NISS

Scaled Link Functions and Generalized Likelihood Inferences for Heterogeneous Binary Response Data

Douglas G. Simpson, Raymond J. Carroll and
Minge Xie

Technical Report Number 24
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.

Scaled Link Functions and Generalized Likelihood Inferences for Heterogeneous Binary Response Data

Douglas G. Simpson, Raymond J. Carroll and Minge Xie *

This paper describes binary regression models in which the link function has scale parameters that may be estimated along with the regression parameters. One motivation is to provide a plausible model for group binary responses. In this case a natural class of scale functions is obtained by treating the group responses as thresholded averages of possibly correlated latent variables. The latent variables represent individual responses within the group. Scaled link functions also arise naturally if the latent response variables are heteroscedastic. We develop generalized likelihood ratio inferences in case the observed binary responses are themselves correlated. Group response data from environmental toxicology illustrate the scaled link function analysis.

KEYWORDS: Generalized estimating equation; Latent variable; Logistic regression; Probit regression; Robust model.

---

* Douglas G. Simpson is Associate Professor of Statistics, University of Illinois Urbana–Champaign, Champaign, IL 61820. Raymond J. Carroll is Professor of Statistics, Texas A&M University, College Station, TX 77843-3143. Minge Xie is Graduate Student, University of Illinois Urbana–Champaign, Champaign, IL 61820. This research was supported by US EPA Cooperative Agreement #CR819638-01-0 with the National Institute of Statistical Sciences, NSF contract DMS 92-07730, NSA contract MDA 904-92-H-3058, and National Cancer Institute grant CA-57030.

1 INTRODUCTION

Data in which the individual responses are binary are frequently modeled by a generalized linear model of the form

$$\Pr(Y_i = 1|x_i) = 1 - \Pr(Y_i = 0|x_i) = G(\beta^T x_i), \quad i = 1, 2, \ldots, n, \tag{1}$$

where Y_i is the response of the ith individual, x_i is a vector of explanatory variables, and G is a cumulative distribution function. Common choices for the link function G are a standard normal distribution, which is used in probit analysis, a logistic distribution, which is the inverse of the logit function, and the Weibull, which has as its inverse the complimentary log–log function.

Our concern is with situations in which the assumed constancy of the link function across heterogeneous observations is untenable. For example, in the toxicological data presented in section 5, we have group binary responses rather than the individual outcomes. The number of individuals in each group is known, but the information on individual incidences is missing. With groups of unequal size, the assumption of a common scale is suspect. In other settings we may have other grounds for expecting different levels of precision for different observations, e.g., if an underlying continuous response Z follows a heteroscedastic linear model, and we observe whether or Z exceeds a threshold.

The method is easily implemented, because the scaling of the link function is equivalent to a rescaling of the regression variables. If the scale function contains unknown parameters, as in our group response example, then the estimation of the regression parameters is easily nested within the maximum likelihood optimization for the scale parameters.

In order to make inferences about the scale parameters, and also to address situations in which correlations exist between the observed binary responses, we develop generalized likelihood confidence regions by treating a working likelihood as an estimating criterion, and applying the large sample theory of likelihood ratio statistics under nonstandard conditions. The resulting likelihood type confidence intervals provide a useful alternative to the commonly used delta–method confidence intervals, because they are invariant to the parameterization and they are range preserving.

An appealing feature of the scaled link function model for group responses is that, although the observed elements are marginal responses, the parameters of the model are defined at the individual level and are therefore comparable across groups of different sizes. Moreover, we can combine group and individual level data by treating an individual as a group of size 1.

The rest of the paper is organized as follows. Section 2 provides some derivations of scaled link function models. Section 3 is concerned with profile likelihood and generalized profile likelihood methods for computing confidence intervals. An application of the methodology in toxicological risk assessment is discussed in Section 4. Section 5 provides some further remarks.

2 DERIVATIONS AND INTERPRETATION

To model group binary responses and other forms of heterogeneous data we replace (1) by a model of the form

$$\Pr(Y_i = 1|x_i) = 1 - \Pr(Y_i = 0|x_i) = G(\beta^T x_i w_i), \quad i = 1, 2, \ldots, n, \tag{2}$$

where the design weights w_i are scaling factors and refer to the precision of latent observations. Effectively, we replace the link function $G(\,\cdot\,)$ by the scaled link function $G(\,\cdot\, w_i)$.

We first discuss how this model arises in general from an assumed weighted likelihood. Then we derive special forms of the weight function under certain hierarchical latent structures. For group binary responses, a latent variable model described Section 2.2 suggests weights of the form

$$w_i = n_i^{1/2}\{1 + (n_i - 1)\rho\}^{-1/2}, \tag{3}$$

where ρ is the intraclass correlation coefficient for the latent responses in a group. In a different vein, a latent random coefficient model, described in section 2.3, yields weights of the form

$$w_i = (1 + x_i^T \Gamma x_i)^{-1/2}. \tag{4}$$

In this case we have a type of robust model that bounds the influence of outlying regression variables.

2.1 WEIGHTED LIKELIHOODS

Within the framework of logistic regression, a scaled link function follows from the assumption that a group binary response has a density of the form

$$\Pr(Y = y|\pi, w) \propto \pi^{wy}(1 - \pi)^{w(1-y)}, \quad y = 0, 1;\ \pi \in [0, 1] \tag{5}$$

where the normalizing constant is $\pi^w + (1-\pi)^w$. A more convenient expression is obtained by reparameterizing in terms of the odds, $\theta = \pi(1 - \pi)^{-1}$. A Bernoulli random variable with odds θ has the density

$$f(y|\theta) = \theta^y (1 + \theta)^{-1}, \quad y = 0, 1.$$

Using this parameterization and normalizing in (5) yields the group model,

$$f(y|\theta, w) = \theta^{wy}(1+\theta^{w})^{-1}, \tag{6}$$

a Bernoulli model with odds θ^w. In logistic regression $\theta = \exp(x^T\beta)$ and $\theta^w = \exp(wx^T\beta)$. In this case (6) is a scaled logistic regression model with link–scale w^{-1}.

2.2 LATENT INDIVIDUAL RESPONSES

One way to derive the probit model for binary response data is to assume that the binary response to stimulus level d is an indicator of the event $(Z \leq d)$, where Z is the minimum level of stimulation required to achieve a positive response. Assuming Z is a Gaussian random variable with mean μ and variance σ^2, one obtains the probit model,

$$\Pr(Y=1|d) = \Pr(Z \leq d) = \Phi(-\mu\sigma^{-1} + \sigma^{-1}d),$$

where $\Phi(t) = (2\pi)^{-1/2}\int_{-\infty}^{t}\exp(-t^2)dt$. Defining the probit regression parameters by $\alpha = -\mu\sigma^{-1}$ and $\beta = \sigma^{-1}$, we see that β relates to the precision of the latent variable Z, and $-\alpha\beta$ is the mean. Prentice (1976) used this kind of quantal response modeling as a starting point in deriving other shapes for the link function. For a comprehensive account of quantal response analysis see Morgan (1992).

Next we note that covariate information may be incorporated by supposing that $\mu = z^T\eta$ for a vector of known covariates z. Propagating the covariate information through the link model yields

$$\Pr(Y=1|x) = \Phi(-z^T\eta\sigma^{-1} + d\sigma^{-1}) = \Phi(x^T\beta), \tag{7}$$

where $x^T = (d, z^T)$ and $\beta^T = (\sigma^{-1}, -\sigma^{-1}\eta^T)$.

To model the response of a group subjected to stress d, we suppose that the response occurs because d exceeds the average stimulus required to achieve a response, that is,

$$(Y_i = 1) = (\bar{Z}_i \leq d), \quad \text{where} \quad \bar{Z}_i = n_i^{-1}\sum_{j=1}^{n_i} Z_{ij}$$

and $Z_{i1}, \ldots, Z_{in_i}$ are the latent responses in the ith group. To allow for within–group correlation between individuals, we suppose the latent variables themselves follow a latent variable model,

$$Z_{ij} = \mu_i + a_i + e_{ij} \quad \text{and} \quad \mu_i = z_i^T\eta, \tag{8}$$

where z_i is an observed vector of covariates, and a_i and e_{ij} are independent random variables such that $a_i \sim N(0, \sigma_a^2)$ and $e_{ij} \sim N(0, \sigma_e^2)$.

Setting $\sigma^2 = \sigma_a^2 + \sigma_e^2$, Z_{ij} is marginally distributed as $N(\mu_i, \sigma^2)$, and the intraclass correlation coefficient is $\rho = \sigma_a^2/\sigma^2$. Model (8) implies that $\bar{Z}_i$ is normally distributed with mean μ_s and variance $(\sigma^2/n_i)\{1 + (n_i - 1)\rho\} = \sigma^2/w_i^2$, where w_i is given by (3). By analogy with (7) we have the scaled probit model,

$$\Pr(Y_i = 1|x) = \Phi\{(-\mu_i\sigma^{-1} + \sigma^{-1}d)w_i\} = \Phi(w_i x^T \beta), \tag{9}$$

with $x^T = (d, z^T)$ and $\beta = (\sigma^{-1}, -\sigma^{-1}\eta^T)$. McCulloch (1994) used a related latent variable model for correlated binary outcomes. The difference here is that (9) is for aggregates of latent variables.

The square of the weight function may be interpreted as an effective group size. If $\rho = 0$ then $w_i^2 = n_i$, whereas a positive correlation shrinks the weight towards one. In the extreme, if $\rho = 1$, then the effective group size is 1. Figure 1 shows how the weight function varies over a range of values of ρ and n_i.

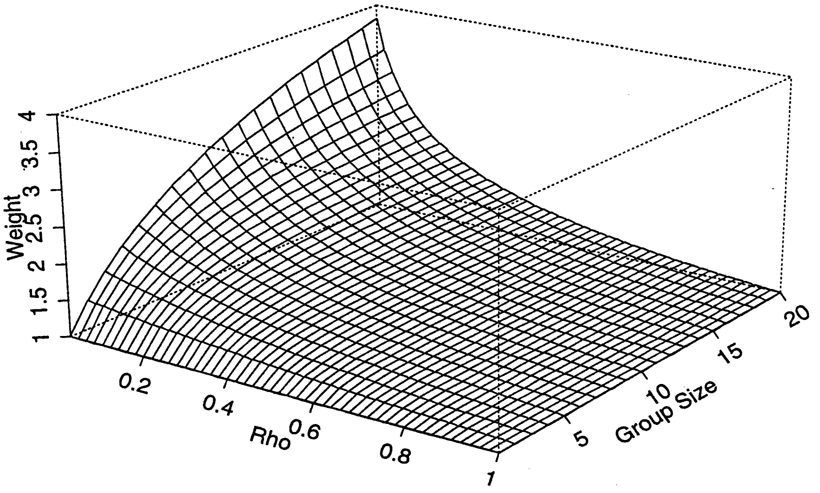

Figure 1: Group weight as a function of group size and latent correlation (ρ).

Identifiability of the parameter ρ requires heterogeneity of the group sizes. If the group sizes were all the same, then the weighting would simply introduce a constant scale factor in the regression parameters. If the group sizes are heterogeneous, then the weights in (3) identify β and γ at the individual level, $n_i = 1$. In this way the scaled model avoids attenuation effects that would otherwise occur in the combined analysis of group responses with heterogeneous group sizes. In practice, the parameter ρ may be difficult to estimate precisely, because the thresholded observations supply only crude information about scaling.

The derivation of (9) extends approximately to other link functions. For instance, the logistic link function, $\Psi(t) = \{1 + \exp(-t)\}^{-1}$, is near to a Gaussian link function in absolute terms, because

$$\sup_t |\Psi(1.702\, t) - \Phi(t)| < 0.01;$$

see, e.g., Baker (1992, p. 16). Thus, the preceding analysis might be used as a heuristic to suggest a weight function for scaled logistic regression.

2.3 HETEROSCEDASTIC LATENT VARIABLES

The model of the preceding section is a special case of a general class of models in which one supposes that a latent variable Z_i follows a heteroscedastic linear model,

$$(Z_i + x_i^T\beta)w_i \,|\, x_i \;\sim\; G, \tag{10}$$

but we observe only $Y_i = 1_{(-\infty,0]}(Z_i)$ rather than Z_i. We then have the weighted link function model,

$$\Pr(Y = 1|x_i) = G(w_i x_i^T\beta) = 1 - \Pr(Y = 0|x_i), \tag{11}$$

where the weights are interpreted as inverse scale factors for the latent responses.

Another interesting special case occurs under a hierarchical setup in which the latent variable follows a linear model with random coefficients, that is,

$$\begin{aligned} Z_i|x_i, b_i \;&\sim\; N(-x_i^T b, 1) \\ b_i &\sim N(\beta, \Gamma) \end{aligned} \tag{12}$$

The marginal distribution of Z_i is then normal with mean $-x_i^T\beta$ and variance $1 + x_i^T\Gamma x_i$. For fully observed Z_i this is a standard growth curve model; see, for instance, Johansen (1984). Such a model might also be useful for observational data. If we observe only the indicator Y_i, then we obtain a binary regression model of the form (11) with $w_i = (1 + x_i^T\Gamma x_i)^{-1/2}$. This is a special case of the general type of latent model considered by Zeger, Liang and Albert (1988) and McCulloch (1994), who focused on the latent structure for modeling correlated binary outcomes. Here the observable binary responses are independent, but heteroscedastic.

In the preceding example the latent distribution for b_i is a way of expressing uncertainty about the linear model. The resulting link function model therefore controls the influence of extreme values of x_i. Note, in particular, that $\|w_i x_i\| \leq \lambda_{\min}^{-1/2}(\Gamma)$, where $\lambda_{\min}$ denotes the smallest eigenvalue. As a result, the maximum likelihood estimator has a bounded influence function in the sense of Hampel, Rousseeuw, Ronchetti and Stahel (1986). For further discussion of robust logistic regression see Pregibon (1982), Copas (1988), Künsch, Stefanski and Carroll (1989), and Carroll and Pederson (1993).

3 GENERALIZED PROFILE LIKELIHOOD

Inferences about ρ in the group model of section 2.2 provide information about the effect of grouping and the correlations between individuals in the same group. If the group effect is the only random effect, then the profile likelihood for ρ is straightforward to compute.

Assume model (2) holds with weights given by (3). The log–likelihood is given by $\ell(\beta, \rho | y, x) = \sum_{i=1}^{N} \{y_i \log(G_i) + (1 - y_i) \log(1 - G_i)\}$, where $G_i = G\{w_i(\rho)\beta^T x_i\}$. Given the weights $w_i(\rho)$ we compute $\hat{\beta}(\rho)$ and the profile likelihood $\ell(\rho) = \ell\{\hat{\beta}(\rho), \rho\}$ as follows:

(i) Set $\tilde{x}_i = w_i(\rho) x_i$;

(ii) Compute $\hat{\beta}(\rho)$ by maximizing $\ell(\beta, 1 \; | y, \tilde{x})$;

(iii) Set $\ell(\rho) = \ell\{\hat{\beta}(\rho), 1 \; | y, \tilde{x}\}$.

The method is computationally convenient because $\ell(\beta, 1 \; | y, \tilde{x})$ is an unweighted log–likelihood for binary regression. Standard software will perform the optimization for each fixed ρ, typically by iterative weighted least squares or the scoring algorithm. Computing $-2\ell(\rho)$ over a grid in $[0, 1]$ provides a deviance profile, which yields a likelihood based confidence region for ρ as illustrated in Figure 2.

It often happens that the response data are clustered in some way. For instance, in combining data from multiple experiments, the experimental run is a natural clustering variable. In this situation it is necessary to account for correlations between clustered observations. In particular, the profile likelihood given above no longer has the standard large sample theory associated with likelihood ratio tests and likelihood confidence intervals. However, a nonstandard theory of likelihood ratios is available; see Huber (1967), Schrader and Hettmansperger (1980), Kent (1982), Ronchetti (1982) and Li and McCullagh (1994). This theory is essentially an extension of the theory of generalized estimating equations, where the estimating equation is assumed to correspond to a criterion function. In our case we begin with a natural criterion, the working likelihood, and proceed from there to develop the working likelihood inferences.

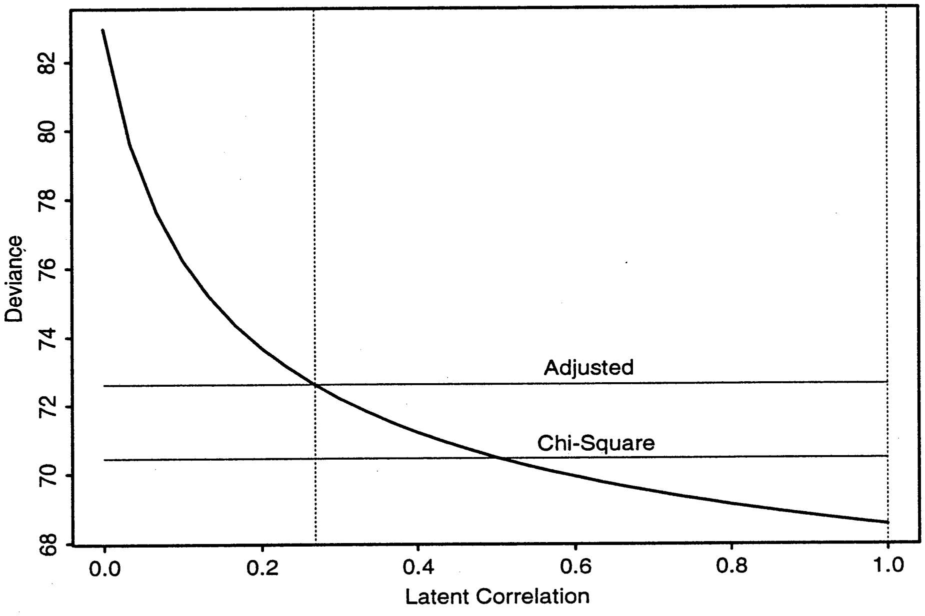

Figure 2: Deviance profile of latent correlation parameter for tetrachloroethylene data. Cutoff values are shown for the likelihood ratio (chi–square) and generalized likelihood (adjusted) tests.

3.1 GENERAL THEORY

In the general theory we consider inferences about a parameter vector θ, and we assume that the estimate $\hat{\theta}$ maximizes an estimating criterion, $\ell(\theta|x,y)$, which is effectively the working log likelihood, although it need not be the logarithm of an actual density function. We refer to e^{ℓ} as the generalized likelihood function. Li and McCullagh (1994) considered a class of generalized likelihoods determined by the first and second moment structure, in which case they called e^{ℓ} the quasilikelihood. Define the score function,

$$\mathcal{U}(x,y|\theta) = \frac{\partial}{\partial\theta}\ell(\theta|x,y), \tag{13}$$

the score covariance,

$$\mathcal{J} = \mathrm{E}\{\mathcal{U}(x,y|\theta)\mathcal{U}(x,y|\theta)^T\}, \tag{14}$$

and the negative expected hessian,

$$\mathcal{H} = -\mathrm{E}\left\{\frac{\partial}{\partial\theta^T}\mathcal{U}(x,y|\theta)\right\}. \tag{15}$$

If ℓ were the true log likelihood, then we would have $\mathcal{H} = \mathcal{J}$, but this equality usually fails for generalized likelihoods. As in the theory of generalized estimating equations, the parameter θ is determined by the equation $E\{\mathcal{U}(x,y|\theta)\} = 0$.

We partition $\theta = (\gamma^T, \eta^T)^T$, where γ is the p–dimensional parameter vector of interest, and η is the vector of nuisance parameters. Partition $\mathcal{H}$ similarly as

$$\mathcal{H} = \begin{pmatrix} \mathcal{H}_{\gamma\gamma} & \mathcal{H}_{\gamma\eta} \\ \mathcal{H}_{\eta\gamma} & \mathcal{H}_{\eta\eta} \end{pmatrix},$$

and define $\mathcal{H}_{\gamma\gamma\cdot\eta} = \mathcal{H}_{\gamma\gamma} - \mathcal{H}_{\gamma\eta}\mathcal{H}_{\eta\eta}^{-1}\mathcal{H}_{\eta\gamma}$.

Let $\hat{\theta}_0 = (\gamma_0^T, \hat{\eta}_0^T)^T$ denote the maximum generalized likelihood estimate subject to $\gamma = \gamma_0$. We need the large sample distribution of the log generalized likelihood ratio,

$$\mathcal{L}(\gamma_0|x,y) = 2\{\ell(\hat{\theta}|x,y) - \ell(\hat{\theta}_0|x,y)\}. \tag{16}$$

The following result is well–known under various regularity conditions. For the basic idea of the proof see Kent (1982).

Theorem: *If $\gamma = \gamma_0$, then, as the number of independent observations increases, $\mathcal{L}(\gamma_0|x,y)$ converges in distribution to $\sum_{k=1}^{p} \lambda_k W_k$, where $W_1, \ldots, W_p$ are independently distributed as χ_1^2, and $\lambda_1, \ldots, \lambda_p$ are the eigenvalues of $\mathcal{H}_{\gamma\gamma\cdot\eta}(\mathcal{H}^{-1}\mathcal{J}\mathcal{H}^{-1})_{\gamma\gamma}$.*

To use this result in practice, either to perform a generalized likelihood ratio test of $H_0 : \gamma = \gamma_0$, or to compute a generalized likelihood confidence set for γ_0, we need to estimate the matrices $\mathcal{H}$ and $\mathcal{J}$. If all data are independent, an obvious approach is to replace the theoretical expectations in (14) and (15) by the analogous empirical averages. The next section discusses the situation in which the data are correlated in batches.

We also need to compute quantiles of the distribution of $\sum_k \hat{\lambda}_k W_k$. Observe that if $p = 1$ the appropriate distribution is simply a scaled χ_1^2 distribution. If $p > 1$, then algorithms given by Marrazzi (1980) and Griffiths and Hill (1985) may be used. A quick and simple way to do the computation is to simulate from the distribution of $\sum_k \hat{\lambda}_k W_k$, since chi–square deviates are easy to generate.

3.2 CORRELATED BINARY RESPONSES

We now suppose the response data are binary and correlated in batches. For instance, in the data of section 4 we have observations from a number of different studies, and the responses within a particular study are naturally assumed to be correlated. We construct a generalized likelihood confidence interval for the latent within–group correlation parameter. The modifications for other parameters are obvious.

Let $\theta = (\beta^T, \rho)^T$, and let n_i be the number of response groups in the ith batch. We consider the generalized likelihood,

$$\ell(\theta|x,y) = \sum_{i=1}^{n}\sum_{j=1}^{n_i}\{y_{ij}\log(G_{ij}) + (1-y_{ij})\log(1-G_{ij})\},$$

with score function, $\mathcal{U}(x,y|\theta) = \sum_{i=1}^{n}\Psi_i$, where

$$\Psi_i = \sum_{j=1}^{n_i}(y_{ij}-G_{ij})\frac{\dot{G}_{ij}}{G_{ij}(1-G_{ij})}\begin{pmatrix} w_{ij}x_{ij} \\ \dot{w}_{ij}\beta^T x_{ij}\end{pmatrix},$$

$G_{ij} = G(w_{ij}\beta^T x_{ij})$, $\dot{G}_{ij} = G'(w_{ij}\beta^T x_{ij})$, $w_{ij} = w_{ij}(\rho) = n_{ij}^{1/2}\{1+(n_{ij}-1)\rho\}^{-1/2}$, $\dot{w}_{ij} = -\frac{1}{2}w_{ij}^3(n_{ij}-1)/n_{ij}$, and n_{ij} is the number of latent responses for the jth group in batch i. As in the theory of generalized estimating equations described, e.g., by Liang and Zeger (1986), the idea here is to treat the correlated observations as a multivariate observation and view Ψ_i as a function of this vector. Thus we estimate $\mathcal{H}$ by $\hat{\mathcal{H}} = \frac{\partial}{\partial\theta}\mathcal{U}(x,y|\theta)$, but use a batched estimate of $\mathcal{J}$, given by $\hat{\mathcal{J}} = \{n/(n-p)\}\sum_{i=1}^{n}\Psi_i\Psi_i^T$, where p is the number of parameters in the model. Note that the asymptotics require the number of batches to be large rather than the number of groups within each batch. Given these estimates, one then constructs estimates $\hat{\lambda}_k$ of the eigenvalues λ_k by insertion into the formula given in the Theorem of section 4.1. This is particularly convenient for a one–dimensional parameter, in which case we have a scalar adjustment.

In the next section we use the generalized likelihood technique both for confidence intervals and for testing within a hierarchical family of models.

4 EXAMPLE: GROUP ANALYSIS OF TOXICOLOGY DATA

Figures 3(a)-(d) show data from multiple studies on the effects of acute inhalation exposure to tetrachloroethylene. The responses are binary indicators for adverse (solid diamond) versus nonadverse (hollow circle) outcomes. These outcomes were determined by a toxicologist during a thorough review of the literature. Figure 3(a) shows all of the outcomes plotted versus log concentration and log duration; Figures 3(b)-(d) break this down by species. The lines in the graphs are effective concentrations estimated to give a 10% response rate, conditional on the duration. We call these the EC_{10} lines.

The data are from 17 studies ranging in size from 1 group to 35 groups, with median of 4 groups per study. These data comprise a subset of a larger database. They include all the group responses for studies in which adverse and nonadverse outcomes can be distinguished. The larger database includes some studies for which only partial response information is available. Carroll, Simpson and Zhou (1994) discussed censored analysis of such studies.

Usually the reason for group responses rather than individual incidences is that the published results are simply descriptions of lesions and tissue damage typical of the group. This leads to data in which the group size is known, and an overall group adversity level can be assigned, but individual incidence information is missing. In addition to the response indicator, we have information on the exposure concentration and duration, the species, and the gender.

We fit a scaled link function model for logistic regression on $\log_{10}$(concentration) and $\log_{10}$(duration) with the intercepts stratified on species × gender and with the concentration parameter stratified on species. Of particular interest is the question of how much information group binary responses provide as compared with individual outcomes. The latent correlation parameter, ρ, allows us to investigate this question.

Figure 2 shows the profile deviance function for ρ, along with the projected 95% generalized likelihood confidence interval computed via the sandwich approximation method described in section 4. The cutoff value was computed by batching the groups within studies as described in section 3.2. We see that the maximum likelihood estimate of ρ is 1, but the confidence interval is rather wide. A naive confidence interval based on the usual chi–square approximation is a bit narrower but suspect because of the likely correlation within studies. The within *group* correlation is significantly different from 0. In fact, a group binary response appears to contain about the same information as a single individual binary response. This is useful knowledge for the analysis of the existing database, and for the design of future toxicological studies. It implies that individual data should be obtained whenever possible.

Another question is how much stratification is necessary in the model. We used the batched generalized likelihood method to test the hypothesis that the concentration parameters for the different species are equal. The log generalized likelihoods are obtained as deviances from standard logistic regression software after weighting the regression variables. The drop in deviance in going from the smaller to the larger model was 2.118. The estimated eigenvalues from the sandwich formula were 1.44 and 0.69. Comparing the log generalized likelihood ratio to quantiles of $1.44W_1 + 0.69W_2$ gives a p-value of 0.35, so there is little evidence of nonparallelism among these data. Figure 3 shows why. Within species the EC_{10} lines show little difference, and the difference could easily be explained by random variation.

The main work was in computing $\hat{\mathcal{H}}$, $\hat{\mathcal{J}}$ and the appropriate eigenvalues. To get quantiles of $\sum \lambda_i \chi_i^2 |_{\lambda_i = \hat{\lambda}_i}$ we generated 5000 observations from the distribution.

A further test of equality of intercepts is highly significant. With concentration entering the model logarithmically, the different intercept parameters for the different species correspond to dose conversions. Because the different species have different inhalation rates and metabolisms, one would expect such an adjustment of the dosage across species.

As the example illustrates, once the machinery is in place for performing these kinds of tests, model selection and confidence intervals may proceed quite analogously to standard likelihood based analysis.

5 DISCUSSION

If both group and individual level data are available, the latent model for the group responses provides the basis for combining these sources of information. The key is that the parameters are defined at the individual level rather than at the group level. In particular (5) implies that the individual ordinal responses follow model (6) with $w_i = 1$. (Observe that the appropriate variance of the individual latent response is σ^2, not σ_e^2.) Therefore the regression parameters of the group and individual response models have the same interpretation.

In applications, however, some caution is required in combining group and individual responses when the individuals are also subject to group correlations. The group–only responses might well appear to have a high within group correlation in part because the group level reporting itself inflates the apparent agreement among elements of the group. Thus, although the latent responses appear to be highly correlated, this need not imply that fully observed individual responses would exhibit the same high correlation. For this reason it would be a good idea to model the within group correlations separately for individual level and group–only data.

The scaled models generalize the unscaled models just as weighted least squares regression generalizes ordinary least squares. The simple technique of weighting the regression variables may be useful in a variety of problems. In the group ordinal model, group size and within group correlation are natural effects on which to model the scale of the link function. In other settings it might be more natural to key on covariate information to model heterogeneity of the scale.

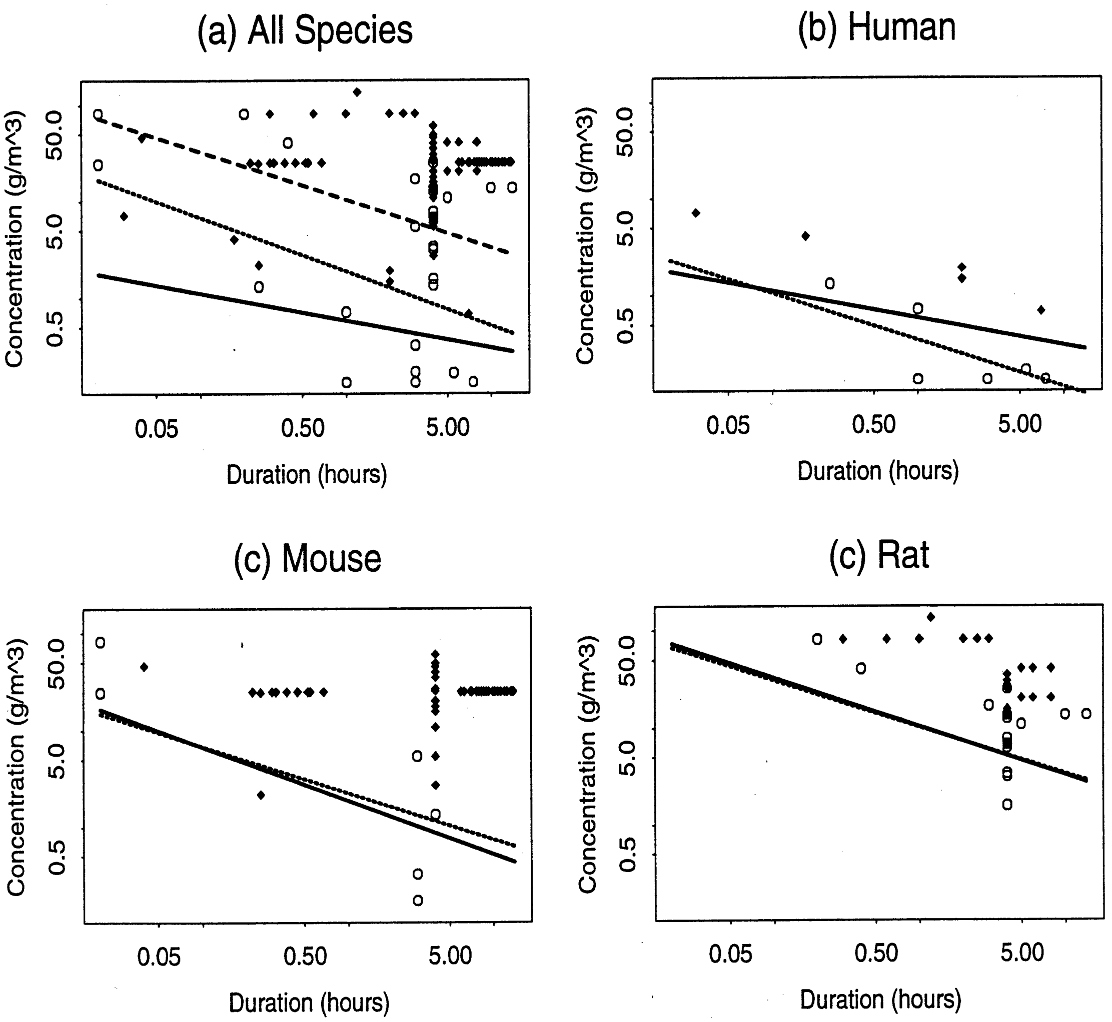

Figure 3: Tetrachloroethylene data showing the binary response (hollow circle = nonadverse, solid diamond = adverse) versus concentration and duration for (a) pooled data, (b) humans only, (c) mice only, and (d) rats only. Lines in (a) are EC10's for humans (solid), mice (dotted) and rats (dashed). Lines in (b)-(d) are EC10's for unconstrained slope model (solid) and parallel slope model (dotted).

Sandwich type confidence intervals, i.e., delta method confidence intervals derived from generalized estimating equations, are often suggested as a means for dealing with correlations among the responses. Although convenient, these confidence intervals have some well–known drawbacks. They lack invariance to reparameterization, they are frequently inaccurate without a transformation of some kind, and they can lead to confidence intervals that extend beyond the boundary of the parameter space, e.g., in the estimation of a correlation parameter. To address these problems we have developed generalized likelihood confidence intervals, which have similar properties as likelihood based intervals. In particular, they are invariant to transformations, and they are range preserving, so they offer an improvement on delta method confidence intervals. In the context of scaled link function modeling, generalized likelihood ratio tests and confidence intervals provide a convenient way to test for structure in the model in the presence of correlations.

ACKNOWLEDGMENT

Daniel Guth of the U.S. Environmental Protection Agency supplied the data on tetrachloroethylene.

REFERENCES

Anderson, D.A. and Aitkin, M. (1985). Variance component models with binary response: interviewer variability. *Journal of the Royal Statistical Society*, Series B, 47, 203-210.

Baker, F.B. (1992) *Item Response Theory.* Marcel Dekkar, New York.

Carroll, R.J. and Pederson, S. (1993). On robustness in the logistic regression model. *Journal of the Royal Statistical Society*, Series B, 55, 693-706.

Carroll, R.J., Simpson, D.G. and Zhou, H. (1994). Interval censoring and marginal analysis in ordinal regression. Technical Report, National Institute of Statistical Sciences, Research Triangle Park, NC.

Copas, J.B. (1988). Binary regression models for contaminated data (with discussion). *Journal of the Royal Statistical Society*, Series B, 225-265.

Griffiths, P. and Hill, I.D. (1985). *Applied Statistics Algorithms.* Horwood, London.

Hampel, F.R., Ronchetti, E.M., Rousseeuw, P.J., and Stahel, W.A. (1986). *Robust Statistics: The Approach Based on Influence Functions.* Wiley, New York.

Huber, P.J. (1967). The behavior of maximum likelihood estimates under nonstandard conditions. *Proceedings of the 5th Berkeley Symposium*, 1, 221-233.

Johansen, S. (1984). *Functional Relations, Random Coefficients, and Nonlinear Regression with Application to Kinetic Data.* Lecture Notes in Statistics, 22, Springer-Verlag, New York.

Kent, J.T. (1982). Robust properties of likelihood ratio tests. *Biometrika*, 69, 19-27.

Künsch, H.R., Stefanski, L.A. and Carroll, R.J. (1989). Conditionally unbiased bounded influence estimation in general regression models, with applications to generalized linear models. *Journal of the American Statistical Association*, 84, 460-466.

Li, B. and McCullagh, P. (1994). Potential functions and conservative estimating functions. *Annals of Statistics*, 22, 340-356.

Liang, K-Y. and Zeger, S.L. (1986). Longitudinal data analysis using generalized linear models. *Biometrika*, 73, 13-22.

Marazzi, A. (1980). ROBETH, a subroutine library for robust statistical procedures. COMPSTAT 1980, Proceedings in Computational Statistics, Physica, Vienna.

McCulloch, C.E. (1994). Maximum likelihood variance components estimation for binary data. *Journal of the American Statistical Association*, 89, 330-335.

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

Pregibon, D. (1982). Resistant fits for some commonly used logistic models with medical applications. *Biometrics*, 38, 485-498.

Prentice, R.L. (1976). A generalization of the probit and logit methods for dose response curves. *Biometrics*, 32, 761-768.

Ronchetti, E. (1982). Robust testing in linear models: the infinitesimal approach. Ph.D. Thesis. ETH, Zurich.

Schrader, R.M. and Hettmansperger, T.P. (1980). Robust analysis of variance based upon a likelihood criterion. *Biometrika*, 67, 93-101.

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