UGMDR: A unified conceptual framework for detection of multifactor interactions underlying complex traits

Xiang-Yang Lou

Department of Biostatistics, University of Alabama at Birmingham, Birmingham, Alabama 31294, USA

xylou@uab.edu

TECHNICAL APPENDICES

Appendix A: Parameter Estimation in Statistical Models

A1. Generalized Linear Models

In generalized linear model (GLM), the response variable is assumed to follow a distribution from the exponential family that includes Normal, Poisson, Bernoulli, Multinomial, Gamma, Beta, Geometric, Negative Binomial and many other distributions. As the univariate can be viewed as a special case of multivariate where the dimensionality of variable is one, without loss of generality, multivariate exponential family of distributions is used here to illustrate the theory of generalized linear model. A GLM is characterized by three parts: stochastic component — the response distribution, systematic component — the linear predictor, and the link function between the linear predictor and the mean of the response variable. Specifically, considering a response vector $\mathbf{Y}$ in a GLM, the density function has such a general form,

$$f(\mathbf{Y};\boldsymbol{\theta},\boldsymbol{\varphi}) = \exp\left[\frac{\mathbf{Y}^T\boldsymbol{\theta} - b(\boldsymbol{\theta})}{a(\boldsymbol{\varphi})} + c(\mathbf{Y},\boldsymbol{\varphi})\right],$$

where $\boldsymbol{\theta}$ is the location parameter vector of interest, also known as the canonical parameter, $\boldsymbol{\varphi}$ is the nuisance scale (dispersion) parameter vector, $a(\cdot)$, $b(\cdot)$, and $c(\cdot)$ are known functions to specify a member of the exponential family. The expectation and variance of $\mathbf{Y}$ are given as, respectively,

$$E(\mathbf{Y}) = \boldsymbol{\mu} = b'(\boldsymbol{\theta}),$$

and

$$Var(\mathbf{Y}) = a(\boldsymbol{\varphi})b''(\boldsymbol{\theta}),$$

where $b'(\boldsymbol{\theta}) = \frac{\partial b(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}^T} = \left(\frac{\partial b(\boldsymbol{\theta})}{\partial \theta_i} \right)$ and $b''(\boldsymbol{\theta}) = \frac{\partial^2 b(\boldsymbol{\theta})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^T} = \left(\frac{\partial^2 b(\boldsymbol{\theta})}{\partial \theta_i \partial \theta_j} \right)$ are the first-order and second-order derivatives of $b(\boldsymbol{\theta})$, respectively. Denote $V(\boldsymbol{\mu}) = b''(\boldsymbol{\theta})$, called variance function, to highlight that $b''(\boldsymbol{\theta})$ depends on $\boldsymbol{\mu}$. The variance of $\mathbf{Y}$ depends on both location and scale parameters.

Assuming the set of explanatory variables influence the outcome only through a linear function, the linear predictor upon which $\boldsymbol{\theta}$ depends can be written as,

$$l(\boldsymbol{\mu}) = \boldsymbol{\eta} = \mathbf{I}\boldsymbol{\beta}_0 + \mathbf{X}_t\boldsymbol{\beta}_t + \mathbf{X}_c\boldsymbol{\beta}_c = \mathbf{X}\boldsymbol{\beta},$$

where $l(\cdot)$ is an invertible link function, $\boldsymbol{\beta}$ is the effect vector probably consisting of $\boldsymbol{\beta}_0$, $\boldsymbol{\beta}_t$ and $\boldsymbol{\beta}_c$ for the intercept(s), the target effects of interest and the covariate effects, respectively, $\mathbf{X}$ is the corresponding design matrix consisting of block matrices $\mathbf{I}$, $\mathbf{X}_t$ and $\mathbf{X}_c$, and $\mathbf{I}$ is a unit matrix. The link function $l(\boldsymbol{\mu}) = \boldsymbol{\eta}$ relates the mean to the linear predictor. When the link function makes the linear predictor $\boldsymbol{\eta}$ the same as the canonical parameter $\boldsymbol{\theta}$ (i.e., $\boldsymbol{\theta} \equiv \boldsymbol{\eta}$), it is called the canonical (or natural) link which has nice mathematical and numerical properties in connection with the estimation process.

GLMs cover a wide range of practical situations by allowing for response variables that follow any probability distribution in the exponential family of distributions (not only

simply normal distributions) and for the link function of the response variable to vary linearly with the predictor (not only assuming that the response itself must vary linearly). For dichotomous phenotypes following a Bernoulli distribution, the natural link function is a log-odds (or logit),

$$\theta = \eta = \log it(\mu) = \log\left(\frac{\mu}{1-\mu}\right),$$

while $a(\phi) = 1$, $b(\theta) = \ln\left(1 + e^{\theta}\right)$, and $c(Y, \phi) = 0$. There are also several popular alternative link choices for binomial data including probit and complementary log-log. For count outcomes following a Poisson distribution, the canonical link is the log function,

$$\theta = \eta = \log(\mu),$$

while $a(\phi) = 1$, $b(\theta) = e^{\theta}$, and $c(Y, \phi) = -\ln(Y!)$. For continuous responses having a univariate normal distribution, $N\left(\mu, \sigma^2\right)$, the natural link is the identity, i.e., $\theta = \eta = \mu$, while $a(\phi) = \sigma^2$, $b(\theta) = \frac{\theta^2}{2}$, and $c(Y, \phi) = -\frac{Y^2}{2\sigma^2} - \ln\left(\sqrt{2\pi}\sigma\right)$.

A1.1. Likelihood and Parameter Estimation in GLMs

The log-likelihood, also known as the support, for a set of independent observations $\mathbf{y}_i$ ($i = 1, 2, \cdots, N$) is,

$$\ln L(\mathbf{y}_1, \mathbf{y}_2, \cdots, \mathbf{y}_N; \boldsymbol{\beta}, \boldsymbol{\varphi}, \mathbf{X}_1, \mathbf{X}_2, \cdots, \mathbf{X}_N) = \sum_{i=1}^{N}\left[\frac{\mathbf{y}_i^T\boldsymbol{\theta}_i - b(\boldsymbol{\theta}_i)}{a(\boldsymbol{\varphi})} + c(\mathbf{y}_i, \boldsymbol{\varphi})\right],$$

where $\mathbf{X}_i$s are the design matrix. The score, Hessian matrix and the expected Hessian matrix are, respectively,

$$U(\boldsymbol{\beta}) = \frac{\partial \ln L}{\partial \boldsymbol{\beta}^T} = \sum_{i=1}^{N}\frac{1}{a(\boldsymbol{\varphi})}\mathbf{X}_i^T\left(\frac{\partial \boldsymbol{\mu}_i}{\partial \boldsymbol{\eta}_i}\right)^T\mathbf{V}_i^{-1}(\mathbf{y}_i - \boldsymbol{\mu}_i)\underline{\underline{\boldsymbol{\eta}_i \equiv \boldsymbol{\theta}_i}}\sum_{i=1}^{N}\frac{1}{a(\boldsymbol{\varphi})}\mathbf{X}_i^T(\mathbf{y}_i - \boldsymbol{\mu}_i),$$

$$\mathbf{H}(\boldsymbol{\beta}) = \frac{\partial^2 \ln L}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}^T} = \sum_{i=1}^{N} -\frac{1}{a(\boldsymbol{\varphi})} \left[\mathbf{X}_i^T \left(\frac{\partial \boldsymbol{\mu}_i}{\partial \boldsymbol{\eta}_i} \right)^T \frac{\partial \boldsymbol{\theta}_i}{\partial \boldsymbol{\mu}_i} \frac{\partial \boldsymbol{\mu}_i}{\partial \boldsymbol{\eta}_i} \mathbf{X}_i - \mathbf{X}_i^T \frac{\partial \left(\frac{\partial \boldsymbol{\theta}_i}{\partial \boldsymbol{\mu}_i} \frac{\partial \boldsymbol{\mu}_i}{\partial \boldsymbol{\eta}_i} \right)^T}{\partial \boldsymbol{\beta}} [(\mathbf{y}_i - \boldsymbol{\mu}_i) \otimes \mathbf{I}] \right],$$

$$\underline{\underline{\boldsymbol{\eta}_i \equiv \boldsymbol{\theta}_i}} \sum_{i=1}^{N} -\frac{1}{a(\boldsymbol{\varphi})} \mathbf{X}_i^T \mathbf{V}_i \mathbf{X}_i$$

and

$$E\mathbf{H}(\boldsymbol{\beta}) = E\left(\frac{\partial^2 \ln L}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}^T} \right) = \sum_{i=1}^{N} -\frac{1}{a(\boldsymbol{\varphi})} \left[\mathbf{X}_i^T \left(\frac{\partial \boldsymbol{\mu}_i}{\partial \boldsymbol{\eta}_i} \right)^T \frac{\partial \boldsymbol{\theta}_i}{\partial \boldsymbol{\mu}_i} \frac{\partial \boldsymbol{\mu}_i}{\partial \boldsymbol{\eta}_i} \mathbf{X}_i \right],$$

where $\mathbf{V}_i = b''(\boldsymbol{\theta}_i) = V(\boldsymbol{\mu}_i)$, $\mathbf{I}$ is a unit matrix and $\otimes$ represents a Kronecker product.

The GLM can be fitted by maximum likelihood (ML) estimation method. Note that estimation of $\boldsymbol{\beta}$ does not require knowledge of $\boldsymbol{\varphi}$, implying that one first estimates $\boldsymbol{\beta}$ and estimates $\boldsymbol{\varphi}$ afterwards on the basis of the estimate $\hat{\boldsymbol{\beta}}$. The ML estimator can be derived by setting the score equal to zero and solving the resulting likelihood equations. In general, there is no closed form of ML estimates available for GLMs, and numerical algorithms are required in fitting parameters to data. The Newton-Raphson method and the Fisher's scoring method are two well worked methods for finding ML estimates in GLMs. In the Newton-Raphson method, the estimate at the $(t+1)$ th cycle of iteration is,

$$\hat{\boldsymbol{\beta}}^{(t+1)} = \hat{\boldsymbol{\beta}}^{(t)} - \left\{ \mathbf{H}\left(\hat{\boldsymbol{\beta}}^{(t)} \right) \right\}^{-1} U(\hat{\boldsymbol{\beta}}^{(t)}),$$

where $\hat{\boldsymbol{\beta}}^{(t)}$ is the estimated effects at the t th cycle of iteration, $U(\hat{\boldsymbol{\beta}}^{(t)})$ and $\mathbf{H}\left(\hat{\boldsymbol{\beta}}^{(t)} \right)$ are, respectively, the estimated score and estimated Hessian matrix in which $\hat{\boldsymbol{\beta}}^{(t)}$ is used in place of $\boldsymbol{\beta}$. This algorithm is repeated until convergence in $\boldsymbol{\beta}$ or the log-likelihood. In the Fisher's scoring method, the estimate at the $(t+1)$ th cycle of iteration is,

$$\hat{\boldsymbol{\beta}}^{(t+1)} = \hat{\boldsymbol{\beta}}^{(t)} - \left\{ E\mathbf{H}\left(\hat{\boldsymbol{\beta}}^{(t)} \right) \right\}^{-1} U(\hat{\boldsymbol{\beta}}^{(t)}),$$

where $\hat{\boldsymbol{\beta}}^{(t)}$ is the estimated effects at the t th cycle of iteration, $U(\hat{\boldsymbol{\beta}}^{(t)})$ and $E\mathbf{H}\left(\hat{\boldsymbol{\beta}}^{(t)}\right)$ are, respectively, the estimated score and the estimated value of the expected Hessian matrix.

Conventionally, either a Newton-Raphson method or a Fisher's scoring method can be implemented with iteratively weighted (or reweighted) least squares method (IWLS) that is formally similar to the familiar solution for weighted least squares estimates in linear models; the Newton-Raphson method and the Fisher's scoring method coincide when the canonical link function is used. Each iteration step for $\hat{\boldsymbol{\beta}}$ in IWLS can be written as,

$$\hat{\boldsymbol{\beta}}^{(t+1)} = \hat{\boldsymbol{\beta}}^{(t)} + \left(\sum_{i=1}^{N} \mathbf{X}_i^T \hat{\mathbf{W}}_i \mathbf{X}_i\right)^{-1} \sum_{i=1}^{N} \mathbf{X}_i^T \hat{\mathbf{W}}_i \hat{\Delta}_i^{-1} (\mathbf{y}_i - \hat{\boldsymbol{\mu}}_i) = \left(\sum_{i=1}^{N} \mathbf{X}_i^T \hat{\mathbf{W}}_i \mathbf{X}_i\right)^{-1} \sum_{i=1}^{N} \mathbf{X}_i^T \hat{\mathbf{W}}_i \mathbf{z}_i ,$$

where $\hat{\boldsymbol{\mu}}_i$ is the estimated mean, $\hat{\mathbf{W}}_i$ is the estimate of the working weight $\mathbf{W}_i$ by using $\hat{\boldsymbol{\beta}}^{(t)}$ in place of $\boldsymbol{\beta}$, $\mathbf{W}_i = \left(\frac{\partial \boldsymbol{\mu}_i}{\partial \boldsymbol{\eta}_i}\right)^T \frac{\partial \boldsymbol{\theta}_i}{\partial \boldsymbol{\mu}_i} \frac{\partial \boldsymbol{\mu}_i}{\partial \boldsymbol{\eta}_i} = \boldsymbol{\Delta}_i^T \mathbf{V}_i^{-1} \boldsymbol{\Delta}_i$ in Fisher's scoring method or $\mathbf{W}_i = -a(\boldsymbol{\varphi}) \frac{\partial^2 \ln L_i}{\partial \boldsymbol{\eta} \partial \boldsymbol{\eta}^T}$ where L_i is the component likelihood of individual i in Newton-Raphson method, $\hat{\boldsymbol{\Delta}}_i$ is the estimate of $\boldsymbol{\Delta}_i = \frac{\partial \boldsymbol{\mu}_i}{\partial \boldsymbol{\eta}_i}$, and $\mathbf{z}_i = \mathbf{X}_i \hat{\boldsymbol{\beta}}^{(t)} + \hat{\boldsymbol{\Delta}}_i^{-1} (\mathbf{y}_i - \hat{\boldsymbol{\mu}}_i)$ is the working response.

Some of the exponential families on which GLMs are based include unknown dispersion parameter $\boldsymbol{\varphi}$. The scale parameter $\boldsymbol{\varphi}$ can be separately estimated from $\boldsymbol{\beta}$ after computing residuals with $\hat{\boldsymbol{\beta}}$. There are various ways of estimating $\boldsymbol{\varphi}$. Although this parameter can, in principle, be estimated by maximum likelihood as well, it is more common to use a "method of moments" estimator. Unbiased estimator of $\boldsymbol{\varphi}$ can be obtained via Pearson's Chi square as,

$$a(\hat{\boldsymbol{\varphi}}) = \frac{1}{N-p} \sum_{i=1}^{N} (\mathbf{y}_i - \hat{\boldsymbol{\mu}}_i)^T \mathbf{V}_i^{-1} (\mathbf{y}_i - \hat{\boldsymbol{\mu}}_i),$$

where p is the number of independent parameters estimated in $\boldsymbol{\beta}$ ($p = 0$ if $\boldsymbol{\beta}$ is known).

A1.2. Residuals under the Null Hypothesis in GLMs

Under the null hypothesis of no target effects (i.e., $\boldsymbol{\beta}_t = \mathbf{0}$), fit $\hat{\boldsymbol{\beta}}_0$ and $\hat{\boldsymbol{\beta}}_c$ as well as $\hat{\boldsymbol{\varphi}}$ to data. Then, the residuals can be computed for forming the statistic. Among several types of residuals considered in GLMs including response residuals, working residuals, Pearson residuals, Anscombe residuals and deviance residuals, the score-contributed residual is suggested to use here, which indicates a subject-specific contribution to the residual score,

$$\mathbf{r}_i = \frac{1}{a(\hat{\boldsymbol{\varphi}})} \hat{\boldsymbol{\Delta}}_i^T \mathbf{V}_i^{-1} \left(\mathbf{y}_i - \hat{\boldsymbol{\mu}}_i \right), \tag{A1}$$

where $\hat{\boldsymbol{\mu}}_i = l^{-1}(\mathbf{I}\hat{\boldsymbol{\beta}}_0 + \mathbf{X}_{ci}\hat{\boldsymbol{\beta}}_c)$ and $\hat{\boldsymbol{\Delta}}_i$ is an estimate of $\boldsymbol{\Delta}_i$ in which $\hat{\boldsymbol{\beta}}_0$, $\mathbf{0}$ and $\hat{\boldsymbol{\beta}}_c$ are used in place of $\boldsymbol{\beta}_0$, $\boldsymbol{\beta}_t$ and $\boldsymbol{\beta}_c$, respectively. When the canonical link is used,

$$\mathbf{r}_i = \frac{1}{a(\hat{\boldsymbol{\varphi}})} \left(\mathbf{y}_i - \hat{\boldsymbol{\mu}}_i \right).$$

The product of this residual and the corresponding value(s) of predictor variable(s) forms the individual contribution(s) to the first partial derivative of the log likelihood, indicating the input of this subject in the corresponding estimating equation with respect to the given covariate. Thus, using this kind of residuals is coherent with the estimating equations.

As an alternative, other residuals may be also used. They can be computed as follows, respectively (Fox, 2008; Gill, 2001; Lindsey, 1997; Pierce and Schafer, 1986). Response (raw fitted value) residual,

$$\mathbf{r}_i^R = \mathbf{y}_i - \hat{\boldsymbol{\mu}}_i .$$

Working residual, computed from the working response in the final IWLS fitting,

$$\mathbf{r}_i^W = \left. \frac{\partial \boldsymbol{\eta}_i}{\partial \boldsymbol{\mu}_i} \right|_{\boldsymbol{\mu}_i = \hat{\boldsymbol{\mu}}_i} \left(\mathbf{y}_i - \hat{\boldsymbol{\mu}}_i \right) = \mathbf{z}_i - \hat{\boldsymbol{\eta}}_i .$$

Pearson residual,

$$\mathbf{r}_i^P = \mathbf{V}^{-\frac{1}{2}}(\hat{\boldsymbol{\mu}}_i)\left(\mathbf{y}_i - \hat{\boldsymbol{\mu}}_i \right).$$

Anscombe (Anscombe, 1961) residual,

$$\mathbf{r}_i^A = \left[\frac{A(y_{ij}) - A(\hat{\mu}_{ij})}{A'(\hat{\mu}_{ij})\sqrt{V(\hat{\mu}_{ij})}} \right],$$

where $A(z) = \int V^{-\frac{1}{3}}(z)dz$. Deviance residual can be calculated by the extension to the multivariate case (Pregibon, 1981),

$$\mathbf{r}_i^D = \sqrt{\frac{d_i}{\sum_j (y_{ij} - \hat{\mu}_{ij})^2}}(\mathbf{y}_i - \hat{\boldsymbol{\mu}}_i),$$

where $d_i = 2\{\ln L(\mathbf{y}_i; \tilde{\boldsymbol{\theta}}_i, \boldsymbol{\varphi}) - \ln L(\mathbf{y}_i; \hat{\boldsymbol{\theta}}_i, \boldsymbol{\varphi})\}$, $\ln L(\mathbf{y}_i; \tilde{\boldsymbol{\theta}}_i, \boldsymbol{\varphi})$ is the likelihood of observation $\mathbf{y}_i$ under the saturated model, and $\ln L(\mathbf{y}_i; \hat{\boldsymbol{\theta}}_i, \boldsymbol{\varphi})$ is the likelihood under the fitted model.

Appendix A2: Quasi-likelihood Models

A quasi-likelihood model (QLM) only specifies the link function and the relationship between the first two moments but does not necessarily specify the complete distribution of the response variable. Consider a response variable Y has expectation μ and variance $a(\phi)V(\mu)$, where ϕ is the nuisance scale parameter, and $a(\cdot)$ and $V(\cdot)$ are some known functions. Suppose there is a known function between μ and a set of predictor variables as,

$$l(\mu) = \eta = \mathbf{x}^T\boldsymbol{\beta},$$

where $l(\mu)$ is an invertible link function, $\boldsymbol{\beta}$ is the effect vector probably consisting of the intercept β_0, the target effects $\boldsymbol{\beta}_t$ and the covariate effects $\boldsymbol{\beta}_c$, and $\mathbf{x}$ is the corresponding explanatory variable vector. The quasi-likelihood function $Q(Y, \mu)$ is defined by the relation,

$$\frac{\partial Q(Y, \mu)}{\partial \mu} = \frac{Y - \mu}{a(\phi)V(\mu)},$$

or equivalently,

$$Q(Y,\,\mu) = \int^{\mu} \frac{Y-\nu}{a(\phi)V(\nu)}\,d\nu + c(Y,\phi),$$

where $c(\cdot)$ are known functions.

A2.1. Quasi-score Equations and Parameter Estimation in QLMs

For a set of independent responses y_i ($i = 1, 2, \cdots, N$), there is the quasi-score function that behaves like the score function in GLMs,

$$\frac{\partial Q}{\partial \beta_j} = \sum\nolimits_{i=1}^{N} x_{ij}\Delta_i \frac{1}{a(\phi)V(\mu_i)}(y_i - \mu_i),$$

where $\Delta_i = \frac{d\mu_i}{d\eta_i}$. The second-order partial derivative is,

$$\frac{\partial^2 Q}{\partial \beta_k \partial \beta_j} = \sum\nolimits_{i=1}^{N} -\frac{1}{a(\phi)}\left[x_{ij}x_{ik}\frac{\Delta_i^2}{V(\mu_i)} - x_{ij}\frac{\partial\left(\frac{\Delta_i}{V(\mu_i)}\right)}{\partial \beta_k}(y_i - \mu_i)\right],$$

where the expectation of the second term on the right side of the equal sign is 0. Setting quasi-score functions to zero leads to a set of quasi-likelihood estimating equations. Then, QLMs can be fitted using a straightforward extension of the algorithms used to fit GLMs.

A2.2. Residuals under the Null Hypothesis in QLMs

Similar to that in GLMs, the quasi-score residual can be computed by,

$$r_i = \hat{\Delta}_i \frac{1}{a(\hat{\phi})V(\hat{\mu}_i)}(y_i - \hat{\mu}_i), \qquad \text{(A2)}$$

where $\hat{\mu}_i$, $\hat{\Delta}_i$, $\hat{\beta}_0$ and $\hat{\boldsymbol{\beta}}_c$, as well as $\hat{\phi}$ when necessary, are the quasi-likelihood estimates under the null hypothesis of no target effects (i.e., $\boldsymbol{\beta}_t = \mathbf{0}$).

Appendix A3: Generalized Estimating Equations Models

Generalized estimating equations (GEE) model requires only to specify a functional form for relationships between the outcome variable and the explanatory factors and between the mean and the variance of the marginal distribution, avoiding the need to model the multivariate distribution for data. Specifically, letting $\mathbf{Y} = (Y_1, Y_2, \cdots, Y_K)^T$ be a group of response variables, suppose that (1) $E(\mathbf{Y}) = (\mu_j) = \boldsymbol{\mu}$ and there is a link function relating the expectation of $\mathbf{Y}$ to a linear predictor,

$$\mathbf{l}(\boldsymbol{\mu}) = \boldsymbol{\eta} = \mathbf{I}\boldsymbol{\beta}_0 + \mathbf{X}_t\boldsymbol{\beta}_t + \mathbf{X}_c\boldsymbol{\beta}_c = \mathbf{X}\boldsymbol{\beta},$$

where $\boldsymbol{\beta}$ is the effect vector that consists of $\boldsymbol{\beta}_0$, $\boldsymbol{\beta}_t$ and $\boldsymbol{\beta}_c$ for the intercept, the target effects and covariate effects, respectively, and $\mathbf{X}$ is the corresponding incidence matrix; and (2) the variance is a function of the mean, $Var(Y_j) = a_j(\phi_j)V_j(\mu_j)$, ϕ_j is the nuisance scale parameter, and $a_j(\cdot)$ and $V_j(\cdot)$ are some known functions. For convenience, denote in the vector form, $\boldsymbol{\varphi} = (\phi_j)$ and $a(\boldsymbol{\varphi}) = \lfloor a_j(\phi_j) \rfloor$ when necessary.

The GEE models for diverse scenarios may take the same general form. For repeated measurements, Y_i may have the same parameters $\boldsymbol{\beta}$, $a(\cdot)$ and $V(\cdot)$, and regressor values $\mathbf{x}$, and thus the design matrix $\mathbf{X} = (\mathbf{x}, \mathbf{x}, \cdots, \mathbf{x})^T$. In a clustered design or a longitudinal study, the components of $\mathbf{Y}$ may share the same $\boldsymbol{\beta}$, $a(\cdot)$ and $V(\cdot)$, but have their own regressors $\mathbf{x}_j$ s, and thus $\mathbf{X} = (\mathbf{x}_1, \mathbf{x}_2, \cdots, \mathbf{x}_K)^T$. For grouped phenotypes, each Y_j has the component-specific predictor values, functions and parameters including $\mathbf{x}_j$, $a_j(\cdot)$, $V_j(\cdot)$, ϕ_j and $\boldsymbol{\beta}^{(j)}$, and thus resulting in the block effect vector and block incidence matrix, respectively, as follows,

$$\boldsymbol{\beta} = \begin{pmatrix} \boldsymbol{\beta}^{(1)} \\ \boldsymbol{\beta}^{(2)} \\ \vdots \\ \boldsymbol{\beta}^{(K)} \end{pmatrix},$$

and

$$\mathbf{X} = \begin{pmatrix} \mathbf{x}_1^T & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{x}_2^T & \cdots & \mathbf{0} \\ \vdots & \vdots & \cdots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{x}_K^T \end{pmatrix}.$$

In application to detection of overall effects and/or pleiotropic effects on multiple traits, the components of $\mathbf{Y}$ may share the same regressor $\mathbf{x}$, but have component-specific parameters $a_j(\cdot)$, $V_j(\cdot)$, ϕ_j and $\boldsymbol{\beta}^{(i)}$.

A3.1. GEE and Parameter Estimation in GEE Models

Considering a set of data $\mathbf{y}^T = (\mathbf{y}_1^T, \mathbf{y}_2^T, \cdots, \mathbf{y}_N^T)$ that is decomposed into N strata and the $\mathbf{y}_i$s are uncorrelated with each other, the estimating function is formed via a set of score or quasi-score functions,

$$U(\boldsymbol{\beta}) = \sum_{i=1}^{N} \mathbf{D}_i^T \boldsymbol{\Sigma}_i^{-1} (\mathbf{y}_i - \boldsymbol{\mu}_i),$$

where $\mathbf{D}_i = \frac{\partial \boldsymbol{\mu}_i}{\partial \boldsymbol{\beta}} = \boldsymbol{\Delta}_i \mathbf{X}_i$, $\boldsymbol{\Delta}_i = \frac{\partial \boldsymbol{\mu}_i}{\partial \boldsymbol{\eta}_i}$, $\boldsymbol{\Sigma}_i = \mathbf{A}_i^{\frac{1}{2}} \mathbf{R}(\boldsymbol{\alpha}) \mathbf{A}_i^{\frac{1}{2}}$ is a working variance with a given correlation structure, $\mathbf{A}_i = diag\left[a_j(\phi_j) V_j(\mu_{ij})\right]$, a diagonal matrix with $a_j(\phi_j) V_j(\mu_{ij})$ as the jth diagonal element, and $\mathbf{R}(\boldsymbol{\alpha})$ is a working correlation matrix that may depend on some unknown parameter vector $\boldsymbol{\alpha}$. $\boldsymbol{\Sigma}_i$ needs not to be equal to the true one, although the better it hits, the better will be the precision of the estimates. There are several choices for the working correlation matrix. The variance structure is commonly specified as independence, exchangeable, autoregressive, stationary *m*-dependent, and unstructured. Parameter estimates from the GEE are consistent even when the covariance structure is incorrectly specified and the variance function $V(\cdot)$ is not the true one, under mild regularity conditions.

The function $U(\boldsymbol{\beta})$ behaves like the derivative of a log-likelihood (i.e., a score function). There is the second derivative as,

$$H(\boldsymbol{\beta}) = -\sum_{i=1}^{N}\left\{\mathbf{D}_i^T\boldsymbol{\Sigma}_i^{-1}\mathbf{D}_i - \frac{\partial\left(\mathbf{D}_i^T\boldsymbol{\Sigma}_i^{-1}\right)}{\partial\boldsymbol{\beta}}\left[(\mathbf{y}_i - \boldsymbol{\mu}_i)\otimes\mathbf{I}\right]\right\},$$

where $\mathbf{I}$ is a unit matrix and $\otimes$ represents a Kronecker product. The estimates of $\boldsymbol{\beta}$ can be found by solving estimating equation $U(\boldsymbol{\beta}) = \mathbf{0}$. Estimation is typically accomplished through a series of iterations between a modified Fisher's scoring algorithm for $\boldsymbol{\beta}$ and moment estimation of correlation parameters $\boldsymbol{\alpha}$ and scale parameter $\boldsymbol{\varphi}$. Given current estimates $\hat{\boldsymbol{\alpha}}$ and $\hat{\boldsymbol{\varphi}}$ of the nuisance parameters, the following modified iterative procedure is for $\boldsymbol{\beta}$,

$$\hat{\boldsymbol{\beta}}^{(t+1)} = \hat{\boldsymbol{\beta}}^{(t)} + \left(\sum_{i=1}^{N}\hat{\mathbf{D}}_i^T\hat{\boldsymbol{\Sigma}}_i^{-1}\hat{\mathbf{D}}_i\right)^{-1}\left[\sum_{i=1}^{N}\hat{\mathbf{D}}_i^T\hat{\boldsymbol{\Sigma}}_i^{-1}(\mathbf{y}_i - \hat{\boldsymbol{\mu}}_i)\right].$$

The working correlation matrix $\mathbf{R}_i(\boldsymbol{\alpha})$ and $\boldsymbol{\varphi}$ are estimated by the moments method. Using the current values of parameters calculates the current Pearson residuals defined as,

$$r_{ij}^{P} = \frac{y_{ij} - \hat{\mu}_{ij}}{\sqrt{V(\hat{\mu}_{ij})}}.$$

Then, $\boldsymbol{\varphi}$ can be estimated by,

$$a(\hat{\boldsymbol{\varphi}}) = \frac{\sum_{i=1}^{N}\sum_{j=1}^{K_i} r_{ij}}{\sum_{i=1}^{N} K_i - p}.$$

The specific estimator for $\boldsymbol{\alpha}$ depends on the choice of $\mathbf{R}(\boldsymbol{\alpha})$; the general approach is by the function of,

$$\hat{R}_{uv} = \frac{\sum_{i=1}^{N} r_{iu} r_{iv}}{N - p}.$$

A3.2. Residuals under the Null Hypothesis in GEE Models

After fitting the model under the null hypothesis, the residuals can be computed for several different purposes, e.g., a repeated measurement study, a longitudinal study, a clustered design, and multivariate analysis,

$$\mathbf{r}_i = \hat{\mathbf{\Delta}}_i^T \hat{\mathbf{\Sigma}}_i^{-1} (\mathbf{y}_i - \hat{\boldsymbol{\mu}}_i). \tag{A3}$$

where $\hat{\boldsymbol{\mu}}_i$, $\hat{\mathbf{\Delta}}_i$ and $\hat{\mathbf{\Sigma}}_i$ are, respectively, the GEE estimates of the mean, matrices $\mathbf{\Delta}_i$ and $\mathbf{\Sigma}_i$ under the null hypothesis, $\boldsymbol{\beta}_t = \mathbf{0}$. When all the components of $\mathbf{y}_i$ have the same target effect parameter $\boldsymbol{\beta}_t$ and the same predictor vector $\mathbf{x}_t$ as in a repeated measurement study, the residuals can be further averaged over these measurements for a better estimation,

$$r_{ij} = \tfrac{1}{K_i} \mathbf{1}^T \hat{\mathbf{\Delta}}_i^T \hat{\mathbf{\Sigma}}_i^{-1} (\mathbf{y}_i - \hat{\boldsymbol{\mu}}_i), \; j = 1, 2, \cdots, K_i, \tag{A4}$$

where $\mathbf{1}$ is a vector of which all components are 1 and K_i is the dimensionality of $\mathbf{y}_i$.

Appendix A4: Multinomial Logistic Models

Considering a multinomial response variable, index response categories by $1, 2, \cdots, K$, and denote the outcome $\mathbf{Y} = (Y_1, Y_2, \cdots, Y_K)^T$ where Y_j is an indicator variable taking value 1 if the observed category is j and 0 otherwise. Then, $E(\mathbf{Y}) = (\mu_j) = \boldsymbol{\mu}$, $Var(Y_j) = \mu_j(1-\mu_j) = V(\mu_j, \mu_j)$, and $Cov(Y_i, Y_j) = -\mu_i \mu_j = V(\mu_i, \mu_j)$. A polytomous logit model can be formed by nominating one of the response categories as a baseline and then formulating a set of $K-1$ logits for all other categories relative to the baseline. Without loss of generality, using category K as the baseline, then, the multinomial density has a multivariate exponential form,

$$f(\mathbf{Y}; \boldsymbol{\theta}, \boldsymbol{\varphi}) = \exp\left[\frac{\mathbf{Y}^T \boldsymbol{\theta} - b(\boldsymbol{\theta})}{a(\boldsymbol{\varphi})} + c(\mathbf{Y}, \boldsymbol{\varphi}) \right],$$

where $\boldsymbol{\theta} = \left[\ln\left(\frac{\mu_j}{\mu_K} \right) \right]$, $b(\boldsymbol{\theta}) = \ln\left(1 + \sum_{j=1}^{K-1} e^{\theta_j}\right)$, $a(\boldsymbol{\varphi}) = 1$, and $c(\mathbf{Y}, \boldsymbol{\varphi}) = 0$. Suppose the logit function to link the expectation to the linear predictor,

$$\theta_j \equiv \eta_j = l(\mu_j) = \mathbf{x}^T \boldsymbol{\beta}^{(j)}, \quad j = 1, 2, \cdots, K-1,$$

where $\boldsymbol{\beta}^{(j)}$ consists of $\beta_0^{(j)}$, $\boldsymbol{\beta}_t^{(j)}$ and $\boldsymbol{\beta}_c^{(j)}$ for the intercept, the target effects and the covariate effects, respectively. The link can be rewritten in the vector-matrix notation,

$$\boldsymbol{\theta} = \begin{pmatrix} \mathbf{x}^T & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{x}^T & \cdots & \mathbf{0} \\ \vdots & \vdots & \cdots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{x}^T \end{pmatrix} \begin{pmatrix} \boldsymbol{\beta}^{(1)} \\ \boldsymbol{\beta}^{(2)} \\ \vdots \\ \boldsymbol{\beta}^{(K-1)} \end{pmatrix} = \left(\mathbf{I} \otimes \mathbf{x}^T\right)\boldsymbol{\beta} = \mathbf{X}\boldsymbol{\beta},$$

where $\mathbf{I}$ is a $(K-1)\times(K-1)$ unit matrix and $\otimes$ represents a Kronecker product. This model is analogous to a logistic regression model, except that the probability distribution of the response is multinomial instead of binomial and there are $K-1$ equations instead of one. The polytomous logit model is a special case of GLMs with a canonical link in Appendix A1.

A4.1. Likelihood and Parameter Estimation in Multinomial Logistic Models

The log-likelihood for a set of independent observations $\mathbf{y}_i$ $(i = 1, 2, \cdots, N)$ is,

$$\ln L(\mathbf{y}_1, \mathbf{y}_2, \cdots, \mathbf{y}_N; \boldsymbol{\beta}, \boldsymbol{\varphi}, \mathbf{X}_1, \mathbf{X}_2, \cdots, \mathbf{X}_N) = \sum_{i=1}^{N} \left[\frac{\mathbf{y}_i^T \boldsymbol{\theta}_i - b(\boldsymbol{\theta}_i)}{a(\boldsymbol{\varphi})} + c(\mathbf{y}_i, \boldsymbol{\varphi}) \right],$$

where $\mathbf{X}_i$ s are the design matrix. The score and Hessian matrix are, respectively,

$$U(\boldsymbol{\beta}) = \sum_{i=1}^{N} \frac{1}{a(\boldsymbol{\varphi})} \mathbf{D}_i^T \mathbf{V}_i^{-1} (\mathbf{y}_i - \boldsymbol{\mu}_i) \underline{\underline{\boldsymbol{\eta}_i \equiv \boldsymbol{\theta}_i}} \sum_{i=1}^{N} \frac{1}{a(\boldsymbol{\varphi})} \mathbf{X}_i^T (\mathbf{y}_i - \boldsymbol{\mu}_i),$$

and

$$\mathbf{H}(\boldsymbol{\beta}) = \sum_{i=1}^{N} \frac{1}{a(\boldsymbol{\varphi})} \left[-\mathbf{X}_i^T \left(\frac{\partial \boldsymbol{\mu}_i}{\partial \boldsymbol{\eta}_i} \right)^T \frac{\partial \boldsymbol{\theta}_i}{\partial \boldsymbol{\mu}_i} \frac{\partial \boldsymbol{\mu}_i}{\partial \boldsymbol{\eta}_i} \mathbf{X}_i + \mathbf{X}_i^T \frac{\partial \left(\frac{\partial \boldsymbol{\theta}_i}{\partial \boldsymbol{\mu}_i} \frac{\partial \boldsymbol{\mu}_i}{\partial \boldsymbol{\eta}_i} \right)^T}{\partial \boldsymbol{\beta}} \left[(\mathbf{y}_i - \boldsymbol{\mu}_i) \otimes \mathbf{I} \right] \right],$$

$$\underline{\underline{\boldsymbol{\eta}_i \equiv \boldsymbol{\theta}_i}} \left(\sum_{i=1}^{N} -\frac{1}{a(\boldsymbol{\varphi})} \mathbf{X}_i^T \mathbf{V}_i \mathbf{X}_i \right)$$

where $\mathbf{D}_i = \frac{\partial \boldsymbol{\mu}_i}{\partial \boldsymbol{\beta}} = \frac{\partial \boldsymbol{\mu}_i}{\partial \boldsymbol{\eta}_i} \mathbf{X}_i$, $\mathbf{V}_i = \left[V(\mu_{ij}, \mu_{ik}) \right]$ is the variance-covariance matrix whose element at row j and at column k is $V(\mu_{ij}, \mu_{ik})$, $\mathbf{I}$ is a $(K-1)\times(K-1)$ unit matrix and $\otimes$ represents a Kronecker product.

Solving the score equations leads to the ML estimation. This requires numerical procedures, and Fisher's scoring or Newton-Raphson often work rather well. As an alternative, the GEE approach in the previous subsection can also be used to fit the polytomous logit model (Sutradhar and Kovacevic, 2000). The polytomous logistic model does not utilize the ordering of response categories. It is applicable to analysis of both unordered and ordered categorical outcomes.

A4.2. Residuals under the Null Hypothesis in Multinomial Logistic Models

Having fitted parameters to data under the null hypothesis where $\boldsymbol{\beta}_t^{(j)} = \mathbf{0}$ ($j = 1, 2, \cdots, K-1$), the residual is the response one calculated by,

$$\mathbf{r}_i = \mathbf{y}_i - \hat{\boldsymbol{\mu}}_i , \tag{A5}$$

where $\hat{\boldsymbol{\mu}}_i$ is the estimated mean.

Appendix A5: Proportional Odds Models

The widely used model for ordinal response is the proportional odds model, often called cumulative logits. The proportional odds model uses logits of cumulative probabilities and assumes an identical effect of the predictors for each cumulative probability (McCullagh,

1980), thus being a more parsimonious model. Consider a ordinal response consisting of K ordered categories, denoted by $1, 2, \cdots, K$. Assuming being in a decreasing order of severity or certainty, we define,

$$Z_j = \begin{cases} 1 & \text{the observed category is equal to or less than } j \\ 0 & \text{otherwise} \end{cases}, \quad j = 1, 2, \cdots, K$$

where $Z_K = 1$. Then, the observation vector, its expectation and its variance are, respectively,

$$\mathbf{Z} = \left(Z_j\right) = \mathbf{LY},$$

$$E\left(\mathbf{Z}\right) = \boldsymbol{\gamma} = \mathbf{L}\boldsymbol{\mu},$$

and

$$Var\left(\mathbf{Z}\right) = \mathbf{LVL}^T,$$

where $\mathbf{Y}$, $\boldsymbol{\mu}$, the expectation of $\mathbf{Y}$, and $\mathbf{V}$, the variance of $\mathbf{Y}$, are defined in Appendix A4, respectively, and the lower triangular matrix,

$$\mathbf{L} = \begin{pmatrix} 1 & 0 & \cdots & 0 \\ 1 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 1 & 1 & \cdots & 1 \end{pmatrix},$$

with the inverse matrix,

$$\mathbf{L}^{-1} = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ -1 & 1 & 0 & \ddots & 0 \\ 0 & -1 & 1 & \ddots & 0 \\ 0 & \ddots & \ddots & \ddots & 0 \\ 0 & 0 & 0 & -1 & 1 \end{pmatrix}.$$

The multinomial density has a multivariate exponential form,

$$f\left(\mathbf{Z};\boldsymbol{\theta},\boldsymbol{\varphi}\right) = \exp\left[\frac{\left(\mathbf{L}^{-1}\mathbf{Z}\right)^T\boldsymbol{\theta} - b(\boldsymbol{\theta})}{a(\boldsymbol{\varphi})} + c(\mathbf{L}^{-1}\mathbf{Z},\boldsymbol{\varphi})\right],$$

where $\boldsymbol{\theta} = \left[\ln\left(\frac{\mu_j}{\mu_K} \right) \right]$, $b(\boldsymbol{\theta}) = \ln\left(1 + \sum_{j=1}^{K-1} e^{\theta_j}\right)$, $a(\boldsymbol{\varphi}) = 1$, and $c((\mathbf{L}^{-1}\mathbf{Z}, \boldsymbol{\varphi}) = 0$. The link in a proportional odds model has the following representation,

$$\eta_j = \log it(\gamma_j) = \ln\left(\frac{\gamma_j}{1-\gamma_j} \right) = \beta_0^{(i)} - \mathbf{x}_t^T \boldsymbol{\beta}_t - \mathbf{x}_c^T \boldsymbol{\beta}_c, \ j = 1, 2, \cdots, K-1.$$

The proportional odds model imposes the restriction that only the intercepts of the regression equations differ. In the vector-matrix form,

$$\boldsymbol{\eta} = \mathbf{I}\boldsymbol{\beta}_0 - \mathbf{X}_t\boldsymbol{\beta}_t - \mathbf{X}_c\boldsymbol{\beta}_c = \mathbf{X}\boldsymbol{\beta},$$

where $\mathbf{I}$ is a unit matrix. The cumulative logits model is a special case of GLMs with a non-canonical link in Appendix A1.

A5.1. Likelihood and Parameter Estimation in Proportional Odds Models

The log-likelihood for a set of independent observations $\mathbf{y}_i$ ($i = 1, 2, \cdots, N$) is,

$$\ln L(\mathbf{z}_1, \mathbf{z}_2, \cdots, \mathbf{z}_N; \boldsymbol{\beta}, \boldsymbol{\varphi}, \mathbf{X}_1, \mathbf{X}_2, \cdots, \mathbf{X}_N) = \sum_{i=1}^{N} \left[\frac{\left(\mathbf{L}^{-1}\mathbf{z}_i\right)^T \boldsymbol{\theta}_i - b(\boldsymbol{\theta}_i)}{a(\boldsymbol{\varphi})} + c(\mathbf{L}^{-1}\mathbf{z}_i, \boldsymbol{\varphi}) \right],$$

where $\mathbf{X}_i$ s are the design matrix. It has the score and Hessian matrix, respectively,

$$U(\boldsymbol{\beta}) = \sum_{i=1}^{N} \frac{1}{a(\boldsymbol{\varphi})} \mathbf{X}_i^T diag\left[\gamma_{ij}(1-\gamma_{ij})\right]\left(\mathbf{L}\mathbf{V}_i\mathbf{L}^T\right)^{-1}(\mathbf{z}_i - \boldsymbol{\gamma}_i),$$

and,

$$H(\boldsymbol{\beta}) = \sum_{i=1}^{N} -\frac{1}{a(\boldsymbol{\varphi})} \begin{bmatrix} \mathbf{X}_i^T diag\left[\gamma_{ij}(1-\gamma_{ij})\right]\left(\mathbf{L}\mathbf{V}_i\mathbf{L}^T\right)^{-1} diag\left[\gamma_{ij}(1-\gamma_{ij})\right]\mathbf{X}_i \\ -\mathbf{X}_i^T \frac{\partial\left\{diag\left[\gamma_{ij}(1-\gamma_{ij})\right]\left(\mathbf{L}\mathbf{V}_i\mathbf{L}^T\right)^{-1}\right\}}{\partial\boldsymbol{\beta}}\left[(\mathbf{z}_i - \boldsymbol{\gamma}_i) \otimes \mathbf{I}\right] \end{bmatrix}.$$

where $\mathbf{I}$ is a unit matrix and $\otimes$ represents a Kronecker product. Fisher's scoring or Newton-Raphson iterative procedure can be employed to find the ML estimates.

A5.2. Residuals under the Null Hypothesis in Proportional Odds Models

The residual is calculated by,

$$r_i = r_{ij} = \tfrac{1}{K-1}\mathbf{1}^T \, diag\left[\hat{\gamma}_{ij}(1-\hat{\gamma}_{ij})\right]\left(\mathbf{L}\hat{\mathbf{V}}_i\mathbf{L}^T\right)^{-1}(\mathbf{z}_i - \hat{\boldsymbol{\gamma}}_i), \; j = 1, 2, \cdots, K-1, \qquad \text{(A6)}$$

where $\hat{\boldsymbol{\gamma}}_i$ and $\hat{\mathbf{V}}_i$ are the estimated mean and the estimated variance under the null hypothesis where $\boldsymbol{\beta}_t = \mathbf{0}$.

Appendix A6: Proportional hazards models

Hazard rate (or instantaneous risk) is a crucial parameter to characterize survival data. Survival models can be viewed as consisting of two components: the underlying hazard function describing how the hazard changes over time and the predictor-related structural function describing how the hazard varies in response to explanatory variables. The best known proportional hazards regression model assumes the hazard as a product of the time-related baseline hazard and the covariates-related component. Typically under such a model, the baseline hazard can "cancel out", and the effects of predictor variables can be estimated by maximizing the remaining partial likelihood, thus being reported as hazard ratios. The Cox proportional hazards model with stationary coefficients are used here to illustrate the proposed method. The further extensions are straightforward although probably rather technical, for example, time-dependent effects (regression coefficients) are allowed in such models and parametric proportional hazards models can also be constructed by specifying a baseline hazard function such as exponential, Weilbull, Gompenz, log-normal and log-logistic.

The Cox proportional hazards model (Cox, 1972) is a semi-parametric model where the dependence of time-to-event on explanatory variables is precisely modeled, but the actual survival distribution, i.e., the baseline hazard function, is not specified and can take any form. Denote by T a random variable either continuous or discrete representing the time to event or the time to lose follow-up. Let $Y(t)$, $O(t)$ and $X(t)$ be, respectively, $0-1$ counting, $0-1$

censoring and covariate processes for a subject at time t. $O(t) = I(t \le T)$ indicates whether the subject is still being followed and thus at risk for event (i.e., under observation), and $Y(t) = I[t \ge T, O(t) = 1]$ does whether the subject experiences the event at the given time, where $I(\cdot)$ is the indicator function whose value is 1 when the argument is true and 0 otherwise. $X(t)$ may be either scalar-valued or vector-valued, and either time-varying or time-fixed; the constant vector is considered here as genotype, sex, ethnicity are unchanged over time. Then the proportional hazards model (Cox, 1972) is written as,

$$\ln h(t, \mathbf{x}) = \ln h_0(t) + \mathbf{x}^T \boldsymbol{\beta},$$

where $h(t, \mathbf{x})$ is the hazard rate at time t, $h_0(t)$ is the baseline hazard rate , $\boldsymbol{\beta}$ consists of $\boldsymbol{\beta}_t$ and $\boldsymbol{\beta}_c$ for the target effects and the covariate effects, respectively, and $\mathbf{x}$ is the predictor vector. The survival function is,

$$S(t, \mathbf{x}) = \exp\left[-\exp(\mathbf{x}^T \boldsymbol{\beta}) \int_0^t h_0(u) du \right] = [S_0(t, \mathbf{x})]^{\exp(\mathbf{x}^T \boldsymbol{\beta})}.$$

A6.1. Partial Likelihood and Parameter Estimation in Proportional hazards models

For a set of survival data, of which subject i ($i = 1, 2, \cdots, N$) has t_i, $y_i(t)$, $o_i(t)$ and $\mathbf{x}_i$ for the actual observed time (event occurring or censoring), and the realized values of counting, censoring and predictor variables, respectively, the partial likelihood can be constructed as,

$$L_P(\boldsymbol{\beta}) = \prod_{i=1}^{N} \left[\frac{\exp(\mathbf{x}_i^T \boldsymbol{\beta})}{\sum_{j=1}^{N} o_j(t_i) \exp(\mathbf{x}_j^T \boldsymbol{\beta})} \right]^{y_i(t_i)}.$$

Although censored individuals are not explicitly used in the likelihood, they are considered through the risk set. Expressed in vector-matrix form, the support is,

$$\ln L_P(\boldsymbol{\beta}) = \mathbf{y}^T (\mathbf{X}\boldsymbol{\beta} - \mathbf{v}),$$

where vector $\mathbf{y} = [y_i(t_i)]$ and vector $\mathbf{v} = \left(v_i = \ln \sum_{j=1}^{N} o_j(t_i) \exp(\mathbf{x}_j^T \boldsymbol{\beta}) \right)$. The score and Hessian matrix are, respectively,

$$U(\boldsymbol{\beta}) = \sum_{i=1}^{N} y_i(t_i)\left[\mathbf{x}_i - \mathbf{X}^T\mathbf{p}_i\right] = \left(\mathbf{X}^T - \mathbf{X}^T\mathbf{P}^T\right)\mathbf{y},$$

and

$$\mathbf{H}(\boldsymbol{\beta}) = -\sum_{i=1}^{N} y_i(t_i)\left[\mathbf{X}^T diag\left(p_{ij}\right)\mathbf{X} - \mathbf{X}^T\mathbf{p}_i\mathbf{p}_i^T\mathbf{X}\right],$$

where $\mathbf{X} = \left(\mathbf{x}_1, \mathbf{x}_2, \cdots, \mathbf{x}_N\right)^T$, $\mathbf{P} = \left(\mathbf{p}_1, \mathbf{p}_2, \cdots, \mathbf{p}_N\right)^T$, and

$$\mathbf{P} = \left[p_{ij} = \frac{o_j(t_i)\exp\left(\mathbf{x}_j^T\boldsymbol{\beta}\right)}{\sum_{m=1}^{N} o_m(t_i)\exp\left(\mathbf{x}_m^T\boldsymbol{\beta}\right)} \right], i,\ j = 1, 2, \cdots, N.$$

Using this score function and Hessian matrix, the partial likelihood can be maximized via the Newton-Raphson algorithm. The estimate at the $(t+1)$ th cycle of iteration is,

$$\hat{\boldsymbol{\beta}}^{(t+1)} = \hat{\boldsymbol{\beta}}^{(t)} - \left\{\mathbf{H}\left(\hat{\boldsymbol{\beta}}^{(t)}\right)\right\}^{-1} U(\hat{\boldsymbol{\beta}}^{(t)}),$$

where $\hat{\boldsymbol{\beta}}^{(t)}$ is the estimated effects at the t th cycle of iteration, $U(\hat{\boldsymbol{\beta}}^{(t)})$ and $\mathbf{H}\left(\hat{\boldsymbol{\beta}}^{(t)}\right)$ are, respectively, the estimated score and estimated Hessian matrix in which $\hat{\boldsymbol{\beta}}^{(t)}$ is used in place of $\boldsymbol{\beta}$. The iteration continues until convergence is reached.

A6.2. Residuals under the Null Hypothesis in Proportional hazards models

In Cox proportional hazards models, there is no obvious analog to the usual "observed minus predicted" residual used in other regression models. Several different residuals can be computed for a fitted null proportional hazards model assuming no target effects. The martingale residual (Barlow and Prentice, 1988; Therneau *et al*, 1990), the difference between the observation and the cumulative hazard, reflecting excess event occurrence, is used in Lee et al. (Lee *et al*, 2012),

$$r_i^M = m_i(\infty) = m_i(t_i) = y_i(t_i) - \int_0^{t_i} h_0(u)y_i(u)\exp(\mathbf{x}_i^T\boldsymbol{\beta})du.$$

However, the martingale residuals are not symmetrically distributed; the upper limit is 1.0 but the lower limit is $-\infty$, potentially heavily skewed towards the "normal" individuals with deficient event occurrence. As an alternative, the score-contributed residual is suggested here,

$$\mathbf{r} = \left(\mathbf{I} - \hat{\mathbf{P}}^T\right)\mathbf{y}, \tag{A7}$$

where $\mathbf{I}$ is a unit matrix and $\hat{\mathbf{P}}$ is the estimated probability matrix under the null hypothesis. This residual may be viewed as the difference between the observed frequency and the cumulative expected frequency over time for an individual to be a case in the context of time-matched case-control sampling. This residual corresponds to Schoenfeld residual (Schoenfeld, 1982), representing a decomposition of the first partial derivative of the log partial likelihood with respect to the given parameter into the given subject.

In addition to the martingale and the score residuals, other two residuals, the score residual and the deviance residual (Therneau *et al*, 1990), can be also used. The deviance residual is computed by,

$$r_i^D = sign\left(\hat{m}_i(t_i)\right)\sqrt{-2\left\{\hat{m}_i(t_i) + y_i(t_i)\ln\left[y_i(t_i) - \hat{m}_i(t_i)\right]\right\}},$$

where $\hat{m}_i(t_i)$ is the estimated martingale residual. The deviance residual is a normalized transform of the martingale residual. The score residual for subject *i* and covariate *k* is computed by,

$$\mathbf{r} = \left(\mathbf{I} - \hat{\mathbf{P}}^T\right)\left[\hat{m}_i(t_i)\right],$$

where $\left[\hat{m}_i(t_i)\right]$ is the vector consisting of $\hat{m}_i(t_i)$.

Appendix B: Principal Components Analysis for Correcting Population Stratification

When the assumption of population homogeneity does not hold true, principal components analysis (PCA) method (Price *et al*, 2006) can be used to correct for population structure. The PCA-based GMDR proceeds as below (Lou *et al*, 2012; Niu *et al*, 2011): Consider there are M markers and N unrelated subjects available for principal components analysis. Let $\mathbf{m}_i = (m_{i1}, m_{i2}, \cdots, m_{iM})^T$ be the coding vector of marker genotypic values for individual *i* ($i = 1, 2, \cdots, N$). The variance-covariance matrix of marker data,

$$\Sigma = \sum_{i-1}^{N} (\mathbf{m}_i - \overline{\mathbf{m}})(\mathbf{m}_i - \overline{\mathbf{m}})^T ,$$

where $\overline{\mathbf{m}}$ is the overall mean of $\mathbf{m}_i$'s. Let $\mathbf{e}_l$ be the eigen vector (factor loading) corresponds to the lth largest eigen value of Σ, $l = 1, 2, \cdots, N$. Denote $t_{il} = (\mathbf{m}_i - \overline{\mathbf{m}})^T \mathbf{e}_l$, the ancestry of individual i along the lth component (projecting $(\mathbf{m}_i - \overline{\mathbf{m}})^T$ down to the reduced space defined by the lth vector). The main pattern of population structure can be accounted for by the few principal components; as showed in Price et al. (Price *et al*, 2006), results are insensitive to the number of axes of variation used.

Let g_i^G be the coding genotypic value at loci of interest that takes 1 when the multilocus genotype is G or 0 otherwise, and r_i be the residual, for individual i ($i = 1, 2, \cdots, N$). Assuming the population structure can be well represented by the first L principal components, the genotypic values and the residual values can be adjusted by fitting the following regression models,

$$g_i^G = \chi_0^G + t_{i1}\chi_1^G + t_{i2}\chi_2^G + \cdots + t_{iL}\chi_L^G + \tau_i^G ,$$

and

$$r_i = \delta_0 + t_{i1}\delta_1 + t_{i2}\delta_2 + \cdots + t_{iL}\delta_L + \varepsilon_i .$$

The adjusted values for the first L principal components can be calculated by,

$$\hat{g}_i^G = g_i^G - \hat{\chi}_0^G - t_{i1}\hat{\chi}_1^G - t_{i2}\hat{\chi}_2^G - \cdots - t_{iL}\hat{\chi}_L^G ,$$

and

$$\hat{r}_i = R_i - \hat{\delta}_0 - t_{i1}\hat{\delta}_1 - t_{i2}\hat{\delta}_2 - \cdots - t_{iL}\hat{\delta}_L ,$$

where $\hat{\chi}$s and $\hat{\delta}$s are the least squares estimates.

In the previous report (Chen *et al*, 2014), we proposed for simplicity of implementation to adjust the genealogical effects only on the phenotypes, but not on the genetic markers. The two adjustment strategies coincide under the null hypothesis. That is,

adjustment only on the phenotypes can assure the projected space is the same when there are no target effects and thus the resulting GMDR is valid in the sense of giving correct type I error rates. However, as shown as follows, the strategy of adjustment both on the phenotypes and on the genetic markers is more theoretically attractive, unbiased and powerful. Consider the following linear model,

$$\mathbf{y} = \mathbf{X}_1\boldsymbol{\beta}_1 + \mathbf{X}_2\boldsymbol{\beta}_2 + \mathbf{e},$$

where $\mathbf{y}$ is the phenotypic vector, $\boldsymbol{\beta}_1$ is the target effect vector, $\boldsymbol{\beta}_2$ is the covariate effect vector (i.e., the effects of the principal components in this context), $\mathbf{X}_1$ and $\mathbf{X}_2$ are the corresponding incidence matrixes, respectively, and $\mathbf{e}$ is the residual effects, $\mathbf{e} \sim \left(0, \sigma_e^2\mathbf{I}\right)$. When using the adjustment strategy on both the sides of the equation for covariates, the resulting equation is,

$$\left[\mathbf{I} - \mathbf{X}_2\left(\mathbf{X}_2^T\mathbf{X}_2\right)^{-}\mathbf{X}_2^T\right]\mathbf{y} = \left[\mathbf{I} - \mathbf{X}_2\left(\mathbf{X}_2^T\mathbf{X}_2\right)^{-}\mathbf{X}_2^T\right]\mathbf{X}_1\boldsymbol{\beta}_1 + \mathbf{e},$$

where $\left(\mathbf{X}_2^T\mathbf{X}_2\right)^{-}$ is the inverse of matrix $\mathbf{X}_2^T\mathbf{X}_2$ when being full rank, and one of the generalized inverses otherwise. Then it can assure that the sequent estimation stays on the same projected space. For example, the estimation of $\boldsymbol{\beta}_1$ is unbiased,

$$E\left(\hat{\boldsymbol{\beta}}_1\right) = E\left\{\left[\mathbf{X}_1^T\mathbf{X}_1 - \mathbf{X}_1^T\mathbf{X}_2\left(\mathbf{X}_2^T\mathbf{X}_2\right)^{-}\mathbf{X}_2^T\mathbf{X}_1\right]^{-1}\mathbf{X}_1^T\left[\mathbf{I} - \mathbf{X}_2\left(\mathbf{X}_2^T\mathbf{X}_2\right)^{-}\mathbf{X}_2^T\right]\mathbf{y}\right\} = \boldsymbol{\beta}_1.$$

The sum of squares due to regression (SSR) is,

$$\begin{aligned}&\hat{\boldsymbol{\beta}}_1^T\mathbf{X}_1^T\left[\mathbf{I} - \mathbf{X}_2\left(\mathbf{X}_2^T\mathbf{X}_2\right)^{-}\mathbf{X}_2^T\right]\mathbf{X}_1\hat{\boldsymbol{\beta}}_1 \\ &= \mathbf{y}^T\left[\mathbf{I} - \mathbf{X}_2\left(\mathbf{X}_2^T\mathbf{X}_2\right)^{-}\mathbf{X}_2^T\right]\mathbf{X}_1\left[\mathbf{X}_1^T\mathbf{X}_1 - \mathbf{X}_1^T\mathbf{X}_2\left(\mathbf{X}_2^T\mathbf{X}_2\right)^{-}\mathbf{X}_2^T\mathbf{X}_1\right]^{-1}\mathbf{X}_1^T\left[\mathbf{I} - \mathbf{X}_2\left(\mathbf{X}_2^T\mathbf{X}_2\right)^{-}\mathbf{X}_2^T\right]\mathbf{y}.\end{aligned}$$

However, when using the adjustment strategy only on the phenotypes, the resulting equation is,

$$\left[\mathbf{I} - \mathbf{X}_2\left(\mathbf{X}_2^T\mathbf{X}_2\right)^{-}\mathbf{X}_2^T\right]\mathbf{y} = \mathbf{X}_1\boldsymbol{\beta}_1 + \mathbf{e}$$

Then, the projected space may be changed when $\mathbf{X}_1$ and $\mathbf{X}_2$ are correlated and the null hypothesis does not hold true. For example, the estimation of $\boldsymbol{\beta}_1$ is not unbiased,

$$E\left(\hat{\boldsymbol{\beta}}_1^*\right) = E\left\{\left(\mathbf{X}_1^T\mathbf{X}_1\right)^{-1}\mathbf{X}_1^T\left[\mathbf{I}-\mathbf{X}_2\left(\mathbf{X}_2^T\mathbf{X}_2\right)^{-}\mathbf{X}_2^T\right]\mathbf{y}\right\} \neq \boldsymbol{\beta}_1 .$$

The SSR is,

$$\hat{\boldsymbol{\beta}}_1^{*T}\mathbf{X}_1^T\mathbf{X}_1\hat{\boldsymbol{\beta}}_1^* = \mathbf{y}^T\left[\mathbf{I}-\mathbf{X}_2\left(\mathbf{X}_2^T\mathbf{X}_2\right)^{-}\mathbf{X}_2^T\right]\mathbf{X}_1\left(\mathbf{X}_1^T\mathbf{X}_1\right)^{-1}\mathbf{X}_1^T\left[\mathbf{I}-\mathbf{X}_2\left(\mathbf{X}_2^T\mathbf{X}_2\right)^{-}\mathbf{X}_2^T\right]\mathbf{y} .$$

Because,

$$\left[\mathbf{x}_1^T\mathbf{x}_1-\mathbf{x}_1^T\mathbf{x}_2\left(\mathbf{x}_2^T\mathbf{x}_2\right)^{-}\mathbf{x}_2^T\mathbf{x}_1\right]^{-1} = \left(\mathbf{x}_1^T\mathbf{x}_1\right)^{-1}+\left(\mathbf{x}_1^T\mathbf{x}_1\right)^{-1}\mathbf{x}_1^T\mathbf{x}_2\left(\mathbf{x}_2^T\mathbf{x}_2-\mathbf{x}_2^T\mathbf{x}_1\left(\mathbf{x}_1^T\mathbf{x}_1\right)^{-1}\mathbf{x}_1^T\mathbf{x}_2\right)^{-}\mathbf{x}_2^T\mathbf{x}_1\left(\mathbf{x}_1^T\mathbf{x}_1\right)^{-1},$$

and $\left(\mathbf{x}_1^T\mathbf{x}_1\right)^{-1}\mathbf{x}_1^T\mathbf{x}_2\left(\mathbf{x}_2^T\mathbf{x}_2-\mathbf{x}_2^T\mathbf{x}_1\left(\mathbf{x}_1^T\mathbf{x}_1\right)^{-1}\mathbf{x}_1^T\mathbf{x}_2\right)^{-}\mathbf{x}_2^T\mathbf{x}_1\left(\mathbf{x}_1^T\mathbf{x}_1\right)^{-1}$ is a non-negative definite matrix, then,

$$\hat{\boldsymbol{\beta}}_1^T\mathbf{X}_1^T\left[\mathbf{I}-\mathbf{X}_2\left(\mathbf{X}_2^T\mathbf{X}_2\right)^{-}\mathbf{X}_2^T\right]\mathbf{X}_1\hat{\boldsymbol{\beta}}_1 \geq \hat{\boldsymbol{\beta}}_1^{*T}\mathbf{X}_1^T\mathbf{X}_1\hat{\boldsymbol{\beta}}_1^*,$$

implying that there is a larger SSR and, correspondingly, a larger statistical power in the adjustment strategy on both the phenotypes and the markers.

Appendix C: Software Note

A GMDR package has been developed based on the open source MDR software originally developed by Dr. Jason Moore's team. This package is developed in Java, making it compatible with various platforms such as MS Windows and Linux. It has two kinds of user-friendly interfaces: Graphical User Interface (GUI) and Command Line Interface (CLI). GUI can be run in majority of desktop systems, and CLI can be run in all the popular shell systems. GUI offers an integrated environment with a series of self-explanatory and easy-to-follow options. All of the options and the running parameters can be set through typing directly, mouse clicking and drag-and-drop actions, as well as the identified interaction models can be also visualized and saved in various image file formats (e.g., JPG, PNG, BMP and EPS). GUI can create and export the configuration file automatically and thereby reduce the

complexity associated with learning the syntax of GMDR, and is particularly beneficial to novice users. CLI provides an alternative means to execute GMDR analysis. CLI can import the configurations that are generated from GUI or edited by users directly. It is more efficient for users to tune up the arguments according to their need and develop their own scripts to perform batch processing, particularly for experienced and secondary development users to run large-scale data analysis and integrate this software into their analysis protocol. As both GUI and CLI share the configuration resources and are capable to import and export the configurations, users can switch freely between the two interfaces.

Presently, diverse analyses such as unrelated-subject design (i.e., the original GMDR) (Lou *et al*, 2007), pedigree-based (Chen *et al*, 2011a; Lou *et al*, 2008) and unified analysis for unrelated and family samples can be implemented for both quantitative and binary phenotypes with and without covariate adjustment by choosing an appropriate combination of running options. GMDR supports both text and binary file formats as in PLINK (Purcell, et al. 2007). The fileset can be converted between text and binary formats. It is applicable to handling genome-wide and other large-scaled data.

The GMDR software is flexible for further development. The modules for analyzing multiple phenotypes, correlated observations, longitudinal data, categorical and ordinal phenotypes, and survival data are under construction and will be added in the future version.

The software is available at http://www.soph.uab.edu/ssg/software.

Appendix D: An Incomplete Summary of GMDR Applications

There is an expanding list of applications of GMDR to a number of complex disorders for detection of gene-gene and/or gene-environment interactions since the appearance of GMDR. They include alcoholism (Du and Wan, 2009), Alzheimer's disease (Lee *et al*, 2010), asthma (Chan *et al*, 2008; Lee *et al*, 2008; Sy *et al*, 2012), blood pressure response to dietary

potassium (Montasser *et al*, 2010), colorectal cancer (Yu *et al*, 2012), prostate cancer (Beuten *et al*, 2009), coronary heart disease (Tu *et al*, 2010), Crohn's disease (Henckaerts *et al*, 2009), major depressive disorder (Lin *et al*, 2009a; Xiao *et al*, 2011; Yang *et al*, 2010), type 1 diabetes (Zhang *et al*, 2010), type 2 diabetes (Lin *et al*, 2009b; Neuman *et al*, 2010; Wu *et al*, 2009), drug responses (Chen *et al*, 2011b; Xu *et al*, 2012), eczema (Wang *et al*, 2012), internalizing disorders (Meng *et al*, 2011), malaria (Atkinson *et al*, 2011), major mood disorders (Pae *et al*, 2010), nicotine dependence (Bergen *et al*, 2009; Li *et al*, 2008), age-related maculopathy (Jakobsdottir *et al*, 2008), obesity (Angeli *et al*, 2011; Ding *et al*, 2012; Pereira *et al*, 2011; Sungyoung *et al*, 2010; Zhou *et al*, 2012), body mass index (Luo *et al*), osteoporosis (Liu *et al*, 2010), rheumatoid arthritis (Mukherjee *et al*, 2009), thrombotic stroke (Liu *et al*, 2009), schizophrenia (Gasso *et al*, 2010), metabolic syndrome in schizophrenic patients (Liou *et al*, 2012), suicide behavior (Tsai *et al*, 2011), survival time of acute myeloid leukemia patients (Lee *et al*, 2012), age at menarche, age at natural menopause and maximal height in women (Zhao *et al*, 2011). These support that GMDR is playing an increasingly important role in tracking down interacting contributors and mapping complex genotype-phenotype relationship. Conceivably, the proposed conceptual framework will pave the way toward more tailored and effective analysis and broaden the use of GMDR approach.

REFERENCES

Angeli CB, Kimura L, Auricchio MT, Vicente JP, Mattevi VS, Zembrzuski VM *et al* (2011). Multilocus Analyses of Seven Candidate Genes Suggest Interacting Pathways for Obesity-Related Traits in Brazilian Populations. *Obesity* **19**(6)**:** 1244-1251.

Anscombe FJ (1961). Examination of Residuals. *Proc Fourth Berkeley Symp on Math Statist and Prob* **1:** 1-36.

Atkinson A, Barbier M, Afridi S, Fumoux F, Rihet P (2011). Evidence for epistasis between hemoglobin C and immune genes in human P. falciparum malaria: a family study in Burkina Faso. *Genes Immun* **12**(6)**:** 481-489.

Barlow WE, Prentice RL (1988). Residuals for Relative Risk Regression. *Biometrika* **75**(1)**:** 65-74.

Bergen AW, Conti DV, Van Den Berg D, Lee W, Liu JH, Li DL *et al* (2009). Dopamine Genes and Nicotine Dependence in Treatment-Seeking and Community Smokers. *Neuropsychopharmacology* **34**(10)**:** 2252-2264.

Beuten J, Gelfond JA, Franke JL, Weldon KS, Crandall AC, Johnson-Pais TL *et al* (2009). Single and multigenic analysis of the association between variants in 12 steroid hormone metabolism genes and risk of prostate cancer. *Cancer Epidemiol Biomarkers Prev* **18**(6)**:** 1869-1880.

Chan IHS, Tang NLS, Leung TF, Huang W, Lam YYO, Li CY *et al* (2008). Study of gene-gene interactions for endophenotypic quantitative traits in Chinese asthmatic children. *Allergy* **63**(8)**:** 1031-1039.

Chen GB, Liu N, Klimentidis YC, Zhu X, Zhi D, Wang X *et al* (2014). A unified GMDR method for detecting gene-gene interactions in family and unrelated samples with application to nicotine dependence. *Hum Genet* **133**(2)**:** 139-150.

Chen GB, Zhu J, Lou XY (2011a). A faster pedigree-based generalized multifactor dimensionality reduction method for detecting gene-gene interactions. *Stat Interface* **4**(3)**:** 295-304.

Chen Q, Yu CQ, Tang X, Chen DF, Tian J, Cao Y *et al* (2011b). Interactions of renin-angiotensin system gene polymorphisms and antihypertensive effect of benazepril in Chinese population. *Pharmacogenomics* **12**(5)**:** 735-743.

Cox DR (1972). Regression Models and Life-Tables. *Journal of the Royal Statistical Society Series B (Methodological)* **34**(2)**:** 187-220.

Ding Y, Guo Z-R, Wu M, Chen Q, Yu H, Luo W-S (2012). Gene-gene Interaction between PPARδ and PPARγ Is Associated with Abdominal Obesity in a Chinese Population. *Journal of Genetics and Genomics* **39**(12)**:** 625-631.

Du YL, Wan YJY (2009). The Interaction of Reward Genes With Environmental Factors in Contribution to Alcoholism in Mexican Americans. *Alcohol Clin Exp Res* **33**(12)**:** 2103-2112.

Fox J (2008). *Applied regression analysis and generalized linear models*, 2nd edn. Sage Publications, Inc: Thousand Oaks, CA, US.

Gasso P, Mas S, Alvarez S, Trias G, Bioque M, Oliveira C *et al* (2010). Xenobiotic metabolizing and transporter genes: gene-gene interactions in schizophrenia and related disorders. *Pharmacogenomics* **11**(12)**:** 1725-1731.

Gill J (2001). *Generalized linear models: A unified approach*. Sage Publications.

Henckaerts L, Van Steen K, Verstreken I, Cleynen I, Franke A, Schreiber S *et al* (2009). Genetic Risk Profiling and Prediction of Disease Course in Crohn's Disease Patients. *Clin Gastroenterol H* **7**(9)**:** 972-980.

Jakobsdottir J, Conley YP, Weeks DE, Ferrell RE, Gorin MB (2008). C2 and CFB Genes in Age-Related Maculopathy and Joint Action with CFH and LOC387715 Genes. *PLoS One* **3**(5).

Lee JH, Moore JH, Park SW, Jang AS, Uh ST, Kim YH *et al* (2008). Genetic interactions model among Eotaxin gene polymorphisms in asthma. *J Hum Genet* **53**(10)**:** 867-875.

Lee JJ, Jo SA, Park JH, Lee SB, Jo I, Kim DK *et al* (2010). Choline acetyltransferase 2384G > A polymorphism and the risk of Alzheimer's disease. *Eur J Neurol* **17:** 74-74.

Lee S, Kwon MS, Oh JM, Park T (2012). Gene-gene interaction analysis for the survival phenotype based on the Cox model. *Bioinformatics* **28**(18)**:** i582-i588.

Li MD, Lou XY, Chen G, Ma JZ, Elston RC (2008). Gene-gene interactions among CHRNA4, CHRNB2, BDNF, and NTRK2 in nicotine dependence. *Biol Psychiatry* **64**(11)**:** 951-957.

Lin E, Chen PS, Chang HH, Gean PW, Tsai HC, Yang YK *et al* (2009a). Interaction of serotonin-related genes affects short-term antidepressant response in major depressive disorder. *Prog Neuro-Psychoph* **33**(7)**:** 1167-1172.

Lin E, Pei D, Huang YJ, Hsieh CH, Wu LSH (2009b). Gene-Gene Interactions Among Genetic Variants from Obesity Candidate Genes for Nonobese and Obese Populations in Type 2 Diabetes. *Genet Test Mol Bioma* **13**(4)**:** 485-493.

Lindsey JK (1997). *Applying Generalized Linear Models*. Springer-Verlag: New York.

Liou YJ, Bai YM, Lin E, Chen JY, Chen TT, Hong CJ *et al* (2012). Gene-gene interactions of the INSIG1 and INSIG2 in metabolic syndrome in schizophrenic patients treated with atypical antipsychotics. *Pharmacogenomics J* **12**(1)**:** 54-61.

Liu JH, Sun K, Bai YY, Zhang WL, Wang XJ, Wang YB *et al* (2009). Association of three-gene interaction among MTHFR, ALOX5AP and NOTCH3 with thrombotic stroke: a multicenter case-control study. *Hum Genet* **125**(5-6)**:** 649-656.

Liu JM, Zhang MJ, Zhao L, Cui B, Li ZB, Zhao HY *et al* (2010). Analysis of Recently Identified Osteoporosis Susceptibility Genes in Han Chinese Women. *J Clin Endocr Metab* **95**(9)**:** E112-E120.

Lou X-Y, Chen G-B, Yan L, Liu N, Klimentidis YC, Zhu X *et al* (2012). A PCA-based Generalized Multifactor ReductionMethod for Correcting Population Stratification. *Genet Epidemiol* **36**(7)**:** 753.

Lou XY, Chen GB, Yan L, Ma JZ, Mangold JE, Zhu J *et al* (2008). A combinatorial approach to detecting gene-gene and gene-environment interactions in family studies. *Am J Hum Genet* **83**(4)**:** 457-467.

Lou XY, Chen GB, Yan L, Ma JZ, Zhu J, Elston RC *et al* (2007). A generalized combinatorial approach for detecting gene-by-gene and gene-by-environment interactions with application to nicotine dependence. *Am J Hum Genet* **80**(6)**:** 1125-1137.

Luo WS, Guo ZR, Wu M, Chen Q, Zhou ZY, Yu H *et al* [Association of both peroxisome proliferator-activated receptor, gene-gene interactions and the body mass index]. *Zhonghua Liu Xing Bing Xue Za Zhi* **33**(7)**:** 740-745.

McCullagh P (1980). Regression Models for Ordinal Data. *Journal of the Royal Statistical Society, Series B (Methodology)* **42**(2)**:** 109-142.

Meng XF, Kou CG, Shi JP, Yu YQ, Huang YQ (2011). Susceptibility genes, social environmental risk factors and their interactions in internalizing disorders among mainland Chinese undergraduates. *J Affect Disorders* **132**(1-2)**:** 254-259.

Montasser ME, Shimmin LC, Gu DF, Chen J, Gu C, Kelly TN *et al* (2010). Blood pressure response to potassium supplementation is associated with genetic variation in endothelin 1 and interactions with E selectin in rural Chinese. *J Hypertens* **28**(4)**:** 748-755.

Mukherjee O, Sanapala KR, Anbazhagana P, Ghosh S (2009). Evaluating epistatic interaction signals in complex traits using quantitative traits. *BMC Proc* **3 Suppl 7:** S82.

Neuman RJ, Wasson J, Atzmon G, Wainstein J, Yerushalmi Y, Cohen J *et al* (2010). Gene-Gene Interactions Lead to Higher Risk for Development of Type 2 Diabetes in an Ashkenazi Jewish Population. *PLoS One* **5**(3).

Niu A, Zhang S, Sha Q (2011). A Novel Method to Detect Gene-Gene Interactions in Structured Populations: MDR-SP. *Ann Hum Genet* **75**(6)**:** 742-754.

Pae CU, Drago A, Forlani M, Patkar AA, Serretti A (2010). Investigation of an Epistastic Effect Between a Set of TAAR6 and HSP-70 Genes Variations and Major Mood Disorders. *Am J Med Genet B* **153B**(2)**:** 680-683.

Pereira TV, Mingroni-Netto RC, Yamada Y (2011). ADRB2 and LEPR Gene Polymorphisms: Synergistic Effects on the Risk of Obesity in Japanese. *Obesity* **19**(7)**:** 1523-1527.

Pierce DA, Schafer DW (1986). Residuals in Generalized Linear Models. *Journal of the American Statistical Association* **81**(396)**:** 977-986.

Pregibon D (1981). Logistic Regression Diagnostics. *The Annals of Statistics* **9**(4)**:** 705-724.

Price AL, Patterson NJ, Plenge RM, Weinblatt ME, Shadick NA, Reich D (2006). Principal components analysis corrects for stratification in genome-wide association studies. *Nat Genet* **38**(8)**:** 904-909.

Schoenfeld D (1982). Partial Residuals for The Proportional Hazards Regression Model. *Biometrika* **69**(1)**:** 239-241.

Sungyoung L, Sohee O, Min-Seok K, Seungyeoun L, Taesung P. (2010). *Bioinformatics and Biomedicine Workshops (BIBMW), 2010 IEEE International Conference on*, pp 353-358.

Sutradhar BC, Kovacevic M (2000). Analysing ordinal longitudinal survey data: Generalised estimating equations approach. *Biometrika* **87**(4)**:** 837-848.

Sy HY, Ko FW, Chu HY, Chan IH, Wong GW, Hui DS *et al* (2012). Asthma and bronchodilator responsiveness are associated with polymorphic markers of ARG1, CRHR2 and chromosome 17q21. *Pharmacogenet Genomics* **22**(7)**:** 517-524.

Therneau TM, Grambsch PM, Fleming TR (1990). Martingale-Based Residuals for Survival Models. *Biometrika* **77**(1)**:** 147-160.

Tsai SJ, Hong CJ, Liou YJ (2011). Recent molecular genetic studies and methodological issues in suicide research. *Prog Neuro-Psychoph* **35**(4)**:** 809-817.

Tu YC, Ding H, Wang XJ, Xu YJ, Zhang L, Huang CX *et al* (2010). Exploring epistatic relationships of NO biosynthesis pathway genes in susceptibility to CHD. *Acta Pharmacol Sin* **31**(7)**:** 874-880.

Wang SS, Hon KL, Sy HY, Kong AP, Chan IH, Tse LY *et al* (2012). Interactions between genetic variants of FLG and chromosome 11q13 locus determine susceptibility for eczema phenotypes. *J Invest Dermatol* **132**(7)**:** 1930-1932.

Wu LS, Hsieh CH, Pei D, Hung YJ, Kuo SW, Lin E (2009). Association and interaction analyses of genetic variants in ADIPOQ, ENPP1, GHSR, PPARgamma and TCF7L2 genes

for diabetic nephropathy in a Taiwanese population with type 2 diabetes. *Nephrol Dial Transplant* **24**(11)**:** 3360-3366.

Xiao ZM, Liu WH, Gao K, Wan QR, Yang C, Wang HL *et al* (2011). Interaction between CRHR1 and BDNF Genes Increases the Risk of Recurrent Major Depressive Disorder in Chinese Population. *PLoS One* **6**(12).

Xu Z, Zhang Z, Shi Y, Pu M, Yuan Y, Zhang X *et al* (2012). Influence and interaction of genetic polymorphisms in the serotonin system and life stress on antidepressant drug response. *J Psychopharmacol* **26**(3)**:** 349-359.

Yang CX, Xu Y, Sun N, Ren Y, Liu ZF, Cao XH *et al* (2010). The combined effects of the BDNF and GSK3B genes modulate the relationship between negative life events and major depressive disorder. *Brain Res* **1355:** 1-6.

Yu Y, Pan Y, Jin M, Zhang M, Zhang S, Li Q *et al* (2012). Association of genetic variants in tachykinins pathway genes with colorectal cancer risk. *Int J Colorectal Dis* **27**(11)**:** 1429-1436.

Zhang DY, Efendic S, Brismar K, Gu HF (2010). Effects of MCF2L2, ADIPOQ and SOX2 genetic polymorphisms on the development of nephropathy in type 1 Diabetes Mellitus. *Bmc Med Genet* **11**.

Zhao L, Cui B, Liu JM, Zhang MJ, Zhao HY, Sun LH *et al* (2011). Interactions of osteoporosis candidate genes for age at menarche, age at natural menopause, and maximal height in Han Chinese women. *Menopause* **18**(9)**:** 1018-1025.

Zhou JB, Liu C, Niu WY, Xin Z, Yu M, Feng JP *et al* (2012). Contributions of renin-angiotensin system-related gene interactions to obesity in a Chinese population. *PLoS One* **7**(8)**:** e42881.