Ann Inst Stat Math (2011) 63:511–533
DOI 10.1007/s10463-009-0239-z

Semiparametric marginal and association regression methods for clustered binary data

Grace Y. Yi · Wenqing He · Hua Liang

Received: 31 July 2008 / Revised: 23 February 2009 / Published online: 26 May 2009

Abstract Clustered data arise commonly in practice and it is often of interest to estimate the mean response parameters as well as the association parameters. However, most research has been directed to inference about the mean response parameters with the association parameters relegated to a nuisance role. There is little work concerning both the marginal and association structures, especially in the semiparametric framework. In this paper, our interest centers on inference on the association parameters in addition to the mean parameters. We develop semiparametric methods for both complete and incomplete clustered binary data and establish the theoretical results. The proposed methodology is illustrated through numerical studies.

Keywords Association · Binary outcomes · Clustered data · Estimating equation · Missing at random · Semiparametric estimation

1 Introduction

Clustered data arise frequently in practice, and their analysis methods typically differ from univariate analysis methods due to possible association within clusters.

G. Y. Yi
Department of Statistics and Actuarial Science, University of Waterloo,
Waterloo, ON N2L 3G1, Canada

W. He
Department of Statistical and Actuarial Sciences, University of Western Ontario,
London, ON N6A 5B7, Canada

H. Liang (✉)
Department of Biostatistics and Computational Biology, University of Rochester Medical Center,
601 Elmwood Avenue, Box 630, Rochester, NY 14642, USA
e-mail: hliang@bst.rochester.edu

Incorporating complex association structure in inferential procedures becomes a major challenge for analysis of clustered data, especially for discrete data. Under the parametric framework, various methods have been developed for binary data (see Prentice 1988; Lipsitz et al. 1991; Carey et al. 1993; Molenberghs and Lesaffre 1994; Yi and Cook 2002, for instance).

Semiparametric models based on generalized estimating equations (GEE) methods and their extensions have become increasingly popular (see, e.g., Severini and Staniswalis 1994; Carroll et al. 1997; Xia et al. 1999; Lin and Carroll 2001a,b; Wang 2003; Fan and Li 2004; Wang et al. 2005; Chen and Jin 2005; Xia and Härdle 2006, among others). These methods mainly concern the marginal mean parameters with the association parameters treated as nuisance. However, in many applications, estimation of the association parameters is a central theme of the study. For example, in familial studies of inherited traits and developmental toxicology studies of laboratory animals (e.g., Hall and Severini 1998), subjects in a family or cluster share common genetic traits or are subject to common environmental factors, and it is of prime scientific interest to study the association between responses.

Under the likelihood formulation, Aerts and Claeskens (1997) and Claeskens and Aerts (2000) explored inference methods for the marginal and association parameters. However, with a marginal formulation for semiparametric regression there is relatively little discussion on featuring both the mean and association structures. To fill up this gap, in this paper we develop semiparametric inference procedures for both marginal and association parameters for clustered binary data, and rigorously establish the asymptotic properties. As missing observations commonly occur, which causes additional difficulty in conducting inference (Little and Rubin 2002), in this paper we also describe methods to handle incomplete data. The proposed methods provide a flexible framework to feature various types of mean and association structures. Such a flexibility, however, comes at the cost of increasing the complexity of inferential procedures. First, the common computing algorithm based on the Newton–Raphson method that applies to parametric models cannot be directly used now due to the unknown function $\theta(.)$. Secondly, and more importantly, existing asymptotic distribution theory under the parametric framework (Prentice 1988; Yi and Cook 2002) breaks down for the current setup. The inclusion of a nonparametric term $\theta(.)$ into the mean model remarkably changes the nature of the model setup, and it is not trivial to establish the asymptotic results for the estimators of the mean and association parameters.

The remainder of the article is organized as follows. The notation and inference framework are introduced in Sect. 2 and estimation procedures are described in Sec. 3. In Sect. 4 we establish the asymptotic properties of the resulting estimators and discuss the issue of parameter interpretation. Numerical studies are given in Sect. 5 to assess the performance and to illustrate the use of the proposed methods. In Sect. 6, we develop the inference methods for handling incomplete data. We conclude the article with a discussion in the last section.

2 Notation and framework

Suppose that there are n clusters and m_i subjects within cluster i, $i = 1, \ldots, n$. Let Y_{ij} be the binary response for subject j in cluster i, $\mathbf{x}_{ij}$ and $\mathbf{z}_{ij}$ be the $p \times 1$ and $q \times 1$

covariate vectors, respectively. Denote $\mathbf{Y}_i = (Y_{i1}, \ldots, Y_{im_i})^{\mathrm{T}}$, $\mathbf{x}_i = (\mathbf{x}_{i1}, \ldots, \mathbf{x}_{im_i})^{\mathrm{T}}$ and $\mathbf{z}_i = (\mathbf{z}_{i1}, \ldots, \mathbf{z}_{im_i})^{\mathrm{T}}$. Define $\mu_{ij} = E(Y_{ij}|\mathbf{x}_i, \mathbf{z}_i)$, and let $\boldsymbol{\mu}_i = (\mu_{i1}, \ldots, \mu_{im_i})^{\mathrm{T}}$, $i = 1, \ldots, n$. Provided the mean of Y_{ij} depends only on the covariate vector for subject j, i.e. $E(Y_{ij}|\mathbf{x}_i, \mathbf{z}_i) = E(Y_{ij}|\mathbf{x}_{ij}, \mathbf{z}_{ij})$ (Pepe and Anderson 1994), we consider the regression model

$$g^{-1}(\mu_{ij}) = \mathbf{x}_{ij}^{\mathrm{T}}\boldsymbol{\beta} + \theta(\mathbf{z}_{ij}^{\mathrm{T}}\boldsymbol{\alpha}) \quad \text{with } \|\alpha\| = 1 \tag{1}$$

where $g(.)$ is a monotone link function, $\boldsymbol{\beta}$ and $\boldsymbol{\alpha}$ are parameter vectors, and $\theta(\cdot)$ is an unknown smoothing function. The requirement $\|\alpha\| = 1$ ensures identifiability of $\boldsymbol{\alpha}$. Common choices of function $g(.)$ include logit, probit, and complementary log–log functions. In some situations, $g(.)$ can assume a flexible but more complex function form. For instance, Kim et al. (2008) proposed a class of link functions based on generalized t-distributions to characterize binary responses.

We assume that Y_{ij} and $Y_{i'j'}$ are independent for different clusters i and i', but within the same cluster, the responses may be correlated. Let ψ_{ijk} be the odds ratio between responses Y_{ij} and Y_{ik} in cluster i ($j < k$), defined by

$$\psi_{ijk} = \frac{P(Y_{ij} = 1, Y_{ik} = 1|\mathbf{x}_i, \mathbf{z}_i) \cdot P(Y_{ij} = 0, Y_{ik} = 0|\mathbf{x}_i, \mathbf{z}_i)}{P(Y_{ij} = 1, Y_{ik} = 0|\mathbf{x}_i, \mathbf{z}_i) \cdot P(Y_{ij} = 0, Y_{ik} = 1|\mathbf{x}_i, \mathbf{z}_i)}.$$

Regression models may be employed to feature various association structures, with the dependence of the association on covariates being explicitly reflected:

$$h^{-1}(\psi_{ijk}) = \mathbf{u}_{ijk}^{\mathrm{T}}\boldsymbol{\phi}, \tag{2}$$

where $h(.)$ is a monotone link function, $\mathbf{u}_{ijk}$ is a vector of covariates which specifies the form of the association between Y_{ij} and Y_{ik}, and $\boldsymbol{\phi}$ is a vector of regression parameters. Letting $\mathbf{u}_{ijk}$ be the scalar one, for instance, leads to the exchangeable association between responses within the same cluster; while setting $\mathbf{u}_{ijk}^{\mathrm{T}}\boldsymbol{\phi} = \phi^{|j-k|}$ results in an autoregressive correlation among responses ($j < k$). Typically, a log-linear regression may be assumed for (2) (Fitzmaurice and Laird 1993).

Let $\mu_{ijk} = P(Y_{ij} = 1, Y_{ik} = 1|\mathbf{x}_i, \mathbf{z}_i)$ be the joint probability for the pair (Y_{ij}, Y_{ik}), given the covariates $\mathbf{x}_i$ and $\mathbf{z}_i$. It is determined by the marginal means and the odds ratio, given by Lipsitz et al. (1991) and Yi and Thompson (2005)

$$\mu_{ijk} = \begin{cases} \dfrac{a_{ijk} - \{a_{ijk}^2 - 4\psi_{ijk}(\psi_{ijk} - 1)\mu_{ij}\mu_{ik}\}^{1/2}}{2(\psi_{ijk} - 1)}, & \text{if } \psi_{ijk} \neq 1 \\ \mu_{ij}\mu_{ik}, & \text{if } \psi_{i;jk} = 1 \end{cases},$$

where $a_{ijk} = 1 - (1 - \psi_{ijk})(\mu_{ij} + \mu_{ik})$.

3 Estimation procedures

In this section, we describe marginal methods for estimation of mean response parameters $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ and association parameters $\boldsymbol{\phi}$. Let $\boldsymbol{\Sigma}_i = [\sigma_{ijk}]$ be the true covariance matrix for the response vector $\mathbf{Y}_i$ for cluster i, with $\sigma_{ijj} = \mu_{ij}(1 - \mu_{ij})$ and $\sigma_{ijk} = \mu_{ijk} - \mu_{ij}\mu_{ik}$ for $j \neq k$, and $\mathbf{V}_i = \text{diag}(\sqrt{\sigma_{ijj}}, j = 1, \ldots, m_i)\mathbf{C}_i\text{diag}(\sqrt{\sigma_{ijj}}, j = 1, \ldots, m_i)$ be a working matrix, where $\mathbf{C}_i$ is an invertible working correlation matrix. Throughout the paper we assume that $\mathbf{C}_i$ may depend on a parameter vector $\boldsymbol{\tau}$ that is distinct from the mean response parameters $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$, and can be estimated by the method of moments.

Let $\mathbf{U}_{i\alpha}(\boldsymbol{\alpha}, \boldsymbol{\beta}, \theta(.)) = (\frac{\partial \boldsymbol{\mu}_i^{\mathrm{T}}}{\partial \boldsymbol{\alpha}})\mathbf{V}_i^{-1}(\mathbf{Y}_i - \boldsymbol{\mu}_i)$ and $\mathbf{U}_{i\beta}(\boldsymbol{\alpha}, \boldsymbol{\beta}, \theta(.)) = (\frac{\partial \boldsymbol{\mu}_i^{\mathrm{T}}}{\partial \boldsymbol{\beta}})\mathbf{V}_i^{-1}(\mathbf{Y}_i - \boldsymbol{\mu}_i)$. It can be seen that both $\mathbf{U}_{i\alpha}(\boldsymbol{\alpha}, \boldsymbol{\beta}, \theta(.))$ and $\mathbf{U}_{i\beta}(\boldsymbol{\alpha}, \boldsymbol{\beta}, \theta(.))$ have zero expectation, i.e., they are unbiased estimating functions for $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$.

To estimate the association parameters $\boldsymbol{\phi}$, we employ the alternating logistic regression discussed in Carey et al. (1993) where the conditional expectation $\xi_{ijk} = E(Y_{ij}|Y_{ik} = y_{ik}, \mathbf{x}_i, \mathbf{z}_i)$ is needed for $j < k$. The conditional expectation ξ_{ijk} is related to the association, marginal and joint probabilities by $\xi_{ijk} = \text{expit}(d_{ijk})$, where $d_{ijk} = (\log \psi_{ijk})y_{ik} + \log(\frac{\mu_{ij} - \mu_{ijk}}{1 - \mu_{ij} - \mu_{ik} + \mu_{ijk}})$, and $\text{expit}(t) = \exp(t)/(1 + \exp(t))$. Let $\mathbf{V}_i^* = \text{diag}\{\xi_{ijk}(1 - \xi_{ijk}), j < k\}$ be the working matrix, then $\mathbf{U}_{i\phi}(\boldsymbol{\alpha}, \boldsymbol{\beta}, \theta(.), \boldsymbol{\phi}) = (\frac{\partial \boldsymbol{\xi}_i^{\mathrm{T}}}{\partial \boldsymbol{\phi}}) \mathbf{V}_i^{*-1}\boldsymbol{\epsilon}_i$ are unbiased estimating functions for $\boldsymbol{\phi}$, where $\boldsymbol{\epsilon}_i = (y_{i1} - \xi_{i12}, \ldots, y_{i1} - \xi_{i1m_i}, y_{i2} - \xi_{i23}, \ldots, y_{i,m_i-1} - \xi_{i,m_i-1,m_i})^{\mathrm{T}}$, and $\boldsymbol{\xi}_i = (\xi_{i12}, \ldots, \xi_{i1m_i}, \xi_{i23}, \ldots, \xi_{i,m_i-1,m_i})^{\mathrm{T}}$.

If $\theta(.)$ is known to be a linear function, then estimation of $\boldsymbol{\alpha}$, $\boldsymbol{\beta}$ and $\boldsymbol{\phi}$ may proceed in a straightforward manner, as outlined in Carey et al. (1993) and Yi and Cook (2002), where the working matrix $\mathbf{V}_i$ may be taken as the true covariance matrix $\boldsymbol{\Sigma}_i$. Since the estimating functions $\mathbf{U}_{i\alpha}(\boldsymbol{\alpha}, \boldsymbol{\beta}, \theta(.))$, $\mathbf{U}_{i\beta}(\boldsymbol{\alpha}, \boldsymbol{\beta}, \theta(.))$ and $\mathbf{U}_{i\phi}(\boldsymbol{\alpha}, \boldsymbol{\beta}, \theta(.), \boldsymbol{\phi})$ involve an unknown smooth function $\theta(.)$, we need to use nonparametric approaches to estimate this function locally in order to estimate $\boldsymbol{\alpha}$, $\boldsymbol{\beta}$, and $\boldsymbol{\phi}$. Assuming $\theta(u)$ has the second derivative, we may approximate $\theta(u)$ by a locally linear function within the neighborhood of u_0 via the Taylor series expansion $\theta(u) \approx \theta(u_0) + \theta^{(1)}(u_0)(u - u_0)$ for a given point u_0, where $d^{(l)}(.)$ represents the lth derivative of function $d(.)$. Let $K(u)$ be a kernel function (or a mean zero symmetric density function) with a compact support and h be a bandwidth. Denote $K_h(t) = K(t/h)/h$, $a_0(u_0) = \theta(u_0)$, $a_1(u_0) = h\theta^{(1)}(u_0)$, and $\mathbf{a}(u_0) = (a_0(u_0), a_1(u_0))^{\mathrm{T}}$. Let $U_{ij} = \mathbf{z}_{ij}^{\mathrm{T}}\boldsymbol{\alpha}$, and $\boldsymbol{\Delta}_i = \text{diag}(\mu_{ij}^{(1)}, j = 1, \ldots, m_i)$, where $\mu_{ij}^{(1)}$ is the first derivative of the function $g(.)$ evaluated at $\mathbf{x}_{ij}^{\mathrm{T}}\boldsymbol{\beta} + \theta(\mathbf{z}_{ij}^{\mathrm{T}}\boldsymbol{\alpha})$. Define $\mathbf{G}_{ij}(u, U_{ij})$ to be an $m_i \times 2$ matrix with the lth column $\mathbf{e}_j \times \{(u - U_{ij})/h\}^{l-1}$ $(l = 1, 2)$, where $\mathbf{e}_j$ is an $m_i \times 1$ vector of 0 except with the jth entry being 1. Below we describe a two-stage algorithm for estimation of mean parameters $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ and association parameter $\boldsymbol{\phi}$.

Stage 1:

Step 1 For a given point u in a selected grid find $\widehat{\theta}(u, \widehat{\boldsymbol{\alpha}}, \widehat{\boldsymbol{\beta}}) = \widehat{a}_0(u)$ by solving

$$\sum_{i=1}^{n}\sum_{j=1}^{m_i} K_h(u - U_{ij})\mu_{ij}^{(1)}(\widehat{\boldsymbol{\alpha}}, \boldsymbol{\beta})\mathbf{G}_{ij}^{\mathrm{T}}(u, U_{ij})\mathbf{V}_i^{-1}(\mathbf{Y}_i - \boldsymbol{\mu}_{i(j)}^*) = \mathbf{0} \quad (3)$$

with respect to $\mathbf{a}(u)$, where the lth element of $\boldsymbol{\mu}_{i(j)}^*$ is $g\{\mathbf{x}_{il}^{\mathrm{T}}\boldsymbol{\beta} + I(l = j) \cdot (a_0(u) + a_1(u) \cdot (u - U_{ij})/h) + I(l \neq j)\widehat{\theta}(U_{il}, \boldsymbol{\alpha}, \boldsymbol{\beta})\}$ with $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ replaced by $\widehat{\boldsymbol{\alpha}}$ and $\widehat{\boldsymbol{\beta}}$, respectively, and $\mu_{ij}^{(1)}(\widehat{\boldsymbol{\alpha}}, \widehat{\boldsymbol{\beta}})$ is the first derivative of the function $g(.)$ evaluated at $\mathbf{x}_{ij}^{\mathrm{T}}\boldsymbol{\beta} + \{a_0(u) + a_1(u) \cdot (u - \mathbf{z}_{ij}^{\mathrm{T}}\widehat{\boldsymbol{\alpha}})/h\}$.

Step 2 Given the estimate $\widehat{\theta}(u, \widehat{\boldsymbol{\alpha}}, \widehat{\boldsymbol{\beta}}) = \widehat{a}_0(u)$ and $\widehat{a}_1(u)$ for points u in the selected grid, update $(\widehat{\boldsymbol{\alpha}}, \widehat{\boldsymbol{\beta}})$ by solving the following equations for $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$:

$$\sum_{i=1}^{n} \frac{\partial \widehat{\boldsymbol{\mu}}_i^{\mathrm{T}}(\boldsymbol{\alpha}, \boldsymbol{\beta})}{\partial \boldsymbol{\alpha}} \mathbf{V}_i^{-1}\{\mathbf{Y}_i - \widehat{\boldsymbol{\mu}}_i(\boldsymbol{\alpha}, \boldsymbol{\beta})\} = \mathbf{0}, \tag{4}$$

$$\sum_{i=1}^{n} \frac{\partial \widehat{\boldsymbol{\mu}}_i^{\mathrm{T}}(\boldsymbol{\alpha}, \boldsymbol{\beta})}{\partial \boldsymbol{\beta}} \mathbf{V}_i^{-1}\{\mathbf{Y}_i - \widehat{\boldsymbol{\mu}}_i(\boldsymbol{\alpha}, \boldsymbol{\beta})\} = \mathbf{0}, \tag{5}$$

where $\widehat{\boldsymbol{\mu}}_i(\boldsymbol{\alpha}, \boldsymbol{\beta}) = \left(\widehat{\mu}_{i1}(\boldsymbol{\alpha}, \boldsymbol{\beta}), \ldots, \widehat{\mu}_{im_i}(\boldsymbol{\alpha}, \boldsymbol{\beta})\right)^{\mathrm{T}}$ with $\widehat{\mu}_{ij}(\boldsymbol{\alpha}, \boldsymbol{\beta}) = g(\mathbf{x}_{ij}^{\mathrm{T}}\boldsymbol{\beta} + \widehat{\theta}(\mathbf{z}_{ij}^{\mathrm{T}}\boldsymbol{\alpha}, \widehat{\boldsymbol{\alpha}}, \widehat{\boldsymbol{\beta}}))$.

Step 3 Repeat steps 1 and 2 until convergence of $(\widehat{\boldsymbol{\alpha}}, \widehat{\boldsymbol{\beta}})$.

Stage 2: To estimate the association parameter $\boldsymbol{\phi}$, we solve the equations:

$$\sum_{i=1}^{n} \mathbf{U}_{i\phi}(\widehat{\boldsymbol{\alpha}}, \widehat{\boldsymbol{\beta}}, \widehat{\theta}(\mathbf{z}_{ij}^{\mathrm{T}}\widehat{\boldsymbol{\alpha}}, \widehat{\boldsymbol{\alpha}}, \widehat{\boldsymbol{\beta}}), \boldsymbol{\phi}) = \mathbf{0} \tag{6}$$

with respect to $\boldsymbol{\phi}$, where $\mathbf{U}_{i\phi}(\widehat{\boldsymbol{\alpha}}, \widehat{\boldsymbol{\beta}}, \widehat{\theta}(\mathbf{z}_{ij}^{\mathrm{T}}\widehat{\boldsymbol{\alpha}}, \widehat{\boldsymbol{\alpha}}, \widehat{\boldsymbol{\beta}}), \boldsymbol{\phi})$ is $\mathbf{U}_{i\phi}(\boldsymbol{\alpha}, \boldsymbol{\beta}, \theta(.), \boldsymbol{\phi})$ with μ_{ij} replaced by $g(\mathbf{x}_{ij}^{\mathrm{T}}\widehat{\boldsymbol{\beta}} + \widehat{\theta}(\mathbf{z}_{ij}^{\mathrm{T}}\widehat{\boldsymbol{\alpha}}, \widehat{\boldsymbol{\alpha}}, \widehat{\boldsymbol{\beta}}))$, and $\theta(.)$ replaced by $\widehat{\theta}(\mathbf{z}_{ij}^{\mathrm{T}}\widehat{\boldsymbol{\alpha}}, \widehat{\boldsymbol{\alpha}}, \widehat{\boldsymbol{\beta}})$. Denote by $\widehat{\boldsymbol{\phi}}$ the resulting estimate of $\boldsymbol{\phi}$.

To implement the algorithm, we need to choose initial values $\boldsymbol{\alpha}_0$ and $\boldsymbol{\beta}_0$ and set $\widehat{\boldsymbol{\alpha}} = \boldsymbol{\alpha}_0/\|\boldsymbol{\alpha}_0\|$ and $\widehat{\boldsymbol{\beta}} = \boldsymbol{\beta}_0$. One may for instance, take initial values as the estimates obtained from the usual generalized linear models with $\theta(.)$ specified as the identity function in (1). Given $\widehat{\boldsymbol{\alpha}}$ and $\widehat{\boldsymbol{\beta}}$, we first set $\mathbf{V}_i$ to be the independent working matrix and apply Stage 1 to obtain estimate $(\widehat{\boldsymbol{\alpha}}, \widehat{\boldsymbol{\beta}})$. Applying Stage 2 leads to an estimate of $\widehat{\boldsymbol{\phi}}$. Then we iterate Stages 1 and 2 until convergence of $(\widehat{\boldsymbol{\alpha}}, \widehat{\boldsymbol{\beta}}, \widehat{\boldsymbol{\phi}})$. Throughout iterations, the working matrix $\mathbf{V}_i$ is taken as the true covariance matrix Σ_i. A similar estimation algorithm is discussed in Yi et al. (2009). However, there is a substantial difference centering on the treatment of the working matrix $\mathbf{V}_i$. In the algorithm of Yi et al. (2009), the independent working matrix is employed throughout all iterations, which may incur some efficiency loss.

4 Theory

4.1 Asymptotic properties

Analogous to Lin and Carroll (2001a,b) and Wang (2003) we assume $m_i \equiv m$ for ease of notation. Covariates $\mathbf{x}_i$ and $\mathbf{z}_i$ are allowed to be correlated. The triples $(\mathbf{Y}_i, \mathbf{x}_i, \mathbf{z}_i)$, $i = 1, 2, \ldots, n$, are assumed independently identically distributed. Let v_i^{jl} and σ_i^{jl} be the (j,l)th element of $\mathbf{V}_i^{-1}$ and $\boldsymbol{\Sigma}_i^{-1}$, respectively. Let $\widehat{\phi}_{\boldsymbol{\alpha}}(u) = -\frac{\partial\widehat{\theta}(u,\boldsymbol{\alpha},\boldsymbol{\beta})}{\partial\boldsymbol{\alpha}^{\mathrm{T}}}$ and $\widehat{\phi}_{\boldsymbol{\beta}}(u) = -\frac{\partial\widehat{\theta}(u,\boldsymbol{\alpha},\boldsymbol{\beta})}{\partial\boldsymbol{\beta}^{\mathrm{T}}}$ with the dependence on $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ suppressed in notation, and $\phi_{\boldsymbol{\alpha}}(u)$ and $\phi_{\boldsymbol{\beta}}(u)$ be their limits defined in Appendix 2. Denote $\widetilde{\mathbf{x}}_{ij} = \mathbf{x}_{ij} - \phi_{\boldsymbol{\beta}}(U_{ij},\boldsymbol{\alpha},\boldsymbol{\beta})$, $\widetilde{\mathbf{z}}_{ij} = \theta^{(1)}(U_{ij})\mathbf{z}_{ij} - \phi_{\boldsymbol{\alpha}}(U_{ij},\boldsymbol{\alpha},\boldsymbol{\beta})$, $\widetilde{\mathbf{x}}_i = (\mathbf{x}_{i1},\ldots,\mathbf{x}_{im})^{\mathrm{T}}$, and $\widetilde{\mathbf{z}}_i = (\mathbf{z}_{i1},\ldots,\mathbf{z}_{im})^{\mathrm{T}}$. In the sequel, we drop the subject index i for ease of notation whenever expressing expectations. Let $\widetilde{\mathbf{A}}(\mathbf{V}) = E\{(\widetilde{\mathbf{z}},\widetilde{\mathbf{x}})^{\mathrm{T}}\boldsymbol{\Delta}\mathbf{V}^{-1}\boldsymbol{\Delta}(\widetilde{\mathbf{z}},\widetilde{\mathbf{x}})\}$, and $\widetilde{\mathbf{B}}(\mathbf{V},\boldsymbol{\Sigma}) = E\{(\widetilde{\mathbf{z}},\widetilde{\mathbf{x}})^{\mathrm{T}}\boldsymbol{\Delta}\mathbf{V}^{-1}\boldsymbol{\Sigma}\mathbf{V}^{-1}\boldsymbol{\Delta}(\widetilde{\mathbf{z}},\widetilde{\mathbf{x}})\}$.

Theorem 1 *Under the conditions in Appendix 1, as $n \to \infty$ and $h \to 0$ at the rate such that $nh^8 \to 0$ and $nh/\log(1/h) \to \infty$, we have*

$$\sqrt{n}\left\{(\widehat{\boldsymbol{\alpha}}-\boldsymbol{\alpha})^{\mathrm{T}}, (\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta})^{\mathrm{T}}\right\}^{\mathrm{T}} \to_d \mathrm{MVN}(\mathbf{0}, \boldsymbol{\Omega}(\mathbf{V},\boldsymbol{\Sigma})),$$

where $\boldsymbol{\Omega}(\mathbf{V},\boldsymbol{\Sigma}) = \{\widetilde{\mathbf{A}}(\mathbf{V})\}^{-1}\cdot\widetilde{\mathbf{B}}(\mathbf{V},\boldsymbol{\Sigma})\cdot\{\widetilde{\mathbf{A}}(\mathbf{V})\}^{-1}$, which is minimized by $\mathbf{V} = \boldsymbol{\Sigma}$ and in this case equals $\{\widetilde{\mathbf{A}}(\mathbf{V})\}^{-1}$.

We note that Theorem 1 does not just apply to binary data under model (1), it applies to continuous data as well. This can be readily seen from the proof in Appendix 2. With univariate data (i.e., $m = 1$), Carroll et al. (1997) established a similar asymptotic result. With $m > 1$, it is more difficult to develop the asymptotic theory as there is complexity in accommodating the working matrix $\mathbf{V}_i$ for multivariate data. Herein, we not only derive the asymptotic distribution for the estimator $\widehat{\boldsymbol{\alpha}}$ and $\widehat{\boldsymbol{\beta}}$, but also identify the scenario to obtain the semiparametric efficient estimator. Furthermore, Theorem 1 generalizes the results in Wang et al. (2005) where only a scalar covariate $\mathbf{z}_{ij}$ is considered. That is, there is no need to estimate parameter $\boldsymbol{\alpha}$ associated with unknown function $\theta(.)$. In addition to these contributions, our work distinguishes from existing methods because of the following development on estimation of association parameters $\boldsymbol{\phi}$.

Let $\mathbf{J} = E(\mathbf{U}_{i\phi}\mathbf{U}_{i\phi}^{\mathrm{T}})$ and $\mathbf{H} = E[-(\partial/\partial\boldsymbol{\phi}^{\mathrm{T}})\mathbf{U}_{i\phi}]$. If $\boldsymbol{\alpha}$, $\boldsymbol{\beta}$ and $\theta(.)$ are all known, estimating functions $\mathbf{U}_{i\phi}$ are regular parametric unbiased estimating functions of $\boldsymbol{\phi}$, and thereby it is straightforward to establish that $\sqrt{n}(\widehat{\boldsymbol{\phi}}-\boldsymbol{\phi}) \sim N(\mathbf{0}, \mathbf{H}^{-1}\mathbf{J}[\mathbf{H}^{-1}]^{\mathrm{T}})$, according to Liang and Zeger (1986). When $\boldsymbol{\alpha}$, $\boldsymbol{\beta}$ and $\theta(.)$ are unspecified and estimated, variation in the estimators $\widehat{\boldsymbol{\alpha}}$, $\widehat{\boldsymbol{\beta}}$ and $\widehat{\theta}(u)$ must be taken into account. If $\theta(.)$ is a known parametric function, one may easily adapt the arguments in Yi and Cook (2002) to work out the asymptotic distribution of $\sqrt{n}(\widehat{\boldsymbol{\phi}}-\boldsymbol{\phi})$. However, here $\theta(.)$ is unknown and it is estimated locally, we need to incorporate this local estimation variability into the asymptotic variance of $\sqrt{n}(\widehat{\boldsymbol{\phi}}-\boldsymbol{\phi})$ as well. This unknown $\theta(.)$ function presents a challenge in establishing the asymptotic distribution for the estimators, and this feature distinguishes the current work from existing results.

Along with the line of the proof of Theorem 1 in Appendix 2, we can derive an approximation of $\widehat{\theta}_h(u) - \theta(u)$ as follows

$$\frac{1}{2}b_*(u)h^2 + W_2^{-1}(u)\frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{m}\left[\mu_{ij}^{(1)}K_h(U_{ij}-u)\left\{\sum_{l=1}^{m}v_i^{jl}(Y_{il}-\mu_{il})\right\}\right.$$
$$\left. + \mu_{ij}^{(1)}Q_{1,*}(u,U_{ij})\left\{\sum_{l=1}^{m}v_i^{jl}(Y_{il}-\mu_{il})\right\} + \mu_{ij}^{(1)}v_i^{jj}Q_{2,*}(u,U_{ij})(Y_{ij}-\mu_{ij})\right],$$

where b_*, W_2, $Q_{1,*}$, and $Q_{2,*}$ are given in Appendix 2. Standard but tedious calculations can show that $\sqrt{nh}\{\widehat{\theta}_h(u) - \theta(u) - b_*(u)h^2\}$ weakly converges to a normal distribution with zero mean. This result may be used for statistical inference. However, we suggest a bootstrap alternative if needed since the asymptotic variance assumes a very complex form.

Let $\widetilde{\mathbf{A}}_i^* = \sum_{j,k} \frac{\partial d_{ijk}}{\partial \boldsymbol{\phi}}[\xi_{ijk}(1-\xi_{ijk})] \cdot (\frac{\partial d_{ijk}}{\partial \theta(U_{ij})}\widetilde{\mathbf{z}}_{ij}^{\mathrm{T}} + \frac{\partial d_{ijk}}{\partial \theta(U_{ik})}\widetilde{\mathbf{z}}_{ik}^{\mathrm{T}}, \frac{\partial d_{ijk}}{\partial \boldsymbol{\beta}^{\mathrm{T}}}, \frac{\partial d_{ijk}}{\partial \boldsymbol{\phi}^{\mathrm{T}}})$ for $i = 1, \ldots, n$. Let $\boldsymbol{\Delta}_i^* = \mathrm{diag}(\frac{\partial \boldsymbol{\mu}_i^{\mathrm{T}}}{\partial(\boldsymbol{\alpha}^{\mathrm{T}}, \boldsymbol{\beta}^{\mathrm{T}})^{\mathrm{T}}}, \frac{\partial \boldsymbol{\xi}_i^{\mathrm{T}}}{\partial \boldsymbol{\phi}})$, and $\mathbf{A}_i^* = ((\widetilde{\mathbf{A}}_i(\mathbf{V}), \mathbf{0})^{\mathrm{T}}, \widetilde{\mathbf{A}}_i^{*\mathrm{T}})^{\mathrm{T}}$. We establish the joint asymptotic distribution of the estimator $(\widehat{\boldsymbol{\alpha}}, \widehat{\boldsymbol{\beta}}, \widehat{\boldsymbol{\phi}})$ in Theorem 2, and its proof is given in Appendix 3.

Theorem 2 *Under the conditions in Theorem 1,*

$$\sqrt{n}\{(\widehat{\boldsymbol{\alpha}} - \boldsymbol{\alpha})^{\mathrm{T}}, (\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta})^{\mathrm{T}}, (\widehat{\boldsymbol{\phi}} - \widehat{\boldsymbol{\phi}})^{\mathrm{T}}\}^{\mathrm{T}} \rightarrow_d \mathrm{MVN}(\mathbf{0}, \boldsymbol{\Omega}^*),$$

where $\boldsymbol{\Omega}^* = E[\mathbf{A}^{*-1}\boldsymbol{\Delta}^*\mathrm{diag}(\mathbf{V}^{-1}, \mathbf{V}^{*-1})\boldsymbol{\Sigma}^*\mathrm{diag}(\mathbf{V}^{-1}, \mathbf{V}^{*-1})\boldsymbol{\Delta}^*\mathbf{A}^{*-1}]$, *and* $\boldsymbol{\Sigma}^*$ *is the covariance matrix of the vector* $(\mathbf{Y}_i^{\mathrm{T}}, \mathbf{Y}_i^{*\mathrm{T}})^{\mathrm{T}}$ *with* $\mathbf{Y}_i^* = \boldsymbol{\epsilon}_i + \boldsymbol{\xi}_i$.

Inferences about parameters $\boldsymbol{\alpha}$, $\boldsymbol{\beta}$ and $\boldsymbol{\phi}$ may be based on Theorem 2, where $\boldsymbol{\Omega}^*$ is replaced by a consistent estimate in which the associated terms may be substituted by the corresponding empirical estimates. However, implementation of these empirical estimates is too complicated to be of practical interest. It is more plausible to apply the simple bootstrap method for a variance estimate, and this is consistent with available research work concerning semiparametric models, such as Lin and Carroll (2001a,b), Liang et al. (2004), and Wang et al. (2005).

4.2 Bandwidth selection

In the implementation of the procedures above, choosing an appropriate bandwidth h is very crucial. As bandwidth h affects both bias and variance estimate, there is a trade-off between suitable bias and variance estimate. Bias correction requires the choice of a relatively small bandwidth, whereas variance estimate needs a large value of bandwidth. In principle, bandwidth selection is data driven, and traditional methods such as cross-validation approach may be applied to select a proper bandwidth h based on available data. However, as pointed out in Fan et al. (1995), this approach could perform poorly in some settings with a large magnitude of sample variation produced, hence it is not regarded as a sensible bandwidth selection rule for practical use. Instead, "plugging in" method may be a promising candidate for bandwidth selection. Fan et al. (1995) discussed this approach to handle local polynomial regression under the framework of generalized linear models. Ruppert et al. (1995) explored this method of bandwidth selection with local least squares regression.

Along with the same line we may derive an optimal bandwidth based on the asymptotic weighted mean integrated squared error (AMISE) of $\widehat{\theta}_h(u)$

$$\mathrm{AMISE}\{\widehat{\theta}_h\} = E\left[\int \{\widehat{\theta}_h(u) - \theta(u)\}^2 f(u)\mathrm{d}u\right].$$

An optimal bandwidth is given by $h_{\text{opt}} = C \times n^{-1/5}$, where C assumes a complex form depending on $b_*(u)$, $Q_{1,*}$, and $Q_{2,*}$. To reduce the computation burden, we specify C by an ad hoc way in our numerical experiments. See details in Sect. 5.

4.3 Interpretation of coefficients α

The proposed models allow for the flexibility of accommodating complex dependence (e.g., curvature) of the responses on covariates. As is true in general statistical modeling, this flexibility is achieved at the cost of less transparent interpretation of the model parameters. In the formulation of the usual generalized linear models, the interpretation of the model is clearly reflected by the link function and the linear form of the coefficients. However, in the current development the inclusion of a nonlinear function $\theta(.)$ could dramatically change the meaning of the model, though nonzero values of $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ may indicate "significant" predictors of the response (Carroll et al. 1997). As coefficients $\boldsymbol{\alpha}$ appear in a nonlinear function $\theta(.)$ whose form is unknown (this function may not even be monotone), interpretation of coefficients $\boldsymbol{\alpha}$ is not so transparent as that of coefficients $\boldsymbol{\beta}$. Here we particularly discuss the parameter interpretation when the link function $g^{-1}(.)$ in (1) is taken as logit.

In the usual setup of a logistic regression model with $\text{logit}(\mu_{ij}) = \mathbf{x}_{ij}^{\mathrm{T}}\boldsymbol{\beta} + \mathbf{z}_{ij}^{\mathrm{T}}\boldsymbol{\alpha}$, it is seen that the rth component of $\boldsymbol{\beta}$ is expressed as $\beta_r = \text{logit}\{\mu_{ij}(x_{ijr}+1)\} - \text{logit}\{\mu_{ij}(x_{ijr})\}$ with other covariates held fixed, where $\mu_{ij}(x_{ijr})$ stresses explicitly the dependence of μ_{ij} on x_{ijr}. That is, the coefficient β_r represents the change in the log-odds with one unit change in covariate x_{ijr}, given other covariates are fixed. Analogous interpretation applies to the coefficient α_r. Alternatively, those coefficients may be represented by means of the partial derivatives. Let $\text{odd}_{ij} = P(Y_{ij} = 1|\mathbf{x}_i, \mathbf{z}_i)/P(Y_{ij} = 0|\mathbf{x}_i, \mathbf{z}_i)$, then $\text{logit}(\mu_{ij}) = \log(\text{odd}_{ij})$, and hence $\beta_r = \frac{\partial}{\partial x_{ijr}}(\log \text{odd}_{ij})$, $\alpha_r = \frac{\partial}{\partial z_{ijr}}(\log \text{odd}_{ij})$. It is apparent that in model (1) $\log \text{odd}_{ij}$ does change linearly in x_{ijr}. However, it does not change linearly in z_{ijr} as $\frac{\partial}{\partial z_{ijr}}(\log \text{odd}_{ij})$ is generally not a constant. Interpretation of $\boldsymbol{\alpha}$ therefore is not so straightforward as that in the usual logistic regression. To understand how this parameter affects the change in the response, we may use average derivatives (Chaudhuri et al. 1997; Huang and Liu 2006) to explain parameters $\boldsymbol{\alpha}$ here. Let $\mathbf{c}_{ij} = (c_{ij1}, \ldots, c_{ijq})^{\mathrm{T}} = E[\frac{\partial}{\partial \mathbf{z}_{ij}}(\log \text{odd}_{ij})]$ be the average or expected changes in the log-odds with respect to covariates $\mathbf{z}_{ij}$ while other covariates are held fixed. It can be seen that, unlike the property of the usual logistic regression such $\mathbf{c}_{ij}$ depends on the underlying distribution of covariates.

Denote $d_{ij} = E[\theta^{(1)}(\mathbf{z}_{ij}^{\mathrm{T}}\boldsymbol{\alpha})]$, $j = 1, \ldots, m$, then $\mathbf{c}_{ij} = d_{ij}\boldsymbol{\alpha}$, i.e., $c_{ijr} = d_{ij}\alpha_r$, $r = 1, \ldots, q$, leading to $\alpha_s/\alpha_t = c_{ijs}/c_{ijt}$. Therefore, the ratio of α_s to α_t indicates the relative average change in the log-odds with respect to z_{ijs} and z_{jit} while other covariates are held fixed. If taking c_{ij1} as the baseline average change of the log-odds (with respect to covariate z_{ij1}), then $\alpha_r = \alpha_1 \cdot \frac{c_{ijr}}{c_{ij1}}$, and hence α_r may be interpreted as α_1 times of the relative average change in the log-odds (related to the change of z_{ijr}) as opposed to the baseline average change, with other covariates remained unchanged.

5 Numerical studies

We now present the results for simulation experiments and real data analysis. In the implementation of the proposed method we use the standard normal density function as the kernel function for the nonparametric procedure in Step 1 of Stage 1. In realizing Step 1 of Stage 1, for given $\widehat{\boldsymbol{\alpha}}$ and $\widehat{\boldsymbol{\beta}}$, we divide the range $[\min_{i,j}(\widehat{\boldsymbol{\alpha}}^T \mathbf{z}_{ij}), \max_{i,j}(\widehat{\boldsymbol{\alpha}}^T \mathbf{z}_{ij})]$ into 50 equally spaced sub-intervals, and take the cutting points as the grids at which $\theta(.)$ is estimated. Typically, we choose a varying bandwidth at each iteration by taking $h = C \times n^{-1/5}$, where C is the sample standard deviation of $\{\widehat{\boldsymbol{\alpha}}^T \mathbf{z}_{ij} : i = 1, \ldots, n; j = 1, \ldots, m\}$ for given $\widehat{\boldsymbol{\alpha}}$ at each iteration.

5.1 Simulation study

We conduct a simulation study to evaluate the performance of the proposed methods. Here we focus on pairwise association with higher order association being constrained as 0. That is, generate binary vector $\mathbf{y}_i = (y_{i1}, y_{i2}, \ldots, y_{im})^{\mathrm{T}}$ from the joint density function (e.g., Yi and Thompson 2005)

$$f(y_{i1}, y_{i2}, \ldots, y_{im}) = \prod_{j=1}^{m} \mu_{ij}^{y_{ij}} (1-\mu_{ij})^{1-y_{ij}} \cdot \left\{ 1 + \sum_{j<k} \rho_{ijk} \frac{y_{ij}-\mu_{ij}}{\sqrt{\sigma_{ijj}}} \cdot \frac{y_{ik}-\mu_{ik}}{\sqrt{\sigma_{ikk}}} \right\} \tag{7}$$

where ρ_{ijk} is the correlation coefficient of Y_{ij} and Y_{ik}, given by $\rho_{ijk} = (\mu_{ijk} - \mu_{ij}\mu_{ik})/\sqrt{\sigma_{ijj}\sigma_{ikk}}$. The mean responses are modeled as $\text{logit}(\mu_{ij}) = \beta x_{ij} + \theta(\alpha_1 z_{i1} + \alpha_2 z_{i2} + \alpha_3 z_{i3})$, where we take $\theta(t) = \sin[\pi(t - 1.355\sqrt{3}/6)/(1.645\sqrt{3}/3)]$ as in Carroll et al. (1997). For $i = 1, \ldots, n$, consider an exchangeable association structure with $\log \psi_{ijk} = \phi$ for $j < k$. Covariates x_{ij} are generated from the binomial distribution Bin$(1, 0.5)$ and covariates z_{ij} are generated from the uniform distribution $U[0, 1]$. Set $\beta = 0.3$ and $\alpha_1 = \alpha_2 = \alpha_3 = 1/\sqrt{3}$ as in Carroll et al. (1997). Various configurations of ψ_{ijk} are considered to reflect different strengths of association. In particular, $\psi_{ijk} = 1$ represents the scenario of independence structure within clusters. Set $m = 4$. Two hundred-fifty simulations are run for each parameter configuration. We conduct simulation on different sample sizes with $n = 100$ and $n = 200$.

Table 1 reports the differences (Bias) between the estimates and the true values, the empirical standard errors (SE) and the mean squared errors (MSE) for the regression and association parameters. The estimators for the mean and association parameters have reasonably small finite sample biases. It appears that finite sample biases for estimation of $\boldsymbol{\beta}$ and $\boldsymbol{\phi}$ tend to become smaller as the sample size increases. It is not surprising that the empirical standard errors for the estimates of the linear coefficient β are smaller than those of $\widehat{\boldsymbol{\alpha}}$. The mean squared errors for the estimators $\widehat{\boldsymbol{\alpha}}$, $\widehat{\boldsymbol{\beta}}$ and $\widehat{\phi}$ appear reasonably small, though the mean squared errors for $\widehat{\boldsymbol{\alpha}}$ tend to be larger than those of $\widehat{\boldsymbol{\beta}}$. As expected, larger sample size leads to smaller standard errors, and then smaller mean squared errors. This simulation demonstrates that the proposed methods give rise to reasonable estimates for both the mean and association parameters.

Table 1 Simulation results from 250 replications

ψ_{ijk}	Bias					SE					MSE				
	β	α_1	α_2	α_3	ϕ	β	α_1	α_2	α_3	ϕ	β	α_1	α_2	α_3	ϕ
$n = 100$															
0.5	0.026	0.015	−0.004	−0.012	0.066	0.178	0.345	0.373	0.397	0.150	0.033	0.119	0.139	0.158	0.027
1.0	0.029	−0.016	0.021	−0.006	−0.012	0.191	0.399	0.395	0.384	0.186	0.038	0.159	0.157	0.147	0.035
4.0	0.029	−0.005	0.000	0.006	0.060	0.179	0.432	0.446	0.478	0.286	0.033	0.187	0.199	0.229	0.085
6.0	0.030	−0.007	−0.02	0.026	0.093	0.177	0.442	0.442	0.417	0.309	0.032	0.196	0.196	0.175	0.104
$n = 200$															
0.5	0.038	0.006	0.002	−0.008	0.077	0.119	0.212	0.218	0.195	0.099	0.016	0.045	0.048	0.038	0.016
1.0	0.024	0.016	−0.006	−0.010	0.009	0.127	0.168	0.213	0.181	0.141	0.017	0.028	0.045	0.033	0.020
4.0	0.015	0.003	0.003	−0.006	0.039	0.120	0.170	0.162	0.171	0.186	0.014	0.029	0.026	0.029	0.036
6.0	0.014	0.003	−0.006	0.003	0.047	0.123	0.184	0.200	0.176	0.218	0.015	0.034	0.040	0.031	0.050

5.2 An example

We apply the proposed methods to analyze a subset of genetic analysis workshops (GAW) 13 data arising from Cohort 2 of the Framingham Heart Study. The Framingham Heart Study is an ongoing prospective study of risk factors for cardiovascular disease (CVD). The objective of the Framingham Heart Study was to identify common factors or characteristics that contribute to CVD by following its development over a long period of time in a large group of participants who had not yet developed overt symptoms of CVD or suffered a heart attack or stroke.

In the analysis here we consider the data set consisting of 203 families each having 4 members with the baseline measurements. High blood pressure is an important risk factor for cardiovascular disease and is a leading cause of mortality in industrialized countries. As high blood pressure is a complex disorder that results from environmental and genetic factors and their interactions, and other study indicates that blood pressure increases with age (Kraft et al. 2003). It is of interest to study how blood pressure is influenced by the risk factors and how individuals within the same family may be associated. The covariates of interest include age, gender, high density lipoprotein (HDL) and body mass index (BMI) (BMI = weight (kg)/height^2 (m^2)). Let $Y_{ij} = 1$ if subject j in family i has high blood pressure, and $Y_{ij} = 0$ otherwise.

We consider a semiparametric regression model for the mean response

$$\text{logit}(\mu_{ij}) = \beta x_{ij} + \theta(\alpha_1 z_{ij1} + \alpha_2 z_{ij2} + \alpha_3 z_{ij3}), \tag{8}$$

where x_{ij} is gender, taking value 1 for male and 0 otherwise, z_{ij1} is age, z_{ij2} is HDL, and z_{ij3} is BMI. z_{ij1}, z_{ij2} and z_{ij3} are standardized as $(z_{ijr} - \bar{z}_{..r})/s_{..r}$, where $\bar{z}_{..r}$ and $s_{..r}$ represent the sample mean and standard deviation of z'_{ijr}s, respectively, $r = 1, 2, 3$. Exchangeable association structure is modeled here with $\log \psi_{i;jk} = \phi, \quad \text{for } j \neq k$.

Here we conduct three analyses which mainly differ in the treatment of the covariance structure in estimation procedures. Analysis 1 takes $\mathbf{V}_i$ as the independence working matrix in both Steps 1 and 2; Analysis 2, following the spirit of Zeger and Diggle (1994), takes $\mathbf{V}_i$ as the independence working matrix in Step 1 but the true covariance structure $\mathbf{\Sigma}_i$ in Step 2; while Analysis 3 is the proposed method which takes $\mathbf{V}_i$ as the true covariance structure $\mathbf{\Sigma}_i$ in both Steps 1 and 2. In Table 2 we report the parameter estimates, standard errors and p-values for the three analyses. The estimates from the three analyses are fairly comparable. Except for HDL, Analysis 3 yields the smallest standard errors, and this agrees with Theorem 1. At significance level 0.05, the three analyses suggest that age plays an important role in predicting high blood pressure. As people get older, they are more prone to have higher blood pressure. There is no evidence that HDL has an effect on having high blood pressure. Strong evidence indicates that BMI has a statistically significant impact on high blood pressure. An individual with a larger BMI has a larger chance to have higher blood pressure. It is noted that there is no evidence to support an existing association among response measurements of family members.

To understand if there is a curvature relationship between the response and covariate variables, we plot $\text{logit}(\mu_{ij})$ against the single index $\theta(.)$ respectively for female and male data with Analysis 3. The patterns are displayed in Fig. 1. It is seen that there are

Table 2 Analyses of a family data set from the Framingham Heart Study

	Analysis 1			Analysis 2			Analysis 3		
Covariate	Est.	SE	P value	Est.	SE	P value	Est.	SE	P value
Gender	0.290	0.231	0.209	0.297	0.239	0.214	0.278	0.217	0.200
Age	0.558	0.272	0.040	0.557	0.261	0.033	0.543	0.251	0.031
HDL	−0.158	0.285	0.579	−0.152	0.273	0.578	−0.160	0.291	0.582
BMI	0.814	0.333	0.015	0.817	0.221	<0.001	0.824	0.189	<0.001
ϕ	0.255	0.235	0.278	0.256	0.239	0.284	0.254	0.232	0.274

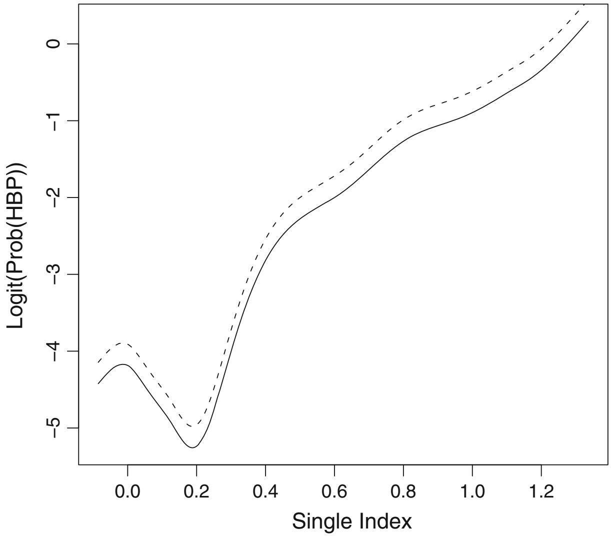

Fig. 1 Estimated nonlinear curves for the family data from the Framingham Heart Study. *Solid curve* is the estimate of logit$\{P$(blood pressure)$\}$ for females, and the *dotted curve* is the estimate of logit$\{P$(blood pressure)$\}$ for males

nonlinear trends for both data sets. Thereby using a nonlinear term in the regression (8) for logit(μ_{ij}) is perhaps more appropriate than using a linear term.

6 Incomplete clustered data

6.1 The missing data process

Let $R_{ij} = I(Y_{ij}$ is observed) and $\mathbf{R}_i = (R_{i1}, R_{i2}, \ldots, R_{im})^{\mathrm{T}}$. Let $\pi_{ij} = P(R_{ij} = 1|\mathbf{Y}_i, \mathbf{x}_i, \mathbf{z}_i)$ be the probability that subject j in cluster i is being observed, given the response and covariates vectors for cluster i. Different missing data mechanisms have been distinguished (Little and Rubin 2002) based on how missing data processes depend on the responses. Here we focus the discussion on missing at random (MAR) mechanisms, where we assume $P(\mathbf{R}_i = 1|\mathbf{Y}_i, \mathbf{x}_i, \mathbf{z}_i) = P(\mathbf{R}_i = 1|\mathbf{Y}_i^{(0)}, \mathbf{x}_i, \mathbf{z}_i)$. Here $\mathbf{Y}_i^{(0)}$ denotes the vector of the observed components of $\mathbf{Y}_i$.

Regression models are typically used to relate a function of $\mathbf{Y}_i^{(0)}$ and the covariates $\mathbf{x}_i$ and $\mathbf{z}_i$ to the probability π_{ij}. Namely, $\pi_{ij} = \eta(\mathbf{Y}_i^{(0)}, \mathbf{x}_i, \mathbf{z}_i; \boldsymbol{\delta})$, where $\eta(.)$ is a known function, and $\boldsymbol{\delta}$ is the vector of regression parameters.

We model the association among missing data indicators in the same manner as we do to the response components. That is, we define the odds ratio for subjects j and k in cluster i as

$$\psi_{ijk}^* = \frac{P(R_{ij} = 1, R_{ik} = 1|\mathbf{Y}_i, \mathbf{x}_i, \mathbf{z}_i) \cdot P(R_{ij} = 0, R_{ik} = 0|\mathbf{Y}_i, \mathbf{x}_i, \mathbf{z}_i)}{P(R_{ij} = 1, R_{ik} = 0|\mathbf{Y}_i, \mathbf{x}_i, \mathbf{z}_i) \cdot P(R_{ij} = 0, R_{ik} = 1|\mathbf{Y}_i, \mathbf{x}_i, \mathbf{z}_i)}.$$

Under MAR let $\boldsymbol{\phi}^*$ be the regression parameters linking the odds ratios to the related covariates and observed responses, $\mathbf{u}_{ijk}^*$, say.

Let $\pi_{ijk} = P(R_{ij} = 1, R_{ik} = 1|\mathbf{Y}_i, \mathbf{x}_i, \mathbf{z}_i)$ be the joint probability for R_{ij} and R_{ik}, conditional on the responses and covariates. It is given by

$$\pi_{ijk} = \begin{cases} \dfrac{a_{ijk}^* - \{a_{ijk}^{*2} - 4\psi_{ijk}^*(\psi_{ijk}^* - 1)\pi_{ij}\pi_{ik}\}^{1/2}}{2(\psi_{ijk}^* - 1)}, & \text{if } \psi_{ijk}^* \neq 1 \\ \pi_{ij}\pi_{ik}, & \text{if } \psi_{ijk}^* = 1 \end{cases},$$

where $a_{ijk}^* = 1 - (1 - \psi_{ijk}^*)(\pi_{ij} + \pi_{ik})$.

Now we describe the estimating equations for the parameters associated with the missing data process. Let $\mathbf{W}_i^* = [w_{ijk}^*]$ be the $m \times m$ matrix with $w_{ijj}^* = \pi_{ij}(1-\pi_{ij})$ and $w_{ijk}^* = \pi_{ijk} - \pi_{ij}\pi_{ik}$ for $j \neq k$. Define $\mathbf{S}_i(\boldsymbol{\delta}, \boldsymbol{\phi}^*) = (\frac{\partial \boldsymbol{\pi}_i^{\mathrm{T}}}{\partial \boldsymbol{\delta}})\mathbf{W}_i^{*-1}(\mathbf{R}_i - \boldsymbol{\pi}_i)$, then $\sum_{i=1}^n \mathbf{S}_i(\boldsymbol{\delta}, \boldsymbol{\phi}^*) = \mathbf{0}$ are unbiased estimating equations for parameters $\boldsymbol{\delta}$.

By the same spirit as in Sect. 3, the estimating functions for $\boldsymbol{\phi}^*$ can be written as $\mathbf{S}_i^*(\boldsymbol{\delta}, \boldsymbol{\phi}^*) = (\frac{\partial \boldsymbol{\xi}_i^{*\mathrm{T}}}{\partial \boldsymbol{\phi}^*})\mathbf{W}_i^{**-1}(\mathbf{R}_i - \boldsymbol{\xi}_i^*)$, where $\mathbf{W}_i^{**} = \mathrm{diag}(\xi_{ijk}^*(1 - \xi_{ijk}^*), j < k)$ is the $m^* \times m^*$ diagonal matrix with $\xi_{ijk}^* = E(R_{ij}|R_{ik} = r_{ik}; \mathbf{Y}_i, \mathbf{x}_i, \mathbf{z}_i)$ given by

$$\xi_{ijk}^* = \mathrm{expit}\left\{(\log \psi_{ijk}^*)r_{ik} + \log\left(\frac{\pi_{ij} - \pi_{ijk}}{1 - \pi_{ij} - \pi_{ik} + \pi_{ijk}}\right)\right\}, \quad \text{for } j < k,$$

and $\boldsymbol{\xi}_i^* = (\xi_{i12}^*, \ldots, \xi_{i1m}^*, \xi_{i23}^*, \ldots, \xi_{i,m-1,m}^*)^{\mathrm{T}}$, and $m^* = m(m-1)/2$.

Setting $\sum_{i=1}^n \mathbf{S}_i(\boldsymbol{\delta}, \boldsymbol{\phi}^*) = \mathbf{0}$ and $\sum_{i=1}^n \mathbf{S}_i^*(\boldsymbol{\delta}, \boldsymbol{\phi}^*) = \mathbf{0}$ leads to the estimator $\widehat{\boldsymbol{\delta}}$ and $\widehat{\boldsymbol{\phi}}$. Denote $\widehat{\pi}_{ij} = \pi_{ij}(\widehat{\boldsymbol{\delta}})$ and $\widehat{\pi}_{ijk} = \pi_{ijk}(\widehat{\boldsymbol{\delta}}, \widehat{\boldsymbol{\phi}}^*)$, and these estimates may enter the estimating functions in Sect. 6.2 below for conducting estimation of $\boldsymbol{\beta}$, $\boldsymbol{\alpha}$, and $\boldsymbol{\phi}$.

6.2 Inference for response parameters

To conduct estimation for mean parameters $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ and association parameters $\boldsymbol{\phi}$, we need to employ weighted estimating functions that accommodate missingness in the estimation procedures. Let $\mathbf{\Pi}_i = \text{diag}(I(R_{ij} = 1)/\widehat{\pi}_{ij}, j = 1, 2, \ldots, m)$ and $\mathbf{\Pi}_i^* = \text{diag}(I(R_{ij} = 1, R_{ik} = 1)/\widehat{\pi}_{ijk}, j < k)$ be the weight matrices. Then unbiased estimating functions for $\boldsymbol{\beta}$, $\boldsymbol{\alpha}$, and $\boldsymbol{\phi}$ are given by

$$\begin{cases} \mathbf{U}_{i\alpha}(\boldsymbol{\alpha}, \boldsymbol{\beta}, \theta(.)) = \left(\frac{\partial \boldsymbol{\mu}_i^{\mathrm{T}}}{\partial \boldsymbol{\alpha}}\right) \mathbf{\Pi}_i \mathbf{V}_i^{-1}(\mathbf{Y}_i - \boldsymbol{\mu}_i), \\ \mathbf{U}_{i\beta}(\boldsymbol{\alpha}, \boldsymbol{\beta}, \theta(.)) = \left(\frac{\partial \boldsymbol{\mu}_i^{\mathrm{T}}}{\partial \boldsymbol{\beta}}\right) \mathbf{\Pi}_i \mathbf{V}_i^{-1}(\mathbf{Y}_i - \boldsymbol{\mu}_i), \quad \text{and} \\ \mathbf{U}_{i\phi}(\boldsymbol{\alpha}, \boldsymbol{\beta}, \theta(.), \boldsymbol{\phi}) = \left(\frac{\partial \boldsymbol{\xi}_i^{\mathrm{T}}}{\partial \boldsymbol{\phi}}\right) \mathbf{\Pi}_i^* \mathbf{V}_i^{*-1} \boldsymbol{\epsilon}_i. \end{cases} \tag{9}$$

Estimation of $\boldsymbol{\alpha}$, $\boldsymbol{\beta}$ and $\boldsymbol{\phi}$ may proceed in the same manner as in Sect. 3 with the modifications to incorporate the weight matrices $\mathbf{\Pi}_i$ and $\mathbf{\Pi}_i^*$. That is, (3),(4), and (5) are modified as

$$\begin{cases} \sum_{i=1}^n \sum_{j=1}^{m_i} K_h(u - U_{ij}) \frac{I(R_{ij}=1)}{\pi_{ij}} \mu_{ij}^{(1)}(\widehat{\boldsymbol{\alpha}}, \boldsymbol{\beta}) \mathbf{G}_{ij}^{\mathrm{T}}(u, U_{ij}) \mathbf{V}_i^{-1}(\mathbf{Y}_i - \boldsymbol{\mu}_{i(j)}^*) = \mathbf{0}, \\ \sum_{i=1}^n \frac{\partial \widehat{\boldsymbol{\mu}}_i^{\mathrm{T}}(\boldsymbol{\alpha}, \boldsymbol{\beta})}{\partial \boldsymbol{\alpha}} \mathbf{\Pi}_i \mathbf{V}_i^{-1} \{\mathbf{y}_i - \widehat{\boldsymbol{\mu}}_i(\boldsymbol{\alpha}, \boldsymbol{\beta})\} = \mathbf{0}, \\ \sum_{i=1}^n \frac{\partial \widehat{\boldsymbol{\mu}}_i^{\mathrm{T}}(\boldsymbol{\alpha}, \boldsymbol{\beta})}{\partial \boldsymbol{\beta}} \mathbf{\Pi}_i \mathbf{V}_i^{-1} \{\mathbf{y}_i - \widehat{\boldsymbol{\mu}}_i(\boldsymbol{\alpha}, \boldsymbol{\beta})\} = \mathbf{0}, \end{cases}$$

respectively, and (6) is replaced by (9) with $\boldsymbol{\alpha}$, $\boldsymbol{\beta}$ and $\theta(\mathbf{z}_{ij}^{\mathrm{T}}\boldsymbol{\alpha})$ replaced by $\widehat{\boldsymbol{\alpha}}$, $\widehat{\boldsymbol{\beta}}$ and $\widehat{\theta}(\mathbf{z}_{ij}^{\mathrm{T}}\widehat{\boldsymbol{\alpha}}, \widehat{\boldsymbol{\alpha}}, \widehat{\boldsymbol{\beta}})$, respectively. Here $\widehat{\boldsymbol{\alpha}}$, $\widehat{\boldsymbol{\beta}}$ and $\widehat{\boldsymbol{\phi}}$ denote the resultant estimators. Analogous to the proof of Theorem 2, we can derive the following asymptotic result.

Theorem 3 *Under the conditions of Theorem 1,* $\sqrt{n}\left\{(\widehat{\boldsymbol{\alpha}} - \boldsymbol{\alpha})^{\mathrm{T}}, (\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta})^{\mathrm{T}}, (\widehat{\boldsymbol{\phi}} - \boldsymbol{\phi})^{\mathrm{T}}\right\}^{\mathrm{T}}$ *has a multivariate normal distribution with mean zero and a sandwich covariance matrix.*

7 Discussion

In this paper we develop semiparametric approaches to analyze clustered data. Interest here lies in estimation of the association coefficients in addition to the marginal mean parameters. The simulation studies demonstrate that the proposed methods work well under various situations. In the current development we focus on modeling the mean response with semiparametric regression but the association structure with parametric regression. More generally, we may use a semiparametric specification to model the association structure

$$h^{-1}(\psi_{ijk}) = \mathbf{u}_{ijk}^{\mathrm{T}}\boldsymbol{\phi} + \eta^*(\mathbf{w}_{ijk}^{\mathrm{T}}\boldsymbol{\delta}^*) \tag{10}$$

with an unknown smooth function $\eta^*(.)$. The extension to including (10) is straightforward by adapting the arguments in the paper, though a more complicated presentation is needed.

Our current work generalizes Carroll et al. (1997) from univariate case to multivariate data, and it also extends Wang et al. (2005) from a scalar covariate $\mathbf{z}$ to multiple covariates. Furthermore, the proposed methods incorporate estimation of association parameters, and apply to incomplete data. These features make the proposed methods attractive as they offer a useful tool to handle problems that are not addressed by existing research. The methods we describe here have applications in a wide variety of settings. For example, they can also be generalized to accommodating data with more complex association structures. In many situations, clustered data may arise from longitudinal studies. Clustered longitudinal data feature both a cross-sectional and a longitudinal correlation structure and interest often lies in the strengths of both types of association (Yi and Cook 2002). The proposed methods may be adapted to handle longitudinal data arising in clusters.

Appendix 1: Conditions

Without exception detailed technical conditions are needed here to guarantee rigorous proofs. Below we just outline several key assumptions with the detailed list of conditions omitted. For more details see Carroll et al. (1997) and Wang et al. (2005).

(a) The density function of $\mathbf{z}_{ij}$ has a continuous second derivative on its support.
(b) The density function of $\mathbf{z}_{ij}^{\mathrm{T}}\boldsymbol{\alpha}$ is positive and uniformly continuous for $\boldsymbol{\alpha}$ in a neighborhood of its true value.
(c) $\theta^{(2)}(u)$ is continuous on its support.
(d) The random vector $\mathbf{x}_{ij}$ is assumed to have a bounded support.
(e) $K(\cdot)$ is a symmetric probability density function with bounded support.

In the following development the identities are valid to the order of $o_p(a_n)$, where $a_n = h^2 + \{\log(n)/nh\}^{1/2} + n^{-1/2}$.

Appendix 2: Proof of Theorem 1

We first introduce the following notation, which is similar to that in Wang et al. (2005). The major change is to replace T_{ij} in Wang et al. (2005) with $U_{ij} = \mathbf{z}_{ij}^{\mathrm{T}}\boldsymbol{\alpha}$. To be specific, let $f_j(u)$ be the marginal density function for U_{ij}, and $f_{jl}(u, v)$ be the joint density function for U_{ij} and U_{il} $(j \neq l)$.

Define

$$W_2(u) = -\sum_{j=1}^{m} E\{\Delta_j^2 v^{jj} | U_j = u\} f_j(u),$$

$$Q(u, v) = \sum_{j=1}^{m} \sum_{l \neq j} E[\Delta_j v^{jl} \Delta_l \{W_2(U_l)\}^{-1} | U_j = u, U_l = v] f_{jl}(u, v),$$

$$\mathcal{A}(B; u, v) = -\sum_{j=1}^{m} \sum_{j \neq l} E[\Delta_j v^{jl} \Delta_l \{W_2(U_l)\}^{-1} B(U_l, v) | U_j = u] f_j(u),$$

and $\phi_{\boldsymbol{\beta}}(u)$ is the solution of

$$\sum_{j=1}^{m}\sum_{l=1}^{m} E\left[\Delta_j \sigma^{jl} \Delta_l \{\mathbf{x}_l - \phi_{\boldsymbol{\beta}}(U_l)\} | U_j = u\right] f_j(u) = 0,$$

where σ^{jl} is the (j, l)th element of $\boldsymbol{\Sigma}^{-1}$ and Δ_j is the (j, j)th element of the diagonal matrix $\boldsymbol{\Delta}$. Further, let

$$b_{[k]}(u) = \theta^{(2)}(u) - W_2^{-1}(u) \sum_{j=1}^{m}\sum_{l\neq j} E\{\Delta_j v^{jl} \Delta_l b_{[k-1]}(U_l)|U_j = u\} f_j(u),$$

$$\begin{aligned}
\widehat{\theta}_{[1]}(u) - \theta(u) &= \frac{1}{2} b_{[1]}(u) h^2 + W_2^{-1}(u) \frac{1}{n} \sum_{i=1}^{n}\sum_{j=1}^{m} \mu_{ij}^{(1)} K_h(U_{ij} - u) \\
&\quad \times \left\{ \sum_{l=1}^{m} v_i^{jl} (Y_{il} - \mu_{il}) \right\} \\
&\quad - W_2^{-1}(u) \frac{1}{n} \sum_{i=1}^{n}\sum_{j=1}^{m} \mu_{ij}^{(1)} v_i^{jj} Q(u, U_{ij})(Y_{ij} - \mu_{ij}),
\end{aligned}$$

$$\begin{aligned}
\widehat{\theta}_{[k]}(u) - \theta(u) &= \frac{1}{2} b_{[k]}(u) h^2 + W_2^{-1}(u) \frac{1}{n} \sum_{i=1}^{n}\sum_{j=1}^{m} \mu_{ij}^{(1)} K_h(U_{ij} - u) \\
&\quad \times \left\{ \sum_{l=1}^{m} v_i^{jl} (Y_{il} - \mu_{il}) \right\} \\
&\quad + W_2^{-1}(u) \frac{1}{n} \sum_{i=1}^{n}\sum_{j=1}^{m} \mu_{ij}^{(1)} Q_{1,[k]}(u, U_{ij}) \left\{ \sum_{l=1}^{m} v_i^{jl} (Y_{il} - \mu_{il}) \right\} \\
&\quad + W_2^{-1}(u) \frac{1}{n} \sum_{i=1}^{n}\sum_{j=1}^{m} \mu_{ij}^{(1)} v_i^{jj} Q_{2,[k]}(u, U_{ij})(Y_{ij} - \mu_{ij}), \qquad (11)
\end{aligned}$$

where $Q_{1,[1]}(u, v) = 0$, $Q_{2,[1]}(u, v) = -Q(u, v)$, $Q_{1,[k]}(u, v) = -Q(u, v) + \mathcal{A}(Q_{1,[k-1]}; u, v)$, and $Q_{2,[k]}(u, v) = \mathcal{A}(Q_{2,[k-1]}; u, v)$. As $k \to \infty$, at convergence, $\widehat{\theta}(u) - \theta(u)$ shares the same asymptotic structure as in (11) except that $b_{[k]}$, $Q_{1,[k]}$, and $Q_{2,[k]}$ are replaced by b_*, $Q_{1,*}$, and $Q_{2,*}$, where $b_*(u) = \theta^{(2)}(u) - W_2^{-1}(u) \sum_{j=1}^{m} \sum_{l\neq j} E\{\Delta_j v^{jl} \Delta_l b_*(U_l)|U_j = u\} f_j(u)$, $Q_{1,*}(u, v) = -Q(u, v) + \mathcal{A}(Q_{1,*}; u, v)$, and $Q_{2,*}(u, v) = \mathcal{A}(Q_{2,*}; u, v)$.

The proof of Theorem 1 can be completed following the spirit of Appendices A.3 and A.4 in Wang et al. (2005). However, great complexity is present in the current development as we have to deal with estimation of an additional parameter vector $\boldsymbol{\alpha}$. We now prove Theorem 1 with two steps. In the first step, we express the derivative of the estimator $\widehat{\theta}(u, \boldsymbol{\alpha}, \boldsymbol{\beta})$ with respect to $\boldsymbol{\alpha}$, and in the second step, we establish the asymptotic distribution of the estimator $(\widehat{\boldsymbol{\alpha}}^{\mathrm{T}}, \widehat{\boldsymbol{\beta}}^{\mathrm{T}})^{\mathrm{T}}$. We note that it is the first step that distinguishes the current development from that in Wang et al. (2005).

First we work on the derivative of $\widehat{\theta}(u, \boldsymbol{\alpha}, \boldsymbol{\beta})$ with respect to $\boldsymbol{\alpha}$. Expressing the first component of (3) in terms of any $(\boldsymbol{\alpha}, \boldsymbol{\beta})$, we obtain

$$
\begin{aligned}
0 = \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{m} K_h(u-U_{ij})\mu_{ij}^{(1)}(v_i^{jj}[Y_{ij} - g\{\mathbf{x}_{ij}^{\mathrm{T}}\boldsymbol{\beta} + \widehat{\theta}(u,\boldsymbol{\alpha},\boldsymbol{\beta})\\
+\widehat{a}_1(u,\boldsymbol{\alpha},\boldsymbol{\beta})\cdot(u-U_{ij})/h\}]\\
+\sum_{l\neq j} v_i^{jl}[Y_{il} - g\{\mathbf{x}_{il}^{\mathrm{T}}\boldsymbol{\beta} + \widehat{\theta}(U_{il},\boldsymbol{\alpha},\boldsymbol{\beta})\}]).
\end{aligned}
$$

Differentiating with respect to $\boldsymbol{\alpha}$ yields

$$
\begin{aligned}
0 &= \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{m}\left\{K_h^{(1)}(u-U_{ij})(\mathbf{z}_{ij}/h)\mu_{ij}^{(1)} + K_h(u-U_{ij})\mu_{ij}^{(2)}\right.\\
&\quad\left.\times\left(\widehat{\phi}_{\boldsymbol{\alpha}}(u) + \frac{\partial\widehat{a}_1(u,\boldsymbol{\alpha},\boldsymbol{\beta})}{\partial\boldsymbol{\alpha}}\cdot(u-U_{ij})/h - \widehat{\theta}^{(1)}(u,\boldsymbol{\alpha},\boldsymbol{\beta})\cdot\mathbf{Z}_{ij}^{\mathrm{T}}\right)\right\}\\
&\quad\times\left\{v_i^{jj}[Y_{ij} - g\{\mathbf{x}_{ij}^{\mathrm{T}}\boldsymbol{\beta} + \widehat{\theta}(u,\boldsymbol{\alpha},\boldsymbol{\beta}) + \widehat{a}_1(u,\boldsymbol{\alpha},\boldsymbol{\beta})\cdot(u-U_{ij})/h\}]\right.\\
&\quad\left.+\sum_{l\neq j} v_i^{jl}[Y_{il} - g\{\mathbf{x}_{il}^{\mathrm{T}}\boldsymbol{\beta} + \widehat{\theta}(U_{il},\boldsymbol{\alpha},\boldsymbol{\beta})\}]\right\} + \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{m} K_h(u-U_{ij})\\
&\quad\times\left[-v_i^{jj}\Delta_{i,j}^2\left(\widehat{\phi}_{\boldsymbol{\alpha}}(u) + \frac{\partial\widehat{a}_1(u,\boldsymbol{\alpha},\boldsymbol{\beta})}{\partial\boldsymbol{\alpha}}\cdot(u-U_{ij})/h \; - \widehat{\theta}^{(1)}(u,\boldsymbol{\alpha},\boldsymbol{\beta})\cdot\mathbf{z}_{ij}^{\mathrm{T}}\right)\right.\\
&\quad\left.-\sum_{l\neq j}\Delta_{i,j} v_i^{jl}\Delta_{i,l}\cdot\widehat{\phi}_{\boldsymbol{\alpha}}(U_{il})\mathbf{z}_{il}\right].
\end{aligned}
\tag{12}
$$

Using the arguments similar to the proof of Theorem 5.1 in Ichimura (1993), we can show that the first, third and fourth terms in the first summand is $o_p(1)$, and the second and third terms in the second summation is of order $o_p(1)$. On the other hand,

$$\frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{m}K_h(u-U_{ij})\cdot(v_i^{jj}\Delta_{i,j}^2-\mu_{ij}^{(2)})\widehat{\phi}_{\boldsymbol{\alpha}}(u)$$
$$=\sum_{j=1}^{m}E(v^{jj}\Delta_j^2-\mu_{ij}^{(2)}|U_j=u)f_j(u)\widehat{\phi}_{\boldsymbol{\alpha}}(u)\{1+o_p(1)\},$$

and

$$\frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{m}K_h(u-U_{ij})\sum_{l\neq j}\Delta_{i,j}v_i^{jl}\Delta_{i,l}\widehat{\phi}_{\boldsymbol{\alpha}}(U_{il})\mathbf{z}_{il}^{\mathrm{T}}$$
$$=\sum_{j=1}^{m}\sum_{l\neq j}\int E(\Delta_j v^{jl}\Delta_l\mathbf{z}_l^{\mathrm{T}}|U_j=u)\widehat{\phi}_{\boldsymbol{\alpha}}(u_l)f_{lj}(u_l,u)du_l\{1+o_p(1)\}.$$

It follows that as $n\to\infty$, $\widehat{\phi}_{\boldsymbol{\alpha}}(u)\to\phi_{\boldsymbol{\alpha}}(u)$ uniformly on u, where $\phi_{\boldsymbol{\alpha}}(u)$ is the solution of the limit form of (12). That is,

$$\phi_{\boldsymbol{\alpha}}(u)=\left\{\sum_{j=1}^{m}E(v^{jj}\Delta_j^2-\mu_j^{(2)}|U_j=u)f_j(u)\right\}^{-1}$$
$$\times\left[\sum_{j=1}^{m}\sum_{l\neq j}\int E(\Delta_j v^{ll}\Delta_l\mathbf{z}_l|U_j=u)\phi_{\boldsymbol{\alpha}}(u_l)f_{lj}(u_l,u)du_l\right]$$
$$\times\{1+o_p(1)\}.$$

Recall (4) and (5):

$$\sum_{i=1}^{n}\frac{\partial\widehat{\boldsymbol{\mu}}_i^{\mathrm{T}}(\boldsymbol{\alpha},\boldsymbol{\beta})}{\partial(\boldsymbol{\alpha}^{\mathrm{T}},\boldsymbol{\beta}^{\mathrm{T}})^{\mathrm{T}}}\mathbf{V}_i^{-1}[Y_i-\widehat{\boldsymbol{\mu}}_i(\boldsymbol{\alpha},\boldsymbol{\beta})]=\mathbf{0}, \tag{13}$$

where $\widehat{\boldsymbol{\mu}}_i(\boldsymbol{\alpha},\boldsymbol{\beta})=(\widehat{\mu}_{i1}(\boldsymbol{\alpha},\boldsymbol{\beta}),\ldots,\widehat{\mu}_{im}(\boldsymbol{\alpha},\boldsymbol{\beta}))^{\mathrm{T}}$ with the jth element $g(\mathbf{x}_{ij}^{\mathrm{T}}\boldsymbol{\beta}+\widehat{\theta}(U_{ij},\widehat{\boldsymbol{\alpha}},\widehat{\boldsymbol{\beta}}))$. To ease the notation, let $\mathrm{vec}_{\alpha,\beta}=\{(\widehat{\boldsymbol{\alpha}}-\boldsymbol{\alpha})^{\mathrm{T}},(\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta})^{\mathrm{T}}\}^{\mathrm{T}}$. Note that elementwisely, we have the Taylor series expansion for $\mu_{ij}(\boldsymbol{\beta},\boldsymbol{\alpha})-\widehat{\mu}_{ij}(\boldsymbol{\alpha},\boldsymbol{\beta})$ as follows.

$$-\mu_{ij}^{(1)}\cdot\{\mathbf{x}_{ij}^{\mathrm{T}}(\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta})+\widehat{\theta}(U_{ij},\widehat{\boldsymbol{\alpha}},\widehat{\boldsymbol{\beta}})-\widehat{\theta}(U_{ij},\boldsymbol{\alpha},\boldsymbol{\beta})+\widehat{\theta}(U_{ij},\boldsymbol{\alpha},\boldsymbol{\beta})-\theta(\mathbf{z}_{ij}^{\mathrm{T}}\boldsymbol{\alpha})\}$$
$$=-\mu_{ij}^{(1)}\cdot\left(\widetilde{\mathbf{z}}_{ij}^{\mathrm{T}},\widetilde{\mathbf{x}}_{ij}^{\mathrm{T}}\right)\mathrm{vec}_{\alpha,\beta}-\mu_{ij}^{(1)}\cdot\{\widehat{\theta}(U_{ij},\boldsymbol{\alpha},\boldsymbol{\beta})-\theta(\mathbf{z}_{ij}^{\mathrm{T}}\boldsymbol{\alpha})\}+o_p(1).$$

Putting this in a matrix form, we obtain

$$\boldsymbol{\mu}_i(\boldsymbol{\beta},\boldsymbol{\alpha})-\widehat{\boldsymbol{\mu}}_i(\boldsymbol{\alpha},\boldsymbol{\beta})=\left\{\boldsymbol{\mu}_i^{(1)}*(\widetilde{\mathbf{z}}_i,\widetilde{\mathbf{x}}_i)\right\}^{\mathrm{T}}\cdot\mathrm{vec}_{\alpha,\beta}+\boldsymbol{\mu}_i^{(1)}*\{\widehat{\boldsymbol{\theta}}(\mathbf{U}_i,\boldsymbol{\alpha},\boldsymbol{\beta})-\boldsymbol{\theta}(\mathbf{U}_i)\},$$

where $\boldsymbol{\mu}_i^{(1)} = (\mu_{i1}^{(1)}, \ldots, \mu_{im}^{(1)})^{\mathrm{T}}$, $\widehat{\boldsymbol{\theta}}(\mathbf{U}_i, \boldsymbol{\alpha}, \boldsymbol{\beta})$ and $\boldsymbol{\theta}(\mathbf{U}_i)$ are stacked vectors with the jth element being $\widehat{\theta}(U_{ij}, \boldsymbol{\alpha}, \boldsymbol{\beta})$ and $\theta(U_{ij})$, respectively, and $\mathbf{b}*\mathbf{c} = (b_1c_1, \ldots, b_rc_r)^{\mathrm{T}}$ denotes the elementwise product of vectors $\mathbf{b} = (b_1, \ldots, b_r)^{\mathrm{T}}$ and $\mathbf{c} = (c_1, \ldots, c_r)^{\mathrm{T}}$.

Note that

$$\frac{\partial g\{\mathbf{x}_{ij}^{\mathrm{T}}\boldsymbol{\beta} + \widehat{\theta}(U_{ij}, \boldsymbol{\alpha}, \boldsymbol{\beta})\}^{\mathrm{T}}}{\partial \boldsymbol{\alpha}} \approx \mu_{ij}^{(1)} \cdot \widetilde{\mathbf{z}}_{ij},$$
$$\frac{\partial g\{\mathbf{x}_{ij}^{\mathrm{T}}\boldsymbol{\beta} + \widehat{\theta}(U_{ij}, \boldsymbol{\alpha}, \boldsymbol{\beta})\}^{\mathrm{T}}}{\partial \boldsymbol{\beta}} \approx \mu_{ij}^{(1)} \cdot \widetilde{\mathbf{x}}_{ij}.$$

After some tedious calculation, (13) can be simplified as

$$\begin{aligned}\frac{1}{n}\sum_{i=1}^{n}\widetilde{\mathbf{A}}_i(\mathbf{V}) \cdot \sqrt{n} \cdot \mathrm{vec}_{\alpha,\beta} = &-\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left(\frac{\partial \widehat{\boldsymbol{\mu}}_i^{\mathrm{T}}}{\partial(\boldsymbol{\alpha}^{\mathrm{T}}, \boldsymbol{\beta}^{\mathrm{T}})^{\mathrm{T}}}\right)\mathbf{V}_i^{-1}(\mathbf{Y}_i - \boldsymbol{\mu}_i)\\ &-\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\widetilde{\mathbf{B}}_i + o_p(1),\end{aligned}$$

where $\widetilde{\mathbf{A}}_i(\mathbf{V}) = \left\{\boldsymbol{\mu}_i^{(1)} * (\widetilde{\mathbf{z}}_i, \widetilde{\mathbf{x}}_i)\right\}^{\mathrm{T}} \mathbf{V}_i^{-1} \left\{\boldsymbol{\mu}_i^{(1)} * (\widetilde{\mathbf{z}}_i, \widetilde{\mathbf{x}}_i)\right\} \sqrt[n]{}$ and $\widetilde{\mathbf{B}}_i = \left\{\boldsymbol{\mu}_i^{(1)} * (\widetilde{\mathbf{z}}_i, \widetilde{\mathbf{x}}_i)\right\}^{\mathrm{T}}$ $\mathbf{V}_i^{-1}\left[\boldsymbol{\mu}_i^{(1)} * \{\widehat{\boldsymbol{\theta}}(\mathbf{U}_i, \boldsymbol{\alpha}, \boldsymbol{\beta}) - \boldsymbol{\theta}(\mathbf{U}_i)\}\right]$. Let $\widetilde{\mathbf{A}}(\mathbf{V}) = E[\widetilde{\mathbf{A}}_i(\mathbf{V})] = E\{(\widetilde{\mathbf{z}}, \widetilde{\mathbf{x}})^{\mathrm{T}}\boldsymbol{\Delta}\mathbf{V}^{-1}\boldsymbol{\Delta}(\widetilde{\mathbf{z}}, \widetilde{\mathbf{x}})\}$, then

$$n^{1/2} \cdot \mathrm{vec}_{\alpha,\beta} = -\widetilde{\mathbf{A}}^{-1}(\mathbf{V})(\mathbf{C}_n + \mathbf{B}_n) + o_p(1), \tag{14}$$

where

$$\begin{aligned}\mathbf{C}_n &= \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left(\frac{\partial \widehat{\boldsymbol{\mu}}_i^{\mathrm{T}}}{\partial(\boldsymbol{\alpha}^{\mathrm{T}}, \boldsymbol{\beta}^{\mathrm{T}})^{\mathrm{T}}}\right)\mathbf{V}_i^{-1}(\mathbf{Y}_i - \boldsymbol{\mu}_i)\\ &= n^{-1/2}\sum_{i=1}^{n}(\widetilde{\mathbf{z}}_i, \widetilde{\mathbf{x}}_i)^{\mathrm{T}}\Delta_i\mathbf{V}_i^{-1}(\mathbf{Y}_i - \boldsymbol{\mu}_i),\end{aligned}$$

and

$$\begin{aligned}\mathbf{B}_n &= n^{-1/2}\sum_{i=1}^{n}\widetilde{\mathbf{B}}_i\\ &= n^{-1/2}\sum_{i=1}^{n}\sum_{j=1}^{m}\sum_{l=1}^{m}(\widetilde{\mathbf{z}}_{ij}^{\mathrm{T}}, \widetilde{\mathbf{x}}_{ij}^{\mathrm{T}})^{\mathrm{T}}\mu_{ij}^{(1)}v_i^{jl}\mu_{il}^{(1)}\\ &\quad\times\left[\frac{1}{2}h^2\{b_*(U_{il}) + hb_{*1}(U_{il}) + O_p(h^2)\}\right.\end{aligned}$$

$$+\left\{W_2^{-1}(U_{il})\frac{1}{n}\sum_{i'=1}^{n}\sum_{j'=1}^{n}\mu_{i'j'}^{(1)}\left[K_h(U_{i'j'}-U_{il})\right.\right.$$
$$\sum_l v_{i'}^{j'l}(Y_{i'l}-\mu_{i'l})+v_{i'}^{j'j'}Q_{2,*}(U_{il},U_{i'j'})(Y_{i'j'}-\mu_{i'j'})$$
$$\left.\left.+Q_{1,*}(U_{il},U_{i'j'})\sum_l v_{i'}^{j'l}(Y_{i'l}-\mu_{i'l})\right]\right\}\Bigg]\{1+o_p(1)\}.$$

Let

$$\mathbf{B}_{1n}=n^{-1/2}\sum_{i=1}^{n}\sum_{j=1}^{m}\sum_{l=1}^{m}(\widetilde{\mathbf{x}}_{ij}^{\mathrm{T}},\widetilde{\mathbf{z}}_{ij}^{\mathrm{T}})^{\mathrm{T}}\mu_{ij}^{(1)}v_i^{jl}\mu_{il}^{(1)}\frac{1}{2}h^2(b_*(U_{il})+hb_{*1}(U_{il})+O_p(h^2)),$$

and

$$\mathbf{B}_{2n}=n^{-1/2}\sum_{i=1}^{n}\sum_{j=1}^{m}\sum_{l=1}^{m}(\widetilde{\mathbf{x}}_{ij}^{\mathrm{T}},\widetilde{\mathbf{z}}_{ij}^{\mathrm{T}})^{\mathrm{T}}\mu_{ij}^{(1)}v_i^{jl}\mu_{il}^{(1)}\cdot\left\{W_2^{-1}(U_{il})\frac{1}{n}\sum_{i'=1}^{n}\sum_{j'=1}^{n}\mu_{i'j'}^{(1)}\right.$$
$$\cdot\left[K_h(U_{i'j'}-U_{il})\sum_l v_{i'}^{j'l}(Y_{i'l}-\mu_{i'l})+v_{i'}^{j'j'}Q_{2,*}(U_{il},U_{i'j'})(Y_{i'j'}-\mu_{i'j'})\right.$$
$$\left.\left.+Q_{1,*}(U_{il},U_{i'j'})\sum_l v_{i'}^{j'l}(Y_{i'l}-\mu_{i'l})\right]\right\}.$$

Adapting the arguments respectively concerning B_n and C_{2n} in Wang et al. (2005), we can show that $\mathbf{B}_{1n}=o_p(1)$ and $\mathbf{B}_{2n}=o_p(1)$. Therefore, using the central limit theorem to (14), we show Theorem 1.

Appendix 3: Proof of Theorem 2

Let $\widehat{\xi}_{ijk}$ be ξ_{ijk} with $\theta(U_{il})$ and $\boldsymbol{\beta}$ being replaced by $\widehat{\theta}(\mathbf{z}_{il}^{\mathrm{T}}\widehat{\boldsymbol{\alpha}},\widehat{\boldsymbol{\alpha}},\widehat{\boldsymbol{\beta}})$ and $\widehat{\boldsymbol{\beta}}$, respectively ($l=j,k$). Note that d_{ijk} is a function of $\theta(U_{ij})$, $\theta(U_{ik})$, $\boldsymbol{\beta}$ and $\boldsymbol{\phi}$. Applying a Taylor series expansion, we express $\xi_{ijk}-\widehat{\xi}_{ijk}$ as

$$-\xi_{ijk}^{(1)}\left\{\frac{\partial d_{ijk}}{\partial\boldsymbol{\beta}^{\mathrm{T}}}(\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta})+\frac{\partial d_{ijk}}{\partial\boldsymbol{\phi}^{\mathrm{T}}}(\widehat{\boldsymbol{\phi}}-\boldsymbol{\phi})\right.$$
$$\left.+\frac{\partial d_{ijk}}{\partial\theta(U_{ij})}(\widehat{\theta}(\mathbf{z}_{ij}^{\mathrm{T}}\widehat{\boldsymbol{\alpha}},\widehat{\boldsymbol{\alpha}},\widehat{\boldsymbol{\beta}})-\theta(\mathbf{Z}_{ij}^{\mathrm{T}}\boldsymbol{\alpha}))+\frac{\partial d_{ijk}}{\partial\theta(U_{ik})}(\widehat{\theta}(\mathbf{z}_{ik}^{\mathrm{T}}\widehat{\boldsymbol{\alpha}},\widehat{\boldsymbol{\alpha}},\widehat{\boldsymbol{\beta}})-\theta(\mathbf{Z}_{ik}^{\mathrm{T}}\boldsymbol{\alpha}))\right\},$$

which can further be expressed as

$$-\xi_{ijk}^{(1)}\left(\frac{\partial d_{ijk}}{\partial\theta(U_{ij})}\widetilde{\mathbf{Z}}_{ij}^{\mathrm{T}}+\frac{\partial d_{ijk}}{\partial\theta(U_{ik})}\widetilde{\mathbf{Z}}_{ik}^{\mathrm{T}},\ \frac{\partial d_{ijk}}{\partial\boldsymbol{\beta}^{\mathrm{T}}},\ \frac{\partial d_{ijk}}{\partial\boldsymbol{\phi}^{\mathrm{T}}}\right)\cdot\mathrm{vec}_{\alpha,\beta,\phi}$$
$$-\xi_{ijk}^{(1)}\left\{\frac{\partial d_{ijk}}{\partial\theta(U_{ij})}(\widehat{\theta}(U_{ij},\boldsymbol{\alpha},\boldsymbol{\beta})-\theta(U_{ij}))+\frac{\partial d_{ijk}}{\partial\theta(U_{ik})}(\widehat{\theta}(U_{ik},\boldsymbol{\alpha},\boldsymbol{\beta})-\theta(U_{ik}))\right\},$$

where $\mathrm{vec}_{\alpha,\beta,\phi}=\left\{(\widehat{\boldsymbol{\alpha}}-\boldsymbol{\alpha})^{\mathrm{T}},(\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta})^{\mathrm{T}},(\widehat{\boldsymbol{\phi}}-\boldsymbol{\phi})^{\mathrm{T}}\right\}^{\mathrm{T}}$. After some algebra, we obtain $\frac{\partial\widehat{\boldsymbol{\xi}}_i^{\mathrm{T}}}{\partial\boldsymbol{\phi}}\mathbf{V}_i^{*-1}(\boldsymbol{\xi}_i-\widehat{\boldsymbol{\xi}}_i)=\widetilde{\mathbf{A}}_i^*\cdot\mathrm{vec}_{\alpha,\beta,\phi}+\widetilde{\mathbf{B}}_i^*$, where $\widetilde{\mathbf{B}}_i^*$ is the vector with the rth element

$$\widetilde{B}_{ir}^*=\sum_{j,k}\xi_{ijk}^{(1)}\frac{\partial d_{ijk}}{\partial\phi_r}[\xi_{ijk}(1-\xi_{ijk})]^{-1}\xi_{ijk}^{(1)}\left\{\frac{\partial d_{ijk}}{\partial\theta(U_{ij})}(\widehat{\theta}(U_{ij},\boldsymbol{\alpha},\boldsymbol{\beta})-\theta(U_{ij}))\right.$$
$$\left.+\frac{\partial d_{ijk}}{\partial\theta(U_{ik})}(\widehat{\theta}(U_{ik},\boldsymbol{\beta})-\theta(U_{ik}))\right\}.$$

Analogously to the arguments in Appendix 2 for $\mathbf{B}_n$, we can show that $B_{ir}^*=o_p(1)$. Now it remains to show the asymptotic distribution. Working on the estimating Eqs. (4)–(6)

$$\mathbf{0}=\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\begin{pmatrix}\frac{\partial\widehat{\boldsymbol{\mu}}_i^{\mathrm{T}}}{\partial(\boldsymbol{\alpha}^{\mathrm{T}},\boldsymbol{\beta}^{\mathrm{T}})^{\mathrm{T}}}\mathbf{V}_i^{-1}(\mathbf{Y}_i-\widehat{\boldsymbol{\mu}}_i)\\ \frac{\partial\widehat{\boldsymbol{\xi}}_i^{\mathrm{T}}}{\partial\boldsymbol{\phi}}\mathbf{V}_i^{*-1}(\mathbf{Y}_i^*-\widehat{\boldsymbol{\xi}}_i)\end{pmatrix},$$

we obtain, after some algebra

$$\frac{1}{n}\sum_{i=1}^{n}\mathbf{A}_i^*\cdot\sqrt{n}\mathrm{vec}_{\alpha,\beta,\phi}=\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\boldsymbol{\Delta}_i^*\mathrm{diag}(\mathbf{V}_i^{-1},\mathbf{V}_i^{*-1})\begin{pmatrix}\mathbf{Y}_i-\boldsymbol{\mu}_i\\ \mathbf{Y}_i^*-\boldsymbol{\xi}_i\end{pmatrix}+o_p(1).$$

Using the central limit theorem, we obtain that

$$\sqrt{n}\{(\widehat{\boldsymbol{\alpha}}-\boldsymbol{\alpha})^{\mathrm{T}},(\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta})^{\mathrm{T}},(\widehat{\boldsymbol{\phi}}-\widehat{\boldsymbol{\phi}})^{\mathrm{T}}\}^{\mathrm{T}}\rightarrow_d MVN(\mathbf{0},\boldsymbol{\Omega}^*),$$

where $\boldsymbol{\Omega}^*=E[\mathbf{A}^{*-1}\boldsymbol{\Delta}^*\mathrm{diag}(\mathbf{V}^{-1},\mathbf{V}^{*-1})\boldsymbol{\Sigma}^*\mathrm{diag}(\mathbf{V}^{-1},\mathbf{V}^{*-1})\boldsymbol{\Delta}^*\mathbf{A}^{*-1}]$, and $\boldsymbol{\Sigma}^*$ is the covariance matrix of the vector $(\mathbf{Y}_i^{\mathrm{T}},\mathbf{Y}_i^{*\mathrm{T}})^{\mathrm{T}}$.

Acknowledgments The authors gratefully acknowledge an associate editor and two referees, whose comments greatly improved the form and content of this work. The research of Yi and He was supported by the Natural Sciences and Engineering Research Council of Canada. The research of Liang was partially supported by the NIAD/NIH grants AI62247 and AI059773 and NSF grant DMS-0806097. The authors thank Boston University and NHLBI for providing the data set from the Framingham Heart Study (No. N01-HC-25195) in the illustration. The Framingham Heart Study is conducted and supported by the National Heart, Lung, and Blood Institute (NHLBI) in collaboration with Boston University. This manuscript was not prepared in collaboration with investigators of the Framingham Heart Study and does not necessarily reflect the opinions or views of the Framingham Heart Study, Boston University, or NHLBI.

References

Aerts, M., Claeskens, G. (1997). Local polynomial estimation in multiparameter likelihood models. *Journal of the American Statistical Association 92*, 1536–1545.

Carey, V., Zeger, S. L., Diggle, P. J. (1993). Modelling multivariate binary data with alternating logistic regressions. *Biometrika 80*, 517–526.

Carroll, R. J., Fan, J., Gijbels, I., Wand, M. P. (1997). Generalized partially linear single-index models. *Journal of the American Statistical Association 92*, 477–489.

Chaudhuri, P., Doksum, K., Samarov, A. (1997). On average derivative quantile regression. *The Annals of Statistics 25*, 715–744.

Chen, K., Jin, Z. (2005). Local polynomial regression analysis of clustered data. *Biometrika 92*, 59–74.

Claeskens, G., Aerts, M. (2000). Bootstrapping local polynomial estimators in likelihood-based models. *Journal of Statistical Planning and Inference 86*, 63–80.

Fan, J., Li, R. (2004). New estimation and model selection procedures for semiparametric modeling in longitudinal data analysis. *Journal of the American Statistical Association 99*, 710–723.

Fan, J., Heckman, N. E., Wand, M. P. (1995). Local polynomial kernel regression for generalized linear models and quasi-likelihood functions. *Journal of the American Statistical Association 90*, 141–150.

Fitzmaurice, G. M., Laird, N. M. (1993). A likelihood-based method for analysing longitudinal binary responses. *Biometrika 80*, 141–151.

Hall, D. B., Severini, T. A. (1998). Extended generalized estimating equations for clustered data. *Journal of the American Statistical Association 93*, 1365–1375.

Huang, J. Z., Liu, L. (2006). Polynomial spline estimation and inference of proportional hazards regression models with flexible relative risk form. *Biometrics 62*, 793–802.

Ichimura, H. (1993). Semiparametric least squares (SLS) and weighted SLS estimation of single-index models. *Journal of Econometrics 58*, 71–120.

Kim, S., Chen, M. H., Dey, D. K. (2008). Flexible generalized t-link models for binary response data. *Biometrika 95*, 93–106.

Kraft, P., Bauman, L., Yuan, J. Y., Horvath, S. (2003). Multivariate variance-components analysis of longitudinal blood pressure measurements from the Framingham heart study. *BMC Genetics 4(Suppl 1)*, S55.

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

Liang, H., Wang, S., Robins, J. M., Carroll, R. J. (2004). Estimation in partially linear models with missing covariates. *Journal of the American Statistical Association 99*, 357–367.

Lin, X. H., Carroll, R. J. (2001a). Semiparametric regression for clustered data using generalized estimating equations. *Journal of the American Statistical Association 96*, 1045–1056.

Lin, X. H., Carroll, R. J. (2001b). Semiparametric regression for clustered data. *Biometrika 88*, 1179–1185.

Lipsitz, S. R., Laird, N. M., Harrington, D. P. (1991). Generalized estimating equations for correlated binary data: using the odds ratio as a measure of association. *Biometrika 78*, 153–160.

Little, R. J. A., Rubin, D. B. (2002). *Statistical analysis with missing data*, 2nd edn. New York: Wiley.

Molenberghs, G., Lesaffre, E. (1994). Marginal modelling of correlated ordinal data using an n-Way plackett distribution. *Journal of the American Statistical Association 89*, 633–644.

Pepe, M. S., Anderson, G. L. (1994). A cautionary note on inference for marginal regression models with longitudinal data and general correlated response data. *Communications in Statistics, Simulation and Computation 23*, 939–951.

Prentice, R. L. (1988). Correlated binary regression with covariates specific to each binary observation. *Biometrics 44*, 1033–1048.

Ruppert, D., Sheather, S. J., Wand, M. P. (1995). An effective bandwidth selector for local least squares regression. *Journal of the American Statistical Association 90*, 1257–1270.

Severini, T. A., Staniswalis, J. G. (1994). Quasilikelihood estimation in semiparametric models. *Journal of the American Statistical Association 89*, 501–511.

Wang, N. (2003). Marginal nonparametric kernel regression accounting for within-subject correlation. *Biometrika 90*, 43–52.

Wang, N., Carroll, R. J., Lin, X. (2005). Efficient semiparametric marginal estimation for longitudinal/clustered data. *Journal of the American Statistical Association 100*, 147–157.

Xia, Y. (2007). A constructive approach to the estimation of dimension reduction directions. *The Annals of Statistics 35*, 2654–2690.

Xia, Y., Härdle, W. (2006). Semi-parametric estimation of partially linear single-index models. *Journal of Multivariate Analysis 97*, 1162–1184.

Xia, Y., Tong, H., Li, W. K. (1999). On extended partially linear single-index models. *Biometrika 86*, 831–842.

Yi, G. Y., Cook, R. J. (2002). Marginal methods for incomplete longitudinal data arising in clusters. *Journal of the American Statistical Association 97*, 1071–1080.

Yi, G. Y., He, W., Liang, H. (2009). Analysis of correlated binary data under partially linear single-index logistic models. *Journal of Multivariate Analysis 100*, 278–290.

Yi, G. Y., Thompson, M. E. (2005). Marginal and association regression models for longitudinal binary data with drop-outs: a likelihood-based approach. *The Canadian Journal of Statistics 33*, 3–20.

Zeger, S. L., Diggle, P. J. (1994). Semi-parametric models for longitudinal data with applications to CD4 cell numbers in HIV seroconverters. *Biometrics 50*, 689–699.

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