Generalized Estimating Equations with Applications

K. P. Singh, D. Bass, A. A. Bartolucci

Department of Biostatistics, Ryals Public Health Building, 1665 University Boulevard,
The University of Alabama at Birmingham, Birmingham, AL 35294-0008, USA

Abstract: Quantal response functions have wide applicability for predicting probabilities of adverse outcome. Recently, emphasis has arisen to extend their utilization to model in teratological and genotoxic phenomena. In this application, however, outcomes within litter may be interdependent, so techniques that incorporate correlation structure are recommended. We endorse the generalized estimating equation (GEE) methodology due to its flexibility in accommodating multivariate responses. In this paper, GEEs are proposed from the Mahalanobis distance viewpoint. Then, via a reasonable premise, we extend the GEEs to include estimation of correlation information. Treatment of inverse prediction inference as it relates to dosage is also considered. The resulting framework is illustrated using published data from a study on radiation-induced defects.

1. Introduction

Toxicity studies using rodent models are often used to provide data to governmental agencies vested with informing and protecting the public with regard to potentially dangerous substances. End products of these data may include synthesized risk assessment models which measure the probability of adverse effect (quantal response) as a function of exposure level. A recent addition to areas exhibiting such interest involves that of developmental toxicity. A primary aim of this particular application seeks determination of the exposure level above background rates that affects a real increase in the risk of adverse teratological or fetatoxic events. These experiments typically involve control and exposure groups, each involving a multiplicity of pregnant dams at similar points in their gestational phase. Just prior to term the dams are sacrificed and their uterine contents enumerated with adverse effects noted.

A difference between developmental toxicity and other risk assessment programs (e.g. cancer) is that outcomes may tend to be correlated within the litter. Several approaches to account for this possibility have been forwarded. The Polya-Eggenberger distribution was used by Williams (1975) and Chen and Kodell (1989) for modeling the beta-binomial log-likelihood function. Ryan (1992a), noting that outcomes could be categorized other than dichotomously, considered the problem as it applies to modeling multinomial responses.

More recently, Ryan (1992b) capitalized on the multivariate nature of the generalized estimating equations (GEEs) (Liang and Zeger 1986; Zeger and Liang 1986), which is a quasi-likelihood (QL) (Wedderburn 1974) technique. An advantage of using GEEs is that the estimated parameters are consistent regardless of the true, but unknown correlation structure. Testing procedures based on binary data are summarized by Fung et at.(1994). Goodness-of-fit tests for GEE modeling with binary responses are discussed by authors including Hosmer and Lemeshow(1997), Barnhart and Williamson(1998).

A simple algorithm for generating binary data is given by Park, Park and Shin(1996). The GEEs provide a flexible mechanism for modeling multivariate outcomes. However, as noted by Ryan (1992b), most of the available computer software implicitly assumes linearity of the parameters within a chosen predictor. To overcome this deficiency, we implement the nonlinear predictor suggested by Chen and Kodell (1989) and Ryan (1992b) into the GEE methodology.

The GEEs provide a flexible mechanism for modeling multivariate outcomes. However, as noted by Ryan (1992b), most of the available computer software implicitly assumes linearity of the parameters within a chosen predictor. To overcome this deficiency, we implement the nonlinear predictor suggested by Chen and Kodell (1989) and Ryan (1992b) into the GEE methodology. The paper begins by first proposing GEEs using the weighted least-squares method as applied to the Mahalanobis distance point of view. This is followed by expansion of the resulting

framework to include estimation of correlation components. Inverse prediction and its inferential consequences are briefly explored. Illustration of these results is provided with an example on radiation-induced teratological defects.

2. Generalized Estimating Equations

The development parallels that of Liang and Zeger (1986) in the original formulation of the GEEs. First, it is postulated that individual responses possess marginal distributions from the exponential family:

$$P(Y=y) = f_Y(y;\lambda,\phi) = \exp[\phi\{y\lambda - a(\lambda)\} + b(y,\phi)\ ] \quad (1)$$

from which the well-known result is obtained (McCullagh and Nelder 1983):

$$\mu_y = a'(\lambda) \quad (2)$$
$$\sigma^2 = a''(\lambda)/\phi \quad (3)$$

where λ and ϕ are the canonical and scale parameters, respectively. The typical approach is to maximize the likelihood (or equivalently its logarithm). In the event that the scale parameter is unknown, this maximum likelihood (ML) technique reverts to quasi-likelihood (QL) estimation (Wedderburn 1974).

The canonical parameter, λ, is often equated to a predictor η, a function of covariate information. This is accomplished, owing to its ensuing simplicity, is provided by generalized linear models (Nelder and Wedderburn 1972) in which $\eta = \underline{X}^T \underline{\xi}$, where $\underline{X}$ is a vector of concomitant information and $\underline{\xi}$ is a vector of structural parameters. However, this need not be the case. Nonlinearity is introduced by setting η to $(\alpha + \beta d^{\lambda})$, where d represents the dose or exposure level of the substance under investigation and $(\alpha, \beta, d^{\lambda})^T$ $(= \underline{\xi})$ is the specific structural parameter vector that is to be estimated. This predictor has been advocated by Chen and Kodell (1989) and Ryan (1992b) for teratological research. It must be emphasized that expectations (Equation 3) are modeled explicitly regardless of the form of the predictor.

The scale parameter, ϕ , as alluded to earlier, may or may not be known. In the event that the latter is true, it must be estimated. We shall establish in the subsequent development a rationale and means by which to accomplish this. Fortunately, its presence or absence does not affect the modeling process.

A covariance matrix $\sum$ may be decomposed into $S^{1/2} R S^{1/2}$, where S is a diagonal matrix of variance elements and R is the correlation matrix. By postulate, variance elements are of the form exhibited by Equation 4. Hence, because ϕ is assumed fixed for a given application (whether known or not), it also may be factored out: $\sum = A^{1/2} R A^{1/2}/\phi = \phi^{-1} V$, where elements of diagonal matrix A are of the form $a''(\lambda)$. Therefore, the Mahalanobis distance may be expressed as:

$$D^2 = \varphi[\sum_{i=1}^{L}\sum_{j=1}^{M_i} (\underline{y}_{ij} - \underline{\mu}_{ij})^T V_{ij}^{-1} (\underline{y}_{ij} - \underline{\mu}_{ij})] = \varphi Q,$$

the product of a fixed scalar and a quadratic form functional in $\underline{\xi}$ through the expectation vectors.

Having written the Mahalanobis distance in this form, we observe that quadratic form Q is free of the scale parameter. Using the least squares approach, we equate the partial derivative of D^2 to zero. Thus, we need only to differentiate Q because ϕ has been exercised from the problem through division.

Note that technically the process of differentiation and setting to zero is not an extreme location exercise in this setting. This is because we model the expectation vectors as having exclusive functional relationships with $\underline{\xi}$. In general, elements of V_{ij}^{-1} share in this property also; but we, following the QL methodology, choose instead to use these matrices as weighting factors which vary through a direct mean-variance association (e.g. Equations 2 and 3). For this reason the current derivation is analogous to weighted least squares. Interested readers may survey alternative estimators based on minimization of the quadratic form (Bass, Singh and Hardin 1998). After taking derivatives and algebric simplifications the following expression was resulted:

$$\sum\sum D_{ij}^T V_{ij}^{-1} (\underline{Y}_{ij} - \underline{\mu}_{ij}) = \underline{0} = g(\underline{\xi}) \quad (4)$$

where $\underline{\mu}_{ij}$ is the n_{ij} x 1 expectation vector and V_{ij} is the n_{ij} x n_{ij} matrix proportional to the covariance of $\underline{Y}_{ij}$, with D_{ij} $(= \partial\underline{\mu}_{ij}/\partial\underline{\xi}^T)$ being the n_{ij} X 3 (in our setting) partial derivative matrix. Equation 6, constitutes the GEEs (also known as gradient equations) and upon solution yields the estimate $\hat{\underline{\xi}}$ of $\underline{\xi}$. We next examine these components in detail.

The expectation vector $\underline{\mu}_{ij}$ consists of elements μ_{ijk} (k = 1 to n_{ij}), which are often modeled marginally by the logistic response function, the canonical choice given independent Bernoulli responses. That function is sigmoidal in shape, as are many continuous cumulative distribution functions. Others, such as the standard normal, Weibull, and extreme value distribution functions, could very well be used in a similar manner. Given this spectrum of possibilities, let us postulate that $\underline{\mu}_{ijk}$ be modeled by $F(\lambda_i)$, where $F(\lambda_i)$ is a continuously differentiable cumulative distribution function.

Upon equating canonical parameter λ_i to predictor η_i, we typify the quantal response modeling approach: $\mu_{ijk} = F(\lambda(d_i)) = F(\alpha + d^{\gamma})$. This last formula is often denoted the risk function, as it expresses the probability of adverse outcome functionally with dosage.

The matrix V_{ij} was seen earlier to be factorable into $A^{1/2} R_{ij} A^{1/2}$, where A_{ij} is composed of diagonal elements of the form $a''(\lambda_{ijk})$, and R_{ij} is a correlation matrix. Because dichotomous responses are observed, a convenient choice for representing elements of A_{ij} is $\mu ijk(1 - \mu_{ijk})$, the Bernoulli variance function under conditions of independence. Our next decisions are not as theoretically based, but remain tenable. First, we postulate that the correlation matrix is constant across all indices: $R_{ij} = R$. Then, cognizant of the fact that cluster members originate under nearly identical conditions, we hypothesize that an exchangeable correlation structure is reasonable. Given these propositions, only the single additional parameter, correlation coefficient, ρ, becomes involved to augment ξ and ϕ for estimation. The partial derivative matrix D_{ij} attains a particularly simple form. Upon equating marginal means to cumulative distribution functions, derived D_{ij} (Bass, Singh and Hardin 1998).

Next, we briefly describe the technique by which the covariance of ξ and is fundamental to the estimation process:

$$H(\xi) = - \sum \sum D_{ij}^{T} V_{ij}^{-1} Dij \qquad (5)$$

GEEs are solved to yield estimates for ξ, ρ and ϕ. The Hessian, $\partial g(\xi)/\partial \xi$ is related to the asymptotic

$$\mathrm{Var}(\hat{\xi}) = \mathrm{Var}(- H^{-1} g) = (-H^{-1}) \mathrm{Var}\{g\} (-H^{-1}) \qquad (6)$$

In the event that the true covariance $\mathrm{Var}\{\underline{y}_{ij}\}$ is specified correctly by $\sum_{ij}$, the above expression simplifies to (Bass, Singh, Hardin 1998):

$$\mathrm{Var}(\xi) = -(\phi.H)^{-1}$$

Equation 6 is referred to as the naive estimated covariance matrix. The details for computing its robust analogue are given in Liang and Zeger (1986).

Estimation of the exchangeable correlation coefficient ρ (jointly with ξ) requires augmentation onto the gradient (Equation 4) and Hessian (Equation 5) quantities described earlier. Components of the joint iterative system are displayed in matrix augmentation in Bass (1993) and Bass, Singh and Hardin (1998).

3. Inference On Dosage - The Inverse Problem

The ED_{01} is the dosage level that effects a one-percent increase in adverse response over the observed background rate. Given the quantal response model the ED_{01} is given as follows:

$$ED_{01} = \{ F^{-1}[F(\lambda(0) + 0.01]/\beta \}^{1/\gamma}$$

Another, but more conservative quantity, is the benchmark dose (BD) (Crump 1984). This is defined as the lower 95% confidence limit (CL) on the ED_{01}. Because dose has been the independent variable throughout the previous development, we must now pursue a strategy for inverting the inference.

Several approaches have been suggested for computing the BD. A likelihood-based method was suggested by Chen and Kodell (1989). A benefit of this technique is that it yields attainable (i.e. positive) values. But, as noted by Ryan (1992b), adaptation of this procedure for GEEs has not yet been accomplished. Until newer techniques evolve, we follow Gart et al. (1986) and Ryan (1992a, 1992b) in using the asymptotically multivariate normal distribution of the predictors to base variance estimates for CL construction. Extrapolation to the BD is direct, but (caveat) negative values loom possible. Two suppositions for the

situations where γ is known or unknown are given in Bass, Singh and Hardin (1998).

3.1. γ known

This assumption allows the predictor to be considered essentially linear. We begin by recalling the definition of the risk function: $\mu = F(\alpha + \beta d^{\gamma})$, where ideally all parameters are known. Upon assuming this function is one-to-one (as is the case for continuous monotonic increasing functions), it is invertible. Thus, $F^{-1}(\mu)$ equals $(\alpha + \beta d^{\gamma})$.

Solving for d^{γ} implies:

$$d^{\gamma} = \frac{F^{-1}(\mu) - \alpha}{\beta}$$

In reality, μ is only estimated through modeling. Hence:

$$d^{\gamma} = \frac{F^{-1}(\hat{\mu}) - \hat{\alpha}}{\hat{\beta}} .$$

Therefore,

$$var\{d^{\gamma}\} = \beta^{-2}[1, d^{\gamma}] \; var\{[\hat{\alpha}, \hat{\beta}]^{T}\}[1, d^{\gamma}]^{T} \qquad (7)$$

In practice, point estimates are inserted for parameters. The BD is approximated by

$$[(ED_{01})^{\gamma} - 1.645\hat{\sigma}]^{1/\gamma}$$

where $\leq$ is the square root of Equation 10 evaluated at $(d=ED_{01})$.

3.2. γ unknown

Unlike above, this assumption requires the predictor to remain nonlinear. Using the derivation above, we may now write by $var\{d^{\hat{\gamma}}\} = \beta^{-2} var\{(\hat{\alpha} + \hat{\beta} d^{\hat{\gamma}})\}$. However, the vector factorization does not directly follow. Therefore, consider a Taylor series expansion of around the true parameter vector ξ. Upon a first-order approximation:

$$var\{d^{\hat{\gamma}}\} = \beta^{-2}[1, d^{\hat{\gamma}}, \beta d^{\hat{\gamma}} \log(d)] \; var\{\hat{\alpha}, \hat{\beta}, \hat{\gamma}]^{T}\}[1, d^{\hat{\gamma}}, \beta d^{\hat{\gamma}} \log(d)]^{T} \qquad (8)$$

As before, the BD is computed by

$$[(ED_{01})^{\gamma} - 1.645\hat{\sigma}]^{1/\gamma}$$

where $\leq$ is the square root of Equation 8 evaluated at $(d=ED_{01})$. But due to the inclusion of variance and covariance information from $\hat{\gamma}$, this BD will most likely differ from that previously derived.

4. Radiation Exposure

Using the derived GEE framework we present tables for the following example. Flexibility is illustrated through fitting the data with logistic $[F(\lambda) = \{1 + \exp(-\lambda)\}^{-1}]$, standard normal $[F(\lambda) = \Phi(\lambda)]$, Weibull $[F(\lambda) = 1 - \exp(-\lambda)]$, and extreme value $[F(\lambda) = 1 - \exp(-\exp(\lambda)]$ distribution functions to model quantal responses. Rai and Van Ryzin (1985) examined the effect on fetuses sired by male CBA mice irradiated within a week before mating. The data set was originally analyzed in Lüning et al. (1966). In this experiment, a control group was utilized along with exposure groups receiving 300 and 600 rads (R). Total impregnations involved 683, 604, and 486 dams, respectively, with a total of 11775 implants conceived.

Following Rai and Van Ryzin (1985), dose levels were rescaled by a factor of 10^{-3} prior to modeling. To reduce selection bias, unlike the antecedent reports, all data is incorporated into the current analysis. Table I furnishes parameter estimates across the modeling distributions.

All structural parameter values are extremely significant ($p < .0001$), and, surprisingly, correlation coefficient and scale estimates agree to the third decimal place. The low values accorded ρ are especially of acknowledgment, in that they support the position of Lüning et al. (1966) by denoting an absence of perceived intralitter correlations.

Table II validates the close agreement between observed and predicted proportions incurring events. Nonetheless, divergence is observed at levels around 650R. This illustrates that further research is necessary to establish model selection policies when dose-response extrapolation beyond ranges provided by the data is desired.

Similarly, but not nearly as obvious, low-dose extrapolation has problems needing resolution. The Weibull fit differs contrastingly in that it has the least slope. A direct consequence is revealed in Table III, wherein the Weibull ED_{01} and BDs are seen to be

several times larger than the others, and hence, are suggestive of more liberal levels of dosing.

On the other hand, Weibull modeling produces lower (ED_{01}) variance estimates, thus providing inferential appeal. These various points highlight the necessity of creating definitive, theoretically based guidelines for teratological analysis methodologies.

5. Discussion

The flexibility of the GEE approach in teratological applications has been illustrated with both theoretical considerations and a demonstrative example. Perhaps the most beneficial attribute afforded by this technique is the relaxation of specific distributional assumptions, an especially welcome virtue in view of the sparsity of multivariate distributions in general. The hallmark of this approach, capitalizing on QL, while remaining vague towards a correlation structure, embodied by the consistency of the GEE estimator (Liang and Zeger, 1986).

As noted earlier, a valid mean-variance relationship is imperative. However, modeling of the mean itself is, for the most part, left to choice. The remaining quantity, a postulated form of the correlation, follows intuitively from the nature of the data in this setting. Although this assumption was made partly for its simplicity and ease of implementation, a more valid approach may be to utilize separate exchangeable correlation structures, paralleling Chen and Kodell (1989), for each dosage level.

A persistent problem, shared with other modeling approaches, concerns inverse prediction inference as it applies to low-dose extrapolation. As we observed via example, individual models may agree closely in other respects yet yield BDs of sizable relative difference. Furthermore, suppositions on which to base variance estimators and CLs often are not well-founded on theory and are therefore open to criticism -- particularly disturbing are the conspicuous usages of power transformations. We agree with Ryan (1992b) that a likelihood-based approach seems more appropriate, and it remains a current topic for investigation.

4. References

Barnhart, H. X., Williamson, J. M. Goodness-of-fit tests for GEE modeling with binary data. *Biometrics* 54, 720-729, 1998.

Bass, D., Singh, K. P., and Hardin, J. M. Generalized estimating equations with nonlinear predictors in teratological applications, *Journal of the Malaysian Statistical Institute*, Accepted.

Chen, J. J. and Kodell, R. L. Quantitative risk assessment for teratological effects', *Journal of the American Statistical Association*, 84, 966 - 971, 1989.

Fung, K. Y., Krewski, D., Rao, J. N. K., Scott, A. J. Tests of trend in developmental toxicity experiments with correlated binary data. *Risk Analysis* 14, 621-630, 1994.

Crump, K. S. 'A new method for determining allowable daily intakes', *Fundamental and Applied Toxicology*, 4, 854-871, 1984.

Gart, J. J., Krewski, D., Lee, P. N., Tarone, R. E., and Wahrendorf, J. *Statistical Methods in Cancer Research, Volume III: The Design and Analysis of Long-Term Animal Experiments*, International Agency for Research on Cancer, Lyon, 1986.

Johnson, R. A. and Wichern, D. W. Applied Multivariate Statistical Analysis, Prentice-Hall, Englewood Cliffs, NJ, 1982.

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

Lüning, K. G., Sheridan, W., Ytterborn, K. H., and Gullberg, U. The relationship between the number of implantations and the rate of intra-uterine death in mice, *Mutation Research*, 3, 444-451, 1966.

McCullagh, P. and Nelder, J. A. *Generalized Linear Models*, Chapman and Hall, London, 1983.

Nelder, J. A. and Wedderburn, R. W. M. Generalized linear models, *Journal of the Royal Statistical Society, Series A*, 135, 370-384, 1972.

Park, C. G., Park, T., Shin, D. W. A simple algorithm for generating correlated binary data. *American Statistician*, 50, 306 - 310, 1996.

Rai, K. and Van Ryzin, J. A dose-response model for teratological experiments involving quantal responses, *Biometrics*, 41, 1-9, 1985.

Ryan, L. Quantitative risk assessment for developmental Toxicity, Biometrics, 40, 163-174, 40, 163-174 1992a.

Ryan, L. The use of generalized estimating equations for risk assessment in developmental toxicity, *Risk Assessment*, 12, 439-447 1992b.

Williams, D. A. The analysis of binary responses from toxicological experiments involving reproduction and teratogenicity, *Biometrics*, 31, 949-952, 1975.

Zeger, S. L. and Liang, K.-Y. Longitudinal data analysis for discrete and continuous outcomes, *Biometrics*, 42, 121-130, 1986.

Table I. Estimates of parameters

Quantal response model	Structural (naive std error)			Nuisance	
	α	β	γ	ρ	ϕ
Probit	-1.283(.026)	1.489(.074)	0.732(.066)	0.011	1.001
Logistic	-2.200(.050)	2.528(.120)	0.679(.063)	0.011	1.001
Weibull	0 .105(.005)	0.711(.054)	1.112(.094)	0.011	1.001
Extreme	-2.254(.048)	2.171(.096)	0.627(.059)	0.011	1.001

Table II. Observed and estimated quantal responses

	Dosage (R)		
	0	300	600
OBSERVED	479/4809 = .09961	1004/3975 = .25258	1191/2991 = .39819
PREDICTED			
Probit	.09974	.25285	.39827
Logistic	.09978	.25287	.39829
Weibull	.09986	.25289	.39830
Extreme	.09968	.25286	.39828

Table III. ED_{01} and BD estimates ($R \times 10^{-3}$)

Quantal response model	ED_{01}	(1) λ known		(2) λ unknown	
		$\hat{\sigma}$	BD	$\hat{\sigma}$	BD
Probit	.009432	.019244	.001223	.017194	.001840
Logistic	.011038	.016812	.001682	.015369	.002292
Weibull	.023894	.007057	.007155	.007187	.006821
Extreme	.007481	.021232	.000816	.018741	.001322