STATISTICS IN MEDICINE
Statist. Med. 2001; **20**:2935–2949 (DOI: 10.1002/sim.889)

Graphical model checking with correlated response data

Wei Pan[1,*,†], John E. Connett[1], Giovanni C. Porzio[2] and Sanford Weisberg[3]

[1]*Division of Biostatistics, School of Public Health, University of Minnesota, A460 Mayo Building, Box 303, 420 Delaware Street SE, Minneapolis, MN 55455-0378, U.S.A.*
[2]*Dipartimento di Economia e Territorio, University of Cassino, Via M. Mazzaroppi, Cassino, I 03043, Italy.*
[3]*School of Statistics, University of Minnesota, 353f Classroom-Office Building, 1994 Buford Avenue, St. Paul, MN 55108, U.S.A.*

SUMMARY

Correlated response data arise often in biomedical studies. The generalized estimation equation (GEE) approach is widely used in regression analysis for such data. However, there are few methods available to check the adequacy of regression models in GEE. In this paper, a graphical method is proposed based on Cook and Weisberg's marginal model plot. A bootstrap method is applied to obtain the reference band to assess statistical uncertainties in comparing two marginal mean functions. We also propose using the generalized additive model (GAM) in a similar fashion. The proposed two methods are easy to implement by taking advantage of existing smoothing and GAM softwares for independent data. The usefulness of the methodology is demonstrated through application to a correlated binary data set drawn from a clinical trial, the Lung Health Study. Copyright © 2001 John Wiley & Sons, Ltd.

1. INTRODUCTION

Model checking is an important component of data analysis. In general the validity of any statistical inference depends on the adequacy of the model being used. For independent data, there is a huge literature on this issue. There are two general approaches. One is to use formal goodness-of-fit tests. The other is less formal through graphics, which we consider in this paper. One commonly used graphical model checking technique is residual analysis (see, for example, references [1] and [2]). However, residual analysis is much less used in examining correlated response data, largely due to the existing within-subject correlation among residuals. The correlation structure is often difficult to model and is often treated as nuisance parameters in GEE [3]. In addition, for some generalized linear models (GLMs) (for example, reference [4]), such as logistic regression for (ungrouped) binary data, the usual Pearson or deviance residuals do not have an asymptotically normal distribution. Recently,

*Correspondence to: Wei Pan, Division of Biostatistics, School of Public Health, University of Minnesota, A460 Mayo Building, MMC 303, 420 Delaware Street SE, Minneapolis, MN 55455-0378, U.S.A.
†E-mail: weip@biostat.umn.edu

Received April 2000
Accepted November 2000

Porzio and Weisberg [5] applied the *marginal model plot* proposed by Cook and Weisberg [6], along with the reference band suggested by Bowman and Young (in comparing two non-parametric curves) [7], to assess the adequacy of a regression model for independent data. This methodology proceeds by comparing two mean functions in a series of two-dimensional scatter plots. One mean function is obtained under the given model (that is, model-based); another is obtained non-parametrically (that is, model-free). If the model is (almost) correct, it is expected that the two (estimated) mean functions in any plot should be close to each other: they should be covered by a reference band, which is centred at the average of the two mean functions and has a width twice the standard error of the difference of the two mean functions. Hence if the two mean functions are out of the reference band, their difference is greater than twice the (estimated) standard error, and we have evidence that the two mean functions are not equal, which implies that the assumed model may not be sufficient to describe the data at hand. This is a simple and intuitively appealing approach in assessing the adequacy of a model. A primary goal of this paper is to extend the method to correlated response data.

In addition, we show that it is more powerful to combine the use of the marginal model plot with another technique, the generalized additive model (GAM). The GAM was first proposed by Hastie and Tibshirani (see reference [8] for a detailed account) for independent data. It has proved to be very useful, especially as an exploratory tool for model building. Wild and Yee [9] have proposed its extension to GEE. The use of this extension needs special software which, however, is not widely available in many commonly used statistical packages. Here we illustrate that, as for the marginal model plot, an extension of GAM to correlated data can be implemented simply by using existing GAM or non-parametric smoothing programs for independent data. The key idea is the validity of the working independence model, under which the correlated response data can be treated as if they were independent when one is estimating the mean parameters either in GEE [3] or in smoothing a scatter plot [10], thus enabling the use of the rich resource of existing computer programs for independent data. However, to correctly assess the variation of the mean parameters, some adjustment is necessary to take account of the correlation present in the data. We propose a simple approach using the bootstrap [11].

The paper is organized as follows. First, we review the marginal model plot for independent data. Then we describe its extension to correlated data. In particular, the reference band is obtained through a bootstrap-based method. Then we present the GAM method. As a case study, we illustrate the use of the marginal model plot with a reference band and of GAM by applying the methodology to a correlated binary data set drawn from a clinical trial. Some implementation and performance issues are further discussed using several simulated data sets with various sizes.

2. MARGINAL MODEL PLOTS

2.1. Marginal model plots with independent data

In regression analysis, we have a (univariate) response variable Y and a p-dimensional covariate (that is, predictor) vector X. The most general goal is to infer the conditional distribution $F(Y|X)$, though often the conditional mean $E_F(Y|X)$ is of primary interest. The parametric approach assumes that $F(Y|X)$ can be modelled as $M(Y|X,\theta)$ with an unknown parameter

vector θ. For instance, in linear regression, M is the linear model

$$Y|X = X'\beta + \sigma\varepsilon$$

where $\theta = (\beta, \sigma)$ and ε is a standard normal random variable. If the model M is correct, often we can obtain consistent estimate $\hat{\theta}$. Of course, a central issue in any modelling process is whether the assumed model is correct. This can be done via formal goodness-of-fit tests. In this paper we approach this issue by using a graphical method to compare the estimated model $M(Y|X, \hat{\theta})$ with the unknown $F(Y|X)$. In the following, for simplicity we write $M(Y|X, \hat{\theta})$ as $M(Y|X)$ or M by suppressing its dependence on $\hat{\theta}$.

If the covariate X is univariate, the conditional distribution $F(Y|X)$ can be summarized nicely by non-parametric methods (for example, reference [8] and references therein). For example, we can smooth the scatter plot of Y versus X to estimate the mean function $E(Y|X)$ and variance function $\text{var}(Y|X)$ [12]. However, in a regression setting, generally the covariate X has a high dimension p, where a direct application of a smoothing method is difficult due to the so-called 'curse of dimensionality'. Accordingly Cook and Weisberg proposed the marginal model plot as a dimension reduction technique to assess the adequacy of the model graphically. Their method rests on the following *marginal model-checking condition*. Let $F(Y|X)$ and $G(Y|X)$ be two conditional cumulative distribution functions. Then $F(Y|X) = G(Y|X)$ for any value X in the sample space Ω if and only if $F(Y|a'X) = G(Y|a'X)$ for all $a \in R^p$ and $X \in \Omega$. Hence, to compare $F(Y|X)$ and $M(Y|X)$ we can restrict our attention to the marginal models $F(Y|a'X)$ and $M(Y|a'X)$, where any $a'X$ is univariate. The first two words of 'marginal model plots' mean that such plots are drawn in a lower-dimensional space, typically two-dimensional. As we will see later, they are not directly related with 'marginal regression models' for correlated data, where it is emphasized that marginal means, not other conditional means (for example, in random-effects models), of the response are modelled in the regression equation.

In regression, the most important aspect of modelling is the agreement between the true (conditional) mean function $E_F(Y|X)$ and the $E_M(Y|X)$ induced from the model M. The true marginal mean function $E_F(Y|a'X)$ can be estimated by smoothing Y against $a'X$. There are many existing techniques to smooth a scatter plot of Y versus $a'X$ [8]. The marginal mean function $E_M(Y|a'X)$ induced from M can be obtained by averaging over the conditional distribution of X given $a'X$

$$E_M(Y|a'X) = E[E_M(Y|X)|a'X]$$

Hence, $E_M(Y|a'X)$ can be estimated by smoothing $E_M(Y|X)$ against $a'X$, where $E_M(Y|X)$ is estimated through the regression model M. Cook and Weisberg discussed how to obtain estimates of the marginal variance functions $\text{var}_F(Y|a'X)$ and $\text{var}_M(Y|a'X)$.

In each marginal plot of Y versus $a'X$, one can display the estimated marginal mean functions $E_F(Y|a'X)$ and $E_M(Y|a'X)$. Any discrepancy (after considering the variability) between the two marginal mean functions $E_F(Y|a'X)$ and $E_M(Y|a'X)$ demonstrates potential inadequacy of the model M.

Cook and Weisberg suggested several standard choices for plotting direction a. One is the linear predictor in the generalized linear model. Selecting $a'X$ to equal each predictor in turn is also very useful.

2.2. Marginal model plots with longitudinal data

Now we describe how to extend the marginal model plots to correlated data. For correlated data, the generalized estimating equation (GEE) approach is now widely used in regression analysis [3]. It is a natural extension of the generalized linear models to dependent response data. An attractive point of the GEE is that one does not need to model the correlation structure among responses; the correlation structure may be treated as a nuisance parameter. In particular, we can treat the dependent responses as if they were independent, and thus obtained GEE estimates of the regression coefficients are still consistent and asymptotically normal. This is the approach to be used later. Note that in some situations the correlation structure is of interest or is to be modelled to improve estimation efficiency [13], but for the purpose of model checking we can still use the working independence model.

Our goal is to assess the adequacy of the regression model $E_M(Y|X)$. This can be achieved using marginal model plots. We can compare the marginal mean functions $E_F(Y|a'X)$ and $E_M(Y|a'X)$ for some chosen direction $a'X$. Note that even though the responses Y's are correlated, we can still apply the usual smoothing methods (as if Y's were independent) to the scatter plot of Y versus $a'X$ [10]. Furthermore, according to the surprising result of Lin and Carroll [14], it is more efficient to smooth correlated data under the working independence assumption. However, it is much harder to estimate the marginal variance functions analytically. In this paper we propose to use the bootstrap [11]. First we need some notation.

To fix the idea, we give notation based on a longitudinal study, where each subject i, $i=1,\ldots,n$, is followed for a period of time and may have more than one measurement. Denote the n_i (continuous or discrete) response measurements arising from the same subject as $y_i=(y_{i1},\ldots,y_{in_i})'$ and $n_i\times p$ matrix of covariates for subject i by $x_i=(x_{i1},\ldots,x_{in_i})'$. This model permits, but does not require, time-varying covariates. As usual we assume that y_i and y_j are independent for $i\neq j$, while in general the components of any y_i are correlated. We assume that the marginal regression model M is in the form of a GLM: $E_M(y_{it}|x_{it})=g^{-1}(\beta' x_{it})$, where g is the given link function (which is the identity function in linear regression), $i=1,\ldots,n$ and $t=1,\ldots,n_i$. Applying the GEE method we obtain the estimate $\hat{\beta}$, and thus estimate of $E_M(y_{it}|x_{it})$, $\hat{E}_M(y_{it}|x_{it})=g^{-1}(\hat{\beta}' x_{it})$. Our goal is to assess the adequacy of the model M. Note that there are many other aspects of modelling, such as choosing the link function g, which can be considered. Here we restrict our discussion to the form of the linear predictor $\eta_{it}=\beta' x_{it}$; in other words, we are concerned with which covariates are to be included, and what are their functional forms. This can be also regarded as an aspect of variable selection.

Now we describe how to draw a marginal model plot for any given direction a. The estimates of the marginal mean functions $\hat{E}_F(Y|a'X)$ and $\hat{E}_M(Y|a'X)$ are obtained by smoothing a scatter plot of y_{it} versus $a'x_{it}$ and that of $\hat{E}_M(y_{it}|x_{it})$ versus $a'x_{it}$, respectively, where $i=1,\ldots,n$ and $t=1,\ldots,n_i$. This is essentially the same as for independent data. Note that here we use Y and X to represent the (marginal) random variables for y_{it} and x_{it}, respectively. To assess the variability for independent data, Porzio and Weisberg [5] proposed adding a reference band to the marginal model plot. The reference band was originally proposed by Bowman and Young [7]. In comparing two mean functions $\hat{E}_F(Y|a'X)$ and $\hat{E}_M(Y|a'X)$, it is helpful to draw a shaded area centred at their average $[\hat{E}_F(Y|a'X)+\hat{E}_M(Y|a'X)]/2$ with half-width $\mathrm{SE}(a'X)$, where $\mathrm{SE}(a'X)$ is the standard error function of $\hat{E}_F(Y|a'X)-\hat{E}_M(Y|a'X)$.

If the two mean functions $E_F(Y|a'X)$ and $E_M(Y|a'X)$ are equal, it is expected that with a high probability their estimates should lie within the shaded area, that is, their difference is within two standard errors. Bowman and Young [7] call this shaded area the reference band. If the response Y is approximately normal, the reference band corresponds to a pointwise test for no difference between the two mean functions at 5 per cent significance level. For independent continuous data and a linear smoother such as local polynomials or splines, Porzio and Weisberg [5] derived a closed form for SE($a'X$). With correlated continuous or discrete data, SE($a'X$) depends on the correlation structure of the response variable, which is difficult to model and estimate, especially when the number of measurements at a time point is in general small compared with the number of subjects, and thus is usually treated as a nuisance parameter in GEE. Therefore we propose to use the bootstrap to circumvent this issue. The resampling unit is taken as the subject (y_i, x_i), rather than the individual measurement (y_{it}, x_{it}), to conserve the within-subject correlation [10, 15].

Specifically, for each $b = 1, \ldots, B$, we draw n random pairs (y_i^b, x_i^b) with replacement from $\{(y_i, x_i)\}$ to form a resampled data set. Then for each resampled data set $\{(y_i^b, x_i^b)\}$, we can estimate the marginal mean functions $\hat{E}_F^b(Y|a'X)$ and $\hat{E}_M^b(Y|a'X)$ as above. Then SE($a'X$) is estimated as the square root of the sample variance of $\{(\hat{E}_F^b(Y|a'X) - \hat{E}_M^b(Y|a'X)\}$. We superimpose the reference band with the two estimated marginal mean functions in the same plot. If the model M can describe the data well, we expect that both $\hat{E}_F(Y|X)$ and $\hat{E}_M(Y|X)$ fall within the reference band; otherwise, there is an evidence of lack-of-fit (since their difference is larger than twice the standard error). A rough idea can be gained by looking at Figure 1, which will be discussed in detail later.

As one referee pointed out, the standard error function SE($a'X$) is not estimated under the null hypothesis that the model M fits well. It is not clear how to take advantage of being under the null hypothesis to estimate SE($a'X$). This may influence the performance of the proposed method. Hence, the proposed method is more in spirit like constructing a confidence interval, which is more general but may be less powerful in rejecting the null hypothesis. The bootstrap can be also implemented in other ways. An alternative to bootstrapping the pairs is to resample residuals, which we do not pursue here.

3. GENERALIZED ADDITIVE MODELS

To relax the modelling assumptions in the marginal model M, one can model the effect of each covariate x_{itj} non-parametrically, rather than parametrically. This leads to the GAM

$$E_M(y_{it}|x_{it}) = g^{-1}[\beta_0 + \sum_{j=1}^{k} f_j(x_{itj})]$$

where f_j's are some smooth functions to be estimated. The GLM can be regarded as a special case of the GAM in the sense that f_j is specified as linear: $f_j(x_{itj}) = \beta_j x_{itj}$. Of course, one can use the usual linear form for some f_j. Hastie and Tibshirani [8] have proposed efficient algorithms to fit the GAM for independent data, and relevant computer programs are readily available in some commonly used statistical packages.

With correlated response data, we can still use the usual computer programs (for independent data) to fit the GAM to estimate $\hat{f}_j$'s under the working independence model [10].

However, the standard error function $\mathrm{SE}(\hat{f}_j)$ of each $\hat{f}_j$ from such programs are not valid any more due to the nature of the correlated data. Again we propose to use the bootstrap as before. A valid estimate of $\mathrm{SE}(\hat{f}_j)$ can be obtained from the bootstrap estimates of f_j. Plotting $\hat{f}_j$ against x_{itj}, along with a shaded area centred at $\hat{f}_j$ with a half-width $2\mathrm{SE}(\hat{f}_j)$, one can visualize possible transformations suggested for the covariate x_{itj}. This is the widely used GAM plot. In this paper, as in many other applications, the GAM plot is used as an exploratory tool for model selection/checking.

4. A REAL EXAMPLE

For illustration we apply the methods to the data from the Lung Health Study (LHS) [16]. The LHS was a multi-centre, randomized controlled clinical trial. One of its goals was to determine whether an intervention program can help participants quit smoking. The participants between 35 and 60 years old were all smokers at the beginning of the study. They were randomized into one of three treatment groups: smoking intervention plus inhaled ipratropium bromide (SIA); smoking intervention and an inhaled placebo (SIP); and usual care (no intervention). A behavioural intervention program was provided to all participants in the two intervention groups to help them quit smoking. The participants were followed for five years. At each annual visit information about changes in smoking habits since last visit was collected along with other relevant information.

To mimic the practical sample size from many studies, we only use a random sample of 100 subjects with complete five follow-up examinations. We started from a model including only the first-order terms of the treatment group ($\mathrm{SIA}=1$ for the SIA group, $=0$ otherwise; $\mathrm{SIP}=1$ for the SIP group, $=0$ otherwise), some baseline characteristics measured at the beginning of the study – age, gender, forced expiratory volume within one second (FEV1), cigarette-smoking pack years and years of education – a linear term of visiting year ($t=1,\ldots,5$), and the weight change (DWT) in pounds between the current year and the previous year. The binary response variable is the smoking status ($\mathrm{Quit}=0$ for a smoker; $=1$ otherwise) at each visiting year. Note that some participants might smoke again after quitting for a while, whereas some others might quit at later times or never quit. Using the backward elimination procedure we found only SIA, SIP, t and DWT are significant at the nominal level 0.1 based on GEE analyses. Hence our initial marginal model is

$$\mathrm{logit}(p_{it}) = \beta_0 + \beta_1 \mathrm{SIA}_i + \beta_2 \mathrm{SIP}_i + \beta_3 \mathrm{DWT}_{it} + \beta_4 t \tag{1}$$

where $p_{it} = E(\mathrm{Quit}_{it}) = \Pr(\mathrm{Quit}_{it}=1)$ is the probability that the ith subject quits smoking at the visiting year t, $t=1,\ldots,5$, and $i=1,\ldots,n$ with $n=100$. Note that in addition to the visiting year t, DWT_{it} is also time dependent.

Figure 1 shows three marginal model plots: $a'X$ is taken as the weight change (DWT), the visiting year t and the linear predictor, respectively. We used $B=40$ bootstrap replications throughout. In the first marginal model plot (Figure 1(a)), the two marginal means are not covered by the reference band. The model-based mean curve is monotone whereas the model-free one is V-shaped. There is a strong discrepancy between them when $\mathrm{DWT} < 0$. In Figure 1(b), the two marginal mean curves are not covered by the reference band at $t=1$ and are nearly covered at $t=2$. The two mean curves are close along the direction of the linear

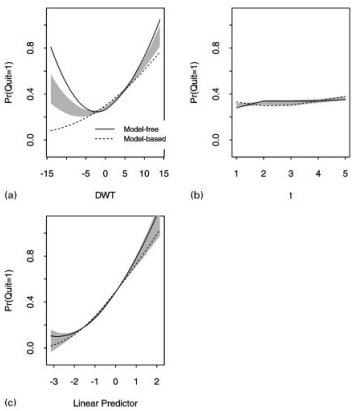

Figure 1. Marginal model plots for model (1).

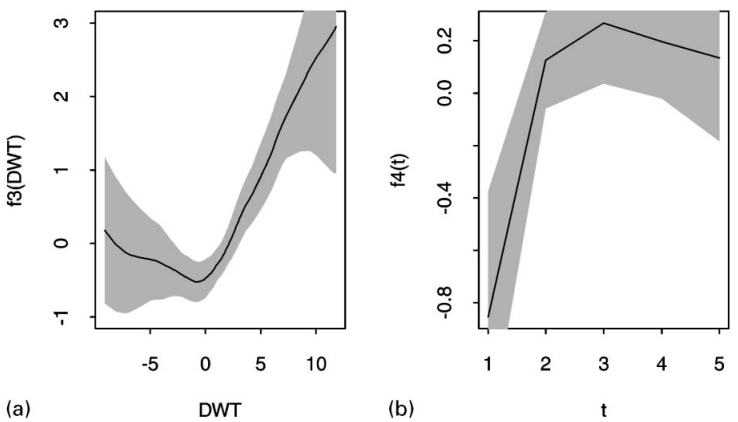

Figure 2. Generalized additive model plots for model (2). The shaded area is plus/minus twice the (pointwise) standard error.

predictor in Figure 1(c). In summary, the first two marginal model plots cast some doubts on the adequacy of model (1).

As a very useful exploratory tool, the GAM can suggest possible transformations for continuous covariates to improve model fitting. We fitted the following additive model and the results are shown in Figure 2:

$$\text{logit}(p_{it}) = \beta_0 + \beta_1 \text{SIA}_i + \beta_2 \text{SIP}_i + f_3(\text{DWT}_{it}) + f_4(t) \tag{2}$$

From Figure 2(a) it is apparent that the effect of DWT is non-linear, and a quadratic transformation is suggested. Hence we fitted the following model:

$$\text{logit}(p_{it}) = \beta_0 + \beta_1 \text{SIA}_i + \beta_2 \text{SIP}_i + \beta_3 \text{DWT}_{it} + \beta_4 \text{DWT}_{it}^2 + \beta_5 t \tag{3}$$

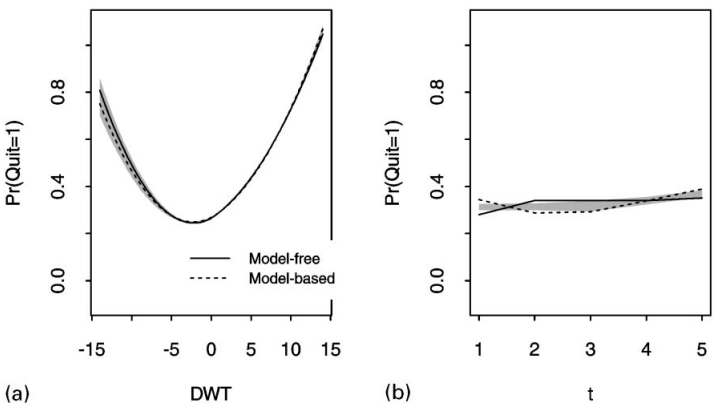

Figure 3. Marginal model plots for model (3).

Table I. Estimated regression coefficients $\hat{\beta}$, their robust standard errors (SE) and p-values based on a two-sided z-test in GEE (with a working AR(1) correlation structure) for model (4).

Term	$\hat{\beta}$	SE	p-value
Intercept	−1.755	0.363	< 0.001
SIA	1.570	0.482	0.001
SIP	0.907	0.463	0.052
DWT	0.024	0.015	0.119
DWT^2	0.003	0.002	0.050
t	0.012	0.054	0.822
$I(t=1)$	−0.210	0.091	0.021

Now the marginal model plot along the direction of the weight change (Figure 3(a)) shows no evidence of the discrepancy between the two marginal mean curves, but in Figure 3(b) the two mean functions are still not covered by the reference band for $t=1$ or 2.

Figure 2(b) clearly suggests that the effect of year 1 is different from that of all following four years whereas a linear term seems to be enough to describe the latter. Hence a binary variable indicating year 1 is added:

$$\text{logit}(p_{it}) = \beta_0 + \beta_1 \text{SIA}_i + \beta_2 \text{SIP}_i + \beta_3 \text{DWT}_{it} + \beta_4 \text{DWT}_{it}^2 + \beta_5 t + \beta_6 I(t=1) \tag{4}$$

Now the two mean functions are all covered by the reference band in each marginal model plot (Figure 4) . Note that a marginal model plot along the direction of age is also presented, though age is not included in the model. Table I shows the model fitting results. It is verified that both added terms in model (4) (compared with model (1)) are statistically significant at the usual 0.05 level.

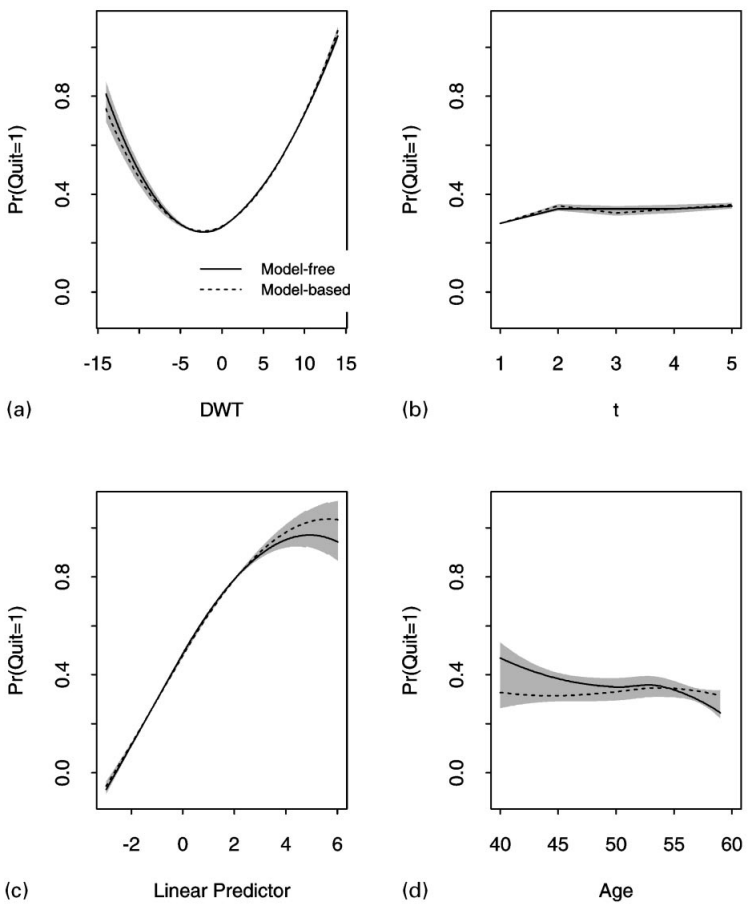

Figure 4. Marginal model plots for model (4).

5. OTHER ISSUES AND MORE SIMULATED EXAMPLES

There are some implementation issues involved. The first is how to determine the bootstrap replication number B. Of course, the larger B, the more accurate the resulting bootstrap standard error estimate. So far we have only used a moderate $B = 40$, which appears to suffice. As suggested by Efron and Tibshirani [11], if the goal is to estimate standard errors, usually it does not need a very large bootstrap replication number. As a verification, we also tried $B = 80$; the resulting marginal model plots (not shown) were essentially the same as those using $B = 40$.

Another issue is the smoothing method used. Following Cook and Weisberg [6] we used the loess smoother [17], though other smoothers can be also applied. Cook and Weisberg commented that although the loess smoother provides a biased estimate of the mean function, the bias depends only on the smoothing parameter and the curvature of the mean function, and not on the density of points. Thus, by smoothing both curves with the same smoothing

parameter, bias will cancel for their difference. We have chosen the smoothing parameter manually. In our implementation, we used the `S-plus` function *loess( )* such that its effective number of parameters (enp) is 3, which was used throughout Section 4 (including in bootstrapping). There may be some concerns about the required sample size, partly due to the non-parametric nature of the smoothing method being used. Next we use simulated data to further explore these issues.

To facilitate generating correlated binary data, we use a logistic mixed-effects model, similar to model (4)

$$\text{logit}(E(y_{it}|b_i)) = \beta_0 + \beta_1 \text{SIA}_i + \beta_2 \text{SIP}_i + \beta_3 \text{DWT}_{it} + \beta_4 \text{DWT}_{it}^2 + \beta_5 t + \beta_6 I(t=1) + b_i$$

where the random effects b_i's are i.i.d. from a mean-zero normal distribution $\text{N}(0, \sigma^2)$. Note that although in general a non-linear random-effects model may not be equivalent to any marginal model, the above logistic-normal mixed-effects model can be well approximated by a corresponding marginal logistic model [18]. Thus the above logistic mixed-effects model should be close to the marginal model (4). Using the SAS macro GLIMMIX [19] and the original data in Section 4, we fitted the model and found that the estimated parameters are

$$\hat{\beta} = (0.4497, 3.3892, 2.0601, 0.1329, 0.0082, 0.0279, -1.6384)', \quad \hat{\sigma}^2 = 13.600$$

For any given sample size n, we took a random sample from the original data set such that there was an equal number of participants in each intervention group. Rather than using the observed response values in the drawn sample, we used simulated ones based on the fitted model. Four random samples with sizes $n = 30$, 60, 90 and 120 were thus obtained, and we refer them as data sets A, B, C and D, respectively.

First, we used our default $enp = 3$ to fit the data to model (1) and model (4), respectively. The resulting marginal plots along the direction of DWT are shown in Figure 5. For data set A or D, there is clear evidence against the adequacy of model (1), whereas for the other two data sets there is no evidence against it. On the other hand, model (4) appears to be fine for all four data sets. A close look at Figure 5 shows that for data sets A and D, marginally the (non-parametrically estimated) effect of DWT on the response seems to be quadratic and V-shaped, whereas it is nearly monotone for the other two data sets. At the same time, the induced effect of DWT based on model (1) is close to being linear, and that based on model 4 is quadratic (and almost linear for data set C). Owing to these reasons the marginal model plot along the direction of DWT fails to detect the departure of model (1) from the true underlying model for data sets B and C. Next we show that due to sampling variation, the quadratic effect of DWT is in fact too weak to be detected in data sets B and C.

Note that the marginal model plot only shows the *marginal* effect of a covariate. One may argue that there may exist strong and different effects of DWT after adjusting for the other covariates in the model. The adjusted effects of covariates can be non-parametrically estimated using the GAM method. The model-fitting results using a GAM with non-parametric terms for DWT and t (that is, model (2) in Section 4) are presented in Figure 6. It is verified that the effect of DWT is closer to being linear for data sets B and C than it is for A and D. Furthermore, if we fit the data to model (4) using GEE, the estimated regression coefficients (robust standard errors) for the square term DWT^2 are 0.0113 (0.0045), 0.0018 (0.0038),

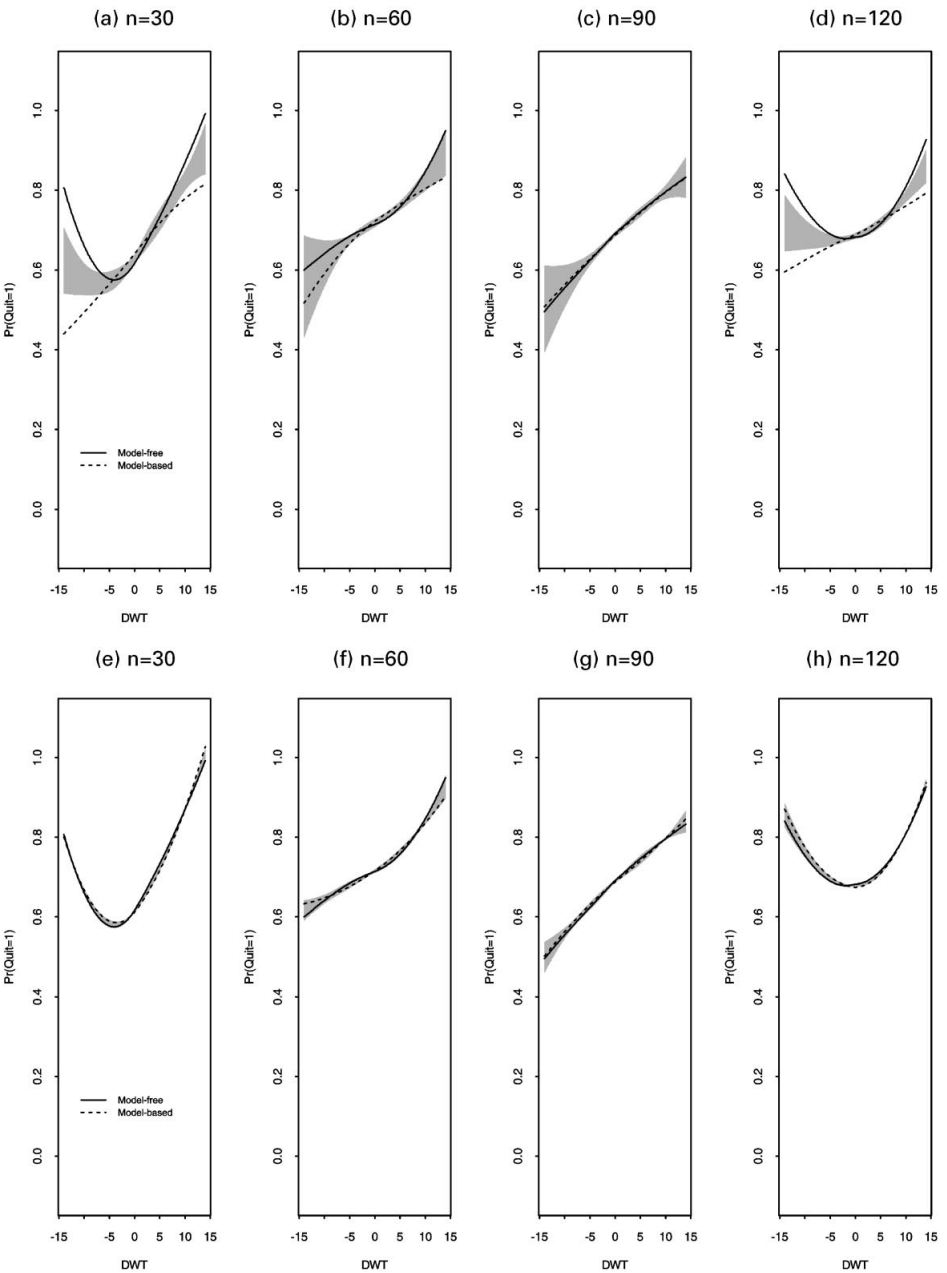

Figure 5. Marginal model plots for four simulated data sets: (a)–(d) are for model (1) and (e)–(h) for model (4). Smoothing parameter used was $enp = 3$.

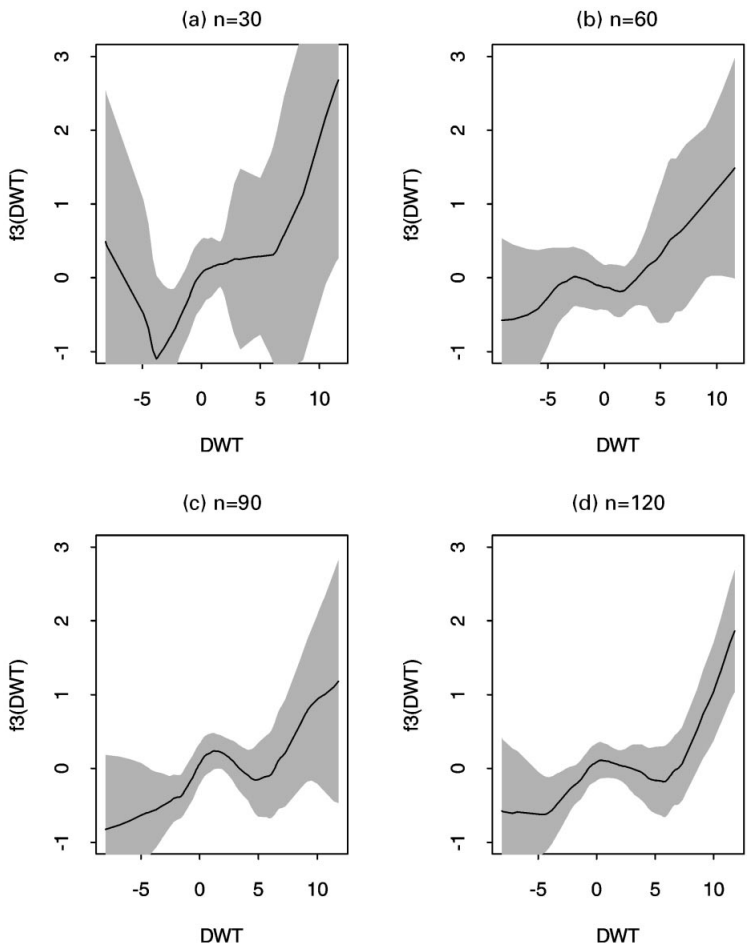

Figure 6. GAM plots for four simulated data sets.

0.0007 (0.0021) and 0.0075 (0.0020), respectively, for the four data sets. Hence the non-linear effect of DWT is significant for data sets A and D, but it is almost negligible for B and C.

Now we investigate whether using different smoothing parameters may lead to completely different results with marginal plots. To be extreme, we used a much larger smoothing parameter $enp = 12$ in the *loess* and the results are presented in Figure 7. It is not surprising to see that the estimated mean curves are much less smooth now, but also note that the reference bands are also much wider. Based on the appearance of the estimated mean curves and the given sample size, it is not hard to judge that a too large *enp* has been used. None the less, the general trend of the marginal effect of DWT does not change much. For instance, in spite of some spurious jumps in the middle, the non-parametrically estimated effect of DWT is still V-shaped for data sets A and D, whereas it is nearly monotonically increasing for B and C. An interpretation similar to that in Figure 5 can be also made for the model-based marginal

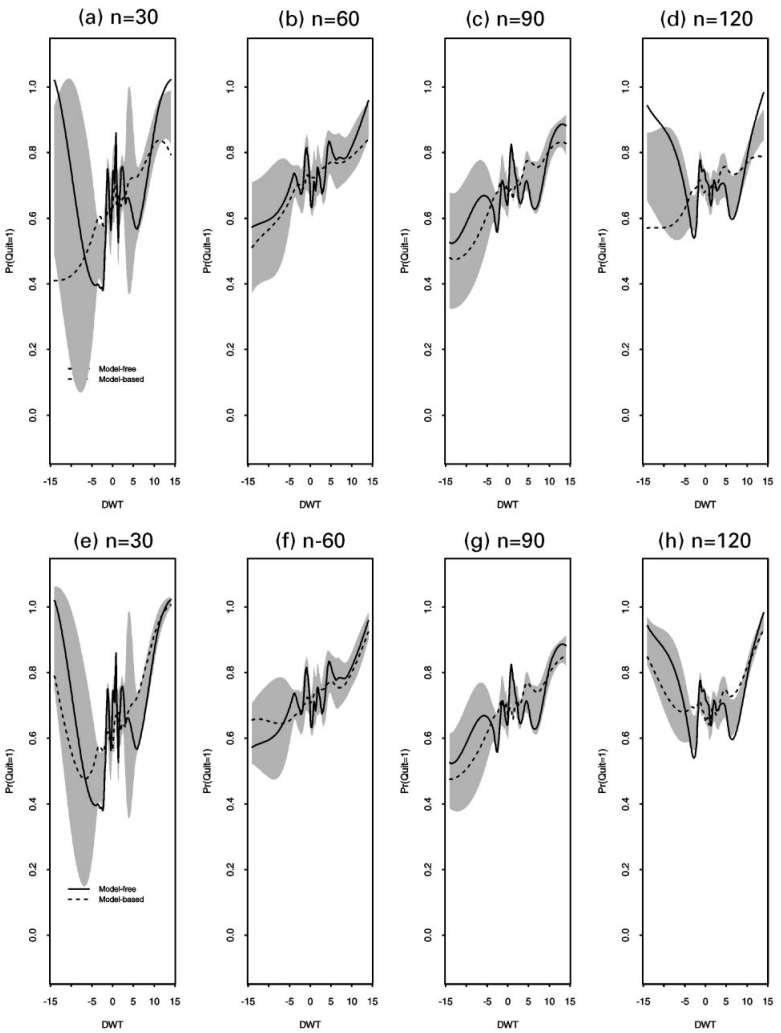

Figure 7. Marginal model plots for four simulated data sets: (a)–(d) are for model (1) and (e)–(h) for model (4). Smoothing parameter used was $enp = 12$.

effects of DWT. In summary, the same conclusion as that from Figure 5 can be also drawn here with regard to the adequacy or inadequacy of the two models.

Hence, even though the choice of smoothing parameter may influence estimated mean curves and reference bands, its effect is not really so critical here. We emphasize that our goal here is not to estimate the mean functions. Of course, in practice, sensitivity analysis by using several smoothing parameter values can be also helpful. A more objective alternative is to use cross-validation to select the smoothing parameter based on the data. However, using cross-validation is more computing intensive. Further, due to its adaptive nature, using cross-validation will make otherwise linear smoothers no longer linear.

6. DISCUSSION

We have demonstrated how to apply two graphical methods, the marginal model plot with a reference band and the GAM plot, to check the marginal regression model for correlated response data. Their effectiveness in uncovering potential modelling inadequacy has been supported in application to the Lung Health Study data set. In particular, by using the working independence model in GEE, the proposed methods can take advantage of existing smoothing and GAM softwares for independent data and thus are easy to implement.

The proposed two graphical techniques can be used in a complementary way. The marginal model plot is more like a general goodness-of-fit test and is less informative in suggesting alternative models if the current model is rejected, whereas the GAM plot can provide more specific information on the functional forms of the continuous covariates in the model. Hence using the GAM plot can help formulate new promising candidate models. However, the GAM plot is not a panacea and cannot always replace the use of the marginal model plot. For instance, using the GAM plot may not discover whether some important interaction terms, especially those involving categorical variables, are missing. On the other hand, the marginal model plot may help uncover some general evidence against the current model. Also, as in any other contexts, model building process is iterative and using a combination of various techniques is helpful.

The proposed methodology here is very general. For instance, the marginal model plot can be applied to check not only the mean functions, but also association models [20] for correlated data. It can be also applied to other more flexible modelling contexts, such as the vector GAM [21] and varying-coefficient models [22].

Correlated response data arise often in medical studies. The GEE has become more and more popular in regression analysis with such data. The properties of the GEE estimates under the *correct* regression model have been well studied in the literature. However, with regard to model checking, there is a surprising lack of relevant studies. Recently two formal goodness-of-fit tests for GEE with binary response data have been proposed [23, 24], but we are not aware of any graphical methods presented in the literature to assess model fitting in GEE, in contrast to many existing ones for independent data. Our current work is only a first step in this direction. For correlated data, especially in the GEE methodology, it appears difficult to derive an analytic form for the standard error of a smoothed marginal mean function, which however may be less computing-consuming than the bootstrap method proposed here. An important class of graphical methods is residual analysis, though it is not clear how to handle the within-subject correlation among residuals and the discrete nature of the residuals if the response variable is not continuous. These are all interesting topics to be studied in the future.

ACKNOWLEDGEMENTS

We thank the two referees for helpful suggestions.

REFERENCES

1. Cook RD, Weisberg S. *Residuals and Influence in Regression*. Chapman & Hall: New York, 1982.
2. Cook RD, Weisberg S. *An Introduction on Regression Graphics*. Wiley: New York, 1994.
3. Liang K-Y, Zeger SL. Longitudinal data analysis using generalized linear models. *Biometrika* 1986; **73**:13–22.
4. McCullagh P, Nelder JA. *Generalized Linear Models*, 2nd edn. Chapman & Hall: London, 1989.

5. Porzio GC, Weisberg S. Tests for lack-of-fit of regression models. Technical Report, University of Minnesota, St. Paul, 1999.
6. Cook RD, Weisberg S. Graphics for assessing the adequacy of regression models. *Journal of the American Statistical Association* 1997; **92**:490–499.
7. Bowman A, Young S. Graphical comparison of nonparametric curves. *Applied Statistics* 1996; **45**:83–98.
8. Hastie TJ, Tibshirani RJ. *Generalized Additive Models*. Chapman & Hall: London, 1990.
9. Wild CJ, Yee TW. Additive extension to generalized estimating equation methods. *Journal of the Royal Statistical Society*, *Series B* 1996; **58**:711–725.
10. Rice JR, Silverman BW. Estimating the mean and covariance structure nonparametrically when the data are curves. *Journal of the Royal Statistical Society*, *Series B* 1991; **53**:233–243.
11. Efron B, Tibshirani RJ. *An Introduction to the Bootstrap.* Chapman & Hall: London, 1993.
12. Ruppert D, Wand M, Holst U, Hossjer O. Local polynomial variance-function estimation. *Technometrics* 1997; **39**:262–273.
13. Prentice RL, Zhao LP. Estimating equations for parameters in means and covariances of multivariate discrete and continuous responses. *Biometrics* 1991; **47**:825–839.
14. Lin X, Carroll RJ. Nonparametric function estimation for clustered data when the predictor is measured without/with error. *Journal of the American Statistical Association* 2000; **95**:520–534.
15. Hoover DR, Rice JA, Wu CO, Yang L-P. Nonparametric smoothing estimates of time-varying coefficient models with longitudinal data. *Biometrika* 1998; **85**:809–822.
16. Connett JE, Kusek JW, Bailey WC, O'Hara P, Wu M for the Lung Health Study Research Group. Design of the Lung Health Study: a randomized clinical trial of early intervention for chronic obstructive pulmonary disease. *Controlled Clinical Trials* 1993; **14**:3S–19S.
17. Cleveland W, Devlin SJ. Locally weighted regression: an approach to regression analysis by local fitting. *Journal of the American Statistical Association* 1988; **83**:596–610.
18. Zeger SL, Liang K-Y, Albert PS. Models for longitudinal data: a generalized estimating equation approach. *Biometrics* 1988; **44**:1049–1060.
19. Littell RC, Milliken GA, Stroup WW, Wolfinger RD. *SAS System for Mixed Models.* SAS Institute Inc, Cary, NC, 1996.
20. Heagerty PJ, Zeger SL. Lorelogram: a regression approach to exploring dependence in longitudinal categorical responses. *Journal of the American Statistical Association* 1988; **93**:150–162.
21. Yee TW, Wild CJ. Vector generalized additive models. *Journal of the Royal Statistical Society*, *Series B* 1996; **58**:481–493.
22. Hastie TJ, Tibshirani RJ. Varying-coefficient models. *Journal of the Royal Statistical Society*, *Series B* 1993; **55**: 757–796.
23. Barnhart HX, Williamson JM. Goodness-of-fit tests for GEE modelling with binary responses. *Biometrics* 1998; **54**: 720–729.
24. Horton NJ, Bebchuk JD, Jones CL, Lipsitz SR, Catalano PJ, Zahner GEP, Fitzmaurice GM. Goodness-of-fit for GEE: an example with mental health service utilization. *Statistics in Medicine* 1999; **18**:213–222.

