STATISTICS IN MEDICINE
Statist. Med. **19**, 389–404 (2000)

A SEMIPARAMETRIC APPROACH TO ANALYSING DOSE–RESPONSE DATA

QUINTON J. NOTTINGHAM[1] AND JEFFREY B. BIRCH[2]*

[1] *Department of Management Science & Information Technology, Virginia Polytechnic Institute and State University, 1007 Pamplin Hall, Blacksburg, VA 24061-0235, U.S.A.*
[2] *Department of Statistics, Virginia Polytechnic Institute and State University, Blacksburg, VA 24061, U.S.A.*

SUMMARY

In the analysis of a quantal dose-response experiment with grouped data, the most commonly used parametric procedure is logistic regression, commonly referred to as 'logit analysis'. The adequacy of the fit by the logistic regression curve is tested using the chi-square lack-of-fit test. If the lack-of-fit test is not significant, then the logistic model is assumed to be adequate and estimation of effective doses and confidence intervals on the effective doses can be made. When the tolerance distribution of the dose–response data is not known and cannot be assumed by the user, one can use non-parametric methods, such as kernel regression or local linear regression, to estimate the dose–response curve, effective doses and confidence intervals. This research proposes another alternative based on semi-parametric regression to analysing quantal dose–response data called model-robust quantal regression (MRQR). MRQR linearly combines the parametric and non-parametric predictions with the use of a mixing parameter. MRQR uses logistic regression as the parametric portion of the model and local linear regression as the non-parametric portion of the model. Our research has shown that the MRQR procedure can improve the fit of the dose–response curve by producing narrower confidence intervals for predictions while providing improved precision of estimates of the effective doses with respect to either logistic or local linear regression results.

1. INTRODUCTION

Indirect quantal biological assays are used for the estimation of the potency of a substance by means of a quantal response variable. In indirect quantal assays, one concern is with the estimation of the distribution of the tolerance (see Finney[1]) of the subjects to a particular drug or chemical being studied, where tolerance is defined as the dose just sufficient to elicit a response from the subject.

In an indirect quantal assay, the response is a Bernoulli random variable with parameter P_i, the probability that a subject will respond to dose x_i. Suppose that an increasing sequence of doses x_1, x_2, $\ldots$, x_d is given to n_i subjects and that r_i subjects respond. Then, $p_i = r_i/n_i$, the sample proportion responding at the ith dose, estimates the true probability, P_i for $i = 1, 2, \ldots, d$.

* Correspondence to: Jeffrey B. Birch, Department of Statistics, Virginia Polytechnic Institute and State University, Blacksburg, VA 24061, U.S.A.

CCC 0277–6715/2000/030389–16$17.50

Received June 1998
Accepted March 1999

The probabilities of response can be modelled parametrically[1] as $P_i = F(\beta_0 + \beta_1 x_i)$ where F is the tolerance cumulative distribution function. The most commonly used parametric procedures are probit analysis, which uses the standard normal CDF, and logit analysis, which uses the standard logistic CDF. This paper will emphasize the use of logistic regression analysis, but the set-up will be such that any general form of F may be used. See Morgan[2] for other parametric models. The procedure for estimating the coefficients, β_0 and β_1, is the method of maximum likelihood, see, for example, Finney,[1] Govindarajulu,[3] Goedhart[4] or Morgan,[2] yielding the estimated equation of $\hat{P}_i = F(\mathbf{x}_i' \hat{\boldsymbol{\beta}})$ where $\mathbf{x}_i' = (1 \ x_i)$ and $\hat{\boldsymbol{\beta}} = (\hat{\beta}_0 \ \hat{\beta}_1)'$. When the user is satisfied with the chosen model, the primary concern then shifts to the estimation of effective doses. The effective dose, denoted by $ED_{100\alpha}$, is the dose where 100α per cent of the subjects in the population show a response. The most commonly used value of α is 0·5. Extreme dose levels, such as the ED_{01}, ED_{10} and the ED_{90}, may be regarded as percentiles of interest for many biological problems. The parametric confidence interval on effective dose is typically computed using Fieller's theorem.[1,5] Other methods that can be used for confidence intervals on effective dose are the delta method[4] and the likelihood-ratio interval.[5]

It may happen that neither the logit nor the probit models provide an adequate fit to the data. In such a case, the researcher may try other parametric models that have been suggested for use in quantal bioassays (Morgan[2]). However, due to the behaviour of the data, it may be difficult to find an adequate parametric model. One possible solution then is to fit the data with a non-parametric method such as kernal regression or local linear regression. These methods provide a model-free estimate of the dose–response curve. One advantage of a non-parametric method over a misspecified parametric model is a reduction in bias of fit. On the other hand, it is well-known that non-parametric regression methods tend to place too much emphasis on random behaviour in the data, resulting in estimates of fits that are imprecise, especially when compared to the fits obtained by the parametric model.

One solution to this bias versus variance dilemma is to use an estimation technique that combines the stability of fits from the parametric model with the reduction in bias provided by non-parametric methods. This can be accomplished by taking a linear combination of the two fits using a mixing parameter, λ. Such a semi-parametric approach has been used successfully by several others, including Rahman *et al.*,[6] Burman and Chaudhuri,[7] Woolridge,[8] Robinson[9] and Ullah and Vinod,[10] for the continuous measurement variable case. The primary concern of several of these papers[6,8,10] is to use λ to test for lack of fit of the parametric model, the implication being that if lack of fit is found, the parametric model will be altered to another more appropriate parametric model. We suggest a slightly different method now adapted to the quantal assay situation as described above. Essentially, our method repairs an inadequate parametric model by incorporating useful information from the non-parametric fit by taking the proper linear combination of these two fits. The proposed method will be called model-robust quantal regression (MRQR) because it 'robustifies' the quantal parametric model by adjusting the parametric predictions with non-parametric predictions through a mixing parameter.

The non-parametric method considered in this paper for estimating the dose–response relationship in the quantal response variable setting is local linear regression (LLR), a procedure first introduced by Cleveland[11] and later studied extensively by Fan[12] and Fan and Gijbels[13], among others. Another non-parametric method, kernel regression, has also been applied previously to quantal bioassays by a number of authors including references 14 to 17. Work by Nottingham,[18] however, has demonstrated that kernel regression is inferior to LLR in the quantal bioassay

Table I. An assay of deguelin (Martin[20])

Dose (log 10)	n	r
0·71	49	16
1·00	48	18
1·31	48	34
1·48	49	47
1·61	50	47
1·70	48	48

setting, a fact consistent with the superiority of LLR over kernel regression in other applications (see, for example, Fan and Gijbels[19]).

To illustrate these ideas, consider the data presented in Table I and plotted in Figure 1 from Martin[20] which is a six-dose quantal assay with 48–50 subjects at each dose. In this case, the substance applied was deguelin, given to groups of *Macrosiphoniella sanborni*, the chrysanthemum aphis. The logistic fit to this data yields a chi-square statistic of 13·37 resulting in a significance level of 0·009, indicating that the logistic model fit to this data is inadequate. The poor fit of the logistic model suggests a non-parametric or semi-parametric procedure may be more pertinent in fitting this data.

Figure 1 shows the logistic regression fit, a second parametric model fit, the Aranda–Ordaz model,[2] and the MRQR (based on a convex combination of the logistic and LLR) fit to the data along with the raw data. The chi-square statistics for the Aranda–Ordaz model and MRQR fit are 6·49 and 4·54, respectively. The Aranda–Ordaz fit uses a three-parameter model, whereas the MRQR uses 2·72 'parameters' (model equivalent degrees of freedom, see Cleveland[11]). Examining Figure 1, it is quite clear that the MRQR fit is superior at four of the six doses to that of the logistic fit as well as the Aranda–Ordaz model. Logistic regression results in an estimate of the ED_{50} of 0·95 with a 95 per cent fiducial interval of (0·60, 1·16), considerably different and less precise from the estimate of 1·07 obtained by MRQR with interval estimate of (0·94, 1·16). This example is interesting in that is appears in Morgan[2] to motivate the use of other more complicated models, such as the Aranda–Ordaz, than the logistic for fitting quantal data. Yet, MRQR allows an improved fit using fewer parameters and without requiring knowledge of a more complicated model.

Details of the MRQR technique will be presented in Section 3, after a brief presentation in Section 2 of LLR applied to quantal regression. Section 4 presents the bias, variance and mean squared error properties of the proposed MRQR method, and Sections 5 and 6 present numerical and simulation results.

2. LOCAL LINEAR REGRESSION APPLIED TO QUANTAL BIOASSAYS

In LLR, the estimated response is obtained at each dose $x = x_i$ by fitting a simple linear regression line at x_i using the method of weighted least squares (WLS). The weights, $h_{i1}^k, h_{i2}^k, \ldots, h_{id}^k$, obtained using kernel regression, are given by

$$h_{ij}^k = K\left(\frac{x_i - x_j}{b}\right) \Big/ \sum_{j=1}^{d} K\left(\frac{x_i - x_j}{b}\right)$$

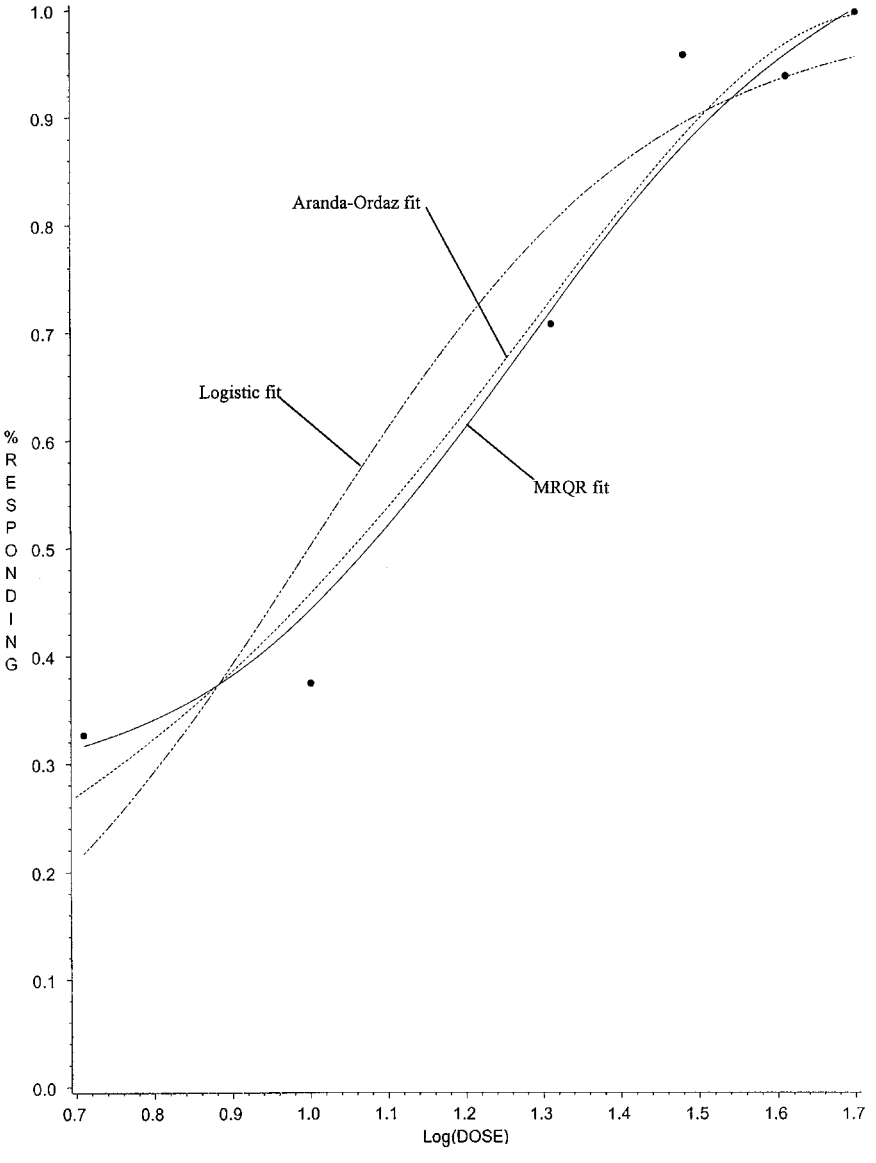

Figure 1. Logistic regression, Aranda–Ordaz model, and MRQR fits to the Martin[20] data

where $K(\,)$ represents some appropriate kernel function and b is the bandwidth. The LLR estimate at dose x_i is given by

$$\hat{\mathbf{P}}_i^{\mathrm{LLR}} = \hat{\beta}_{0i} + \hat{\beta}_{1i} x_i = \mathbf{x}_i'(\mathbf{X}'\mathbf{H}_i^k\mathbf{X})^{-1}\,\mathbf{X}'\mathbf{H}_i^k\,\mathbf{p} = \sum_{j=1}^{d} \mathbf{x}_i'(\mathbf{X}'\mathbf{H}_i^k\mathbf{X})^{-1}\,\mathbf{x}_j h_{ij}^k p_j = \sum_{j=1}^{d} h_{ij}^{\mathrm{LLR}}\, p_j \qquad (1)$$

where $\hat{\beta}_{0i}$, $\hat{\beta}_{1i}$ are the WLS estimates of the intercept and slope, respectively, $\mathbf{H}_i^k$ is the $d \times d$ diagonal matrix of kernel weights $(h_{i1}^k \ \ldots \ h_{id}^k)$, $\mathbf{X}$ is the $d \times 2$ model matrix of the form $(\mathbf{x}_1' \ \ldots \ \mathbf{x}_d')'$, and $\mathbf{p}$ is the $d \times 1$ vector of observed proportions. We note that $\hat{\mathbf{P}}_i^{\mathrm{LLR}}$ can also be thought of as a weighted average of the d sample proportions, with LLR weights of $h_{ij}^{\mathrm{LLR}} = \mathbf{x}_i'(\mathbf{X}'\mathbf{H}_i^k\mathbf{X})^{-1}\,\mathbf{x}_j h_{ij}^k$. One can express the $d \times 1$ vector of predictions, $\hat{P}^{\mathrm{LLR}}$, at the d doses as $\hat{\mathbf{P}}^{\mathrm{LLR}} = \mathbf{H}^{\mathrm{LLR}}\,\mathbf{p}$ where the ith row of $\mathbf{H}^{\mathrm{LLR}}$ is expressed as $\mathbf{x}_i'(\mathbf{X}'\mathbf{H}_i^k\mathbf{X})^{-1}\mathbf{X}'\mathbf{H}_i^k$.

One form for the approximate variance of the LLR estimate, assuming that the bandwidth is fixed, is

$$\mathrm{var}\,(\hat{P}_i^{\mathrm{LLR}}) = \mathbf{x}_i'(\mathbf{X}'\mathbf{H}_i^k\mathbf{X})^{-1}\,\mathbf{X}'\mathbf{H}_i^k\mathbf{V}\mathbf{H}_i^k\mathbf{X}(\mathbf{X}'\mathbf{H}_i^k\mathbf{X})^{-1}\,\mathbf{x}_i \tag{2}$$

where $\mathbf{V}$ is the $d \times d$ variance matrix of $\mathbf{p}$ where the ith diagonal element is $P_i(1-P_i)/n_i$, $i = 1, \ldots, d$. Using the variance expression in (2), one can compute confidence intervals for the LLR estimates of ED_α via inverse regression. The model degrees of freedom associated with the LLR fit is obtained following Cleveland[11] as the trace of the LLR weight matrix, $\mathrm{tr}[\mathbf{H}^{\mathrm{LLR}}]$.

Thus, there are at least two methods of analysis for quantal dose–response data: logistic regression (the parametric procedure) and non-parametric regression. If the observed data conform closely to the prescribed parametric model one would expect that the parametric approach offers the best results. On the other hand, if the model is misspecified, then the model is a poor descriptor of the data and the resulting logistic method is no longer optimal. In this case, the estimates of parameters, mean response and effective doses will be biased with increased variances. A non-parametric method such as local linear regression may perform better than logistic analysis in the misspecified model case with reduced bias and perhaps reduced variance, in the estimation of the response curve and of effective doses. The semi-parametric method proposed here, model-robust quantal regression, utilizes the best of both of these procedures; the lower variance of the parametric method and the lower bias of the non-parametric method under model misspecification.

3. MODEL-ROBUST QUANTAL REGRESSION APPLIED TO QUANTAL BIOASSAYS

Model-robust regression (MRR), proposed by Einsporn,[21] Einsporn and Birch[23] and Mays and Birch,[23] attempts to improve predictions in the linear regression setting by combining parametric and non-parametric predictions using a mixing parameter λ. That is, the MRR predictions at the n data points are obtained by $\hat{\mathbf{y}}^{\mathrm{MRR}}(\lambda) = \lambda\hat{\mathbf{y}}^{\mathrm{NP}} + (1-\lambda)\hat{\mathbf{y}}^{\mathrm{P}}$ where $\hat{\mathbf{y}}^{\mathrm{NP}}$ is the $n \times 1$ vector of non-parametric predictions and $\hat{\mathbf{y}}^{\mathrm{P}}$ is the $n \times 1$ vector of parametric predictions. The unknown parameter λ must be estimated from the data. A similar approach has been taken by Olkin and Spiegelman[24] for density estimation. Speckman[25] introduced 'partial linear regression' models in which the non-parametric and parametric portions of the model are additive. Work by others[6–10] addresses the test of lack of fit of the parametric portion of mixture models similar to the convex combination of MRR. The mixing parameter in these papers is estimated by a closed-form expression based on conditional least squares. We take a different approach in estimating λ, as will be demonstrated in Section 7. Additionally, semi-parametric regression has been used in conjunction with generalized linear models by Severini and Staniswalis[26] and Hastie and Tibshirani.[27]

Model-robust regression can be extended to quantal bioassays in an attempt to combine the parametric logistic regression analysis (or any other parametric fit) with non-parametric regression methods in fitting dose–response data. We will demonstrate that model-robust quantal

regression (MRQR) provides an improved analysis of dose–response data by improved estimation of effective doses, and by providing narrower confidence bands for mean response, thus resulting in narrower confidence intervals on the $ED_{100\alpha}$ values. The model-robust method for fitting quantal dose–response data is

$$\hat{P}_i^{MRQR} = \lambda \hat{P}_i^{NP} + (1 - \lambda)\hat{P}_i^{P} \tag{3}$$

where $\hat{P}_i^{MRQR}$ is the model-robust estimate of probability of response at dose x_i, λ is the mixing parameter, $\hat{P}_i^{NP}$ and $\hat{P}_i^{P}$ represent the estimated probability that the subject will respond to dose x_i obtained through non-parametric and parametric methods, respectively. In this paper, LLR is the non-parametric method and logistic regression is the parametric method. However, any parametric method and/or non-parametric method for quantal data may be used. The mixing parameter λ, defined so that $0 \leqslant \lambda \leqslant 1$, determines the degree to which the parametric predictions are adjusted. The basic motivation of λ is as follows. If the parametric fit is adequate, then use of the non-parametric fit would increase the variance of the overall fit. A $\lambda \approx 0$ would control for this. On the other hand, if the parametric model has been misspecified, then the non-parametric fit should be used to improve upon it. The amount of misspecification, and thus the amount of correction needed from the non-parametric fit, is reflected in the size of λ.

Once the MRQR estimate has been computed, confidence bands on the MRQR curve can be calculated. From these, the $ED_{100\alpha}$ value of interest can be computed along with their confidence intervals. The effective doses will be obtained iteratively by the method of inverse regression. That is, iterate over the range of the doses until x_α is found such that

$$\hat{P}^{MRQR}(x_\alpha) = \lambda \hat{P}^{NP}(x_\alpha) + (1 - \lambda)\hat{P}^{P}(x_\alpha) = \alpha. \tag{4}$$

An approximate 95 per cent pointwise confidence interval for the true probability of response at any dose x_0 can be obtained by $\hat{P}_0^{MRQR} \pm 1{\cdot}96\sqrt{\{\mathrm{var}(\hat{P}_0^{MRQR})\}}$ where one proposed expression for $\mathrm{var}(\hat{P}_0^{MRQR})$ is

$$\mathrm{var}(\hat{P}_0^{MRQR}) = [\lambda \mathbf{h}_0^{NP} + (1 - \lambda)\mathbf{B}_0]' \; V \; [\lambda \mathbf{h}_0^{NP} + (1 - \lambda)\mathbf{B}_0] \tag{5}$$

where V is defined in (2) and $\mathbf{h}_0^{NP}$ is the $d \times 1$ vector of non-parametric weights given by $\mathbf{h}_0^{NP} = (h_{01}^{NP} h_{02}^{NP} \ldots h_{0d}^{NP})'$ where, in our application here, h_{0j}^{NP} is h_{0j}^{LLR} from (1), $j = 1,2, \ldots, d$. The matrix $\mathbf{B}_0$ in (5) is defined as $\mathbf{B}_0 = f(\mathbf{x}_0'\boldsymbol{\beta})\,\mathbf{x}_0'(\mathbf{X}'\mathbf{W}\mathbf{X})^{-1}\mathbf{X}'\mathbf{W}\langle f(\mathbf{X}\boldsymbol{\beta})\rangle^{-1}$ where $\langle f(\mathbf{X}\boldsymbol{\beta})\rangle$ is a diagonal matrix with elements $(f(\mathbf{x}_1'\boldsymbol{\beta})\, f(\mathbf{x}_2'\boldsymbol{\beta}) \ldots f(\mathbf{x}_d'\boldsymbol{\beta}))$ *where* $f()$ is the PDF associated with the CDF, F. $\mathbf{W}$ is a $d \times d$ diagonal matrix of 'true weights', $w_1, \ldots, w_d$, where $w_i = n_i f^2(\mathbf{x}_i'\boldsymbol{\beta})^* P_i Q_i$ with $P_i = F(\mathbf{x}_i'\boldsymbol{\beta})$ *and* $Q_i = 1 - P_i$. Then $\widehat{\mathrm{var}}(\hat{\mathbf{P}}_0^{MRQR})$ is found by replacing $\boldsymbol{\beta}$ by $\hat{\boldsymbol{\beta}}$, obtained by the parametric method.

4. BIAS, VARIANCE AND MEAN SQUARED ERROR PROPERTIES

The fits obtained by the parametric, the non-parametric, and the MRQR procedures can be compared by evaluating their bias, variance and mean squared error properties. In this section, a brief summary of our results is given. Derivations of the bias, variance and MSE formulae can be found in Nottingham and Birch.[28]

For the parametric procedure, for example, logistic regression, it will be assumed that the true probability of response at the d doses can be written as

$$\mathbf{P} = \mathbf{G}(\mathbf{x}) = \mathbf{F}(\mathbf{X}\boldsymbol{\beta}) + \mathbf{H}(\mathbf{x}) \tag{6}$$

where $\mathbf{P}$ is the $d \times 1$ vector of probabilities of response at doses $x_1, \ldots, x_d$, $\mathbf{G}$ is the true cumulative distribution function, a function of the d doses, $\mathbf{F}$ is the user's assumed model, and $\mathbf{H}(\mathbf{x})$ denotes the difference between the user's postulated model and the true model. The function $\mathbf{H}(\mathbf{x})$ represents the amount of model misspecification.

The bias, variance and mean squared error formulae that follow are all asymptotic in nature, and, consequently, they are only approximate formulae for finite samples. Using the model given in (6), the bias in estimation of d probabilities of response for the parametric procedure can be expressed as

$$\text{bias}(\hat{\mathbf{P}}^{\mathrm{P}}) \approx [\langle f(\mathbf{X}\boldsymbol{\beta})\rangle \mathbf{X}(\mathbf{X}'\mathbf{W}\mathbf{X})^{-1}\mathbf{X}'\mathbf{W}\langle f(\mathbf{X}\boldsymbol{\beta})\rangle^{-1} - \mathbf{I}]\,\mathbf{H}(\mathbf{x}).$$

The variance expression for the parametric prediction is given by

$$\text{var}(\hat{\mathbf{P}}^{\mathrm{P}}) \approx \langle f(\mathbf{X}\boldsymbol{\beta})\rangle \mathbf{X}\mathbf{V}_{\hat{\boldsymbol{\beta}}}\mathbf{X}'\langle f(\mathbf{X}\boldsymbol{\beta})\rangle \tag{7}$$

where $V_{\hat{\beta}}$ *is an expression for the variance of the coefficient vector* $\boldsymbol{\beta}$ written as

$$\text{var}(\hat{\boldsymbol{\beta}}) = \mathbf{V}_{\hat{\boldsymbol{\beta}}} \approx (\mathbf{X}'\mathbf{W}\mathbf{X})^{-1}\mathbf{X}'\mathbf{W}\left\langle\frac{\mathbf{G}}{\mathbf{F}}\right\rangle\mathbf{X}(\mathbf{X}'\mathbf{W}\mathbf{X})^{-1} \tag{8}$$

where

$$\left\langle\frac{\mathbf{G}}{\mathbf{F}}\right\rangle = \left\langle\frac{G(x_1)(1-G(x_1))}{F(\mathbf{x}_1'\boldsymbol{\beta})(1-F(\mathbf{x}_1'\boldsymbol{\beta}))}\cdots\frac{G(x_d)(1-G(x_d))}{F(\mathbf{x}_d'\boldsymbol{\beta})(1-F(\mathbf{x}_d'\boldsymbol{\beta}))}\right\rangle$$

is a diagonal matrix. A noteworthy observation is that if the user's model is correct, that is, if $\mathbf{F} = \mathbf{G}$, then $\mathbf{H}(\mathbf{x}) = \mathbf{0}$ and $\langle(\mathbf{G}/\mathbf{F})\rangle = \mathbf{I}$, which yields zero bias for the parametric procedure; expression (8) reduces to $\text{var}(\hat{\boldsymbol{\beta}}) = (\mathbf{X}'\mathbf{W}\mathbf{X})^{-1}$, a well-known result, and the variance expression of (7) is equivalent to the mean squared error expression for the parametric procedure.

For the non-parametric procedure, the mean squared error properties will be developed for the model $\mathbf{P} = \mathbf{G}(\mathbf{x})$ where $\mathbf{G}$ is any arbitrary CDF. The bias for the non-parametric prediction, based on a fixed bandwidth, is given by $\text{bias}(\hat{\mathbf{P}}^{\mathrm{NP}}) = (\mathbf{H}^{\mathrm{NP}} - \mathbf{I})\,\mathbf{P}$. Similarly, the variance can be expressed as $\text{var}(\hat{\mathbf{P}}^{\mathrm{NP}}) = \mathbf{H}^{\mathrm{NP}}\mathbf{V}_G\mathbf{H}^{\mathrm{NP}}$ where

$$V_{\mathrm{G}} = \text{var}(\mathbf{p}) = \left\langle\frac{G(x_1)(1-G(x_1))}{n_1}\cdots\frac{G(x_d)(1-G(x_d))}{n_d}\right\rangle.$$

For the model-robust quantal regression procedure, the vector of estimated probabilities, $\hat{\mathbf{P}}^{\mathrm{MRQR}}$, is written as

$$\hat{\mathbf{P}}^{\mathrm{MRQR}} = \lambda\,\hat{\mathbf{P}}^{\mathrm{NP}} + (1-\lambda)\,\hat{\mathbf{P}}^{\mathrm{P}}.$$

The variance expression is approximated by

$$\text{var}(\hat{\mathbf{P}}^{\mathrm{MRQR}}) = \text{var}[\lambda\hat{\mathbf{P}}^{\mathrm{NP}} + (1-\lambda)\,\hat{\mathbf{P}}^{\mathrm{P}}] \approx [\lambda\mathbf{H}^{\mathrm{NP}} + (1-\lambda)\,\mathbf{B}]\,\mathbf{V}_G\,[\lambda\mathbf{H}^{\mathrm{NP}} + (1-\lambda)\,\mathbf{B}]'$$

where $\mathbf{B} = \langle f(\mathbf{X}\boldsymbol{\beta})\rangle\mathbf{X}(\mathbf{X}'\mathbf{W}\mathbf{X})^{-1}\mathbf{X}'\mathbf{W}\langle f(\mathbf{X}\boldsymbol{\beta})\rangle^{-1}$. The bias for the MRQR procedure can be expressed as

$$\text{bias}(\hat{\mathbf{P}}^{\mathrm{MRQR}}) \approx \lambda[\mathbf{H}^{\mathrm{NP}} - \mathbf{I}]\,\mathbf{P} + (1-\lambda)\,[\langle\mathbf{f}(\mathbf{X}\boldsymbol{\beta})\rangle\mathbf{X}(\mathbf{X}'\mathbf{W}\mathbf{X})^{-1}\mathbf{X}'\mathbf{W}\langle\mathbf{f}(\mathbf{X}\boldsymbol{\beta})\rangle^{-1} - \mathbf{I}]\mathbf{H}(\mathbf{x})$$

which turns out to be a linear combination of the biases of the parametric and non-parametric procedures. Note that under the correctly specified parametric model $\mathbf{H}(\mathbf{x}) = \mathbf{0}$ and if $\lambda \approx 0$, as it should, then the mean squared error is comprised of primarily the parametric variance. On the other hand, if the user's postulated model has been grossly misspecified, $\mathbf{H}(\mathbf{x}) \neq \mathbf{0}$ and if $\lambda \approx 1$ then the mean squared error will be made up of mainly the non-parametric fit to the data. Adding the squared bias to the variance gives the desired mean-squared error (MSE) formulae.

5. NUMERICAL COMPARISONS

The three procedures, logistic regression, LLR, and MRQR with LLR as the non-parametric portion, will be evaluated by comparing the mean squared error (MSE) of fit at any dose value or the average mean squared error across several dose values. As is readily seen from the bias and variance formulae above, the mean squared errors are dependent on several factors including the model matrix, $\mathbf{X}$, which in turn reflects the number of doses, d, the dose values, $x_1, \ldots, x_d$, and the sample size at each doses, $n_1, \ldots, n_d$, the true CDF, $\mathbf{G}(\mathbf{x})$, and the CDF specified by choice of the parametric method, $\mathbf{F}(\mathbf{x})$. These parameters must be established in order to evaluate at MSE properties.

Additionally, the LLR procedure and MRQR require values for the bandwidth, and, for MRQR, the value of λ, the mixing parameter. Given the properties derived in the previous section, the optimal values of the bandwidth (b) and mixing parameter (λ), denoted by b_0 and λ_0, respectively, for the non-parametric and model-robust procedures, were computed by minimizing the average mean squared error across the design points. In evaluating the MSE formulae, the bandwidth and mixing parameters are fixed at their optimal values, following the approach taken by Speckman.[25]

To enable $\mathbf{G}(\mathbf{x})$ to be expressed in the form (6) we need $F(\mathbf{X}\boldsymbol{\beta})$. Assuming that $\mathbf{G}$ is the true cumulative distribution function, the true responses, P, can be obtained at each of the dose levels. Having the true responses at the d dose levels, the coefficient vector, $\boldsymbol{\beta}$, is now obtained via the method of maximum likelihood. Using this value of $\boldsymbol{\beta}$, $\mathbf{G}(\mathbf{x})$ can be partitioned into the components $\mathbf{F}(\mathbf{X}\boldsymbol{\beta})$ and $\mathbf{H}(\mathbf{x})$. For convenience, the domain x, of the function $\mathbf{G}$ will be scaled to be between zero and one. The dose levels (or design points) will be evenly spaced to support the entire curve.

The true CDF used in our evaluation of the procedures was chosen as $G(x) = (1 - \gamma)L(x; 0{\cdot}5, 0{\cdot}1) + \gamma[\delta L(x; 0{\cdot}25, 0{\cdot}05) + (1 - \delta)L(x; 0{\cdot}75, 0{\cdot}05)]$, which, depending on the value of γ, is a mixture of logistic cumulative distribution functions. The notation $L(x; \mu, \tau)$ represents the logistic CDF written as $L(x; \mu, \tau) = \{1 + \exp(-((x - \mu)/\tau))\}^{-1}$, with μ and τ being the location and scale parameters of the logistic distribution, respectively. The value of δ will be 0·5 throughout this paper. With the value of $\delta = 0{\cdot}5$, $G(x)$ is a mixture of a logistic CDF and a symmetric bimodal CDF. The value of γ, which will denote the degree of model misspecification, will range from zero to one. As the value of γ increases to one, the degree of model-misspecification increases.

Finney[1] suggests that the number of doses (or design points) be equally spaced and that the same number of subjects (or replicates) be assigned to each dose. Thus, along with the range of values indicated for γ, there will also be three dose values used ($d = 3$, 5 and 7) and three sets of replicates at each dose level ($n = 10$, 20 and 50). The dose levels and replicates at each dose level were chosen to represent typical quantal bioassay situations found in practice.[1,2]

Table II. Theoretical integrated MSE efficiencies with respect to the logistic regression procedure for $d = 5$ doses using the optimal bandwidth and mixing parameters. Bold values indicate simulated MSE efficiencies using bandwidth and mixing parameters selected using PRESS*

γ	$n = 10$		$n = 20$		$n = 50$	
	LLR	MRQR	LLR	MRQR	LLR	MRQR
0·0	0·8955	1·1111	0·7317	1·0000	0·5714	1·0000
	0·8154	**1·1042**	**0·8049**	**1·0000**	**0·5000**	**1·0000**
0·1	0·9677	1·1539	0·7500	1·0714	0·5714	1·0000
	0·9672	**1·1800**	**0·7895**	**1·1111**	**0·6500**	**1·0000**
0·2	1·1273	1·2653	0·8889	1·1852	0·7000	1·1667
	1·1636	**1·3333**	**0·8889**	**1·0667**	**0·6191**	**1·0769**
0·3	1·3542	1·4444	1·1290	1·4000	0·9444	1·4167
	1·4444	**1·5476**	**1·2000**	**1·4400**	**0·9444**	**1·4167**
0·4	1·6429	1·6429	1·5000	1·6957	1·4000	1·7500
	1·7727	**1·7727**	**1·6800**	**1·8261**	**1·4000**	**1·9091**
0·5	1·8750	1·8750	1·9565	2·0455	2·2500	2·4546
	1·9756	**1·9286**	**2·0000**	**2·0000**	**2·3333**	**2·5455**
0·7	2·0000	2·0000	2·2857	2·2857	2·7059	2·7059
	2·1064	**2·1064**	**2·3103**	**2·3103**	**2·7059**	**2·7059**
1·0	1·7284	1·7284	2·0556	2·0556	2·8182	2·8182
	1·4554	**1·4554**	**2·1346**	**2·1346**	**2·7647**	**2·7647**

The measure by which the procedures will be compared is the approximate integrated MSE (IMSE) statistic computed over the range of the data for the entire curve over 100 evenly spaced doses. The IMSE for each procedure represents that procedure's ability to fit the curve $\mathbf{G}(\mathbf{x})$ by taking into account both average bias and average variance of the fitted responses over the entire curve. Naturally, the smaller the IMSE the better a procedure is able to fit $\mathbf{G}(\mathbf{x})$.

To compare procedures across all model parameters, the integrated MSE efficiency (IMSE efficiency) for any non-parametric or model-robust procedure ('other') with respect to the logistic model is computed as IMSE − efficiency = IMSE(logistic)/IMSE(other)

Although three values of d were examined (3, 5 and 7) in our study, due to the similarity of results and for the sake of brevity, only the $d = 5$ results will be presented here. For a complete accounting of all results, see Nottingham.[18] Table II contains the IMSE efficiencies for each of the ($d = 5$, $n = 10$, 20, 50) combinations for the values of γ given above. The important observations are:

1. the LLR procedure has efficiencies less than one for small values of γ, where there is slight misspecification, and efficiencies greater than one for larger values of γ, representing greater misspecification;
2. the MRQR procedure always has efficiencies greater than or equal to the LLR procedure, and always has a mean squared error smaller than the logistic method, except when the model is correct, where the efficiency is one.

We note that the formulae used to compute the IMSEs rely on Taylor series expansions and are asymptotic in nature and their appropriateness for finite sample size problems must be considered. This issue will be addressed in the next section. It is apparent from these results that MRQR performs as anticipated, improving the fits over those obtained by either the parametric or non-parametric method. These results, based on optimal fits, show the *potential* theoretical benefits of the MRQR method.

6. SIMULATION RESULTS

For the first simulations obtained here, the LLR and MRQR methods use a fixed bandwidth and mixing parameter set equal to the optimal values utilized to obtain the efficiencies in Table II. This simulation is designed, in part, to investigate the appropriateness and accuracy of the theoretical MSE formulaes. Additionally, the procedures were also evaluated in their ability to estimate other aspects typically desired in a quantal bioassay such as estimation of the probability of response at a given dose and confidence intervals for these estimated responses. Estimates of effective doses are also of interest along with confidence intervals on these estimates. Only the $d = 5$, $n = 20$, $\gamma = 0$ and 0·5 results are shown here as they are representative of the results obtained for all other combinations of d, n and γ. Results based on data driven values of b and λ will be discussed in Section 7.

The mixed logistic model was simulated using SAS IML for each n (10, 20 and 50), d (3, 5 and 7) and γ (0, 0·1, 0·2, 0·3, 0·4, 0·5, 0·7 and 1·0) combination. To judge the accuracy of the theoretical MSE formulae, the fitted response curves obtained for each of the three methods were compared to the true curve, $G(x)$, for each combination of parameters and each Monte Carlo repetition at 100 equally spaced doses (the support for $G(x)$, is rescaled to be between 0 and 1) by computing the sum of squared error (SSE) of fit given by $\mathrm{SSE} = \sum_{i=1}^{100} (P_i - \hat{P}_i)^2/100$. This process is repeated 500 times and the Monte Carlo average squared error (ASE) is obtained as the average SSE over the 500 Monte Carlo repetitions. We view these ASE values as representing a measure of the ability of a procedure to fit the true curve and the number of which the theoretical (and asymptotic) IMSES of Sections 4 and 5 are attempting to represent. Though these ASEs values will not be reproduced here, we learned through examination of them that the MSE formulae of Section 5 are reasonable approximations to the true ASEs for d as small as 3. In fact, if $d \geqslant 5$ and $n \geqslant 10$ the ASEs, obtained through simulation, and the IMSEs, obtained through the formulae of Section 4, are in close agreement. At $d = 5$ and $n = 10$ or 20, the relative error between actual and approximated results is usually less than 10 per cent, while at $d = 5$, $n = 50$, the maximum observed relative error was less than 8 per cent, with most values agreeing to four decimal places. The relative errors are even smaller at $d = 7$. Thus, it is clear that for d as small as 5 and n as small as 10 the asymptotic formulae for the MSE from Section 4 provide accurate representations of the true IMSEs.

Table III presents the effective dose estimation results, again for the $d = 5$, $n = 20$, $\gamma = 0$, representing no model misspecification, and $\gamma = 0{\cdot}5$, representing moderate model misspecification, respectively. The inverse estimation technique was used to obtain estimates of the ED_α for α equal to 0·2, 0·5 and 0·8 and corresponding 95 per cent confidence intervals for the non-parametric and the MRQR methods. Fieller's theorem was used to obtain 95 per cent confidence (fiducial[1]) intervals for the logistic method. The true ED_αs are also presented in the tables for comparison. At $\gamma = 0$ the logistic method gives superior results over the LLR procedure as expected, with the most accurate estimates of effective doses and the closest observed coverage

Table III. Effective dose estimation summary for the ($d = 5$, $n = 20$, $\gamma = 0$ and 0·5) combination using optimal values of bandwidth and mixing parameter

γ	Method	α for $ED_{100\alpha}$	True $ED_{100\alpha}$	Mean $ED_{100\alpha}$	Mean width	Observed coverage probability (%)
0	MLE	0·2	0·3615	0·366	0·161	90·8
		0·5	0·5	0·502	0·118	92·8
		0·8	0·6385	0·638	0·161	91·6
	LLR	0·2	0·3615	0·328	0·158	88·4
		0·5	0·5	0·504	0·141	93·4
		0·8	0·6385	0·676	0·149	84·0
	MRQR	0·2	0·3615	0·366	0·161	90·8
		0·5	0·5	0·502	0·118	92·8
		0·8	0·6385	0·638	0·161	91·6
0·5	MLE	0·2	0·2730	0·3229	0·2072	88·8
		0·5	0·5	0·5051	0·1450	92·8
		0·8	0·7270	0·6873	0·2083	90·8
	LLR	0·2	0·2730	0·2601	0·1301	93·2
		0·5	0·5	0·5042	0·1344	94·2
		0·8	0·7270	0·7479	0·1340	91·4
	MRQR	0·2	0·2730	0·2663	0·1359	93·8
		0·5	0·5	0·5044	0·1358	94·2
		0·8	0·7270	0·7420	0·1300	92·2

proportions to the nominal 95 per cent. Note that the MRQR method results are identical because the optimal value of the mixing parameter was $\lambda = 0$. The coverage proportions for all of the effective doses are near the nominal 95 per cent, with the ED_{50} having the highest coverage at 92.8 per cent. At $\gamma = 0{\cdot}5$, the MRQR results are far superior to those of the logistic and slightly better than the LLR method. The coverage proportions are still low for $\alpha = 0{\cdot}2$ and 0·8 for the logistic procedure, but much higher for the MRQR method.

Other aspects of the analysis of quantal dose–response data of interest are the estimation of the probability of response at an arbitrary dose and the properties of the precision of this estimate. Table IV presents the simulated mean response based on 500 repetitions at the $d = 5$, $n = 20$ situation for $\gamma = 0$ and 0·5, respectively. The true probability of response at each dose is given in the tables as well. Also included are the average widths of the 95 per cent confidence intervals for probability of response at each of the dose values along with the proportion of the 500 intervals that contained the true probability of response, the observed coverage.The confidence interval for the logistic method was obtained by straightforward application of maximum likelihood principles as $F(\mathbf{x}_0'\hat{\boldsymbol{\beta}}) \pm 1{\cdot}96\,\mathrm{SE}(F(\mathbf{x}_0'\hat{\boldsymbol{\beta}}))$, where $\mathrm{SE}(F(\mathbf{x}_0'\hat{\boldsymbol{\beta}})) = \sqrt{\{f^2(\mathbf{x}_0'\hat{\boldsymbol{\beta}})\,\mathbf{x}_0'(\mathbf{X}'\mathbf{W}\mathbf{X})^{-1}\,\mathbf{x}_0\}}$ with $\mathbf{x}_0' = (1 \;\; \mathbf{x}_0)$.

The key comparisons are those between the logistic and MRQR methods. At $\gamma = 0$, the model is correctly specified so that one would expect the logistic method to give the most accurate estimates of the probability of response. The Monte Carlo coverage proportions are equal for both of these procedures with nearly equal average 95 per cent confidence interval widths. It is

Table IV. Summary of mean response at the ($d = 5$, $n = 20$, $\gamma = 0$ and 0·5) combination for 500 Monte Carlo simulations and the optimal values of the bandwidth and mixing parameter

γ	Method	Dose	Mean response	True response	Mean width	Observed coverage probability (%)
0	MLE	0·1	0·019	0·018	0·065	86·0
		0·3	0·116	0·119	0·227	89·2
		0·5	0·495	0·500	0·372	95·0
		0.7	0.881	0.881	0.232	91.2
		0·9	0·980	0·982	0·068	86·6
	LLR	0·1	0·020	0·018	0·093	100·0
		0·3	0·164	0·119	0·195	89·4
		0·5	0·492	0·500	0·296	93·4
		0·7	0·833	0·881	0·194	91·6
		0·9	0·981	0·982	0·091	100·0
	MRQR	0·1	0·019	0·018	0·065	86·0
		0·3	0·116	0·119	0·227	89·2
		0·5	0·495	0·500	0·372	95·0
		0·7	0·881	0·881	0·232	91·2
		0·9	0·980	0·982	0·068	86·6
0·5	MLE	0·1	0·047	0·021	0·140	98·6
		0·3	0·175	0·242	0·287	75·6
		0·5	0·490	0·500	0·352	95·2
		0·7	0·814	0·758	0·294	81·4
		0·9	0·950	0·979	0·146	98·4
	LLR	0·1	0·020	0·021	0·149	99·6
		0·3	0·249	0·242	0·157	94·4
		0·5	0·495	0·500	0·164	94·2
		0·7	0·741	0·758	0·162	93·2
		0·9	0·974	0·979	0·157	99·2
	MRQR	0·1	0·023	0·021	0·142	99.6
		0·3	0·241	0·242	0·162	92·4
		0·5	0·494	0·500	0·176	94·2
		0·7	0·749	0·758	0·167	93·2
		0·9	0·972	0·979	0·150	99·2

interesting to note that the 95 per cent nominal coverage probability is met only at the middle dose for the methods, and all methods, including the logistic procedure, have coverage proportions less than the nominal value for all other doses.

At $\gamma = 0{\cdot}5$, with moderate model misspecification, one would expect that the logistic method would have problems. The logistic method rather severely underestimates the probability of response at $x = 0{\cdot}3$ and overestimates the probability at $x = 0{\cdot}7$. This represents the bias problem discussed in Section 2. The MRQR method, is remarkably accurate in estimating the probability of response throughout the entire dose range. The coverage proportions for the MRQR method are very close to the nominal values at the middle range of doses with average widths far smaller than those obtained by the logistic method.

7. DISCUSSION AND SUMMARY

The MRQR procedure shows promise as a method that can be used alongside the parametric analysis of quantal dose–response data and as a tool for curve fitting and effective dose estimation when the user's logistic model is inadequate. The theoretical results indicate that the model-robust procedure improves the integrated mean squared error with respect to the parametric (logistic) and non-parametric (LLR or kernel) methods.

Based on our theoretical and simulation results, where only a small proportion of our extensive simulation results has been presented here, the model-robust procedure offers improved analysis of quantal dose–response data when the logistic model has been misspecified. Summarizing the theoretical and simulation results, the following is a list of the more pertinent features of MRQR:

1. With a small number of doses, such as $d = 3$, the MRQR procedure yields as small or smaller average mean squared error than the other procedures, with the exception of the logistic method when $\gamma = 0$, for $n = 10$, 20 and 50 replicates at each dose level when using the optimal bandwidth and mixing parameter.
2. Using the optimal bandwidth and mixing parameters, the MRQR procedure had higher mean squared error efficiencies than the other procedures for the ($d = 5$, 7; $n = 10$, 20, 50) combinations across almost all values of γ. This implies that with the optimal values of the bandwidth and mixing parameter, the MRQR procedure performs better than the logistic and LLR procedures.
3. Using the optimal values of the bandwidth, MRQR mixes appropriately. That is, as the degree of model-misspecification increases, the value of the optimal mixing parameter increases as well, giving more weight to the non-parametric procedure.
4. Based on 500 Monte Carlo repetitions, the asymptotic theoretical properties of the bias, variance and mean squared error formulae for the non-parametric and model-robust procedures are valid, especially when $d \geqslant 5$.
5. Although the estimated median effective dose was approximately the same for all the procedures presented, as the degree of model-misspecification increased, the model-robust procedure estimated the extreme doses more accurately, and with coverage probabilities closer to the nominal 95 per cent when using the optimal bandwidth and mixing parameter.
6. Using the optimal bandwidth and mixing parameter for the non-parametric and model-robust procedures, respectively, the true response, P, is also more accurately estimated by MRQR.

It is evident from the above summary that the model-robust procedure applied to quantal dose–response data is capable of improving the fit obtained by the either the parametric and non-parametric methods with the proper values of the bandwidth and mixing parameter. Although there are several areas for future research with respect to the model-robust procedures, the primary area is that of bandwidth and mixing parameter selection. Many techniques are available for bandwidth selection in non-parametric regression.[19] The cross-validation procedure and several versions of penalized cross-validation procedures have been applied to model-robust quantal regression with promising results. Current work by Mays and Birch[29] and Ruppert *et al.*[30] are also appealing and are currently being evaluated by the authors.

In particular, a version of a generalized cross-validation procedure, termed PRESS*, has been studied extensively for selecting the bandwidth in LLR. Here, the bandwidth b is found by

minimizing

$$\text{PRESS*} = \frac{\sum_{i=1}^{d} w_i (p_i - \hat{P}_{i,-i}^{\text{LLR}}(b))^2}{d - \text{tr}(\mathbf{H}^{\text{LLR}})}$$

where $w_i = n_i / p_i(1 - p_i)$, serving as the weight, is the reciprocal of the estimated variance of response at each dose, $\hat{P}_{i,-i}^{\text{LLR}}(b)$ is the 'minus-i' predicted proportion of subjects responding to the ith dose x_i for the current value of b with the ith observation removed, and $\text{tr}(\mathbf{H}^{\text{LLR}})$ is the trace of the $d \times d$ LLR weight matrix $\mathbf{H}^{\text{LLR}}$. Since $\text{tr}(\mathbf{H}^{\text{LLR}})$ reflects the LLR fits' 'model degrees of freedom',[9] it is seen that the denominator of PRESS* penalizes the weighted PRESS statistic (the numerator of PRESS*) for choosing b too small.

Empirical studies[29] show the PRESS* can be superior to other forms of penalized cross-validation criteria. The mixing parameter for MRQR may be chosen in a similar manner. As can be expected, use of b and λ chosen by PRESS* results in some loss of MSE efficiency when compared to using the optimal values. However, the general features stated above as 1 to 6 still hold. Justification of this statement is readily apparent from Table II where the IMSE efficiencies given in bold result from the average IMSE over 500 Monte Carlo runs where b and λ are chosen by the PRESS* method for each run. We note that the MRQR method is never less efficient that logistic regression, even when the model is correct and can be much more efficient when the model is misspecified. MRQR is also never less efficient than LLR, even when the model is badly misspecified and can be much more efficient when the model is correct (see Nottingham and Birch[28] for more details).

In the Martin example, the bandwidth was chosen as 0·3 and the mixing parameter as 0·95, as determined by PRESS*. The large value of the mixing parameter implies that majority of the MRQR fit (approximately 95 per cent) be composed of the non-parametric fit (LLR). This seems appropriate since the logistic regression model was so poor, resulting in a highly significant goodness-of-fit statistic. Presumably, using the MRQR fit composed of 95 per cent from the LLR fit greatly improves the bias in fits caused by the inadequate logistic model, while the 5 per cent from the logistic model reduces the variance off its that would be obtained had we used solely LLR to determine the fits. The authors have prepared a macro in SAS IML to perform the analysis demonstrated by this example. It is available from the authors upon request.

Work in non-parametric regression by Hastie and Loader,[31] Fan[12] and Cleveland and Devlin[32] have found that the local linear and local quadratic fits are most useful. Therefore, another possibility for the non-parametric procedure is to use local quadratic regression in conjunction with a parametric method. The local quadratic regression method may provide a better fit in the tails of the data.

In addition, it seems feasible to fit a local logistic regression as well. This would entail fitting a logistic regression at each dose, just as in local linear regression, where a linear regression model is used at each dose. This can also be generalized so that any CDF can be used to fit the dose–response curve locally. Thus, instead of using logistic regression as the parametric function in MRQR, one can also use any parametric model, such as the Aranda–Ordaz, a finite mixture logistic, or any generalized linear model using error distributions such as the Cauchy, Gompertz and Weibull, to name a few. Using any generalized linear model would make the model-robust procedure more applicable across a variety of fields. This issue is currently being addressed by the authors. Other examples of current semi-parametric regression research include Severini and

Staniswalis[26] who use quasi-likelihood methods to fit semi-parametric models and Fan *et al.*[33] who use local polynomial regression for generalized linear models.

An additional area for future research includes quantifying λ. It should be possible to construct a 'goodness-of-fit' test of the user's specified parametric model by using a test statistic based on λ, extending the work of Rahman *et al.*[6] and others[8–11] from the measurement variable case to the quantal regression setting.

REFERENCES

1. Finney, D. J. *Statistical Method in Biological Assay*, 3rd edn, Oxford University Press, New York, 1978.
2. Morgan, B. J. T. *Analysis of Quantal Response Data*, Chapman and Hall, London, 1992.
3. Govindarajula, Z. *Statistical Techniques in Bioassay*, Karger, Basel, Switzerland, 1988.
4. Goedhart, P. 'Statistical analysis of quantal-assay data', MSc dissertation, University of Kent, Canterbury, Kent, U.K. 1985.
5. Cox, C. 'Fieller's Theorem, the likelihood and the delta method', *Biometrics*, **46**, 709–718 (1990).
6. Rahman, M., Gokhale, D. V. and Ullah, A. 'A note on combining parametric and nonparametric regression', *Communications in Statistics – Simulation and Computation*, **26**, 519–529 (1997).
7. Burman, P. and Chaudhuri, P. 'A hybrid approach to parametric and nonparametric regression', Technical Report No. 243, Division of Statistics, University of California Davis, 1992.
8. Wooldridge, M. 'A test for functional form against nonparametric alternatives', *Econometric Theory*, **8**, 452–475 (1992).
9. Robinson, P. 'Root-N-consistent semiparametric regression', *Econometrica*, **56**, 931–954 (1988).
10. Ullah, A. and Vinod, H. D. 'General nonparametric regression estimation and testing in econometrics', in Maddala, G. S., Rao, C. R. and Vinod, H. D. (eds), *Handbook of Statistics*, *Vol.* 11, Elsevier Science Publishers, New York, 1993.
11. Cleveland, W. S. 'Robust locally weighted regression and smoothing scatterplots', *Journal of the American Statistical Association*, **74**, 368, 829–836 (1979).
12. Fan, J. 'Local linear regression smoothers and their minimax efficiency', *Annals of Statistics*, **21**, 196–216 (1993).
13. Fan, J. and Gijbels, I. 'Variable bandwidth and local linear regression smoothers', *Annals of Statistics*, **20**, 2008–2036 (1992).
14. Staniswalis, J. G. and Cooper, V. 'Kernel estimates of dose response', *Biometrics*, **44**, 4, 1103–1120 (1988).
15. Müller, H. -G. and Schmitt, T. 'Kernel and probit estimates in quantal bioassay', *Journal of the American Statistical Association*, **83**, 403, 750–759 (1988).
16. Kappenman, R. F. 'Nonparametric estimation of dose-response curves with applications to ED_{50} estimation', *Journal of Statistical Computation and Simulation*, **28**, 1–13 (1987).
17. Kelley, C. and Rice, J. 'Monotone smoothing with applications to dose-response curves and the assessment of synergism', *Biometrics*, **46**, 1071–1085 (1990).
18. Nottingham, Q. J. 'Model-robust quantal regression', PhD dissertation, Virginia Polytechnic Institute and State University, Blacksburg, Virginia, 1995.
19. Fan, J. and Gijbels, I. *Local Polynomial Modelling and Its Applications*, Chapman and Hall, London, 1996.
20. Martin, J. T. 'The problem of the evaluation of rotenone-containing plants. VI. The toxicity of *l*-elliptone and of poisons applied jointly, with further observations on the rotenone equivalent method of assessing the toxicity of derris root', *Annals of Applied Biology*, **29**, 69–81 (1942).
21. Einsporn, R. 'A link between least squares regression and nonparametric curve estimation', PhD dissertation, Virginia Polytechnic Institute and State University, Blacksburg, Virginia, 1987.
22. Einsporn, R. L. and Birch, J. B. 'Model-robust regression: Using nonparametric regression to improve parametric regression analyses', Technical Report Number 93-5, Virginia Polytechnic Institute and State University, Blacksburg, Virginia, 1993.
23. Mays, J. and Birch, J. 'Model-robust regression combining parametric, nonparametric, and semiparametric methods', Technical Report Number 98-2, Virginia Polytechnic Institute and State University, Blacksburg, Virginia, 1996.

24. Olkin, I. and Spiegelman, C. H. 'A semiparametric approach to density estimation', *Journal of the American Statistical Association*, **82**, 399, 858–865 (1987).
25. Speckman, P. 'Kernel smoothing in partial linear models', *Journal of the Royal Statistics Society*, *Series B*, **50**, 3, 413–436 (1988).
26. Severini, T. A. and Staniswalis, J. G. 'Quasi-likelihood estimation in semiparametric models', *Journal of the American Statistical Association*, **89**, 426, 501–511 (1994).
27. Hastie, T. and Tibshirani, R. 'Generalized additive models', *Statistical Science*, **1**, 3, 297–318 (1987).
28. Nottingham, Q. J. and Birch, J. B. 'Model-robust quantal regression', Technical Report Number 96-6, Virginia Polytechnic Institute and State University, Blacksburg, Virginia, 1996.
29. Mays, J. and Birch, J. B. 'Smoothing considerations in nonparametric and semiparametric regresson', Technical Report Number 96-7, Virginia Polytechnic Institute and State University, Blacksburg, Virginia, 1998.
30. Ruppert, D., Sheather, S. J. and Ward, M. P. 'An effective bandwidth selector for local linear regression', *Journal of the American Statistical Association*, **90**, 432, 1257–1270 (1995).
31. Hastie, T. and Loader, C. 'Local regression: automatic kernel carpentry', *Statistical Science*, **8**, 2, 120–143 (1993).
32. Cleveland, W. S. and Devlin, S. J. 'Locally weighted regression: An approach to regression analysis by local fitting', *Journal of the American Statistical Association*, **83**, 403, 596–610 (1988).
33. Fan, J., Heckman, N. E. and Wand, M. P. 'Local polynomial kernel regression for generalized linear models and quasi-likelihood functions', *Journal of the American Statistical Association*, **90**, 429, 141–150 (1995).

