to use as the comparison group substantially impacts the slopes of the predicted RRs when the model fit is forced to go through one at zero exposure. (The Supplementary Material contains an illustrative example.)

Therefore, when the RRs are the only available data (as opposed to the individual data), the biggest problem in the exposure–response modeling is fixing the intercept (i.e., forcing the fitted model to pass through a specified RR at a specified exposure) instead of estimating the intercept.

5. Conclusion

There are both general pitfalls and important current practical concerns associated with exposure–response modeling of human epidemiological data based on RRs. When individual data are available, exposure–response modeling is best based on that data rather than being based on summaries of that data such as the RRs. Because the data for the controls (or the lowest exposure interval) are random and not known with certainty a priori, any exposure–response modeling should estimate the intercept in the RR model rather than fixing the intercept equal to one. Summary RRs from a categorical model may be based on a different background hazard rate than the background hazard rate in a continuous model. Evaluation of a model's goodness-of-fit to the individual data should not be based on the assumption that summary RRs describe the true underlying exposure–response relationship. That these pitfalls are a practical concern is illustrated by the need for U.S. EPA to reconsider its most recent evaluation of ethylene oxide.

Conflict of interest statement

The authors are exposure-response assessment consultants to both EO chemical and sterilant trade groups. Funding for this research and its publication was received from the Ethylene Oxide Sterilant Association (EOSA) and the American Chemistry Council (ACC).

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.yrtph.2013.07.011.

References

Allison, P.D., 2010. Survival Analysis Using SAS: A Practical Guide, Second ed. SAS Institute, Inc., Cary, NC, USA.
BMDS. 2012. Benchmark Dose Software. United States Environmental Protection Agency. Available online at: http://www.epa.gov/ncea/bmds/.
Berlin, J.A., Longnecker, M.P., Greenland, S., 1993. Meta-analysis of epidemiologic dose-response data. Epidemiology 4, 218–228.
Breslow, N.E., Day, N.E., 1980. Statistical methods in cancer research. The Analysis of Case-Control Studies, vol. I. International Agency for Research Cancer, Lyon, France.
Breslow, N.E., Day, N.E., 1987. Statistical methods in cancer research. The Design and Analysis of Cohort Studies, vol. II. International Agency for Research Cancer, Lyon, France.
Checkoway, H., Pearce, N.E., Crawford-Brown, D.J., 1989. Research Methods in Occupational Epidemiology. Oxford University Press, New York, New York.
Cleves, M.A., Gould, W.W., Gutierrez, R.G., 2004. An Introduction to Survival Analysis Using STATA, Revised ed. STATA Corporation, College Station, TX, USA.
Crump, K.S., Hoel, D.G., Langley, C.H., Peto, R., 1976. Fundamental carcinogenic processes and their implications for low dose risk assessment. Cancer Res. 36, 2973–2979.
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.).
Easton, D.F., Peto, J., Babiker, A.G.A., 1991. Floating absolute risk: an alternative to relative risk in survival and case-control analysis avoiding an arbitrary reference group. Stat. Med. 10, 1025–1035.
Greenberg, H.L., Ott, M.G., 1990. Men assigned to ethylene oxide production or other ethylene oxide related chemical manufacturing: a mortality study. Br. J. Ind. Med. 47, 221–230.
Greenland, S., Longnecker, M.P., 1992. Methods for trend estimation from summarized dose-response data, with applications to meta-analysis. Am. J. Epidemiol. 135, 1301–1309.
Greife, A.L., Hornung, R.W., Stayner, L.T., Steenland, K., 1988. Development of a model for use in estimating exposure to ethylene oxide in a retrospective cohort mortality study. Scand. J. Work Environ. Health 14, 29–30.
Haney Jr., J.T., McCant, D.D., Sielken Jr., R.L., Valdez-Flores, C., Grant, R.L., 2012. Development of a unit risk factor for nickel and inorganic nickel compounds based on an updated carcinogenic toxicity assessment. Regul. Toxicol. Pharmacol. 62, 191–201.
Hartemink, N., Boshuizen, H.C., Nagelkerke, N.J.D., Jacobs, M.A.M., van Houwelingen, H.C., 2006. Combining risk estimates from observational studies with different exposure cutpoints: a meta-analysis on body mass index and diabetes type 2. Am. J. Epidemiol. 163, 1042–1052.
Hornung, R.W., Greife, A.L., Stayner, L.T., Steenland, 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. Am. J. Ind. Med. 25, 825–836.
Rothman, Kenneth.J., 1986. Modern Epidemiology. Little, Brown and Company, Boston, MA/Toronto, Canada, pp. 341–344.
Steenland, K., 2010. Re-analyses and interpretation of ethylene oxide exposure-response data. Report submitted to EPA on May 27, 2010. Published as Appendix D in EPA 2011.
Steenland, K., Stayner, L., Griefe, A., Halperin, B., Hayes, R., Hornung, R., Nowlin, S., 1991. A cohort mortality study of workers exposed to ethylene oxide. N. Eng. J. Med. 324 (20), 1402–1407.
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. Occup. Environ. Med. 61, 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. Cancer Causes Control 14, 531–539.
Swaen, G.M., Burns, C., Teta, M.J., Bodner, K., Keenan, D., Bodnar, C.M., 2009. Mortality study update of ethylene oxide workers in chemical manufacturing: a 15 year update. J. Occup. Environ. Med. 51 (6), 714–723.
Teta, M.J., Benson, L.O., Vitale, J.N., 1993. Mortality study of ethylene oxide workers in chemical manufacturing: a mortality study. Br. J. Ind. Med. 50, 704–709.
Texas Commission on Environmental Quality (TCEQ). 2012. TCEQ Guidelines to Develop Toxicity Factors. Toxicology Division, Texas Commission on Environmental Quality. RG-442, Revised October 2012.
United States Environmental Protection Agency (EPA). 2006. Evaluation of the Carcinogenicity of Ethylene Oxide. External Review Draft, EPA/635/R-06/003, August 2006, U.S. Environmental Protection Agency, Washington, DC.
United States Environmental Protection Agency (EPA). 2011. Evaluation of the Inhalation Carcinogenicity of Ethylene Oxide. In Support of Summary Information on the Integrated Risk Information System (IRIS), Final Agency Review/Interagency Science Discussion draft, EPA/600/P-03/007B, July 2011, U.S. Environmental Protection Agency, Washington, DC. (Downloaded on August 23, 2012; subsequently withdrawn).
United States Environmental Protection Agency (EPA). 2013. *Evaluation of the Inhalation Carcinogenicity of Ethylene Oxide*. In Support of Summary Information on the Integrated Risk Information System (IRIS), EPA/635/R-13/007B, July 2013, U.S. Environmental Protection Agency, Washington, DC.
Van Wijngaarden, E., Hertz-Picciotto, I., 2004. A simple approach to performing quantitative cancer risk assessment using published results from occupational epidemiology studies. Sci. Total Environ. 332, 81–87.