



# Generalized linear models

John Neuhaus and Charles McCulloch\*

The class of generalized linear models (GLMs) extends the classical linear model for continuous, normal responses to describe the relationship between one or more predictor variables  $x_1, \dots, x_p$  and a wide variety of nonnormally distributed responses  $Y$  including binary, count, and positive-valued variates. GLMs expand the class of response densities from the normal to an exponential family that contains the normal, Poisson, binomial, and other popular distributions as special cases. The models produce estimated expected values that conform to response constraints and allow nonlinear relationships between predictors and expected values. It is straightforward to construct the likelihood for a set of data so that maximum likelihood and related likelihood-based methods are popular techniques for parameter estimation and inference. A key point with GLMs is that many of the considerations in model construction are the same as for standard linear regression models as the models have many common features. © 2011 John Wiley & Sons, Inc.

WIREs Comp Stat 2011 3 407–413 DOI: 10.1002/wics.175

## INTRODUCTION

The classical linear regression model specifies that a continuous response  $Y$  follows a normal distribution with variance  $\sigma^2$  and with its mean,  $\mu$ , a function of one or more predictor variables  $x_1, \dots, x_p$ :

$$Y \sim N(\mu, \sigma^2),$$

$$\mu = \beta_0 + \beta_1 x_1 + \dots + \beta_p x_p. \quad (1)$$

The class of generalized linear models (GLMs)<sup>1,2</sup> extends the model (1) for continuous, normal responses to describe the relationship between the vector of predictors  $\mathbf{x} = (x_1, \dots, x_p)$  and a wide variety of nonnormal response distributions for  $Y$ , able to accommodate responses that are binary indicators, counts and positive-valued variates. The class contains popular models such the logistic, probit, Poisson, and some survival models as special cases.

## EXAMPLE

A study of back pain<sup>3</sup> motivates the need for models that generalize the classical regression linear model and illustrates the features such models should have.

Von Korff et al.<sup>3</sup> were interested in several outcomes related to back pain including the cost of care for back pain, a binary indicator of whether the patient understood how to care for their back, and the number of days back pain limited their activities. A linear model with symmetric, normally distributed errors does not describe any of these outcomes well. Medical costs are positive and tend to have a distribution that is highly skewed to the right. One could obtain a more symmetric and approximately normally distributed cost outcome by transforming cost using a function such as the logarithm. But interest focused on obtaining estimates directly of the average cost, which is cumbersome to estimate after transforming the response.

More specifically, with a log transformed response, a linear regression model would provide ready estimates of the average log cost,  $E[\log(\text{cost})]$ . The temptation would be to exponentiate estimates of the average log cost to obtain cost estimates. However, using Jensen's inequality,

$$E[\log(\text{cost})] \geq \log[E(\text{cost})],$$

leading to

$$\exp\{E[\log(\text{cost})]\} \geq E(\text{cost}),$$

so that one cannot immediately calculate means on the original scale from means on the log scale. With GLMs the expected value of the response (in this

\*Correspondence to: chuck@biostat.ucsf.edu

Division of Biostatistics, University of California, San Francisco, San Francisco, CA, USA

DOI: 10.1002/wics.175

case, cost itself) can be modeled directly as a function of predictors and it is straightforward to calculate estimates of expected values.

A linear model does not provide a good description of the patient understanding variable as the expected value of this binary random variable is a probability and must lie between 0 and 1. Linear models can produce fitted probabilities outside of the (0, 1) range. In addition, as toxicological studies from as far back as Bliss<sup>4</sup> show, the proportion responding tends to follow a sigmoidal curve as a function of a continuous predictor, rather than a line. GLMs produce estimated expected values that conform to response constraints and allow nonlinear relationships between predictors and expected values.

The Poisson distribution for counts provides a natural distribution for the number of pain-limited days, rather than a distribution such as the normal with support on the entire real line. GLMs expand the class of response densities from the normal to an exponential family that contains the normal, Poisson, binomial, and other popular distributions as special cases.

## STRUCTURE OF THE MODEL

Constructing a GLM involves three separate specifications:

1. The random component: the distribution of the response  $Y$ .
2. The systematic component: the function of the covariates  $\mathbf{x} = x_1, \dots, x_p$  that relate  $\mathbf{x}$  to the expected value of the response. With GLMs the covariates enter the statistical model via a function  $\eta = \beta_0 + \beta_1 x_1 + \dots + \beta_p x_p$ , known as the linear predictor.
3. The connection between the random and systematic components which is assumed to be through  $\mu = E(Y)$ . It is assumed that this takes the form  $\mu = g^{-1}(\eta)$ , where  $g$  is a known function.

We consider each of these specifications in more detail.

### Distribution of $Y$

While a classical linear regression model specifies that the response  $Y$  follows a normal distribution, GLMs allow the response to follow a distribution with density from the exponential family or similar to

the exponential family:

$$y_i \sim \text{indep. } f_{Y_i}(y_i),$$

$$f_{Y_i}(y_i) = \exp\{[y_i \theta_i - b(\theta_i)]/\tau^2 - c(y_i, \tau)\}, \quad (2)$$

where  $\theta_i$  is known as the canonical parameter,  $c$  is a known function, and  $\tau^2$  is a dispersion parameter. For example, for binary data

$$f_{Y_i}(y_i) = p_i^{y_i} (1 - p_i)^{1-y_i}, \quad (3)$$

where  $p_i$  is the probability of a success. Writing  $f_{Y_i}(y_i)$  in the format of Eq. (2) shows that  $\theta_i = \log[p_i/(1 - p_i)]$  (or  $p_i = 1/(1 + e^{-\theta_i})$ ),  $b(\theta_i) = \log(1 + e^{\theta_i})$ , and  $c(y_i, \tau) = 1$ . Most commonly used distributions can be written in the form of Eq. (2).

### Function of Covariates $\mathbf{x}$

As in the classical linear model, GLMs assume that the covariates  $x_1, \dots, x_p$  relate to the expected value of the response through a linear predictor  $\eta$  with

$$\eta = \beta_0 + \beta_1 x_1 + \dots + \beta_p x_p. \quad (4)$$

This key feature of GLMs, namely that the predictors affect Eq. (4) exactly the same way as they do in a linear regression model, Eq. (1), means that many of the features of the latter carry over to the former and many of the considerations in modeling are the same. As examples, just like linear regression, GLMs can handle a mix of continuous and categorical predictors. Issues of how to represent predictors and model nonlinear relationships are similar (e.g., using polynomial functions of predictors or splines). Furthermore, analysts can employ similar strategies to assess whether to retain a predictor in a model. And random factors can be incorporated in exactly the same way with linear models and GLMs.

### Link Function

Rather than assume that  $\mu = E(Y)$  is a linear function of covariates, as in linear models, GLMs assume that a function  $g$  of  $\mu$  is equal to the linear predictor,  $\eta$ :

$$g(\mu) = \eta, \quad (5)$$

where  $g(\cdot)$  is a known function, called the *link function* because it links together the mean of  $y$  and the linear form of predictors given in Eq. (4). Table 1 lists several distributions related to commonly used GLMs along with typical settings where an analyst might apply them. An important feature of GLMs

**TABLE 1** | Common GLMs

Distribution of $Y$	Variance to mean relation	Link	Used for
Normal	Constant $\sigma^2$	Identity	Linear regression
Binomial	$\sigma^2 = n\mu(1 - \mu)$	Logit	Logistic regression
		Probit	Probit regression
		Complementary log–log	Complementary log–log regression
Overdispersed binomial	$\sigma^2 \propto n\mu(1 - \mu)$	Logit	Clustered logistic regression
Poisson	$\sigma^2 = \mu$	Log	Count data, Poisson regression
Overdispersed Poisson	$\sigma^2 \propto \mu$	Log	Count data, $\sigma^2$ proportional to mean
Negative binomial	$\sigma^2 = \mu + \mu^2/k$	Log	Count data, $\sigma^2$ quadratic in the mean
Gamma	$\sigma \propto \mu$	Reciprocal	Continuous data, $\sigma$ proportional to mean

is that regression coefficients in Eq. (4) retain the nice interpretation of linear regression coefficients as changes in the function of the expected value of the response,  $g(\mu)$ , associated with a one unit increase in a predictor, for any value of the predictor. The function  $g$  is the identity for the linear regression model but takes other forms for other models. For example, for the probit model,<sup>5</sup> for binary data, in which  $\mu$  is the probability of a success and is related to the canonical parameter via  $\mu = 1/(1 + \exp[-\theta])$ , the link function is  $g(\mu) = \Phi^{-1}(\mu)$ . In general,  $g$  can be any monotonic, differentiable function but should be appropriate for the situation. For example, log link models for binary responses are problematic because for some covariate values the model will produce estimates of the expected value of the binary response, a probability, that exceed 1. Link functions based on inverse cumulative probability functions such as the logistic do not have this defect and always produce expected values in the interval (0, 1). A potential drawback of a standard GLM is that predictors relate to the expected value of the response solely through the linear predictor  $g(\mu) = \eta$ , whereas predictors may affect the response distribution in a more complicated manner.

For some special models known as canonical link models, the canonical parameter of Eq. (2),  $\theta_i$ , is equal to the linear predictor  $\eta_i$ ,  $\theta_i = \eta_i$ . Canonical link models include identity link models for normal responses, logistic models for binary responses, and Poisson models for counts. For these special models sufficient statistics for model parameters exist.

## ESTIMATION AND INFERENCE

Using Eq. (2), it is straightforward to construct the likelihood based on a set of  $n$  independent observations so that maximum likelihood and related likelihood-based methods are popular techniques for

parameter estimation and inference. These methods possess several optimal properties such as smallest asymptotic variance<sup>6</sup> but require strong assumptions about the density of the responses  $Y$ . In some settings data analysts may not have sufficient information to make strong assumptions about the density of responses but are able to specify models for the expected value and variance of  $Y$ . Wedderburn<sup>7</sup> showed that the specification of just a few features such as the expected value and variance of a random variable allows the construction of a likelihood-like quantity known as a quasilielihood which can perform as well as the full likelihood or else nearly as well. The following sections discuss estimation and inference by full likelihood, quasilielihood, and conditional likelihood methods.

### Estimation by Maximum Likelihood

For a sample of  $n$  independent observations, indexed by  $i = 1, \dots, n$ , the log likelihood,  $l$ , from Eq. (2) is

$$l = \sum_{i=1}^n [y_i \theta_i - b(\theta_i)] / \tau^2 - \sum_{i=1}^n c(y_i, \tau), \quad (6)$$

which we maximize as a function of  $\beta$  and  $\tau$ , if it is unknown, to obtain estimates of the parameters of interest. Before discussing parameter estimation further it is useful to relate the parameters of Eq. (6) to the moments of  $Y$ . It is well known that

$$E \left\{ \frac{\partial l}{\partial \theta_i} \right\} = E[\{y_i - b'(\theta_i)\} / \tau^2] = 0, \quad (7)$$

and

$$\text{var} \left\{ \frac{\partial l}{\partial \theta_i} \right\} = -E \left\{ \frac{\partial^2 l}{\partial \theta_i^2} \right\} = b''(\theta_i) / \tau^2 \quad (8)$$

under regularity conditions.<sup>6</sup> Standard texts such as McCullagh and Nelder<sup>2</sup> show that

$$E[y_i] = \mu_i = b'(\theta_i), \quad (9)$$

and

$$\text{var}(y_i) = \tau^2 b''(\theta_i) \equiv \tau^2 \nu(\mu_i), \quad (10)$$

where  $\nu(\mu_i)$  is known as the variance function.

Using Eqs (9) and (10), as well as the chain rule, standard texts such as McCullagh and Nelder<sup>2</sup> show that

$$\frac{\partial l}{\partial \beta_k} = \frac{1}{\tau^2} \sum (y_i - \mu_i) w_i g_\mu(\mu_i) \mathbf{x}'_{ik}, \quad k = 0, \dots, p, \quad (11)$$

where  $w_i = [v(\mu_i)g_\mu^2(\mu_i)]^{-1}$ . Solving  $\partial l / \partial \beta = 0$  yields maximum likelihood estimators  $\hat{\beta}$ . The expected value of the matrix of second derivatives of Eq. (2) provides asymptotic variances of the estimators  $\hat{\beta}$ ,  $\text{var}_\infty(\hat{\beta})$ . Replacing parameters in this matrix by their maximum likelihood estimates provides estimated variances of  $\hat{\beta}$ .

An extension of least squares estimation known as the iteratively re-weighted least squares algorithm provides a common approach to compute the maximum likelihood estimators of the parameters of GLMs. Given the current value of the linear predictor  $\hat{\eta}_0$  and associated fitted value  $\hat{\mu}_0 = g^{-1}(\hat{\eta}_0)$ , one obtains updated estimates  $\hat{\beta}$  by regressing an adjusted dependent variable  $z_0$  on the covariates  $\mathbf{x}$  with weight  $w_0$ . The adjusted dependent variable  $z_0$  is

$$z_0 = \hat{\eta}_0 + (y - \hat{\mu}_0)g'(\hat{\mu}_0), \quad (12)$$

while the weight is defined by

$$w_0^{-1} = [g'(\hat{\mu}_0)]^2 \nu(\hat{\mu}_0), \quad (13)$$

where  $\nu$  is the variance function. One iterates the above algorithm to convergence at the maximum likelihood estimate.

We can test hypotheses about the associations of covariates with the response, that is, test hypotheses about the parameters  $\beta_k$  in Eq. (4) using likelihood ratio, score, or Wald tests. Note that these tests provide asymptotic inferences and all three have the same limiting  $\chi^2$  distribution. We can construct large sample confidence intervals for parameters by inverting any of these three tests. For example, using the notation of Eq. (4), suppose that we wish to test the hypothesis  $H_0 : \beta_1 = \beta_{10}$ , where  $\beta_{10}$  is a hypothesized

value of  $\beta_1$ . To conduct the test, we form the Wald statistic

$$W = (\hat{\beta}_1 - \beta_{10})' [\widehat{\text{var}}_\infty(\hat{\beta}_1)]^{-1} (\hat{\beta}_1 - \beta_{10}), \quad (14)$$

which, under  $H_0$ , has a large sample  $\chi^2$  distribution with one degree of freedom. Inverting the Wald test, a  $1 - \alpha$  Wald-based confidence interval for  $\beta_1$  includes all values  $\beta_1$  such that

$$(\hat{\beta}_1 - \beta_1)' [\widehat{\text{var}}_\infty(\hat{\beta}_1)]^{-1} (\hat{\beta}_1 - \beta_1) \leq \chi_{1,1-\alpha}^2. \quad (15)$$

We can follow a similar approach to construct likelihood- and score-based confidence intervals.

## Quasilikelihood

To implement the maximum likelihood procedures of the previous section, a data analyst must specify the density given in Eq. (2) of  $Y$  and there may not be sufficient information about  $Y$  to make this specification; an analyst may be able to specify a model for the mean and variance of the response but not the full density. Quasilikelihood methods provide parameter estimation and inference in such settings. The approach involves a likelihood-like quantity whose construction involves fewer assumptions. To construct the quasilikelihood, we note that the function

$$q_i = \frac{y_i - \mu_i}{\tau^2 \nu(\mu_i)} \quad (16)$$

has important properties in common with the log likelihood derivative  $\partial l / \partial \theta_i$ . Namely  $E(q_i) = 0$ , as in Eq. (7), and  $\text{var}(q_i)$  is equal to the negative of the expected value of the derivative of  $q_i$ , as in Eq. (8). Analogous to the relationship of  $\partial l / \partial \theta_i$  to the log likelihood, we define the log quasilikelihood via the contribution  $y_i$  makes to it:

$$Q_i = \int_{y_i}^{\mu_i} \frac{y_i - t}{\tau^2 \nu(t)} dt, \quad (17)$$

which, by definition, has derivative with respect to  $\mu_i$  equal to  $q_i$ . Maximizing  $\sum Q_i$  with respect to  $\beta$  yields maximum quasilikelihood estimators. Evaluating derivatives with respect to  $\beta$ , we obtain estimators by solving

$$\sum \frac{y_i - \mu_i}{\tau^2 \nu(\mu_i)} \frac{\partial \mu_i}{\partial \beta} = \sum \frac{y_i - \mu_i}{\tau^2 \nu(\mu_i) g(\mu_i)} \mathbf{x}'_i = 0, \quad (18)$$

where we use the fact that

$$\frac{\partial \mu_i}{\partial \beta} = \mathbf{x}'_i / g(\mu_i) \quad (19)$$

for GLMs.

It is important to note that the construction of  $Q_i$  only requires specifying how the variance changes with the mean. Also, it is often the case that if we specify a mean-to-variance relationship, we obtain maximum quasiliquid equations which are exactly the same as those corresponding to a legitimate likelihood. Directly modeling the mean-to-variance relationship frees data analysts from some constraints imposed by fully parametric GLMs. For example, the quasiliquid approach provides methods to analyze count data where the response variance substantially exceeds the mean, a feature that a standard Poisson regression model cannot accommodate. Settings where the actual variability exceed model-based values are termed ‘overdispersed’. Common statistical packages estimate the degree of overdispersion using moment-based methods applied to model goodness-of-fit measures such as the Pearson  $\chi^2$ .<sup>2</sup>

Inference for the parameters  $\beta$  using quasiliquid methods proceeds much as with full maximum likelihood. This follows from the work of McCullagh<sup>8</sup> that shows that maximum quasiliquid estimators consistently estimate  $\beta$  and have an asymptotic normal distribution.

## Conditional Likelihood

Settings commonly arise where experimental units are gathered in clusters, matched sets, or other fine strata. Gathering costs and pain-related outcomes on the same patients (=cluster) during several time periods in the back pain study would be an example. Cluster-specific means of responses typically exhibit substantial variability between clusters and it would be natural to consider describing such data by fitting a GLM with a parameter for each cluster, along with other predictors of interest. The model will typically contain many cluster-specific parameters and relatively few parameters corresponding to the predictors of interest. It is well known that standard full likelihood estimation methods can yield inconsistent estimators of the effects of the predictors of interest.<sup>9</sup> Intuitively, the inconsistent estimation results from the fact that the parameter set grows with the size of the sample with little new information about these additional parameters.

Instead of trying to simultaneously estimate both the parameters associated with the predictors of interest and all the cluster-specific parameters, conditional likelihood methods focus on the parameters of interest and eliminate all the cluster-specific parameters from the likelihood by conditioning on sufficient statistics which exist for canonical link models.

Let  $i = 1, \dots, m$  denote clusters or strata,  $j = 1, \dots, n_i$  denote units within clusters, and  $\alpha_i$  denote

the cluster-specific intercepts in  $\eta_i = \theta_i$ . For canonical link models the sufficient statistics are  $\sum_{j=1}^{n_i} y_{ij}$  and the conditional likelihood has terms

$$f\left(y_{i1}, \dots, y_{in_i} \middle| \sum_{j=1}^{n_i} y_{ij}\right), \quad (20)$$

which depend on the parameters of interest,  $\beta$  but not the  $\alpha_i$ .

We can use the conditional likelihood as we would use a standard likelihood. For example, we obtain parameter estimates by maximizing the likelihood built up from terms in Eq. (20) and obtain estimated large sample standard errors from the Hessian of the conditional likelihood. We can also use standard likelihood-based inference methods such as likelihood ratio tests.

Although conditional likelihood methods are only readily available for canonical link model, Neuhaus and Kalbfleisch<sup>10</sup> and Neuhaus and McCulloch<sup>11</sup> show that one can obtain conditional likelihood-like inference for noncanonical link models. These authors show that the conditional likelihoods (Eq. (20)), depend on the covariates  $x_{ij}$  only through the deviations  $(x_{ij} - \bar{x}_i)$ . This suggests a strong connection between conditional likelihood estimates and estimates from generalized linear mixed models (GLMMs)<sup>12</sup> that model covariate effects in terms of these deviations. Such models replace  $\beta_1 x_{ij}$  in the linear predictor by

$$\beta_B \bar{x}_i + \beta_W (x_{ij} - \bar{x}_i), \quad (21)$$

so that  $\beta_B$  measures the change in  $E[y_{ij}]$  associated with between-cluster differences in covariate means while  $\beta_W$  measures the change in  $E[y_{ij}]$  associated with within-cluster covariate differences. The connection between covariate decomposition and conditional likelihood approaches suggests that GLMMs that include separate between- and within-cluster covariate components can provide conditional likelihood-like inference even for noncanonical link models that do not support conditional likelihood methods.

## DIAGNOSTICS

After specifying and fitting a statistical model of interest, it is important for the data analyst to assess the quality of the fit to the data in order to validate model assumptions, to identify observations where the model fits the data particularly poorly, and to identify observations which influence the fit far more than others. Methods to assess the quality of the fit of a linear regression model, termed regression diagnostics, are well developed.<sup>13</sup> The close connection of the



linear predictor (Eq. (4)) to the classical linear model allows the extension of linear regression diagnostics to GLMs in a straightforward manner. For example, Pregibon<sup>14</sup> replaces the continuous response in diagnostic methods of Cook and Weisberg<sup>13</sup> with the adjusted dependent variable  $z$  in form of Eq. (12) to develop deletion diagnostics and leverage plots for GLMs. In addition, Wang<sup>15</sup> developed residual plots for GLMs. Common statistical computing packages have implemented these diagnostics so that data analysts can apply common model assessment methods for any GLM.

## COMMENTS

This article has focused on GLMs for independent observations, but as the Conditional Likelihood Section indicates, settings often arise where observations are gathered in clusters or groups and analysts need to accommodate within-cluster dependence of the responses. Longitudinal data are an important case of such dependent data. GLMMs<sup>12</sup> extend standard GLMs to settings with dependent data by introducing random effects into the linear predictor,  $\eta$  to model the dependence. The assumptions underlying GLMMs allow the construction of a likelihood and analysts can use likelihood-based methods for parameter estimation and inference. GLMMs are usually more difficult to fit than standard GLMs, often requiring numerical integration methods, but provide estimates of the associations of changes in predictors within clusters with changes in response, typically the object of scientific interest with longitudinal data.

Lee and Nelder<sup>16</sup> describe a related class of models, known as hierarchical GLMs, which combine GLMs with latent variables in the linear predictor. Rather than estimating parameters by maximizing the marginal likelihood obtained by integrating out the random effects, as with GLMMs, Lee and Nelder<sup>16</sup>

advocate the use of a hierarchical likelihood, known as  $h$ -likelihood, which avoids the integration but can yield biased estimators.

Generalized estimating equations (GEE) methods<sup>17</sup> are another extension of GLMs to dependent data settings. GEE is essentially a quasilielihood approach where the data analysts specifies a GLM for each unit in the sample along with some working assumptions about the within-cluster correlation structure of the responses. Like the methods of the Quasilielihood Section, GEE approaches do not specify the joint distribution of the responses so that standard likelihood procedures are not available. Instead, one constructs an estimating equation whose solution is a consistent estimator of the parameters of the unit-specific models.<sup>18</sup> The parameters of the models underlying the GEE approach measure the same covariate effects that one would with a single observation per subject and do not involve the cluster structure of the data.

## CONCLUSION

GLMs are extensions of the classical linear regression model for continuous, normal responses that allow the regression analysis of a variety of nonnormal responses such as binary indicators, counts, and positively valued random variables. As the covariates of GLMs enter the statistical model via a linear predictor, as in linear regression models, many of the modeling decisions and strategies of linear models carry over exactly or with a minor modification to GLMs. Analysis with GLMs is typically likelihood-based and data analysts can use a common algorithm to obtain maximum likelihood estimators for any model in the class. Like linear regression models for continuous, normal-like data, GLMs provide useful assessments of the associations of predictors with a wide variety of responses.

## REFERENCES

1. Nelder JA, Wedderburn RWM. Generalized linear models. *J R Stat Soc, Ser A* 1972, 135:370–384.
2. McCullagh P, Nelder JA. *Generalized Linear Models*. 2nd ed. London: Chapman & Hall; 1989.
3. Von Korff M, Barlow W, Cherkin D, Deyo R. Effects of practice style in managing back pain. *Ann Internal Med* 1994, 121:187–195.
4. Bliss C. The calculation of the dose-mortality curve. *Ann Appl Biol* 1935, 22:134–167.
5. Finney DJ. *Probit Analysis*. 3rd ed. Cambridge: Cambridge University Press; 1971.
6. Cox DR, Hinkley DV. *Theoretical Statistics*. London: Chapman & Hall; 1974.
7. Wedderburn RWM. Quasi-likelihood functions, generalized linear models, and the Gauss–Newton method. *Biometrika* 1974, 61:439–447.
8. McCullagh P. Quasi-likelihood functions. *Ann Stat* 1983, 11:59–67.
9. Neyman J, Scott EL. Consistent estimates based on partially consistent observations. *Econometrica* 1948, 16:1–32.

10. Neuhaus JM, Kalbfleisch JD. Between- and within-cluster covariate effects in the analysis of clustered data. *Biometrics* 1998, 54:638–645.
11. Neuhaus JM, McCulloch CE. Separating between and within-cluster covariate effects using conditional and partitioning methods. *J R Stat Soc, Ser B* 2006, 68:859–872.
12. McCulloch CE, Searle SR, Neuhaus JM. *Generalized, Linear and Mixed Models*. 2nd ed. New York: John Wiley & Sons; 2008.
13. Cook RD, Weisberg S. *Residuals and Influence in Regression*. New York: Chapman & Hall; 1982.
14. Pregibon D. Logistic regression diagnostics. *Ann Stat* 1981, 9:705–724.
15. Wang P. Residual plots for detecting nonlinearity in generalized linear models. *Technometrics* 1987, 29:435–438.
16. Lee Y, Nelder J. Hierarchical generalized linear models (with discussion). *J R Stat Soc, Ser B* 1996, 58:619–678.
17. Diggle PJ, Heagerty PJ, Liang K-Y, Zeger SL. *Analysis of Longitudinal Data*. 2nd ed. Oxford: Oxford University Press; 2002.
18. Liang K-Y, Zeger SL. Longitudinal data analysis using generalized linear models. *Biometrika* 1986, 73:13–22.