NISS

# Interval Censoring and Marginal Analysis in Ordinal Regression

R.J. Carroll, D.G. Simpson and H. Zhou

Technical Report Number 25
September, 1994

National Institute of Statistical Sciences
19 T. W. Alexander Drive
PO Box 14006
Research Triangle Park, NC 27709-4006
www.niss.org

Although the information in this document has been funded wholly or in part by the United States Environmental Protection Agency under assistance agreement #CR819638-01-0 to the National Institute of Statistical Sciences, it may not necessarily reflect the views of the Agency and no official endorsement should be inferred.

# INTERVAL CENSORING AND MARGINAL ANALYSIS IN ORDINAL REGRESSION

R. J. Carroll
Department of Statistics
Texas A&M University
College Station TX 77843–3143

D. G. Simpson
Department of Statistics
University of Illinois
Champaign IL 61820

H. Zhou
NISS
P. O. Box 14162
Research Triangle Park NC 27709

## Abstract

This paper develops methodology for regression analysis of ordinal response data subject to interval censoring. This work is motivated by the need to analyze data from multiple studies in toxicological risk assessment. Responses are scored on an ordinal severity scale, but not all responses can be scored completely. For instance, in a mortality study, information on nonfatal but adverse outcomes may be missing. In order to address possible within–study correlations we develop a generalized estimating approach to the problem, with appropriate adjustments to uncertainty statements. We develop expressions relating parameters of the implied marginal model to the parameters of a conditional model with random effects, and, in a special case, we note an interesting equivalence between conditional and marginal modeling of ordinal responses. We illustrate the methodology in an analysis of of a toxicological database. [1]

[1]Research supported by the U.S. Environmental Protection Agency under Cooperative Agreement #CR819638-01-0, the National Cancer Institute under grant CA–57030, the National Science Foundation under contract DMS 92-07730 and the National Security Agency under contract MDA 904-92-H-3058.

# 1 Introduction

This paper is concerned with the analysis of ordinal categorical response data subject to interval censoring. Our work was motivated by the toxicological risk assessment of a widely used chemical solvent, perchloroethylene (PERC).

The Clean Air Act Amendments of 1990 require the Environmental Protection Agency to develop emission standards for 189 pollutants and to set standards for other substances "to provide an ample margin of safety to protect public health." Available data come from all relevant studies in the literature. Different studies may include different toxicological endpoints, and multiple endpoints appear within a given study. Moreover, the database for each chemical includes multiple species, and there is tremendous variation in experimental protocol from one study to the next. The heterogeneity of the available data poses a serious challenge to the risk assessor.

Interval censored categorical response data are an important part of the analysis of acute inhalation exposure to PERC. A major cornerstone of the approach to analyzing PERC is the reduction of diverse endpoints to a common ordinal scale of severity categories; see section 6 for more details. Use of ordinal severity scores in this context has been suggested by Hertzberg and Miller (1985), Dourson, Hertzburg, Hartung and Blackburn (1985), Hertzberg (1989), and Guth, Jarabek, Wymer and Hertzberg (1991), who used ordinal logistic regression across studies, in some cases adjusting for species differences via an empirically derived "human equivalent concentration." A distinct advantage of adverse outcome modeling is that it provides a way to put very different quantitative measurements on a common scale.

The outline of this paper is as follows. In section 2, we describe a general model for an interval censored categorical response, deriving the likelihood function for it. Specific models discussed include the proportional odds model and a conditional model based on continuation ratios. Of special interest in the context of PERC is estimation of the dose at a given duration which leads to a fixed probability of response, the so-called "effective dose" (ED).

An interesting facet of the applications that motivated this investigation is that the data consist of a number of observations within each of a number of independent studies. This structure induces correlations in the observations, and it forces us to adjust the usual information-based standard errors. Section 3 describes how to make the adjustment, using the technology of generalized estimating equations.

Simulation studies presented in section 5 compare interval censored estimation and estimation using only the complete data. The simulations also contrast the reported inferences for naive likelihood and generalized estimating equation methods when correlations are present in the data.

In section 6, we describe the PERC data set in some detail and apply the methods developed in earlier sections. A striking conclusion of this analysis is that humans appear to be an order of magnitude more sensitive to PERC than do mice or rats.

Final remarks are given in section 7.

## 2 Models for Censored Outcome Data

The purpose of this section is to propose a general model for censored categorical outcomes, along with estimation of the effective dose which, at a given duration, leads to any any given probability of response. We will then examine some specific models in detail.

The response $Y$ is categorical taking on the values $s = 0, 1, ..., S \geq 1$. We let $X$ be the covariate of primary interest, in our example concentration (dose), and let $Z$ be all the other covariates, e.g., duration of exposure, species, gender, etc. Denote by $\Theta$ all the unknown parameters, and write the model in general form:

$$\begin{aligned} \text{pr}(Y \geq s|X, Z) &= \mathcal{H}(s, X, Z, \Theta) \text{ if } s = 1, ..., S; \\ &= 1 \text{ if } s = 0; \\ &= 0 \text{ if } s > S. \end{aligned} \tag{1}$$

The response is censored into the interval $[s, s + m]$ if the only thing known about it is that it falls into this interval (and no smaller interval). The probability that an observation falls into the interval $[s, s + m]$ is simply $\mathcal{H}(s, X, Z, \Theta) - \mathcal{H}(s + m + 1, X, Z, \Theta)$.

We assume that the censoring mechanism is defined intrinsically and is hence ignorable in the usual sense (Little & Rubin, 1987). In effect, each response $Y$ follows a conditional multinomial distribution, but the multinomial cells may differ for different observations. Related estimation problems for partially classified multinomial observations have been studied by various authors, including Hartley (1958), Koch, Imrey and Reinfurt (1972), and Chen and Feinberg (1976). The work most closely related to ours is Shipp, Howe, Watson and Hogg (1991), which briefly discussed ordinal regression in which some of the data are classified only according to whether or not $Y$ is in the highest severity category.

In the present framework, all that is observable about $Y_i$ is that it lies in one of the intervals $[0, c_{i1})$, $[c_{i1}, c_{i2})$, ..., $[c_{ik_i}, \infty)$, where the constants $0 < c_{i1} < \cdots < c_{ik_i}$ form an ordered subset of $\{1, 2, \ldots, S\}$. Let $\delta(Y|s,t) = 1$ if $Y$ lies in the interval $[s,t]$, and let it equal zero otherwise. The likelihood function based on a sample of size $n$ is then

$$\ell(\Theta) = \prod_{i=1}^{n} \prod_{t=0}^{k_i} \{\mathcal{H}(c_{it}, X_i, Z_i, \Theta) - \mathcal{H}(c_{i,t+1}, X_i, Z_i, \Theta)\}^{\delta(Y_i|c_{it}, c_{i,t+1})}, \tag{2}$$

where we set $c_{i0} = 0$ and $c_{i,k_i+1} = \infty$. For given $Z$, the *effective dose*, $\mathrm{ED}_{100p}(s, Z)$, for category $s = 1, ..., S$ is that value of $x$ for which the probability is $p$ that $Y \geq s$, i.e,

$$p = \mathcal{H}\{s, ED_{100p}(s, Z), Z, \Theta)\}. \tag{3}$$

For any given set $(s, p, \Theta, Z)$, one finds the $\mathrm{ED}_{100p}(s, Z)$ by solving (3).

The general formulation in (1) encompasses a variety of strategies for modeling ordinal response data. We present several examples.

## 2.1 Proportional Odds Models

The proportional odds model, described by McCullagh (1980), is for polytomous logistic regression assuming parallel effects for different severity levels. To capture the distinction between a primary scalar variable $X$ and a vector of covariates $Z$ we formulate the model as

$$\mathrm{pr}(Y \geq s|X, Z) = H(\alpha_s + \beta^T AX + \gamma^T Z), \quad s = 1, ..., S, \tag{4}$$

where $A$ is a design vector accommodating stratification of the $X$ effect, scale conversions and so on, and $H(v) = \{1 + \exp(-v)\}$ is the logistic distribution function. One could, of course, replace the logistic distribution by other distributions such as the standard normal employed in probit regression. In the general framework of (1), $\Theta = (\alpha_1, ..., \alpha_S, \beta, \gamma)$.

Define the logit function, $\mathrm{logit}(p) = H^{-1}(p) = \log\{p/(1-p)\}$. The assumption that neither $\beta$ nor $\gamma$ depends on the category $s$ means that the logits,

$$\mathrm{logit}\{\mathrm{pr}(Y \geq s|X, Z)\} = \log\left\{\frac{\mathrm{pr}(Y \geq s|X, Z)}{\mathrm{pr}(Y < s|X, Z)}\right\} = \alpha_s + \beta^T AX + \gamma^T Z,$$

are parallel to one another as functions of $X$ and $Z$. For this model equation (3) has the explicit solution

$$ED_{100p}(s, Z) = \frac{\mathrm{logit}(p) - \alpha_s - \gamma^T Z}{\beta^T A}. \tag{5}$$

## 2.2 Conditional Probability Models

Direct extension of the proportional odds model is problematic. The model in (4) implies that if we were to perform logistic regression with "success" being $Y \geq 1$ or logistic regression with "success" being $Y \geq 2$, then the two regression lines would be parallel. The assumption of parallelism may be untenable in some applications. If we were to address this problem by letting $\beta$ in (4) depend on $s$, then it would be possible to get contradictions such as $\text{pr}(Y \geq 1|X,Z) < \text{pr}(Y > 2|X,Z)$ for some $(X,Z)$. Moreover, enforcing monotonicity of probabilities for all $(X,Z)$ leads to a set of constraints on the model which can cause difficulties in computation and in inferences.

An alternative to the proportional odds model, which provides a cleaner treatment of nonparallelism, is the conditional model given by

$$\text{pr}(Y = s|Y \leq s, X, Z) = H(\alpha_s + \beta_s^T AX + \gamma_s^T Z). \tag{6}$$

Feinberg (1980) and Agresti (1984) provided some details for cross–classified data, referring to the conditional probabilities as "continuation ratios." For our purposes it is easy to reconstruct the marginal probabilities by recursion, and we find that

$$\begin{aligned} \text{pr}(Y \leq s|X,Z) &= \prod_{t=s+1}^{S} \left\{1 - H(\alpha_t + \beta_t^T AX + \gamma_t^T Z)\right\}; & (7) \\ \text{pr}(Y = s|X,Z) &= \text{pr}(Y = s|Y \leq s, X, Z)\text{pr}(Y \leq s|X,Z) \\ &= H(\alpha_s + \beta_s^T AX + \gamma_s^T Z) \prod_{t=s+1}^{S} \left\{1 - H(\alpha_t + \beta_t^T AX + \gamma_t^T Z)\right\}. & (8) \end{aligned}$$

It follows that if a response $Y$ is censored into $[s, s+m]$, then its contribution to the likelihood is the multiplicative factor

$$\begin{aligned} &\text{pr}(s \leq Y \leq s+m|X,Z) \\ &= \left[\prod_{t=s+m+1}^{S} \left\{1 - H(\alpha_t + \beta_t^T AX + \gamma_t^T Z)\right\}\right] \left[1 - \prod_{t=s}^{s+m} \left\{1 - H(\alpha_t + \beta_t^T AX + \gamma_t^T Z)\right\}\right] \end{aligned} \tag{9}$$

The effective dose, $\text{ED}_{100p}(s,Z)$, is obtained by solving

$$1 - p = 1 - \text{pr}(Y \leq s-1|X,Z) = \prod_{t=s}^{S} \left[1 - H\left\{\alpha_t + (\beta_t^T A)ED_{100p}(s,Z) + \gamma_t^T Z\right\}\right]. \tag{10}$$

## 2.3 Adjacent Category Odds Models

Another possibility is to model the adjacent category odds in the log–linear model

$$\log\left\{\frac{\text{pr}(Y = s|X,Z)}{\text{pr}(Y = s-1|X,Z)}\right\} = \alpha_s + \beta_s^T AX + \gamma_s^T Z, \quad s = 1, ..., S; \tag{11}$$

see, for example, Clogg and Shihadeh (1994). As in the conditional model of section 2.2, the severity regressions in (11) may be modeled without constraints on the parameters. Using (11) and the fact that the category probabilities sum to one, we obtain the marginal category probabilities

$$\text{pr}(Y = s|X, Z) = \frac{\exp\left\{\sum_{j=1}^{s}(\alpha_j + \beta_j^T A \cdot X + \gamma_j^T Z)\right\}}{1 + \sum_{t=1}^{S}\exp\left\{\sum_{u=1}^{t}(\alpha_u + \beta_u^T A \cdot X + \gamma_u^T Z)\right\}}, \qquad s = 0, \ldots, S, \tag{12}$$

where the sum in the numerator vanishes if $s = 0$. Substituting (12) into (2) and (3) yields the likelihood and the defining equation for the effective dose. In particular, the effective dose solves

$$p = \frac{\sum_{j=s}^{S}\exp\left[\sum_{k=1}^{j}\{\alpha_k + \beta_k^T A \cdot ED_{100p}(s, Z) + \gamma_k^T Z\}\right]}{1 + \sum_{t=1}^{S}\exp\left[\sum_{u=1}^{t}\{\alpha_u + \beta_u^T A \cdot ED_{100p}(s, Z) + \gamma_u^T Z\}\right]}. \tag{13}$$

# 3 Marginal Analysis

The structure of the data in the applications that motivated this research is such that there are a number of groups within each study, see Table 5. This clearly has the potential to introduce correlations among the responses within a study, and hence to bias estimated standard errors.

If one thinks of the studies as clusters, one sees that we are in the typical framework of what is now called generalized estimating equations. That is, we have specified a marginal model (1) for the responses given the observed covariates, but responses within a cluster are correlated. The resulting parameter estimates are consistent and asymptotically normally distributed. We first describe how to adjust the usual formulas for standard errors and large sample confidence intervals, and then we contrast marginal and conditional modeling for a correlated proportional odds model.

## 3.1 Statistical Uncertainty Estimates

In the absence of a model for the correlation structure, the standard device is to compute standard errors using the sandwich formula. To so this, we proceed as follows. Let $(Y_{ij}, X_{ij}, Z_{ij})$ refer to the $j$th observation within the $i$th study (cluster), where $i = 1, ..., n$ and $j = 1, ..., n_i$. Referring to (2), if we compute the logarithm of the likelihood and differentiate with respect to $\Theta$, we see that our estimate solves

$$0 \;=\; \sum_{i=1}^{n}\psi_i\left(\widehat{\Theta}\right), \tag{14}$$

where

$$\psi_i(\Theta) \;=\; \sum_{j=1}^{n_i}\sum_{t=0}^{k_{ij}}\delta(Y_{ij}|c_{ijt}, c_{ij,t+1})\frac{\partial}{\partial\Theta}\log\left\{\mathcal{H}(c_{ijt}, X_{ij}, Z_{ij}, \Theta) - \mathcal{H}(c_{ij,t+1}, X_{ij}, Z_{ij}, \Theta)\right\}. \tag{15}$$

The sandwich estimator of the covariance matrix of $\widehat{\Theta}$ is $\widehat{\text{cov}}(\widehat{\Theta}) = A_n^{-1} B_n A_n^{-T}$, where $A_n = \sum_{i=1}^n \frac{\partial}{\partial \Theta^T} \psi_i \left(\widehat{\Theta}\right)$ and $B_n = \sum_{i=1}^n \psi_i \left(\widehat{\Theta}\right) \psi_i^T \left(\widehat{\Theta}\right)$. In computing standard errors and confidence intervals we suggest two further adjustments to reflect the extra variability due to estimating $p$ parameters: (1) inflate the estimate of the covariance matrix by the multiplier $c(n,p) := n/(n-p)$; and (2) replace the standard normal quantile in the confidence interval by the corresponding quantile of the Student's t distribution with $n-p$ degrees of freedom. These adjustments produce confidence intervals analogous to the t-type intervals commonly used linear regression analysis. The simulations presented in section 5 contrast the the sandwich type intervals with the naive delta method, which ignores the effects of within group correlations.

### 3.2 Marginal versus Conditional Parameters

As in any problem with correlations induced by clustering, an alternative to the sandwich estimator is to model the correlation structure directly, e.g., by a random effects analysis. We describe some relationships between the two approaches, focusing on a conditional proportional odds model with random intercepts, specifically,

$$\text{pr}(Y_{ij} \geq s | U_i, V_{ij}) \quad = \quad H(\alpha_s + \sigma U_i + \eta^T V_{ij}), \quad s = 0, 1, \ldots, S \tag{16}$$

for $j = 1, \ldots, n_i$, $i = 1, \ldots, N$, $\sigma > 0$, and $U_i$ has a distribution $G$ with mean equal to zero. The random variables $U_i$ are independent but unobservable, and the responses $Y_{ij}$ are conditionally independent given $U_i$. The latent variables $U_i$ induce correlations between groups of responses. Binary models ($S = 1$) of this form have been considered by various authors including Anderson and Aitkin (1985), Stiratelli, Laird and Ware (1984), and Preisler (1989).

It is important to realize that the parameters of the marginal model implied by (14) are distinct from the parameters of the conditional model (16). This can be seen as follows. First, define the smoothed link function,

$$H_\sigma(t) = \int_{-\infty}^{\infty} H(t + \sigma u) dG(u), \tag{17}$$

and, letting $\theta = (\alpha^T, \eta^T)^T$, the interval probabilities

$$\pi_{ijt}(\sigma, \theta) = H_\sigma(\alpha_{c_{ijt}} + \eta^T V_{ij}) - H_\sigma(\alpha_{c_{ij,t+1}} + \eta^T V_{ij}), \quad t = 1, \ldots, k_{ij}. \tag{18}$$

Next, let $\tilde{\theta} = (\tilde{\alpha}^T, \tilde{\eta}^T)^T$ be the vector of marginal model parameters. Then $\tilde{\theta}$ solves the theoretical

estimating equation

$$0 = \sum_{i=1}^{n} \sum_{j=1}^{n_i} \sum_{t=0}^{k_{ij}} \pi_{ijt}(\sigma, \theta) \frac{\partial}{\partial \tilde{\theta}} \log \left\{ H(\tilde{\alpha}_{c_{ijt}} + \tilde{\eta}^T V_{ij}) - H(\tilde{\alpha}_{c_{ij,t+1}} + \tilde{\eta}^T V_{ij}) \right\}. \tag{19}$$

If it were to happen that $\pi_{ijt}(\sigma, \theta) \equiv \pi_{ijt}(0, \theta)$, then we would get $\tilde{\theta} = \theta$, because (19) reduces to the theoretical estimating equation of the unconditional proportional odds model. Except for the trivial case $\sigma = 0$, however, we cannot expect this equivalence to hold in general.

A more enlightening simplification occurs if $H_\sigma$ is a scale transformation of $H$, that is, for some constant $\tau$ we have $H_\sigma(x) = H(\tau x)$ for all $x$. For example, if $H = G = \Phi$, where $\Phi$ is the standard normal distribution function, then the scale relationship holds with $\tau = (1 + \sigma^2)^{-1/2}$. The scale relationship implies that $\pi_{ijt}(\sigma, \theta) = \pi_{ijt}(0, \tau\theta)$ for each combination of $i$, $j$ and $t$. Consequently, equation (19) coincides with the theoretical estimating equation of the unconditional proportional odds model ($\sigma = 0$), but with $\theta$ replaced by $\tau\theta$. We therefore have

$$\tilde{\theta} = \tau\theta. \tag{20}$$

The multiplier $\tau$ is the attenuation effect in going from a model for the conditional distribution of $Y$ given $U$ to a model for the marginal distribution of $Y$ integrated over $U$. The effect is similar to the attenuation that occurs in measurement error problems, as described by Carroll, Ruppert and Stefanski (1995).

For related discussion of marginal versus conditional modeling in the context of longitudinal data see Zeger, Liang and Albert (1988), who used the evocative terms "population–averaged" and "subject–specific."

### 3.3 Marginal Effective Dose

Assume now that $\eta^T V = \beta^T A X + \gamma^T Z$, where $X$ is the "dose" and $Z$ is the vector of covariates. Under the conditional model (16), the effective dose giving a *marginal* probability of $p$ for a response severity $s$ or higher is given by

$$p = \mathrm{pr}(Y \geq s | X = ED_{100p}(s, Z), Z) = H_\sigma(\alpha_s + \beta^T A \cdot ED_{100p}(s, Z) + \gamma^T Z). \tag{21}$$

If $H_\sigma(\cdot) = H(\tau\cdot)$, then (20) and (21) imply that

$$p = H(\tau\alpha_s + \tau\beta^T A \cdot ED_{100p}(s, Z) + \tau\gamma^T Z) = H(\tilde{\alpha}_s + \tilde{\beta}^T A \cdot ED_{100p}(s, Z) + \tilde{\gamma}^T Z). \tag{22}$$

Hence, we have established an interesting equivalence between marginal modeling and conditional modeling, which is summarized in the following result.

**Theorem:** *Assume that conditional model (16) holds. If the convolution of $H$ and $G$ is a scale transform of $H$, then the marginal effective dose for the conditional model is the same as the effective dose of the marginal model obtained by the method of generalized estimating equations.*

An example in which this equivalence holds is probit regression with a Gaussian prior on the intercept. More generally we expect that the marginal effective dose is less sensitive to the random effect than the regression parameters are. For instance, if $G$ is Gaussian and $H$ is logistic, then $H$ may be closely approximated by the $N(0, 1.7^2)$ distribution, and the convolution of the two Gaussian distributions is a scale transformed Gaussian, which is approximately a scale transform of the logistic link function. We would therefore expect the conditional and marginal models to lead to similar values for the marginal effective dose.

# 4 Computational Method

For the proportional odds model, it is possible to use standard software to obtain starting values for the parameters, using what we call "pseudo-strata". We illustrate the idea for the case that the response takes on the three values $s = 0, 1, 2 = S$. The possible responses then are $0, 1, 2, \{0, 1\}, \{1, 2\}$. The pseudo-strata correspond to the outcomes $s \geq 1$ and $s \geq 2$, which we will call stratum 1 and stratum 2, respectively. We first create two data sets. For stratum $j = 1, 2$, the information available for a logistic regression has a "success" defined as $Y \geq j$. The data set for stratum 1 thus consists of all data except those for which the assignment to a "success" is ambiguous, i.e., those for which it is only known that $Y = 0, 1$. Similarly, the data set for stratum 2 consists of all data except those for which the assignment to a "success" is ambiguous, i.e., those for which it is only known that $Y = 1, 2$. One then pools the two data sets, and runs a logistic regression with stratum-specific intercepts $\alpha_1$ and $\alpha_2$. Note that many observations appear twice in the newly constructed data base. This data reuse does not affect the consistency of the parameter estimates, and hence they serve as legitimate and easily computed starting values; data reuse does mean that the standard errors computed from logistic regression software are typically incorrect. The pseudo-strata method is easily extended to general problems.

Given pseudo–values, we then iterate to convergence using Newton–Raphson. In the proportional odds model we did not explicitly enforce the constraint that the $\alpha$ parameters remain ordered,

but in all cases the estimates were properly ordered. Violations of the ordering constraint could possibly occur if the model fit were very poor.

We have implemented the censored proportional odds computations in S-Plus (Statistical Sciences, Inc.), including sandwich type variance estimates. In the absence of censoring the SAS (SAS Institute, Inc.) procedure "Logistic" will fit the proportional odds model. Additional programming will be required for the sandwich estimates of variance.

# 5 Simulations

We ran two small simulation studies to illustrate the impact of censoring and correlation in the data, using the conditional proportional odds model (16) with $H$ = logistic and $G$ = normal, i.e., a conditional logistic model with Gaussian random effects. As shown in section 3, because the logistic distribution function can be approximated by a normal distribution function with standard deviation 1.7, the marginal regression parameters are *approximately* attenuated by the factor $\tau = (1+\sigma^2/1.7^2)^{-1/2} \approx 2$ relative to the conditional model parameters, but the effective dose from the marginal model is the same as the marginal effective dose from the conditional model. We report results for regression parameters as well as effective doses.

We set $S = 2$, and the conditional model parameters were set as follows: $\alpha_1 = 5$, $\alpha_2 = 0$, $\beta = 4$ and $\gamma = 2$. We simulated under both the independent response model with $\sigma = 0$ and the random intercept model with $\sigma = 3$.

In all cases, there were $n = 90$ observations, in 18 groups of size 5, so that $N = 18$ and $n_i = 5$ for $i = 1, ..., N$. Each experiment was repeated 500 times. In Tables 1 and 2, we constructed the censored data as follows. Four groups of five observations were selected at random so that only whether $Y \leq 1$ or $Y = 2$ was observed, while another four groups of five observations were selected at random so that only whether $Y = 0$ or $Y \geq 1$ was observed; thus a total of 40 observations were subject to censoring. A complete data analysis is one in which only the fifty observations which are not subject to censoring are retained.

Marginal parameter estimates were obtained by computing maximum likelihood estimates as if the data were independent, both for the reduced set of complete data, and for the full data including censored observations. Variances and confidence intervals were computed using both the naive Fisher information and the sandwich method. The sandwich intervals incorporated the empirical adjustments described in section 3.

Table 1 focuses on estimation of the model parameters, as well as estimating their standard errors and constructing confidence intervals. Table 2 considers estimation and inference for the ED10 for any effect, defined by (5) with $p = .10$ and $s = 1$, computed for three values of $z$; we are thus estimating $\mathrm{ED}_{p=.10}(s = 1, z)$. The overall conclusions formed from these tables are as follows. Generally, the interval censored estimates and the complete data estimates show no serious differences in their means, but the former is more efficient in that it is less variable. When $\sigma^2 = 0$, there are no group random effects, so that all observations are independent and both the Fisher information and sandwich methods yield asymptotically correct inferences. In this case, both the Fisher information and sandwich standard errors yield coverage probabilities at or near the nominal level, although the sandwich estimator tends to have coverages closer to the nominal; this is probably an outgrowth of our empirical adjustments.

When there are group random effects ($\sigma^2 = 9$), the attenuations in the parameter estimates are nearly what one expects from our approximate analysis, in the sense that the simulation means for $\sigma^2 = 9$ are approximate 50% of the simulated means when $\sigma^2 = 0$. The $\mathrm{ED}_{10}$ estimates tends to be more extreme here than is expected from our approximations. Because the approximations are just that, approximations, for this model we took as the "true" parameters the simulated mean of the estimates. This is equivalent to the standard marginal analysis convention of performing inference about the value to which the estimates converge, rather than some theoretical parameter to which the estimate converge only approximately. Here the naive Fisher information standard errors are asymptotically incorrect because of the correlations within the groups; we use the term "naive" here to mean that the group correlations were ignored. In the simulations, the naive Fisher information standard deviations are much too small. This is reflected in coverage probabilities, where the sandwich intervals have nearly nominal coverage, while the naive Fisher information intervals have decreased coverages.

We repeated this simulation, but with a different pattern of censoring, namely that 40 observations were censored so that only whether $Y \leq 1$ or $Y = 2$ was observed. Table 3 presents the results for the parameters estimates, while Table 4 presents results for estimation of effective dose. The results are much in line with those of Tables 1 and 2. The major difference is that the interval censored $\mathrm{ED}_{10}$ estimates are not much less variable than the complete data estimates. Presumably, this is because of the fact that for this pattern of censoring, the censored data provide little information about the intercept $\alpha_1$ for an adverse effect, but this intercept is crucial in the $\mathrm{ED}_{10}$

definition.

It is important for the reader to appreciate what we mean by a "complete" data analysis. As we have envisioned it, interval censoring is an integral part of the study, e.g., some studies are designed to report only an adverse effect ($Y \geq 1$) on many study participants, with finer gradations among adverse effects occurring only on a randomly selected subset. We have defined as a "complete data analysis" one which eliminates those observations which *could* have been censored. In the lexicon of Little & Rubin (1987), a complete data analysis is a missing data analysis in which missingness occurs at random and is ignorable.

Another possible definition of complete data analysis is one which retains all uncensored observations. Thus, in the second simulation, we might have retained any observation for which $Y = 2$, but eliminated those for which it is only known that $Y \leq 1$. A complete data analysis done in this was is also a missing data analysis, but now the missing data are not missing at random, they are not ignorable, and the resulting parameter estimates are inconsistent. Thus, we have chosen not to report results from this type of analysis.

# 6 Analysis of Perchloroethylene Data

Perchloroethylene (PERC) is a widely used solvent, and its effects have been investigated in a number of small studies. Information collected from these studies was assembled into a data base; the details of the data-base construction will be described in a separate report. In broad outline, initially a literature search was done to find all available data from published sources, proceedings and technical reports. The initial set was screened to remove poorly documented studies. Since PERC is widely used, there exist human studies at low levels of exposure. Test species were mice, rats, rabbits, dogs and humans. The rabbit and dog groups were omitted from the analysis due to their scarcity in the data base.

A profile of these studies is given in Tables 5–6. In the first table, we distinguish between a *severity* study, which is clearly aimed at risk assessment for acute exposures, and a *mortality* study, which is aimed at fatal exposures and which has outcomes reported merely as survival or death. There were no pure mortality studies in the PERC data base, but a number of studies included lethal levels of exposure. In one of these studies a portion of the animal responses were reported simply as lethal or nonlethal, with no information on nonlethal adverse outcomes. In this case the nonlethal outcomes were censored, see Table 6

Each study consisted of a number of groups of test subjects, with concentration and duration of exposure reported. The number of test subjects in each group varied, and we have reported the average number in each species-sex combination.

## 6.1 Ordinal Response Scoring

The response variable in the analysis is a severity score with 0 = "No adverse effect," 1 = "Moderately adverse effect," and 2 = "Severe or lethal effect." The severity scoring was performed by a toxicologist based on biological considerations. Substantial censoring occurs because, in addition to the mortality censoring mentioned above, there is often insufficient information to determine the biological significance of a particular response. This type of ordinal severity scoring allows us to treat different endpoints on a common scale.

The severity judgements were based on biological rather than statistical considerations because the use of statistical tests of significance at this stage would bias the subsequent ordinal regression analysis.

Some data were reported at the group level only. Thus, for example, a study might report the results on a group of six rats, and as our response we used a summary severity category for the whole group. Simpson, Carroll and Xie (1994) described latent variable models for group responses. In a preliminary binary analysis of the PERC data, they found a group response score had essentially the same information as a single individual response due to the high imputed correlation between individuals in a group. In the present analysis we analyze group response data only, treating them as equivalent to single individuals. Further work is underway on combining group and individual incidence data, and the analysis presented here is somewhat preliminary.

## 6.2 Log Concentration and Control Data

We use log concentration and log duration of exposure as the primary independent regression variables, and determine effective log–concentration as a function of duration. This is based on several considerations. First, the concentrations and durations vary over orders of magnitude, so an analysis based on the raw scale would allow certain observations to be extremely influential on the results. Second, attempts to fit the model with either the duration or the concentration (or both) on the original scale lead to poorer model fit as measured by comparing likelihoods. Third, the log transform is range preserving in the sense that effective doses are computed to be nonnegative.

With the dose and duration variables entering the model on the logarithmic scale, the proper treatment of control data requires some care. The log–dose model implies that controls have response probabilities equal to zero. Thus, control responses are uninformative under this model (provided they are all zero). The proper likelihood analysis deletes the null control responses from the analysis. In the PERC studies all controls had null responses.

Any nonnull responses among the controls would invalidate the log–dose model. Several approaches have been proposed in the literature to deal with nonzero control probabilities for binary response data when the dose effect is expressed on the logarithmic scale. A relatively clean analysis is to replace the dose response $P(d)$, where $d$ is "dose," by the modified response function $p_0 + (1 - p_0)P^*(d)$, where $p_0$ is the spontaneous response probability, and $P^*(0) = 0$. Further details and references may be found in Morgan (1992, Chapter 3).

### 6.3 Stratified Analysis

If we pool all the data across species and gender, we fit the model (4). However, this model can be extended to allow for stratum-specific effects. For example, if we have strata $j = 1, ..., J$, then a model with stratum-specific intercepts is

$$\Pr(Y \geq s \mid X, Z, \text{ stratum } = j) = H(\alpha_s + \beta_1 x_1 + \beta_2 x_2 + \gamma_{0,j}), \tag{23}$$

where it is understood that $\gamma_{0,J} = 0$, a convention necessary to make the model statistically identifiable.

The model (23) can be further expanded, for example to allow different strata to have different log–concentration parameters. One such model is

$$\Pr(Y \geq s \mid X, Z, \text{ stratum } = j) = H(\alpha_s + \beta_{x,j} X + \beta_z Z + \gamma_j). \tag{24}$$

In our first analysis, we pooled all the data and fit the model (4). Figure 1 contains two important features; (a) a plot indicating the severity category, concentration and duration for each group in the PERC data base; and (b) the estimated $ED_{10}$ and its associated 95% confidence band when pooling all the data. The "censored" category in (a) always refers to interval censoring obtained by pooling the nonadverse and adverse categories. Part (b) shows the negative slope that one would expect; as duration increases, the estimated $ED_{10}$ should decrease. The parameter estimates using interval censored and complete data analyses, along with their estimated standard errors, are given in Tables 7–8.

We next performed a stratified analysis, stratified on the basis of species. We fit in turn models (23) and (24). The former model allows for stratum effects but assumes that the effects of concentration and duration do not depend on the strata. The latter model allows for stratum–specific concentration effects. Model (23) provided a statistically significant improvement over model (4) ($p < 0.001$), while model (24) provided a statistically significant improvement over model (23) ($p < 0.04$). There was little evidence of an important stratum–specific duration effect. The parameter estimates for the model (23) are given in Tables 9–10.

The addition of stratification variables in the model was statistically significant, and the effects appear to be practically significant as well. Figure 2 shows the $ED_{10}$ lines for the different species based upon the model (24), which has stratum–specific concentration effects. In contrast to Figure 1, Figure 2 shows the species associated with each point. It is apparent that in the PERC data base the humans had much lower exposure levels than mice and especially rats. The $ED_{10}$'s are typically an order of magnitude smaller for humans than they are for rats, and similarly for mice at low durations of exposure. Note also that the rat and human lines are very nearly parallel on the log–log scale, with the mice line being only somewhat nonparallel. The line for mice is estimated with much less precision than the human and rat lines, and the observed nonparallelism must not be overinterpreted. Because the duration parameter is shared by all species, this parallelism among the species is a reflection of the similarity of the concentration slope parameters across species. Thus, the major differences between rats and humans especially, and to a less extent the mice, appear to be explained by differences in uptake rather than differences in mechanism. In particular, the differences might be addressed by scaling up the concentrations, because the scale is reflected in the intercept rather than the slope if concentrations enter the model logarithmically. For mice, there may be a difference in mechanism, but the lack of precision in the line for mice makes such a conclusion tenuous at best.

In Figure 3, we combined the results of the pooled and stratified analyses. The vertical lines represent the pooled $ED_{10}$ and its 95% confidence interval and the median duration of exposure observed in the PERC data base. The horizontal lines are the stratum–specific $ED_{10}$'s. There are major differences here. For example, the estimated $ED_{10}$ for rats falls outside the pooled confidence interval. Note the previously mentioned fact that the $ED_{10}$ for mice is poorly estimated relative to that for humans and rats.

We have run further analyses on these data, by allowing for species/gender stratum effects.

While we observed some statistically significant effects in such analyses, in general the practical effects were not striking. To illustrate this, consider Figure 4, which is the analogue to Figure 3 when gender and species are used to form strata. One observes in Figure 4 some small gender effects, but these do not appear to be of major consequence.

# 7 Discussion

We have considered the analysis of ordinal categorical response data subject to interval censoring. We have described general models for interval censored responses, and their associated likelihood functions. Specific models discussed included the proportional odds model and a conditional model based on continuation ratios.

The problem arises as an important part of the analysis of acute inhalation exposure to PERC. Of special interest in the context of PERC is estimation of the dose at a given duration which leads to a fixed probability of response, the so-called "effective dose" (ED).

An interesting facet of the applications that motivated this investigation is that the data consist of a number of observations within each of a number of independent studies. This structure induces correlations in the observations, and it forces us to adjust the usual information-based standard errors. We have described in section 3 how to make the adjustment, using the technology of generalized estimating equations.

For estimating the effective doses, we showed that in some important special cases that there is an equivalence between conditional and marginal modeling of ordinal responses, even though the methods estimate different model parameters.

In section 6, we described the PERC data set in some detail. A striking conclusion of this analysis is that humans appear to be an order of magnitude more sensitive to PERC than are rats.

**Acknowledgments**

Daniel Guth introduced us to the idea of using categorical response categories for risk assessment, and he provided access to the PERC database. Jerome Sacks was instrumental in putting together the research team and contributing valuable discussions.

## REFERENCES

Agresti, A. (1984). *Analysis of Ordinal Categorical Data.* Wiley, New York.

Anderson, D.A. & Aitkin, M. (1986). Variance component models with binary response: interviewer variability. *Journal of the Royal Statistical Society*, Series B, 47, 203-210.

Carroll, R. J., Ruppert, D. and Stefanski, L. A. (1995). *Nonlinear Measurement Error Models.* Chapman & Hall, New York.

Chen, T. and Feinberg, S.E. (1976). The analysis of contingency tables with incompletely classified data. *Biometrics*, 32, 133-144.

Clogg, C.C. & Shihadeh, E.S. (1994). *Statistical Models for Ordinal Variables.* Advanced Quantitative Techniques in the Social Sciences Series, Volume 4, SAGE Publications, Thousand Oaks, California.

Dourson, M.L., Hertzberg, R.C., Hartung, R. & Blackburn, K. (1985). Novel methods for the estimation of acceptable daily intake. *Toxicology and Industrial Health*, 1, 23-33.

Feinberg, S.E. (1980). *The Analysis of Cross–Classified Categorical Data.* MIT Press, Cambridge, MA.

Guth, D.J., Jarabek, A.M., Wymer, L. & Hertzberg, R.C. (1991). Evaluation of risk assessment methods for short–term inhalation exposure. Proceedings of 84th Annual Meeting of the Air & Waste Management Association.

Hartley, H.O. (1958). Maximum likelihood estimation from incomplete data. *Biometrics*, 14, 174-194.

Hertzberg, R.C. (1989). Fitting a model to categorical response data with application to species extrapolation of toxicity. *Health Physics*, 57, 405-409.

Hertzberg, R.C. & Miller, M. (1985). A statistical model for species extrapolation using categorical response data. *Toxicology and Industrial Health*, 1, 43-63.

Hertzberg, R.C. & Wymer, L. (1991). Modeling the severity of toxic effects. Proceedings of 84th Annual Meeting of the Air & Waste Management Association.

Koch, Imrey and Reinfort (1972). Linear model analysis of categorical data with incomplete response vectors. *Biometrics*, 28, 663–692.

Little, R. J. A. & Rubin, D. B. (1987). *Analysis of Missing Data.* John Wiley & Sons, New York.

McCullagh, P. (1980). Regression models for ordinal data. *Journal of the Royal Statistical Society,* Series B, 42, 109-142.

Morgan, B.J.T. (1992). *Analysis of Quantal Response Data.* Chapman & Hall, London.

Preisler, H.K. (1989). Analysis of a toxicological experiment using a generalized linear model with nested random effects. *International Statistical Review*, 57, 145-159.

Shipp, A., Howe, R., Watson, C. and Hogg, M.L. (1991). Health effects and dose–response assessment for hydrogen chloride following short-term exposure. Technical Report, Office of Air Quality Planning and Standards and Office of Health Evaluation Assessment, U.S. Environmental Protection Agency.

Simpson, D.G., Carroll, R.J. & Xie, M. (1994). Scaled link functions and quasilikelihood inferences for grouped binary data. Technical Report, National Institute of Statistical Sciences, Research Triangle Park, NC.

Stiratelli, R., Laird, N. and Ware, J.H. (1984). Random–effects models for serial observations with binary response. *Biometrics*, 40, 961-971.

Zeger, S.L., Liang, K.-Y. and Albert, P.S. (1988). Models for longitudinal data: a generalized estimating equation approach. *Biometrics*, 44, 1049-1060.

Zwart, A. & Woutersen, R.A. (1988). Acute inhalation toxicity of chlorine in rats and mice: time–concentration–mortality relationships and effects of respiration. *Journal of Hazardous Materials*, 19, 195-208.

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | Mean | | | | | | | |
| | $\sigma^2$=0 | | | | $\sigma^2$=9 | | | |
| | $\widehat{\alpha}_1$ | $\widehat{\alpha}_2$ | $\widehat{\beta}_x$ | $\widehat{\beta}_z$ | $\widehat{\alpha}_1$ | $\widehat{\alpha}_2$ | $\widehat{\beta}_x$ | $\widehat{\beta}_z$ |
| Complete | 6.090 | 0.001 | 4.884 | 2.469 | 2.976 | 0.012 | 2.378 | 1.173 |
| I.C. | 5.717 | 0.007 | 4.572 | 2.302 | 2.796 | 0.007 | 2.235 | 1.108 |
| | Standard Deviation | | | | | | | |
| | $\sigma^2$=0 | | | | $\sigma^2$=9 | | | |
| | $\widehat{\alpha}_1$ | $\widehat{\alpha}_2$ | $\widehat{\beta}_x$ | $\widehat{\beta}_z$ | $\widehat{\alpha}_1$ | $\widehat{\alpha}_2$ | $\widehat{\beta}_x$ | $\widehat{\beta}_z$ |
| Complete | 2.192 | 0.663 | 1.708 | 0.989 | 1.142 | 0.723 | 0.723 | 0.511 |
| I.C. | 1.573 | 0.502 | 1.232 | 0.727 | 0.810 | 0.553 | 0.550 | 0.392 |
| | Mean Estimated Standard Deviation, Naive | | | | | | | |
| | $\sigma^2$=0 | | | | $\sigma^2$=9 | | | |
| | $\widehat{\alpha}_1$ | $\widehat{\alpha}_2$ | $\widehat{\beta}_x$ | $\widehat{\beta}_z$ | $\widehat{\alpha}_1$ | $\widehat{\alpha}_2$ | $\widehat{\beta}_x$ | $\widehat{\beta}_z$ |
| Complete | 1.679 | 0.581 | 1.352 | 0.803 | 0.706 | 0.429 | 0.591 | 0.431 |
| I.C. | 1.269 | 0.461 | 1.018 | 0.611 | 0.547 | 0.343 | 0.456 | 0.336 |
| | Mean Estimated Standard Deviation, Sandwich | | | | | | | |
| | $\sigma^2$=0 | | | | $\sigma^2$=9 | | | |
| | $\widehat{\alpha}_1$ | $\widehat{\alpha}_2$ | $\widehat{\beta}_x$ | $\widehat{\beta}_z$ | $\widehat{\alpha}_1$ | $\widehat{\alpha}_2$ | $\widehat{\beta}_x$ | $\widehat{\beta}_z$ |
| Complete | 1.957 | 0.703 | 1.571 | 0.934 | 1.045 | 0.765 | 0.757 | 0.535 |
| I.C. | 1.339 | 0.503 | 1.081 | 0.648 | 0.746 | 0.548 | 0.534 | 0.378 |
| | Coverage of Nominal 95% Interval, Naive | | | | | | | |
| | $\sigma^2$=0 | | | | $\sigma^2$=9 | | | |
| | $\widehat{\alpha}_1$ | $\widehat{\alpha}_2$ | $\widehat{\beta}_x$ | $\widehat{\beta}_z$ | $\widehat{\alpha}_1$ | $\widehat{\alpha}_2$ | $\widehat{\beta}_x$ | $\widehat{\beta}_z$ |
| Complete | 0.978 | 0.958 | 0.982 | 0.982 | 0.814 | 0.792 | 0.904 | 0.928 |
| I.C. | 0.970 | 0.970 | 0.972 | 0.972 | 0.846 | 0.808 | 0.900 | 0.928 |
| | Coverage of Nominal 95% Interval, Sandwich | | | | | | | |
| | $\sigma^2$=0 | | | | $\sigma^2$=9 | | | |
| | $\widehat{\alpha}_1$ | $\widehat{\alpha}_2$ | $\widehat{\beta}_x$ | $\widehat{\beta}_z$ | $\widehat{\alpha}_1$ | $\widehat{\alpha}_2$ | $\widehat{\beta}_x$ | $\widehat{\beta}_z$ |
| Complete | 0.980 | 0.984 | 0.982 | 0.978 | 0.936 | 0.972 | 0.960 | 0.97 |
| I.C. | 0.976 | 0.970 | 0.976 | 0.964 | 0.938 | 0.960 | 0.944 | 0.96 |

Table 1: *Monte-Carlo study of the proportional odds model. Both log(concentration) and log(duration) are generated as standard normal random variables. Here* $(\alpha_1, \alpha_2, \beta_x, \beta_z) = (5, 0, 4, 2)$. *The model has a random intercept effect with mean zero and variance* $\sigma^2$. *"Complete" stands for using the complete data only, while "I.C." stands for the interval-censored mle. The sample size is* $n = 90$, *with 20 observations censored to either* $Y \leq 1$ *or* $Y = 2$, *and 20 observations censored to* $Y = 0$ *or* $Y \geq 1$. *Marginal parameter estimates were obtained by computing maximum likelihood estimates as if the data were independent. Variances and confidence intervals were computed using both the naive Fisher information and the sandwich method.*

| | ED10 for $\sigma^2 = 0$ | | | ED10 for $\sigma^2 = 9$ | | |
|---|---|---|---|---|---|---|
| Duration | 25th | 50th | 75th | 25th | 50th | 75th |
| | | | Mean | | | |
| Complete | -1.404 | -1.748 | -2.091 | -1.954 | -2.293 | -2.632 |
| I.C. | -1.426 | -1.767 | -2.109 | -1.982 | -2.320 | -2.659 |
| | | | Standard Deviation | | | |
| Complete | 0.239 | 0.249 | 0.283 | 0.576 | 0.596 | 0.640 |
| I.C. | 0.198 | 0.206 | 0.232 | 0.469 | 0.484 | 0.517 |
| | | | Mean Estimated Standard Deviation, Naive | | | |
| Complete | 0.267 | 0.280 | 0.314 | 0.448 | 0.477 | 0.534 |
| I.C. | 0.196 | 0.207 | 0.235 | 0.369 | 0.392 | 0.438 |
| | | | Mean Estimated Standard Deviation, Sandwich | | | |
| Complete | 0.277 | 0.292 | 0.330 | 0.605 | 0.63 | 0.686 |
| I.C. | 0.214 | 0.226 | 0.255 | 0.460 | 0.48 | 0.523 |
| | | | Coverage of Nominal 95% Interval, Naive | | | |
| Complete | 0.892 | 0.902 | 0.904 | 0.864 | 0.864 | 0.890 |
| I.C. | 0.916 | 0.912 | 0.918 | 0.876 | 0.874 | 0.892 |
| | | | Coverage of Nominal 95% Interval, Sandwich | | | |
| Complete | 0.952 | 0.934 | 0.944 | 0.952 | 0.964 | 0.970 |
| I.C. | 0.942 | 0.936 | 0.938 | 0.936 | 0.930 | 0.948 |

Table 2: *Monte-Carlo study of the proportional odds model. Both log(concentration) and log(duration) are generated as standard normal random variables. Here* $(\alpha_1, \alpha_2, \beta_x, \beta_z) = (5, 0, 4, 2)$. *The model has a random intercept effect with mean zero and variance* $\sigma^2$. *Displayed are results for estimating the ED*10. *"Complete" stands for using the complete data only, while "I.C." stands for the interval-censored mle. The sample size is* $n = 90$, *with 20 observations censored to either* $Y \leq 1$ *or* $Y = 2$, *and 20 observations censored to either* $Y = 0$ *or* $Y \geq 1$. *Durations are chosen at the 25th, 50th and 75th percentiles of the standard normal distribution. The true ED10 for an adverse effect* $(Y \geq 1)$ *for these durations with* $\sigma^2 = 0$ *are are* $-1.462$, $-1.799$ *and* $-2.136$, *respectively. Via simulations, the true ED10 for an adverse effect* $(Y \geq 1)$ *for these durations with* $\sigma^2 = 9$ *are are set to their mean estimated values. Marginal parameter estimates were obtained by computing maximum likelihood estimates as if the data were independent. Variances and confidence intervals were computed using both the naive Fisher information and the sandwich method.*

| | $\widehat{\alpha}_1$ | $\widehat{\alpha}_2$ | $\widehat{\beta}_x$ | $\widehat{\beta}_z$ | $\widehat{\alpha}_1$ | $\widehat{\alpha}_2$ | $\widehat{\beta}_x$ | $\widehat{\beta}_z$ |
|---|---|---|---|---|---|---|---|---|
| | Mean | | | | | | | |
| | $\sigma^2$=0 | | | | $\sigma^2$=9 | | | |
| Complete | 6.090 | 0.001 | 4.884 | 2.469 | 2.976 | 0.012 | 2.378 | 1.173 |
| I.C. | 5.647 | 0.007 | 4.519 | 2.259 | 2.823 | 0.008 | 2.221 | 1.098 |
| | Standard Deviation | | | | | | | |
| | $\sigma^2$=0 | | | | $\sigma^2$=9 | | | |
| Complete | 2.192 | 0.663 | 1.708 | 0.989 | 1.142 | 0.723 | 0.723 | 0.511 |
| I.C. | 1.447 | 0.424 | 1.089 | 0.694 | 0.925 | 0.494 | 0.544 | 0.375 |
| | Mean Estimated Standard Deviation, Naive | | | | | | | |
| | $\sigma^2$=0 | | | | $\sigma^2$=9 | | | |
| Complete | 1.679 | 0.581 | 1.352 | 0.803 | 0.706 | 0.429 | 0.591 | 0.431 |
| I.C. | 1.264 | 0.402 | 0.959 | 0.575 | 0.603 | 0.302 | 0.442 | 0.324 |
| | Mean Estimated Standard Deviation, Sandwich | | | | | | | |
| | $\sigma^2$=0 | | | | $\sigma^2$=9 | | | |
| Complete | 1.957 | 0.703 | 1.571 | 0.934 | 1.045 | 0.765 | 0.757 | 0.535 |
| I.C. | 1.325 | 0.443 | 1.015 | 0.610 | 0.791 | 0.497 | 0.527 | 0.368 |
| | Coverage of Nominal 95% Interval, Naive | | | | | | | |
| | $\sigma^2$=0 | | | | $\sigma^2$=9 | | | |
| Complete | 0.978 | 0.958 | 0.982 | 0.982 | 0.814 | 0.792 | 0.904 | 0.928 |
| I.C. | 0.968 | 0.954 | 0.96 | 0.954 | 0.838 | 0.812 | 0.910 | 0.934 |
| | Coverage of Nominal 95% Interval, Sandwich | | | | | | | |
| | $\sigma^2$=0 | | | | $\sigma^2$=9 | | | |
| Complete | 0.980 | 0.984 | 0.982 | 0.978 | 0.936 | 0.972 | 0.960 | 0.970 |
| I.C. | 0.966 | 0.978 | 0.972 | 0.968 | 0.94 | 0.964 | 0.950 | 0.966 |

Table 3: *Monte-Carlo study of the proportional odds model. Both log(concentration) and log(duration) are generated as standard normal random variables. Here* $(\alpha_1, \alpha_2, \beta_x, \beta_z) = (5, 0, 4, 2)$. *The model has a random intercept effect with mean zero and variance* $\sigma^2$. *"Complete" stands for using the complete data only, while "I.C." stands for the interval-censored mle. The sample size is* $n = 90$, *with 40 observations censored to either* $Y \leq 1$ *or* $Y = 2$. *Marginal parameter estimates were obtained by computing maximum likelihood estimates as if the data were independent. Variances and confidence intervals were computed using both the naive Fisher information and the sandwich method.*

| | ED10 for $\sigma^2 = 0$ | | | ED10 for $\sigma^2 = 9$ | | |
|---|---|---|---|---|---|---|
| Duration | 25th | 50th | 75th | 25th | 50th | 75th |
| | | | Mean | | | |
| Complete | -1.422 | -1.763 | -2.104 | -2.028 | -2.368 | -2.708 |
| I.C. | -1.46 | -1.795 | -2.130 | -2.002 | -2.34 | -2.678 |
| | | | Standard Deviation | | | |
| Complete | 0.227 | 0.240 | 0.275 | 0.518 | 0.534 | 0.569 |
| I.C. | 0.202 | 0.204 | 0.212 | 0.489 | 0.500 | 0.530 |
| | | | Mean Estimated Standard Deviation, Naive | | | |
| Complete | 0.223 | 0.235 | 0.266 | 0.423 | 0.450 | 0.499 |
| I.C. | 0.192 | 0.196 | 0.208 | 0.389 | 0.411 | 0.452 |
| | | | Mean Estimated Standard Deviation, Sandwich | | | |
| Complete | 0.265 | 0.279 | 0.314 | 0.565 | 0.587 | 0.637 |
| I.C. | 0.185 | 0.190 | 0.202 | 0.478 | 0.497 | 0.536 |
| | | | Coverage of Nominal 95% Interval, Naive | | | |
| Complete | 0.920 | 0.923 | 0.923 | 0.873 | 0.887 | 0.880 |
| I.C. | 0.946 | 0.944 | 0.944 | 0.856 | 0.874 | 0.898 |
| | | | Coverage of Nominal 95% Interval, Sandwich | | | |
| Complete | 0.980 | 0.970 | 0.967 | 0.970 | 0.963 | 0.970 |
| I.C. | 0.928 | 0.924 | 0.924 | 0.928 | 0.934 | 0.940 |

Table 4: *Monte-Carlo study of the proportional odds model. Both log(concentration) and log(duration) are generated as standard normal random variables. Here* $(\alpha_1, \alpha_2, \beta_x, \beta_z) = (5, 0, 4, 2)$. *The model has a random intercept effect with mean zero and variance* $\sigma^2$. *Displayed are results for estimating the ED*10. *"Complete" stands for using the complete data only, while "I.C." stands for the interval-censored mle. The sample size is* $n = 90$, *with 40 observations censored to either* $Y \leq 1$ *or* $Y = 2$. *Durations are chosen at the 25th, 50th and 75th percentiles of the standard normal distribution. The true ED10 for an adverse effect* $(Y \geq 1)$ *for these durations with* $\sigma^2 = 0$ *are are* $-1.462$, $-1.799$ *and* $-2.136$, *respectively. Via simulations, the true ED10 for an adverse effect* $(Y \geq 1)$ *for these durations with* $\sigma^2 = 9$ *are is set to the mean estimated values. Marginal parameter estimates were obtained by computing maximum likelihood estimates as if the data were independent. Variances and confidence intervals were computed using both the naive Fisher information and the sandwich method.*

| | Severity Studies | | | Mortality Studies | | |
|---|---|---|---|---|---|---|
| | # of Studies | Total # of Groups | Average # per Group | # of Studies | Total # of Groups | Average # Per Group |
| Mice - F | 5 | 47 | 5.87 | 1 | 8 | 8 |
| Mice - M | 3 | 16 | 12.62 | 0 | 0 | 0 |
| Mice - B | 2 | 3 | 13.33 | 0 | 0 | 0 |
| Rats - F | 4 | 19 | 5.37 | 0 | 0 | 0 |
| Rats - M | 2 | 14 | 11.43 | 0 | 0 | 0 |
| Rats - B | 1 | 22 | 17.09 | 0 | 0 | 0 |
| Human - M | 3 | 5 | 5.60 | 0 | 0 | 0 |
| Human - B | 2 | 6 | 5.33 | 0 | 0 | 0 |

Table 5: *Information on Perchloroethylene (PERC) Here "–F" means females, "–M" means males and "–B" indicates that gender was either unspecified or the group was mixed.*

| | Response Type | | | | |
|---|---|---|---|---|---|
| Species | [0] | [1] | [2] | [0,1] | [1,2] |
| Mice | 10 | 23 | 28 | 13 | 0 |
| Rats | 17 | 5 | 21 | 12 | 0 |
| Human | 6 | 5 | 0 | 0 | 0 |

Table 6: *Censoring Information on Perchloroethylene (PERC)*

| | Complete | Interval Censored |
|---|---|---|
| $\alpha_1$ | -10.55 | -11.90 |
| | (2.61) | (2.11) |
| | [4.65] | [4.38] |
| $\alpha_2$ | -13.75 | -14.29 |
| | (2.91) | (2.29) |
| | [5.10] | [4.87] |
| Log duration ($\beta_x$) | 2.84 | 3.05 |
| | (.66) | (.51) |
| | [1.14] | [1.09] |
| Log concentration ($\beta_z$) | 1.26 | 1.58 |
| | (.41) | (.35) |
| | [.78] | [.48] |

Table 7: *Regression coefficient estimates for PERC group response data, proportional odds model. Values in parentheses are estimated standard errors using information and ignoring possible within study correlations. Values in square brackets are estimated standard errors using the sandwich formula as explained in the text. The model pooled all the data across gender and species.*

| | Minimum (-1.699) | Median (0.062) | Maximum (1.146) |
|---|---|---|---|
| Complete, $s = 1$ | 3.69 | 2.67 | 2.43 |
| | (0.32) | (0.29) | (0.34) |
| | [0.60] | [0.50] | [0.51] |
| Interval Censored, $s = 1$ | 4.06 | 2.87 | 2.58 |
| | (0.24) | (0.20) | (0.24) |
| | [0.25] | [0.37] | [0.41] |
| Complete, $s = 2$ | 4.81 | 3.80 | 3.56 |
| | (0.33) | (0.16) | (0.20) |
| | [0.42] | [0.17] | [0.20] |
| Interval Censored, $s = 2$ | 4.84 | 3.65 | 3.37 |
| | (0.23) | (0.13) | (0.16) |
| | [0.17] | [0.23] | [0.27] |

*

Table 8: *ED10 estimate for PERC in the proportional odds model, for $s = 1, 2$ and three values of log-duration, namely the minimum, median and maximum. Values in parentheses are estimated standard errors, obtained using the sandiwch formula. Values in square brackets are estimated standard errors using the sandwich formula as explained in the text. The model pooled all the data across gender and species.*

| | Complete | Interval Censored |
|---|---|---|
| $\alpha_1$ | -26.51 | -28.44 |
| | (7.59) | (5.14) |
| | [8.15] | [3.41] |
| $\alpha_2$ | -31.59 | -31.55 |
| | (8.10) | (5.40) |
| | [9.13] | [3.63] |
| $\gamma_{\text{humans}}$ | 3.29 | 5.75 |
| | (11.28) | (7.82) |
| | [4.89] | [4.45] |
| $\gamma_{\text{mice}}$ | 7.88 | 10.68 |
| | (7.65) | (5.46) |
| | [8.67] | [5.91] |
| $\beta_{x,\text{rats}}$ (Concentration) | 6.18 | 6.64 |
| | (1.82) | (1.17) |
| | [1.99] | [0.80] |
| $\beta_{x,\text{mice}}$ (Concentration) | 4.93 | 4.53 |
| | (1.21) | (0.90) |
| | [1.68] | [1.04] |
| $\beta_{x,\text{humans}}$ (Concentration) | 7.74 | 7.31 |
| | (3.19) | (2.23) |
| | [2.36] | [1.14] |
| $\beta_z$ (Duration) | 2.76 | 2.77 |
| | (0.67) | (0.47) |
| | [0.81] | [0.48] |

Table 9: *Regression coefficient estimates for PERC group response data, proportional odds model. Values in parentheses are estimated standard errors using information and ignoring possible within study correlations. Values in square brackets are estimated standard errors using the sandwich formula as explained in the text. The model has different intercepts and concentration slopes for the different species.*

| | Minimum (-1.699) | Median (0.062) | Maximum (1.146) |
|---|---|---|---|
| Rats | | | |
| Complete, $s = 1$ | 4.695 | 3.665 | 3.422 |
| | (0.253) | (0.161) | (0.211) |
| | [0.241] | [0.168] | [0.200] |
| Interval Censored, $s = 1$ | 4.661 | 3.701 | 3.474 |
| | (0.127) | (0.114) | (0.139) |
| | [0.212] | [0.127] | [0.133] |
| Complete, $s = 2$ | 5.518 | 4.489 | 4.245 |
| | (0.406) | (0.164) | (0.148) |
| | [0.392] | [0.239] | [0.231] |
| Interval Censored, $s = 2$ | 5.130 | 4.170 | 3.943 |
| | (0.144) | (0.068) | (0.089) |
| | [0.226] | [0.123] | [0.122] |
| Mice | | | |
| Complete, $s = 1$ | 4.284 | 2.994 | 2.689 |
| | (0.225) | (0.241) | (0.289) |
| | [0.236] | [0.422] | [0.486] |
| Interval Censored, $s = 1$ | 4.471 | 3.064 | 2.732 |
| | (0.188) | (0.214) | (0.268) |
| | [0.229] | [0.334] | [0.392] |
| Complete, $s = 2$ | 5.316 | 4.025 | 3.720 |
| | (0.277) | (0.128) | (0.160) |
| | [0.219] | [0.225] | [0.284] |
| Interval Censored, $s = 2$ | 5.158 | 3.752 | 3.419 |
| | (0.225) | (0.112) | (0.163) |
| | [0.219] | [0.191] | [0.250] |
| Humans | | | |
| Complete, $s = 1$ | 3.324 | 2.502 | 2.307 |
| | (0.168) | (0.292) | (0.352) |
| | [0.109] | [0.159] | [0.178] |
| Interval Censored, $s = 1$ | 3.448 | 2.575 | 2.369 |
| | (0.133) | (0.221) | (0.271) |
| | [0.105] | [0.118] | [0.131] |
| Complete, $s = 2$ | 3.981 | 3.159 | 2.965 |
| | (0.313) | (0.155) | (0.180) |
| | [0.109] | [0.111] | [0.126] |
| Interval Censored, $s = 2$ | 3.874 | 3.002 | 2.796 |
| | (0.197) | (0.132) | (0.172) |
| | [0.067] | [0.071] | [0.088] |

Table 10: *ED10 estimate for PERC in the proportional odds model, for $s = 1, 2$ and three values of log-duration, namely the minimum, median and maximum. Values in parentheses are estimated standard errors obtained using information and ignoring possible within study correlations. Values in square brackets are estimated standard errors using the sandwich formula as explained in the text. The model has different intercepts and concentration slopes for the different species.*

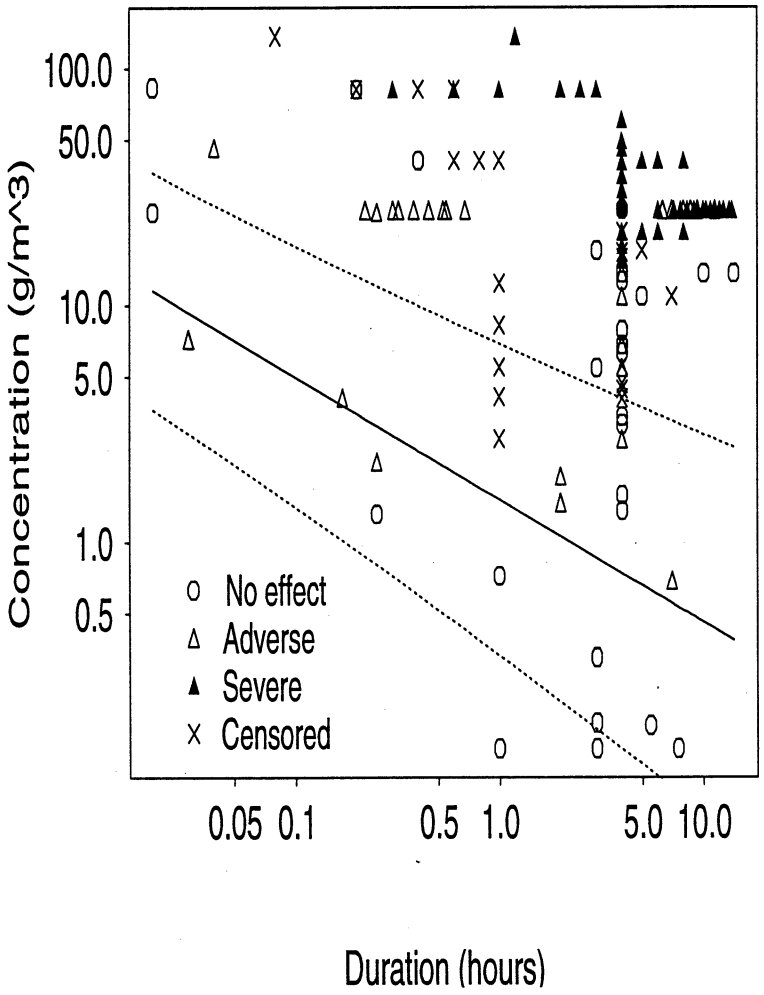


Figure 1: *PERC data, ED10 line (solid line) when pooling all studies, with associated 95% confidence bands (dashed lines).*

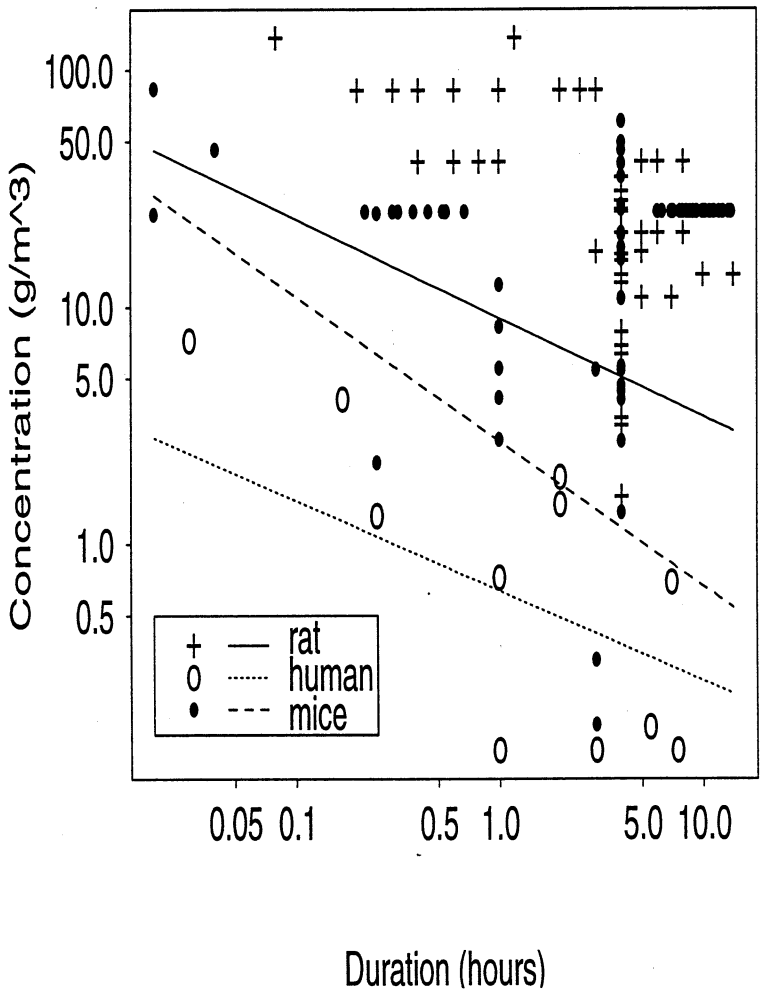


Figure 2: *PERC Data, ED10 lines when intercepts and slopes are stratified by species.*

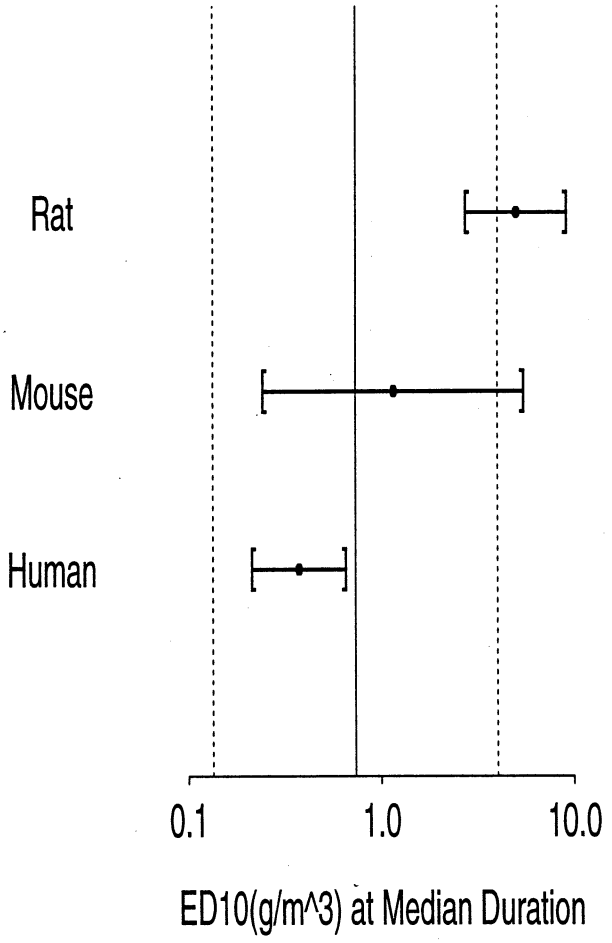


Figure 3: *PERC data ED10's for different species combinations at the median duration of all studies. The solid verticle line is the log (base 10) ED10 when all studies are pooled, while the dashed vertical lines are the associated 95% confidence intervals.*

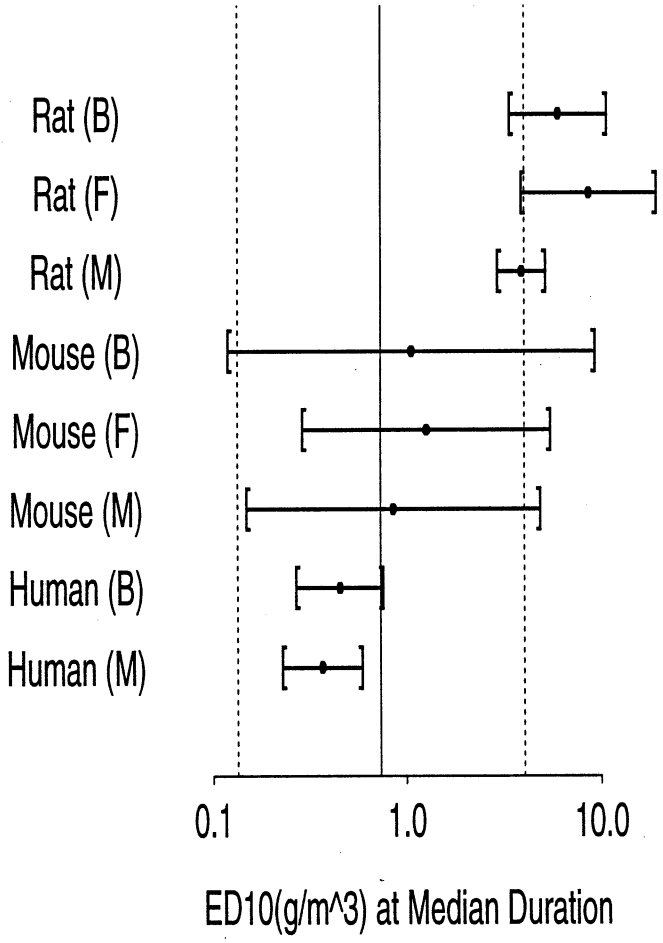


Figure 4: *PERC data ED10's for different sex and species combinations at the median duration of all studies. The solid verticle line is the log (base 10) ED10 when all studies are pooled, while the dashed vertical lines are the associated 95% confidence intervals.*