Statistics and Its Interface Volume 12 (2019) 265–274

Variable selection for correlated bivariate mixed outcomes using penalized generalized estimating equations

Ved Deshpande, Dipak K. Dey, and Elizabeth D. Schifano*

We propose a penalized generalized estimating equations framework to jointly model correlated bivariate binary and continuous outcomes involving multiple predictor variables. We use sparsity-inducing penalty functions to simultaneously estimate the regression coefficients and perform variable selection on the predictors, and use cross-validation to select the tuning parameters. We further propose a method for tuning parameter selection that can control a desired false discovery rate. Using simulation studies, we demonstrate that the proposed joint modeling approach performs better in terms of accuracy and variable selection than separate penalized regressions on the binary and the continuous outcomes. We demonstrate the application of the method on a medical expenditure data set.

AMS 2000 subject classifications: Primary 62J07; secondary 62J12.
Keywords and phrases: False discovery rate, Generalized estimating equations, Medical expenditure, Penalized estimation.

1. INTRODUCTION

The task of modeling multivariate outcomes on sets of covariates is becoming increasingly common across research disciplines. Often, these outcomes are *non-commensurate*, i.e., they are on different scales of measurement. Such multivariate outcomes are likely to be correlated; for example, they may be measured from the same individual. A common subcase of multivariate outcomes is when exactly two outcomes per individual are measured, with one outcome measured on a continuous scale and the other outcome measured on a binary scale. This subcase, which we refer to as *bivariate mixed outcomes*, is the focus of our attention in this paper.

Joint modeling of such correlated bivariate mixed outcomes is preferable over separate modeling as we may be able to obtain more efficient parameter estimates [10]. However, specifying a joint model for bivariate mixed outcomes is challenging due to the lack of appropriate multivariate distributions for mixed outcomes. Likelihood-based approaches that aim to circumvent this problem include the factorization approach, in which the joint distribution of the outcomes is factorized into the marginal distribution of one outcome and the conditional distribution of the other outcome given the first outcome, and the latent variable approach, in which unobserved shared latent variables account for the correlation between the outcomes. See Teixeira Pinto and Normand [10] for a survey of these methods. A drawback of the factorization approach is the arbitrary choice of the conditioning outcome. Disadvantages of the latent variable approach include sensitivity to misspecification of the covariance structure, and arbitrary and untestable distributional assumptions on the latent variables [8].

On the other hand, generalized estimating equations (GEEs) provide an indirect approach to modeling bivariate mixed outcomes [8, 9, 7]. GEEs provide a convenient way to obtain consistent estimates of the marginal parameters even if the correlation structure between the outcomes is misspecified. This approach is primarily used when the correlation between the outcomes is a nuisance parameter and the marginal parameters are of primary interest in estimation.

If the number of covariates is large, regularization of the regression coefficients and variable selection may be important tasks to perform. These can often be achieved simultaneously through penalization, using penalties such as the least absolute shrinkage and selection operator (LASSO) [11], the elastic net (EN) [17], the smoothly clipped absolute deviation (SCAD) penalty [3], the minimax concave penalty (MCP) [15], and others. To apply penalization techniques to GEEs, Fu [4] and Johnson, Lin and Zeng [6] laid the framework for *penalized generalized estimating equations* (PGEEs), while Wang, Zhou and Qu [12] gave the form of PGEEs for commensurate longitudinal outcomes. Analogous to penalized regression, PGEEs perform simultaneous parameter estimation and variable selection through the incorporation of a sparsity-inducing penalty term in GEEs.

In this paper, we provide the framework to apply PGEEs in the (non-longitudinal) bivariate mixed outcome case. Our method differs from the method of Wang, Zhou and Qu [12] in two important ways. First, their method uses one tuning parameter for all outcomes. Because continuous and binary outcomes are on fundamentally different scales, two tuning

*Corresponding author.

parameters are necessary for correct estimation, and we allow for this in our proposed method. Second, their method assumes that a single set of regression coefficients can be used across all outcomes. This assumption is reasonable for their case, as they are concerned with modeling longitudinal outcomes, where there is really only one outcome, measured at different time points. However, when there are true multiple outcomes (even non-mixed), a different set of regression coefficients needs to be estimated for each outcome, because it is not reasonable to assume that a covariate has the same effect on all of the outcomes. Our algorithm is general enough to estimate two different sets of regression coefficients.

Through simulation studies, we show that gains can be made in both estimation and variable selection by using joint analysis rather than by separate marginal analyses of the outcomes. In the context of variable selection, controlling the false discovery rate (FDR) [1] is often of importance as well. Breheny [2] and Yi et al. [14] showed how to estimate and control the FDR for penalized regression. We generalize this method to the PGEE framework, and through simulations, demonstrate that our method is able to control the FDR at a desired level. An R package, `pgee.mixed`, is available at https://cran.r-project.org/ for implementation.

We illustrate the application of our PGEE framework and FDR control methodology to data from the Medical Expenditure Panel Survey (MEPS). MEPS provides a nationally representative sample of health care data at the individual level, and contains information on medical spending, health status, demographics, health conditions, access to care, health insurance coverage, income, and employment. Our analysis is inspired by the work done in Zimmerman [16], who sought to jointly model annual drug spending (modeled as a continuous variable) and health status (modeled as a binary variable) for Medicare enrollees in 2004 and 2005, the two years before Medicare Part D became active. While the primary goal of that analysis was to investigate the strength of association between these two outcomes, our goal is to identify important covariates that affect drug spending and health status. With our penalized GEE framework, we are able to consider a larger set of covariates than Zimmerman [16]. Then, by borrowing information from total drug spending, we are able to identify important covariates for health status that may not be detectable from a marginal analysis on the latter outcome. We also estimate the false discovery rate to reassure ourselves that we are detecting additional signal, rather than noise.

The rest of the paper is organized as follows. In Section 2, we provide the framework for applying PGEEs to bivariate mixed outcomes. We also provide an iterative algorithm to solve the PGEEs and a method to control the FDR. Section 3 contains results from simulation experiments. In Section 4, we apply the PGEE framework to the MEPS data and discuss our findings. Section 5 concludes the paper with some discussions.

2. PENALIZED GENERALIZED ESTIMATING EQUATIONS FOR BIVARIATE MIXED OUTCOMES

2.1 Notation

From the ith individual, we observe a continuous outcome y_{ic}, a binary outcome y_{ib}, a p-dimensional covariate vector $\boldsymbol{x_i}$ corresponding to the continuous outcome y_{ic}, and a q-dimensional covariate vector $\boldsymbol{z_i}$ corresponding to the binary outcome y_{ib}, $\quad i = 1 \ldots n$. It is common to assume $\boldsymbol{x_i} = \boldsymbol{z_i}$ (i.e., use the same set of covariates to model both outcomes), but this need not be so. Let $\boldsymbol{y_i} = (y_{ic}, y_{ib})^T$ denote the bivariate vector of outcomes from the ith individual. We assume that outcomes from the same individual are correlated, but outcomes from different individuals are independent.

We specify the link functions $g_c(\mu_{ic}) = \boldsymbol{x_i}^T\boldsymbol{\beta_c}$ and $g_b(\mu_{ib}) = \boldsymbol{z_i}^T\boldsymbol{\beta_b}$, where $\mu_{ic} = \mathrm{E}(y_{ic})$ and $\mu_{ib} = \mathrm{E}(y_{ib})$. Denote $\boldsymbol{\mu_i} = (\mu_{ic}, \mu_{ib})^T$ and $\boldsymbol{\beta} = (\boldsymbol{\beta_c}^T, \boldsymbol{\beta_b}^T)^T$. The variances of the outcomes can be specified as $v_c(y_{ic}) = \psi_c h_c(\mu_{ic})$, and $v_b(y_{ib}) = \psi_b h_b(\mu_{ib})$, where ψ_c and ψ_b are dispersion parameters, and $h_c(\cdot)$ and $h_b(\cdot)$ are the usual variance functions from generalized linear models. For illustration, we take $g_c(\cdot)$ to be the identity link and $g_b(\cdot)$ to be the logit link. For simplicity, we further assume that $\psi_c = \psi_b = 1$.

2.2 Generalized estimating equations for bivariate mixed outcomes

Rochon [9] gave the setup for generalized estimating equations for bivariate mixed outcomes:

$$\text{(1)} \qquad \boldsymbol{S}(\boldsymbol{\beta}) = n^{-1}\sum_{i=1}^{n} \boldsymbol{D_i}^T \boldsymbol{V_i}^{-1}(\boldsymbol{y_i} - \boldsymbol{\mu_i}) = \boldsymbol{0},$$

where

$$\text{(2)} \qquad \boldsymbol{D_i}^T = \frac{\partial \boldsymbol{\mu_i}(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}^T} = \begin{pmatrix} \partial\mu_{ic}/\partial\boldsymbol{\beta_c}^T & \boldsymbol{0} \\ \boldsymbol{0} & \partial\mu_{ib}/\partial\boldsymbol{\beta_b}^T \end{pmatrix},$$

and $\boldsymbol{V_i}$ is the variance-covariance matrix of $\boldsymbol{y_i}$, given by $\boldsymbol{V_i} = (\boldsymbol{A_i}^T)^{1/2}\boldsymbol{R}\boldsymbol{A_i}^{1/2} = \boldsymbol{A_i}^{1/2}\boldsymbol{R}\boldsymbol{A_i}^{1/2}$, where

$$\boldsymbol{A_i} = \begin{pmatrix} h_c(\mu_{ic}) & 0 \\ 0 & h_b(\mu_{ib}) \end{pmatrix}, \quad \boldsymbol{R} = \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}.$$

Here $\boldsymbol{R}$ is a *working correlation matrix* and ρ measures the strength of association between the continuous and binary outcomes. Note that ρ, which we refer to as the association parameter, is assumed to be fixed across i. Without loss of generality, we also assume that ρ is non-negative, because if the outcomes are negatively correlated, we can simply flip the sign of the continuous variable.

Wang, Zhou and Qu [12] showed that if the marginal density of each outcome can be assumed to come from a canonical exponential family, then $\boldsymbol{S}(\boldsymbol{\beta})$ in (1) can be simplified to

$$\boldsymbol{S}(\boldsymbol{\beta}) = n^{-1}\sum_{i=1}^{n} \boldsymbol{X_i}^T \boldsymbol{A_i}^{1/2}(\boldsymbol{\beta})\hat{\boldsymbol{R}}^{-1}\boldsymbol{A_i}^{-1/2}(\boldsymbol{\beta})\boldsymbol{\epsilon}(\boldsymbol{\beta}), \tag{3}$$

where $\boldsymbol{\epsilon}(\boldsymbol{\beta}) = \boldsymbol{y_i} - \boldsymbol{\mu_i}(\boldsymbol{\beta})$ and $\boldsymbol{X_i}$ is the covariate matrix for the ith individual. In the bivariate mixed outcome case, $\boldsymbol{X_i}$ reduces to the block-diagonal structure

$$\boldsymbol{X_i} = \begin{pmatrix} \boldsymbol{x_i}^T & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{z_i}^T \end{pmatrix}. \tag{4}$$

$\hat{\boldsymbol{R}}$ is the estimated working correlation matrix, in which the association parameter ρ is replaced by an estimate $\hat{\rho}$. We compute $\hat{\rho}$ using the biserial correlation between the binary outcomes and the residuals of the continuous outcomes.

2.3 Penalized generalized estimating equations for bivariate mixed outcomes

A sparsity-inducing penalty term can be incorporated into (3) if we wish to perform simultaneous estimation and variable selection with the GEEs. The PGEEs for bivariate mixed outcomes are given as

$$\boldsymbol{U}(\boldsymbol{\beta}) = \boldsymbol{S}(\boldsymbol{\beta}) - \boldsymbol{q_\lambda}(|\boldsymbol{\beta}|)sign(\boldsymbol{\beta}), \tag{5}$$

where $\boldsymbol{S}(\boldsymbol{\beta})$ is defined in (3),

$$\begin{aligned} \boldsymbol{q_\lambda}(|\boldsymbol{\beta}|) = [&q_{\lambda_c}(\beta_{c1}), q_{\lambda_c}(\beta_{c2}), \dots, q_{\lambda_c}(\beta_{cp}), \\ &q_{\lambda_b}(\beta_{b1}), q_{\lambda_b}(\beta_{b2}), \dots, q_{\lambda_b}(\beta_{bq})]^T \end{aligned} \tag{6}$$

is a $(p+q)$-dimensional vector of the first derivatives of penalty functions, where λ_c and λ_b are the tuning parameters for the penalty functions associated with continuous regression coefficients and binary regression coefficients, respectively, and

$$\begin{aligned} sign(\boldsymbol{\beta}) = [&sign(\beta_{c1}), \dots, sign(\beta_{cp}), \\ &sign(\beta_{b1}), \dots, sign(\beta_{bq})]^T, \end{aligned} \tag{7}$$

where $sign(t) = I(t > 0) - I(t < 0)$. Note that the product of $\boldsymbol{q_\lambda}(\cdot)$ and $sign(\cdot)$ in (5) is component-wise. Unlike previous frameworks for the PGEEs such as in Johnson, Lin and Zeng [6] and Wang, Zhou and Qu [12], we require two tuning parameters λ_c and λ_b, because the continuous and binary outcomes are on fundamentally different scales. Restricting the model to a single tuning parameter would necessarily lead to over-penalization or under-penalization in at least one component of $(\boldsymbol{\beta_c}, \boldsymbol{\beta_b})$.

Although a variety of sparsity-inducing penalties can be chosen in (5), we restrict our attention to the SCAD penalty $q_\lambda(\theta) = \lambda\{I(\theta \le \lambda) + (a-1)^{-1}\lambda^{-1}(a\lambda - \theta)_+ I(\theta > \lambda)\}$, for $\theta \ge 0$ and for fixed $a > 2$, where $(t)_+ = \max(t, 0)$. We fix $a = 3.7$ as recommended in Fan and Li [3].

2.4 Algorithm to solve PGEEs

Wang, Zhou and Qu [12] provided a Newton-Raphson type iterative scheme to solve PGEEs. Their algorithm assumes a single set of regression coefficients for all outcomes, and a single tuning parameter. We generalize their algorithm to estimate two sets of regression coefficients with two tuning parameters, to solve PGEEs for bivariate mixed outcomes:

$$\hat{\boldsymbol{\beta}}^{k+1} = \hat{\boldsymbol{\beta}}^k + [\boldsymbol{H}(\hat{\boldsymbol{\beta}}^k) + \boldsymbol{E}(\hat{\boldsymbol{\beta}}^k)]^{-1}[\boldsymbol{S}(\hat{\boldsymbol{\beta}}^k) - \boldsymbol{E}(\hat{\boldsymbol{\beta}}^k)\hat{\boldsymbol{\beta}}^k], \tag{8}$$

where

$$\boldsymbol{H}(\hat{\boldsymbol{\beta}}^k) = n^{-1}\sum_{i=1}^{n} \boldsymbol{X_i}^T \boldsymbol{A_i}^{1/2}(\hat{\boldsymbol{\beta}}^k)\hat{\boldsymbol{R}}^{-1}\boldsymbol{A_i}^{1/2}(\hat{\boldsymbol{\beta}}^k)\boldsymbol{X_i}, \tag{9}$$

$$\begin{aligned} \boldsymbol{E}(\hat{\boldsymbol{\beta}}^k) = \operatorname{diag}\Bigg\{&\frac{q_{\lambda_c}(|\hat{\beta}^k_{c1}|+)}{\varepsilon + |\hat{\beta}^k_{c1}|}, \dots, \frac{q_{\lambda_c}(|\hat{\beta}^k_{cp}|+)}{\varepsilon + |\hat{\beta}^k_{cp}|}, \\ &\frac{q_{\lambda_b}(|\hat{\beta}^k_{b1}|+)}{\varepsilon + |\hat{\beta}^k_{b1}|}, \dots, \frac{q_{\lambda_b}(|\hat{\beta}^k_{bq}|+)}{\varepsilon + |\hat{\beta}^k_{bq}|}\Bigg\}, \end{aligned} \tag{10}$$

where ε is a small fixed positive number, which we set to 10^{-6}. This algorithm has close connections to the local quadratic approximation algorithm of Fan and Li [3] and the minorization-maximization (MM) algorithm of Hunter and Li [5] for solving penalized regression problems.

The two tuning parameters λ_c and λ_b are chosen using four-fold cross-validation over a two-dimensional grid. The loss function used for the cross-validation is the sum of a squared error loss for the estimated continuous regression coefficient vector $\widehat{\boldsymbol{\beta_c}}$:

$$L_c(\boldsymbol{y_c}, \widehat{\boldsymbol{\eta_c}}) = \sum_{i=1}^{n}(y_{ic} - \widehat{\eta_{ic}})^2, \tag{11}$$

where $\widehat{\eta_{ic}} = \boldsymbol{x_i}^T\widehat{\boldsymbol{\beta_c}}$, and a deviance loss for the estimated binary regression coefficient vector $\widehat{\boldsymbol{\beta_b}}$:

$$L_b(\boldsymbol{y_b}, \widehat{\boldsymbol{\eta_b}}) = \frac{1}{\log(2)}\sum_{i=1}^{n}\log[1 + \exp\{-2\widehat{\eta_{ib}}(2y_{ib} - 1)\}], \tag{12}$$

where $\widehat{\eta_{ib}} = \boldsymbol{z_i}^T\widehat{\boldsymbol{\beta_b}}$. Note that $y_{ib} \in \{0, 1\}$. Convergence of the algorithm is declared if either of two conditions are satisfied: $\|\hat{\boldsymbol{\beta}}^{k+1} - \hat{\boldsymbol{\beta}}^k\|_1 < \epsilon_1$, or $\|\boldsymbol{U}(\hat{\boldsymbol{\beta}}^{k+1})\|_1 < \epsilon_2$, where $\|\boldsymbol{\theta}\|_1 = \sum_{i=1}^{n}|\theta_i|$ is the L_1-norm of an n-dimensional vector $\boldsymbol{\theta}$, and $\boldsymbol{U}(\boldsymbol{\beta})$ are the penalized estimating functions from (5).

From the Newton-Raphson scheme, analogous to Wang, Zhou and Qu [12], we can obtain the asymptotic covariance matrix of $\widehat{\boldsymbol{\beta}}$, given by

$$\operatorname{Cov}(\widehat{\boldsymbol{\beta}}) \approx [\boldsymbol{H}(\widehat{\boldsymbol{\beta}}) + \boldsymbol{E}(\widehat{\boldsymbol{\beta}})]^{-1}\boldsymbol{M}(\widehat{\boldsymbol{\beta}})[\boldsymbol{H}(\widehat{\boldsymbol{\beta}}) + \boldsymbol{E}(\widehat{\boldsymbol{\beta}})]^{-1}, \tag{13}$$

where $\boldsymbol{H}$ and $\boldsymbol{E}$ are defined in (9) and (10), and

$$\boldsymbol{M}(\widehat{\boldsymbol{\beta}}) = \sum_{i=1}^{n}\boldsymbol{S_i}(\hat{\boldsymbol{\beta}})\boldsymbol{S_i}^T(\hat{\boldsymbol{\beta}}), \tag{14}$$

where $\boldsymbol{S_i}(\hat{\boldsymbol{\beta}}) = n^{-1}\boldsymbol{X_i}^T\boldsymbol{A_i}^{1/2}(\widehat{\boldsymbol{\beta}})\widehat{\boldsymbol{R}}^{-1}\boldsymbol{A_i}^{-1/2}(\widehat{\boldsymbol{\beta}})[\boldsymbol{y_i} - \boldsymbol{\mu_i}(\widehat{\boldsymbol{\beta}})]$. Note that we do not use (13) to conduct post-selection inference. We only use it for comparing the efficiencies of the methods we consider in our simulation experiments (Section 3).

2.5 Controlling the false discovery rate

In this section, we propose a method to estimate and thus control the false discovery rate (FDR) in the PGEE setting by selecting appropriate values for the penalty parameters λ_c and λ_b. Breheny [2] and Yi et al. [14] proposed such a method to control the FDR for penalized linear regression and penalized logistic regression. We generalize this method to PGEEs for mixed outcomes.

The FDR can be expressed as

$$\text{FDR} = \frac{\text{E(F)}}{\text{S}}, \tag{15}$$

where S is the total number of covariates selected by the variable selection procedure and F is the number of false discoveries. Under sparsity-inducing penalty functions like SCAD, the jth covariate is selected if its regression coefficient β_j is estimated as non-zero, i.e., $\hat{\beta}_j \neq 0$. We say that the jth covariate is *null* if $\beta_j = 0$. Thus, a false discovery is a null covariate that is selected by the variable selection procedure. Note that since F is unknown in practice, it is replaced with its expectation in (15).

Next, letting $\alpha_j = P(\hat{\beta}_j \neq 0|\beta_j = 0)$ be the probability of making a false discovery on the jth covariate, the numerator of (15) can be estimated by

$$\widehat{\text{E(F)}} = \sum_{j=1}^{J} \alpha_j, \tag{16}$$

where J is the number of covariates being considered in the variable selection procedure. This approach to estimating the FDR is conservative (overestimates the FDR), since the sum in (16) is over all covariates and not just the null covariates. However, we do not know which covariates are null in practice.

We rewrite the estimating functions of the unpenalized GEEs from (1) as

$$\begin{aligned}\boldsymbol{S}(\boldsymbol{\beta}) &= n^{-1}\sum_{i=1}^{n}\boldsymbol{D_i}^T\boldsymbol{V_i}^{-1}(\boldsymbol{y_i} - \boldsymbol{\mu_i})\\ &= n^{-1}\sum_{i=1}^{n}\boldsymbol{W_i}^T\boldsymbol{r_i}\\ &= n^{-1}\boldsymbol{W}^T\boldsymbol{r},\end{aligned}$$

where $\boldsymbol{W_i}^T = \boldsymbol{D_i}^T\boldsymbol{V_i}^{-1}$, $\boldsymbol{r_i} = (\boldsymbol{y_i} - \boldsymbol{\mu_i})$, $\boldsymbol{W}^T = [\boldsymbol{W_1}^T, \cdots, \boldsymbol{W_n}^T]$, $\boldsymbol{r} = [\boldsymbol{r_1}^T, \cdots, \boldsymbol{r_n}^T]$. Note that each $\boldsymbol{r_i} = [r_{ic}, r_{ib}]^T$ is a 2-dimensional vector; hence $\boldsymbol{r}$ is a $2n$-dimensional vector. Denoting $\boldsymbol{w^{(j)}}$ as the jth column vector of $\boldsymbol{W}$, $j = 1, \cdots, (p+q)$, we can express the jth component of $\boldsymbol{S}(\boldsymbol{\beta})$ as $S_j(\boldsymbol{\beta}) = n^{-1}\boldsymbol{w^{(j)}}^T\boldsymbol{r}$.

Wolfson [13] mentions that although estimating equations may not correspond to the gradient of some (unknown) loss function, they can be obtained as the modification of such a gradient, and can be expected to have similar behavior as the gradient. Hence, at the solution, the Karush-Kuhn-Tucker optimality conditions should hold, which give the following conditions for PGEEs:

$$n^{-1}\boldsymbol{w^{(j)}}^T\boldsymbol{r} = \lambda_j sign(\hat{\beta}_j) \qquad \forall\hat{\beta}_j \neq 0, \tag{17a}$$

$$n^{-1}|\boldsymbol{w^{(j)}}^T\boldsymbol{r}| \leq \lambda_j \qquad \forall\hat{\beta}_j = 0, \tag{17b}$$

where λ_j is λ_c or λ_b, depending on whether β_j corresponds to the continuous outcome or to the binary outcome, respectively. Note that the conditions in (17) are derived assuming the LASSO penalty, but as mentioned in Breheny [2], the same conditions can be applied to the SCAD penalty, which we use.

We show in Section 1 of the Supplementary Materials http://intlpress.com/site/pub/files/_supp/sii/2019/0012/0002/SII-2019-0012-0002-s002.pdf that the conditions in (17) further imply the conditions

$$n^{-1}|\boldsymbol{w^{(j)}}^T\boldsymbol{r^{(-j)}}| > \lambda_j \qquad \forall\hat{\beta}_j \neq 0, \tag{18a}$$

$$n^{-1}|\boldsymbol{w^{(j)}}^T\boldsymbol{r^{(-j)}}| \leq \lambda_j \qquad \forall\hat{\beta}_j = 0, \tag{18b}$$

where the $-j$ superscript indicates quantities calculated without using the jth covariate. Hence, we have

$$\begin{aligned}\alpha_j &= P(\hat{\beta}_j \neq 0|\beta_j = 0)\\ &= P(n^{-1}|\boldsymbol{w^{(j)}}^T\boldsymbol{r^{(-j)}}| > \lambda_j|\beta_j = 0).\end{aligned} \tag{19}$$

In general, the distribution of the $\boldsymbol{r^{(-j)}}$'s is complex, hence obtaining an analytical expression for (19) is difficult. However, analogous to Breheny [2], we can make an approximation:

$$\boldsymbol{r^{(-j)}} \overset{approx}{\sim} \boldsymbol{N}_{2n}(\boldsymbol{0}, \tilde{\boldsymbol{V}}), \tag{20}$$

where $\tilde{\boldsymbol{V}} = diag(\boldsymbol{V}, \ldots, \boldsymbol{V})$, and

$$\boldsymbol{V} = \begin{pmatrix} \sigma_c^2 & \rho\sigma_c\sigma_b \\ \rho\sigma_c\sigma_b & \sigma_b^2 \end{pmatrix},$$

where the variance parameters σ_c^2 and σ_b^2 can be estimated from the data as $\hat{\sigma_c^2} = n^{-1}\|\boldsymbol{r_c}\|_2^2$, $\hat{\sigma_b^2} = n^{-1}\|\boldsymbol{r_b}\|_2^2$, with $\boldsymbol{r_c} = [r_{c1}, \ldots, r_{cn}]^T$ and $\boldsymbol{r_b} = [r_{b1}, \ldots, r_{bn}]^T$. The association parameter ρ is already estimated from the algorithm that solves the PGEEs. Note that the block-diagonal structure of the variance-covariance matrix of $\boldsymbol{r^{(-j)}}$ from (20) reflects the assumption that the bivariate outcomes from a single individual are correlated, but outcomes between individuals are independent.

Using (20), we can approximate (19) as:

$$\widehat{\alpha_j} = 2\Phi\left(\frac{-n\lambda_j}{\sqrt{\boldsymbol{w^{(j)T}}\tilde{\boldsymbol{V}}\boldsymbol{w^{(j)}}}}\right). \tag{21}$$

To estimate the total FDR across both continuous and binary outcomes, we can use (15), (16), and (21), with $J = p + q$. Alternatively, we can estimate the FDR separately for the continuous and the binary outcomes using:

$$\widehat{\mathrm{FDR}_c} = \frac{\widehat{\mathrm{E(F}_c)}}{\mathrm{S}_c}, \quad \widehat{\mathrm{E(F}_c)} = \sum_{j=1}^{p}\widehat{\alpha_j},$$

$$\widehat{\mathrm{FDR}_b} = \frac{\widehat{\mathrm{E(F}_b)}}{\mathrm{S}_b}, \quad \widehat{\mathrm{E(F}_b)} = \sum_{j=p+1}^{p+q}\widehat{\alpha_j},$$

where S_c is the total number of continuous outcome covariates selected and S_b is the total number of binary outcome covariates selected.

Note that in general, there will be multiple pairs of tuning parameters (λ_c, λ_b) that can control the FDR at a desired level. Hence, in practice, we choose λ_c and λ_b as the pair with the lowest cross-validated error amongst all pairs that control the FDR at the desired level.

3. SIMULATION EXPERIMENTS

We conducted simulation experiments to compare our method of modeling the bivariate outcomes jointly versus modeling each outcome separately using two unrelated PGEEs. We also conducted a simulation experiment to investigate the effectiveness of our FDR control method.

3.1 Data generation

3.1.1 Comparing the joint PGEEs method versus the separate PGEEs method

We generated 1000 data sets, each consisting of $n = 500$ pairs of correlated bivariate mixed outcomes, with $p = q = 50$ covariates per outcome. Marginally, the continuous outcomes follow normal distributions with the identity link to covariates and the binary outcomes follow Bernoulli distributions with the logit link to covariates. Denote the covariate matrices as $\boldsymbol{X} = [\boldsymbol{x^{(1)}}, \boldsymbol{x^{(2)}}, \ldots, \boldsymbol{x^{(p)}}]$ and $\boldsymbol{Z} = [\boldsymbol{z^{(1)}}, \boldsymbol{z^{(2)}}, \ldots, \boldsymbol{z^{(q)}}]$ for the continuous and the bivariate responses, respectively, where $\boldsymbol{x^{(j)}}$ is the jth column of $\boldsymbol{X}$, and similarly for $\boldsymbol{Z}$. We assumed intercepts for both outcomes, so $\boldsymbol{x^{(1)}} = \boldsymbol{z^{(1)}} = \boldsymbol{1_n}$, whose coefficients are not penalized. Covariates were generated from a multivariate normal distribution with a zero-mean vector, unit marginal variances, and an AR(1) correlation structure with a correlation of 0.25. Three situations for these covariate matrices were considered: (i) All covariates are shared (between the bivariate outcomes), i.e., $\boldsymbol{X} = \boldsymbol{Z}$, (ii) Some but not all covariates are shared, in which case we set $\boldsymbol{z^{(j)}} = \boldsymbol{x^{(j)}}$ for $j = 2, 3$, and (iii) No covariates are shared, in which case we generated $\boldsymbol{X}$ and $\boldsymbol{Z}$ independently.

Next, the true regression coefficient vectors are chosen as $\boldsymbol{\beta_{0c}} = (0.2, 2.0, 0, \ldots, 0, 3.0, -1.5, 2.0)^T$ and $\boldsymbol{\beta_{0b}} = (1.2, 0.8, 0.6, -0.4, 0, \ldots, 0)^T$. This setup lets us consider the case that when the two covariate matrices are identical, exactly one of the covariates (i.e. $j = 2$) is associated with both of the outcomes, while all other covariates are associated with at most one of the outcomes. The correlated bivariate mixed responses are then generated as follows. For $i = 1, \ldots, n$:

$$\begin{aligned}
&(u_i, v_i) \sim C(\cdot, \cdot|\theta), \\
&\quad y_{ic} \sim \Phi^{-1}(u_i|\mu = \boldsymbol{x_i}^T\boldsymbol{\beta_{0c}}, \sigma = 1), \\
&\quad w_i \sim F^{-1}(v_i|\mu = \boldsymbol{z_i}^T\boldsymbol{\beta_{0b}}, s = 1), \\
&\quad y_{ib} = I(w_i > 0),
\end{aligned}$$

where $C(\cdot, \cdot|\theta)$ is a two-dimensional Gaussian copula with correlation parameter θ, $\Phi^{-1}(\cdot|\mu, \sigma)$ is the inverse cumulative distribution function of a normal distribution with mean μ and standard deviation σ, and $F^{-1}(\cdot|\mu, s)$ is the inverse cumulative distribution function of the logistic distribution with location μ and scale s. Thus θ, the parameter of the copula, is the correlation between the continuous outcome and the latent variable that generates the binary outcome. We feel that specifying a correlation between the continuous outcome y_{ic} and the latent logistic variable w_i is more natural than specifying a direct correlation between the continuous outcome y_{ic} and the binary outcome y_{ib}. Note that the copula parameter, θ, and the association parameter in the PGEEs, ρ, are different quantities. Finally, note that to generate the binary response with a logistic link, we used the fact that generating $y \sim \mathit{Bernoulli}(p = e^{\phi}/(1 + e^{\phi}))$ is equivalent to generating $w \sim \mathit{Logistic}(\mu = \phi, s = 1)$, $y = I(w > 0)$. We considered scenarios with $\theta = 0.2$, $\theta = 0.4$, $\theta = 0.6$, and $\theta = 0.8$, corresponding to varying strengths of association between the continuous and binary outcomes.

3.1.2 FDR control

We generated 1000 data sets of correlated bivariate mixed outcomes. For brevity, we only considered the case when all covariates are shared between the bivariate outcomes. The design of the simulation is largely the same as the one described above, with the exception that we set both the continuous and binary regression coefficients to $(1, -1, 1, -1, 1, -1, 0, \ldots, 0)^T$, as in Breheny [2].

3.2 Simulation results

3.2.1 Comparing the joint PGEEs method versus the separate PGEEs method

Here, we compare the joint and the separate PGEEs methods in terms of accuracy and variable selection metrics. For each of the 1000 data sets generated under each

Table 1. Accuracy and variable selection metrics comparing the joint and the separate PGEEs methods, for $\theta = 0.2, 0.6$, with all covariates shared *between the continuous and binary outcomes. Maximum TP is 9 and maximum FP is 91.*

(a) $\theta = 0.2$

Method	MSE	U	O	E	TP	FP
Joint	0.2473	0.55	0.22	0.23	8.21	1.72
Separate	0.2514	0.56	0.20	0.24	8.20	1.63

(b) $\theta = 0.6$

Method	MSE	U	O	E	TP	FP
Joint	0.1996	0.40	0.28	0.32	8.48	1.86
Separate	0.2584	0.56	0.17	0.27	8.18	1.68

scenario, we applied our iterative algorithm to solve the PGEEs and obtained estimates of the regression coefficients $\boldsymbol{\beta} = (\boldsymbol{\beta}_{\boldsymbol{c}}^T, \boldsymbol{\beta}_{\boldsymbol{b}}^T)^T$. As described in Section 2.4, the tuning parameters λ_c and λ_b were selected using four-fold cross-validation over a two-dimensional grid, equally spaced on the log scale. We also applied separate PGEEs to the continuous and the binary outcomes and estimated the regression coefficients. For each of the separate estimations, the tuning parameter was selected using four-fold cross-validation over a one-dimensional grid, equally spaced on the log scale. To evaluate the accuracy of these estimates, we computed the mean squared error (MSE) as $(1000)^{-1} \sum_{i=1}^{1000} \|\hat{\boldsymbol{\beta}}^{(i)} - \boldsymbol{\beta}_0\|_2^2$, where $\hat{\boldsymbol{\beta}}^{(i)}$ is the estimate for the true regression coefficient vector $\boldsymbol{\beta}_0$ from the ith data set. We also computed the absolute bias and the sandwich-formula based standard error for each true non-zero regression coefficient. To compare performance in variable selection, we computed the proportion of data sets in which the methods under-selected (U), over-selected (O) and exactly selected (E) the covariates with true non-zero regression coefficients. (A good variable selection method should have small U and O metrics, and a large E metric). Finally, we calculated the average number of true positives per data set (TP) and the average number of false positives per data set (FP) for both the methods.

Table 1 shows the MSE and variable selection metrics for the joint and the separate methods for $\theta = 0.2$ and $\theta = 0.6$, with all covariates shared between the bivariate outcomes. We observe that the joint method has smaller MSE than the separate method, with a more noticeable difference for $\theta = 0.6$. Under-selection is usually considered worse than over-selection in variable selection, and we observe that the joint method has smaller U metrics than the separate method. The joint method also has a larger E metric than the separate method for $\theta = 0.6$, and its E metric is smaller than that the separate method by only 0.01 for $\theta = 0.2$. The joint method has larger TP and than the separate method for both $\theta = 0.2$ and $\theta = 0.6$, which it trades off with its larger FP metric. Note, however, that although the joint method shows larger FP, because it has smaller MSE, the magnitudes of the estimates of the false positive must be small. Similar behavior in terms of these metrics is observed across the other scenarios considered (see Tables S1, S2 and S3 in the Supplementary Materials). Also, Table S7 in the Supplementary Materials shows these metrics split by outcome type, for the case where no covariates are shared between the bivariate outcome. From that table, we can see that the MSE is smaller for the joint method for both outcomes, and that the joint method's larger TP is driven by the binary outcomes. Corresponding tables for the other covariate cases are not provided, but the trends are similar.

Table 2 shows the absolute bias and the standard errors for the true non-zero coefficients for the scenarios corresponding to Table 1. We observe that the absolute bias and the standard errors under both methods are similar for the continuous outcome coefficients, but are almost always smaller for the joint method for the binary outcome coefficients. The other scenarios considered (see Tables S4, S5 and S6 in the Supplementary Materials) show similar trends, with the effects usually magnified for larger values of ρ.

Overall, we see that the joint method makes gains over the separate method in estimation and variable selection metrics for the binary outcome coefficients, especially for larger values of θ. Intuitively, this makes sense, as the binary outcome coefficients–which are harder to estimate due to the smaller information content of binary outcomes–benefit from *borrowing information* from the continuous outcomes via the correlation between the outcomes. The benefit increases as the strength of the correlation increases.

3.2.2 FDR control

As mentioned previously, there will be multiple pairs of tuning parameters that can control the FDR at a desired level, so we choose λ_c and λ_b as the pair with the lowest cross-validated error amongst all pairs that control the FDR at the desired level. To assess this method's performance, first we computed the estimated FDR for each pair in the two-dimensional grid using the entire data set. Then, we specified an FDR threshold and performed four-fold cross-validation, restricting the grid to pairs that have estimated FDR less than the FDR threshold. We refer to the resulting estimated coefficients as "restricted" coefficients. We computed restricted coefficients for FDR thresholds in $\{0.01, 0.05, 0.1\}$. Similar to the previous section, for comparison, we also estimated the regression coefficients using four-fold cross validation over the entire two-dimensional grid of tuning parameter values, without restricting the FDR at any level. We refer to these estimated coefficients as "unrestricted" coefficients. Using the unrestricted and restricted coefficients of the 1000 generated data sets, we computed the MSE and variable selection metrics as in Table 1. In addition, we computed the average true FDR over the data sets.

Table 3 reports these metrics for $\theta = 0.2$ and $\theta = 0.6$. The FDR is controlled at each threshold. As we would expect, O increases and E decreases as FDR increases. Applying

Table 2. Absolute bias (AB) and sandwich-formula based standard errors (SE) of estimates of true non-zero regression coefficients (excluding intercept) for the joint and the separate PGEEs methods, for $\theta = 0.2, 0.6$, with all covariates shared *between the continuous and binary outcomes.*

(a) $\theta = 0.2$

	Method	Continuous Outcome β_2	β_{48}	β_{49}	β_{50}	Binary Outcome β_{52}	β_{53}	β_{54}
AB	Joint	0.002	0.001	0.001	0.000	0.017	0.138	0.179
AB	Separate	0.002	0.001	0.001	0.000	0.018	0.142	0.180
SE	Joint	0.045	0.046	0.045	0.046	0.144	0.275	0.251
SE	Separate	0.044	0.046	0.045	0.047	0.144	0.281	0.252

(b) $\theta = 0.6$

	Method	Continuous Outcome β_2	β_{48}	β_{49}	β_{50}	Binary Outcome β_{52}	β_{53}	β_{54}
AB	Joint	0.001	0.000	0.001	0.001	0.010	0.099	0.138
AB	Separate	0.001	0.000	0.000	0.000	0.013	0.145	0.181
SE	Joint	0.045	0.042	0.045	0.044	0.130	0.243	0.243
SE	Separate	0.044	0.045	0.049	0.048	0.144	0.288	0.255

Table 3. Accuracy and variable selection metrics comparing various levels of FDR restriction, for $\theta = 0.2, 0.6$. Maximum TP is 12 and maximum FP is 88. A value of "None" for Max. FDR means that no FDR restriction was applied.

(a) $\theta = 0.2$

Max. FDR	MSE	U	O	E	TP	FP	FDR
0.01	0.1449	0.00	0.06	0.94	12.00	0.07	0.005
0.05	0.1452	0.00	0.18	0.82	12.00	0.23	0.017
0.10	0.1453	0.00	0.25	0.75	12.00	0.37	0.027
None	0.1463	0.00	0.32	0.68	12.00	1.15	0.066

(b) $\theta = 0.6$

Max. FDR	MSE	U	O	E	TP	FP	FDR
0.01	0.1441	0.00	0.06	0.94	12.00	0.08	0.006
0.05	0.1443	0.00	0.16	0.84	12.00	0.20	0.015
0.10	0.1441	0.00	0.24	0.76	12.00	0.34	0.025
None	0.1450	0.00	0.32	0.68	12.00	1.27	0.071

stricter FDR controls also seems to benefit estimation accuracy and other variable selection metrics, as evidenced by the smaller MSE, larger E, and smaller FP, for smaller levels of the threshold. This suggests that the cross-validation procedure may be overfitting, and specifying an FDR restriction has a regularizing effect on the model. Finally, we note that because we are reporting the results for the best (λ_c, λ_b) pair for the given FDR level, it is not surprising that the U values are all 0. Table S8 in the Supplementary Materials show the results for $\theta = 0.4$ and $\theta = 0.8$. The same trends are observed there as well.

For each of the 1000 data sets generated, we also noted the estimated FDR and the true FDR over the two-dimensional grid. The smoothed average of these FDRs are plotted in the contour plots in Figure S1 in the Supplementary Materials. Each level of the contour plots shows the various (λ_c, λ_b) combinations which result in the same FDR. These plots indicate that our method is able to control the FDR at the desired level. We also note that the FDR estimates are more conservative, i.e. they overestimate the true FDR, at higher levels of association between the mixed outcomes.

4. MEPS DATA ANALYSIS

In this section, we demonstrate the application of our PGEE framework and FDR control methodology to data from the Medical Expenditure Panel Survey (MEPS) (https://meps.ahrq.gov/). Our goal is to identify covariates on demographics, medical conditions, income, employment, health insurance coverage, and access to care that are associated with total annual drug spending and health status. We used the 2005 data and restricted attention to Medicare enrollees, 65 years of age and older, with an annual drug spending of $100 or more. We used the natural logarithm of total drug spending as our continuous outcome. As done in Zimmerman [16], we dichotomized health status into *fair or poor* (1) and *better than fair* (0), which formed our binary outcome. We considered a total of 40 covariates, and we used the same set of covariates to model both total drug spending and health status. The complete list of covariates with descriptions can be found in Table S7 in the Supplementary Materials. The data set also provides sampling weights for each observation, which we incorporated into the estimation methods. The final data set contains data for 2,953

Table 4. Estimated regression coefficients for log(drug spending) and health status outcomes under the joint and the separate PGEEs methods. A dot indicates that the covariate was not selected by that method, for that outcome.

	log(drug spending)		health status: fair or poor	
Covariate	Joint method	Separate method	Joint method	Separate method
Intercept	6.330	6.274	5.346	0.053
AGEX	.	.	.	.
SEX	-0.057	-0.065	0.259	.
RACE_WHITE	.	0.007	-0.387	.
MARRIED	-0.017	-0.021	.	.
LN_INCOME	.	.	-0.489	-0.004
LOW_INC_FAM	.	.	.	.
LANG_ENG	.	.	-1.017	.
TMTK_MORE_ONEHR	.	-0.004	.	.
DIFF_USC_TRAVEL	.	.	0.541	.
DIFF_USC_PHONE	.	.	.	.
MDUNAB	-0.127	-0.159	.	.
DNUNAB	.	.	0.891	.
PMUNAB	.	.	.	.
MDDLAY	-0.002	.	.	.
DNDLAY	.	.	.	.
PMDLAY	.	.	.	.
MCDEV	0.028	0.049	.	.
PRVEV	.	.	-0.273	.
TRIEV	.	.	.	.
DENTCK_LESS_ONEYR	.	.	0.556	.
CHOLCK_MORE_5YR	.	.	.	.
CHECK_MORE_1YR	-0.105	-0.132	.	.
FLUSHT_MORE_1YR	-0.111	-0.125	.	.
NOTEETH	0.006	0.009	.	.
STOOL	.	.	.	.
BOWEL	0.044	0.051	.	.
PHYACT	-0.060	-0.066	-0.893	-1.156
BMI	0.007	0.009	.	.
SEATBELT_NOT_ALWAYS	.	.	0.573	.
CANCER	-0.001	-0.009	.	.
DIABETES	0.412	0.407	0.624	.
COPD	0.060	0.087	0.034	.
CARDIOVASCULAR	0.418	0.425	.	.
ARTHRITIS	.	.	.	.
ASTHMA	0.155	0.201	.	.
STOMACH_ULCERS	.	.	.	.
MENTAL	0.368	0.365	0.383	.
KIDNEY	0.090	0.124	0.020	.
PRIO	0.090	0.087	0.176	0.019
EMPLOYED	-0.033	-0.038	-0.646	.

individuals, who represent 30,146,029 individuals of the U.S. population.

We applied both our joint PGEEs method as well as separate PGEEs to the responses. Similar to the simulations, four-fold cross-validation was used to select the optimal tuning parameters. Table 4 shows the estimated regression coefficients under the joint method and under the separate method. Sandwich-formula based standard errors for the regression coefficients from the joint model can be found in Table S8 in the Supplementary Materials.

For the continuous outcome–the logarithm of total drug spending–we observe that the joint and separate methods perform similarly in terms of both variable selection and estimation. For the joint model, the covariates with the largest coefficients are CARDIOVASCULAR, DIABETES, and MENTAL, which are binary indicators for the presence of a cardiovascular disease, some form of diabetes, and a mental disease, respectively. Intuitively this makes sense, as pre-existing medical conditions should have strong associations with drug spending.

For the binary outcome–the indicator of *fair or poor* health status–the joint model is able to detect signal from more covariates than the separate model. This is consistent with the results from our simulation studies, in which the gains in variable selection metrics through joint modeling are primarily made for the binary outcome coefficients. Of course, false discoveries could be a concern here. Hence, we estimated the FDR using the method described in Section 2.5 and found it to be 0.07. Because this level of FDR was acceptable to us, we decided to not perform cross-validation with explicit FDR control. However, we do acknowledge that given the benefits of FDR-controlled cross-validation that we noted in Section 2.5, our estimates might improve if we performed it. Interestingly, among all covariates selected by the joint method for the binary outcome, LANG_ENG has the largest coefficient in absolute value. The negative coefficient indicates that individuals whose language of comfort is English report better health status than other individuals. The moderate positive coefficient of DIFF_USC_TRAVEL indicates that individuals who find it difficult to travel to their Usual Source of Care (USC) provider report worse health statuses. Both the joint model and the separate model emphasize the importance of regular physical activity to good health, as seen in the large negative coefficient of PHYACT. In the joint model, the effect of dental health on health status can be seen via the coefficients of DNUNAB (individual was unable to receive dental treatment when it was required) and DENTCK_LESS_ONEYR (frequency of dental checkups are less than once a year). Next, income and employment are positively associated with good health as seen through the negative coefficients of LN_INCOME and EMPLOYED. Interestingly, other than DIABETES, most of the variables related to prior medical conditions have relatively small coefficients. SEATBELT_NOT_ALWAYS has a moderate positive coefficient, indicating that some individuals may have experienced poor health status due to a motor vehicle accident.

Finally, our joint method estimated the association parameter, ρ, to be 0.13. Our simulations indicate that the difference between the copula parameter, θ, and ρ, is roughly 0.10, so θ may be roughly regarded as 0.23.

5. DISCUSSION

We have provided a framework to perform simultaneous estimation and variable selection with correlated bivariate mixed outcomes using PGEEs. The simulation experiments and the MEPS data analysis indicate that the major gains in estimation and variable selection when outcomes are analyzed jointly occur in the binary outcome coefficients. Binary outcome regression coefficients are generally harder to estimate due to the smaller information content in binary data. Thus, by borrowing strength from the continuous outcomes through the correlation, joint estimation is able to outperform separate estimation for the binary outcome regression coefficients, while providing equivalent or better performance for the continuous outcome coefficients.

An obvious extension of our method is to allow for multivariate (more than two) mixed outcomes. Another useful extension would be to allow for longitudinal data for both outcomes. The challenge in each of these extensions lies in the estimation of the correlation structure. A challenge to estimating the FDR with general multivariate outcomes is that the proof of the KKT conditions in (18) does not hold for more than two outcomes. In the case of longitudinal outcomes, the iterative algorithm to solve the PGEEs would have to be modified as well. These are areas of research that we are currently investigating.

ACKNOWLEDGEMENTS

The authors wish to thank the two referees and associate editor for their helpful comments and suggestions, which greatly improved the content and clarity of the manuscript. This work was supported by the University of Connecticut Research Foundation to EDS and VD.

Received 14 September 2017

REFERENCES

[1] Benjamini, Y. and Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society. Series B (Methodological)* **57** 289–300. MR1325392

[2] Breheny, P. J. (2009). Regularized methods for high-dimensional and bi-level variable selection PhD thesis, University of Iowa, Iowa City, IA. MR2713542

[3] Fan, J. and Li, R. (2001). Variable selection via nonconcave penalized likelihood and its oracle properties. *Journal of the American Statistical Association* **96** 1348–1360.

[4] Fu, W. J. (2003). Penalized estimating equations. *Biometrics* **59** 126–132.

[5] Hunter, D. R. and Li, R. (2005). Variable selection using MM algorithms. *Annals of Statistics* **33** 1617–1642. MR2166557

[6] Johnson, B. A., Lin, D. and Zeng, D. (2008). Penalized estimating functions and variable selection in semiparametric regression models. *Journal of the American Statistical Association* **103** 672–680.

[7] Liu, J., Pei, Y., Papasian, C. J. and Deng, H.-W. (2009). Bivariate association analyses for the mixture of continuous and binary traits with the use of extended generalized estimating equations. *Genetic Epidemiology* **33** 217–227.

[8] Prentice, R. L. and Zhao, L. P. (1991). Estimating equations for parameters in means and covariances of multivariate discrete and continuous responses. *Biometrics* **47** 825–839. MR1141951

[9] Rochon, J. (1996). Analyzing bivariate repeated measures for discrete and continuous outcome variables. *Biometrics* **52** 740–750.

[10] Teixeira Pinto, A. and Normand, S.-L. T. (2009). Correlated bivariate continuous and binary outcomes: issues and applications. *Statistics in Medicine* **28** 1753–1773.

[11] Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society. Series B (Methodology)* **58** 267–288.

[12] Wang, L., Zhou, J. and Qu, A. (2012). Penalized Generalized Estimating Equations for High-Dimensional Longitudinal Data Analysis. *Biometrics* **68** 353–360.

[13] Wolfson, J. (2011). EEBoost: A general method for prediction and variable selection based on estimating equations. *Journal of the American Statistical Association* **106** 296–305.

[14] Yi, H., Breheny, P. J., Imam, N., Liu, Y. and Hoeschele, I. (2015). Penalized multimarker vs. single-marker regression methods for genome-wide association studies of quantitative traits. *Genetics* **199** 205–222.

[15] Zhang, C. H. (2007). Penalized linear unbiased selection Technical Report No. 2007–003, Department of Statistics and Bioinformatics, Rutgers University, Piscataway, NJ.

[16] Zimmerman, D. (2013). Analysis of mixed outcomes in econometrics: Applications in health economics. In *Analysis of mixed data: methods & applications* (A. R. de Leon and K. C. Chough, eds.) 11, 157–172. CRC Press.

[17] Zou, H. and Hastie, T. (2005). Regularization and variable selection via the elastic net. *Journal of the Royal Statistical Society: Series B (Methodology)* **67** 301–320. MR2137327

Ved Deshpande
Department of Statistics
University of Connecticut
215 Glenbrook Road, U-4120
Storrs, CT 06269-4120
USA
E-mail address: ved.deshpande@uconn.edu

Dipak K. Dey
Department of Statistics
University of Connecticut
215 Glenbrook Road, U-4120
Storrs, CT 06269-4120
USA
E-mail address: dipak.dey@uconn.edu

Elizabeth D. Schifano
Department of Statistics
University of Connecticut
215 Glenbrook Road, U-4120
Storrs, CT 06269-4120
USA
E-mail address: elizabeth.schifano@uconn.edu