Biostatistics (2017), **0**, 0, *pp.* 1–18
doi:10.1093/biostatistics/"covact biostat supp material 2-24-17"

Covariate Adjusted Classification Trees (COVACT)

(Supplementary Materials)

JOSEPHINE K. ASAFU-ADJEI*

Department of Biostatistics, School of Nursing, University of North Carolina at Chapel Hill, Chapel Hill, NC 27599, USA

jasafuad@email.unc.edu

ALLAN R. SAMPSON

Department of Statistics, University of Pittsburgh, Pittsburgh, PA 15260, USA

APPENDIX

A. Covariate Effects in Discrimination - Simulation Example

To help illustrate the improvement in discrimination ability that results from examining the distribution of covariate adjusted feature values, we consider a brief simulated example using bivariate normal feature and covariate data in two groups. Specifically, we assume that our feature variable and covariate of interest in the i^{th} group ($i = 1, 2$), denoted by (Y_i, X) , are jointly normal with mean $\boldsymbol{\mu}_{Y,X;1} = (10, 0)$ in group 1 and $\boldsymbol{\mu}_{Y,X;2} = (8, 0)$ in group 2, and covariance matrix $\boldsymbol{\Sigma}_{Y,X} = \begin{pmatrix} 4 & 1 \\ 1 & 0.3 \end{pmatrix}$ that is common to both groups. In this example, we note that unlike in Figure 1a of Tu *and others* (1997), the distribution of covariate X does not depend on group.

*To whom correspondence should be addressed.

Based on these assumed distributions for (Y_i, X), the conditional distribution of Y_i given $X = x$ is normal with mean $\boldsymbol{\mu}_{Y|X;1} = 10 + (10/3)x$ in group 1 and $\boldsymbol{\mu}_{Y|X;2} = 8 + (10/3)x$ in group 2, and variance $\boldsymbol{\Sigma}_{Y|X} = 2/3$ in each group. As described in Section 4.2 of the main paper, we let $\tilde{Y}_i = Y_i - (10/3)x$ for a given x denote the adjusted feature variable in the i^{th} group. We then have that $\tilde{Y}_1$ and $\tilde{Y}_2$ are normal with means 10 and 8, respectively, and a common variance of 2/3. Therefore, we observe that the group mean difference is 2 for both the unadjusted feature variable Y and the adjusted feature variable $\tilde{Y}$, but that the variance of $\tilde{Y}$ is smaller than that of Y. In each group, we used these distributions to simulate 20 observed feature and covariate values, which we then used to generate 20 adjusted feature values.

In Figure 1, we examine the distributions of the observed feature values (on the center line of Figure 1(a)) and the adjusted feature values (on the center line of Figure 1(b)) in both groups. In doing so, we see that even though the group mean difference is the same among the observed and adjusted feature values, the adjusted feature values have much lower variation, thereby making it easier to correctly distinguish between the 2 groups.

B. Decrease of Node Impurity Upon Splitting

Proposition 1 Let $\phi(p_1, \ldots, p_g)$ be a strictly concave function such that $p_i \geqslant 0$ $(i = 1, \ldots, g)$ and $\sum_{i=1}^{g} p_i = 1$. In addition, suppose G is defined as a random variable denoting group membership for a particular subject with measurement values for a random feature vector $\mathbf{Y}$. For $M(t) = \phi\left(P(1|t), \ldots, P(g|t)\right)$, where $P(G = i|\mathbf{Y} \in t) = P(i|t)$,

$$M(t) - P_L M(t_L) - P_R M(t_R) \geqslant 0, \tag{B.1}$$

where $P_L = P(\mathbf{Y} \in t_L|\mathbf{Y} \in t)$ and $P_R = P(\mathbf{Y} \in t_R|\mathbf{Y} \in t)$. Equality in (B.1) holds if and only if $P(i|t_L) = P(i|t_R) = P(i|t)$ $(i = 1, \ldots, g)$. If $\mathbf{Y}$ is continuous, the inequality in (B.1) is strict.

Proof. Since ϕ is strictly concave,

$$\begin{aligned} P_L M(t_L) + P_R M(t_R) &= P_L\, \phi\left(P(1|t_L), \ldots, P(g|t_L)\right) + P_R\, \phi\left(P(1|t_R), \ldots, P(g|t_R)\right) \\ &\leqslant \phi\left(P_L P(1|t_L) + P_R P(1|t_R), \ldots, P_L P(g|t_L) + P_R P(g|t_R)\right), \end{aligned} \tag{B.2}$$

with equality holding in (B.2) if and only if $P(i|t_L) = P(i|t_R) = P(i|t)$ $(i = 1, \ldots, g)$. Since

$$\begin{aligned} P_L &= \frac{P(\mathbf{Y} \in t_L, \mathbf{Y} \in t)}{P(\mathbf{Y} \in t)} = \frac{P(\mathbf{Y} \in t_L)}{P(\mathbf{Y} \in t)} \quad (t_L \subset t) \quad \text{and} \\ P_R &= \frac{P(\mathbf{Y} \in t_R, \mathbf{Y} \in t)}{P(\mathbf{Y} \in t)} = \frac{P(\mathbf{Y} \in t_R)}{P(\mathbf{Y} \in t)} \quad (t_R \subset t), \end{aligned}$$

it follows that

$$\begin{aligned} P_L P(i|t_L) + P_R P(i|t_R) &= \frac{[P(G = i, \mathbf{Y} \in t_L) + P(G = i, \mathbf{Y} \in t_R)]}{P(\mathbf{Y} \in t)} \\ &= \frac{P(G = i, \mathbf{Y} \in t)}{P(\mathbf{Y} \in t)} \\ &= P(i|t). \end{aligned} \tag{B.3}$$

From (B.3), the right hand side of the inequality in (B.2) is equal to $\phi\left(P(1|t), \ldots, P(g|t)\right) = M(t)$ and, thus, $M(t) - P_L M(t_L) - P_R M(t_R) \geqslant 0$. Equality holds if and only if $P(i|t_L) = P(i|t_R) = P(i|t)$ $(i = 1, \ldots, g)$. If $\mathbf{Y}$ is continuous, this condition will never hold and $M(t) - P_L M(t_L) - P_R M(t_R)$ will be positive. □

C. Monotone Invariance Property

Result 1 For a chosen GOS criterion, let T' be the classification tree based on the prior probabilities $\pi_1, \ldots, \pi_g$ and the distribution functions $F_{\mathbf{Y}|1}(\cdot)$, …, $F_{\mathbf{Y}|g}(\cdot)$ $(g \geqslant 2)$. Further, let $\mathbf{Z} = (Z_1, \ldots, Z_P)' = (\zeta_1(Y_1), \ldots, \zeta_P(Y_P))' \equiv \boldsymbol{\zeta}(\mathbf{Y})$, where $\zeta_p(Y_p)$ is a strictly increasing function of Y_p $(p = 1, \ldots, P)$. Let $T'_{\mathbf{Z}}$ be the classification tree based on the priors $\pi_1, \ldots, \pi_g$ and the distribution functions $G_{\mathbf{Z}|1}(\cdot)$, …, $G_{\mathbf{Z}|g}(\cdot)$. Then, T' and $T'_{\mathbf{Z}}$ have the same structure, i.e., the set of splitting variables for T', $\mathbf{Y}_{T'}$, are related to those of $T'_{\mathbf{Z}}$, $\mathbf{Z}_{T'_{\mathbf{Z}}}$, by $\mathbf{Z}_{T'_{\mathbf{Z}}} = \boldsymbol{\zeta}(\mathbf{Y}_{T'})$ and the set of cutpoints for T', $\mathbf{c}_{T'}$, are related to those of $T'_{\mathbf{Z}}$, $\mathbf{c}_{T'_{\mathbf{Z}}}$, by $\mathbf{c}_{T'_{\mathbf{Z}}} = \boldsymbol{\zeta}(\mathbf{c}_{T'})$.

Proof. Suppose Y_ν and c_ν are chosen to split the root node t_0 into descendant nodes t_L and t_R, such that the selected GOS criterion defined by the split $Y_\nu \leqslant c_\nu$ is maximized. Recall that

both the impurity measure based GOS criterion and the twoing criterion can be expressed as functions of π_i $(i = 1, \ldots, g)$, $P(\mathbf{Y} \in t_0) = 1$, $P(\mathbf{Y}_i \in t_L) = P(Y_{i,\nu} \leqslant c_\nu) = P(\zeta_\nu(Y_{i,\nu}) \leqslant \zeta_\nu(c_\nu))$ $= P(Z_{i,\nu} \leqslant \zeta_\nu(c_\nu))$, and $P(\mathbf{Y}_i \in t_R) = P(Y_{i,\nu} > c_\nu) = P(\zeta_\nu(Y_{i,\nu}) > \zeta_\nu(c_\nu)) = P(Z_{i,\nu} > \zeta_\nu(c_\nu))$. For example, based on the Gini index $M_G(t)$, we seek to minimize

$$
\begin{aligned}
&2\pi_1\pi_2 \left\{ \frac{F_{Y_\nu|1}(c_\nu)F_{Y_\nu|2}(c_\nu)}{\pi_1 F_{Y_\nu|1}(c_\nu) + \pi_2 F_{Y_\nu|2}(c_\nu)} + \frac{\bar{F}_{Y_\nu|1}(c_\nu)\bar{F}_{Y_\nu|2}(c_\nu)}{\pi_1 \bar{F}_{Y_\nu|1}(c_\nu) + \pi_2 \bar{F}_{Y_\nu|2}(c_\nu)} \right\} \\
&= 2\pi_1\pi_2 \left\{ \frac{G_{Z_\nu|1}\left(\zeta_\nu(c_\nu)\right) G_{Z_\nu|2}\left(\zeta_\nu(c_\nu)\right)}{\pi_1 G_{Z_\nu|1}\left(\zeta_\nu(c_\nu)\right) + \pi_2 G_{Z_\nu|2}\left(\zeta_\nu(c_\nu)\right)} + \frac{\bar{G}_{Z_\nu|1}\left(\zeta_\nu(c_\nu)\right) \bar{G}_{Z_\nu|2}\left(\zeta_\nu(c_\nu)\right)}{\pi_1 \bar{G}_{Z_\nu|1}\left(\zeta_\nu(c_\nu)\right) + \pi_2 \bar{G}_{Z_\nu|2}\left(\zeta_\nu(c_\nu)\right)} \right\}
\end{aligned}
$$

in the case of two groups, where $F_{Y|i}(c) = P(Y_i \leqslant c)$, $G_{Z|i}(\zeta(c)) = P(Z_i \leqslant \zeta(c))$, $\bar{F}_{Y|i}(c) = 1 - F_{Y|i}(c)$, and $\bar{G}_{Y|i}(\zeta(c)) = 1 - G_{Y|i}(\zeta(c))$. Therefore, if Y_ν and c_ν are first chosen to split the feature space $\mathcal{Y}$ in the construction of tree T', then Z_ν and $\zeta_\nu(c_\nu)$ are first chosen to split $\mathcal{Z}$, the monotonic transformation of $\mathcal{Y}$, in the construction of tree $T'_{\mathbf{Z}}$.

Suppose now that we are in a step of the algorithm where there are m descendant nodes or subsets of $\{\mathcal{Y} : Y_\nu \leqslant c_\nu\}$ in T' and $\{\mathcal{Z} : Z_\nu \leqslant \zeta_\nu(c_\nu)\}$ in $T'_{\mathbf{Z}}$, as well as the m' descendant subsets of $\{\mathcal{Y} : Y_\nu > c_\nu\}$ in T' and $\{\mathcal{Z} : Z_\nu > \zeta_\nu(c_\nu)\}$ in $T'_{\mathbf{Z}}$. By the induction setup, we assume that if the split $Y_\upsilon \leqslant c_\upsilon$ is used for a particular node t in T', then the split $Z_\upsilon \leqslant \zeta_\upsilon(c_\upsilon)$ is used for the corresponding node $t_{\mathbf{Z}}$ in $T'_{\mathbf{Z}}$.

Let Y_κ and c_κ be chosen to split the $(m+1)^{st}$ descendant node t of $\{\mathcal{Y} : Y_\nu \leqslant c_\nu\}$, into daughter nodes t_L and t_R, such that the selected GOS criterion defined by the split $Y_\kappa \leqslant c_\kappa$ is maximized. Using the same procedure as that used to split t_0, we can conclude that if Y_κ and c_κ are chosen to split the $(m+1)^{st}$ descendant node of $\{\mathcal{Y} : Y_\nu \leqslant c_\nu\}$ in the construction of tree T', then Z_κ and $\zeta_\kappa(c_\kappa)$ are chosen to split the $(m+1)^{st}$ descendant node of $\{\mathcal{Z} : Z_\nu \leqslant \zeta_\nu(c_\nu)\}$ in the construction of tree $T'_{\mathbf{Z}}$. The same result holds if we wish to split the $(m'+1)^{st}$ descendant node of $\{\mathcal{Y} : Y_\nu > c_\nu\}$.

Thus, by induction, we have that $\mathbf{Z}_{T'_{\mathbf{Z}}} = \boldsymbol{\zeta}(\mathbf{Y}_{T'})$ and $\mathbf{c}_{T'_{\mathbf{Z}}} = \boldsymbol{\zeta}(\mathbf{c}_{T'})$. □

D. Linear Invariance Property

Result 2 Let $\mathbf{Y}_i$ and $\mathbf{Y}_{\mathbf{x},i}$ have CDFs $F_{\mathbf{Y}|i}(\centerdot)$ and $F_{\mathbf{Y}|\mathbf{x},i}(\centerdot)$, respectively, in the i^{th} group. Suppose that $\mathbf{Y}_{\mathbf{x},i}$ is equal in distribution to $\mathbf{Y}_i + \boldsymbol{\xi}(\mathbf{x})$ (i.e., $\mathbf{Y}_{\mathbf{x},i} \overset{d}{=} \mathbf{Y}_i + \boldsymbol{\xi}(\mathbf{x})$), where $\boldsymbol{\xi}(\mathbf{x}) = (\xi_1(\mathbf{x}), \ldots, \xi_P(\mathbf{x}))'$ is a known function of $\mathbf{x}$ that does not depend on group. For a chosen GOS criterion, let $T'^{(\mathbf{x}_a)}$ be the classification tree based on $F_{\mathbf{Y}|\mathbf{x}_a,1}(\centerdot), \ldots, F_{\mathbf{Y}|\mathbf{x}_a,g}(\centerdot)$ for covariate value $\mathbf{x}_a$ and $T'^{(\mathbf{x}_b)}$ be the classification tree based on $F_{\mathbf{Y}|\mathbf{x}_b,1}(\centerdot), \ldots, F_{\mathbf{Y}|\mathbf{x}_b,g}(\centerdot)$ for covariate value $\mathbf{x}_b$. Then, $T'^{(\mathbf{x}_a)}$ and $T'^{(\mathbf{x}_b)}$ have the same set of splitting variables and the set of cutpoints for $T'^{(\mathbf{x}_a)}$, $\mathbf{c}_{T'^{(\mathbf{x}_a)}}$, are related to those of $T'^{(\mathbf{x}_b)}$, $\mathbf{c}_{T'^{(\mathbf{x}_b)}}$, by $\mathbf{c}_{T'^{(\mathbf{x}_b)}} = \mathbf{c}_{T'^{(\mathbf{x}_a)}} - \boldsymbol{\xi}(\mathbf{x}_a) + \boldsymbol{\xi}(\mathbf{x}_b)$.

Proof. For a given $\mathbf{x}_a$ and $\mathbf{x}_b$, $\mathbf{Y}_{\mathbf{x}_b} = \mathbf{Y}_{\mathbf{x}_a} - \boldsymbol{\xi}(\mathbf{x}_a) + \boldsymbol{\xi}(\mathbf{x}_b)$, regardless of group. In other words, $\mathbf{Y}_{\mathbf{x}_b} = \boldsymbol{\zeta}(\mathbf{Y}_{\mathbf{x}_a})$ is an increasing linear function of $\mathbf{Y}_{\mathbf{x}_a}$, i.e., a monotonic transformation of $\mathbf{Y}_{\mathbf{x}_a}$, where $\boldsymbol{\zeta}(\mathbf{Y}) = \mathbf{Y} - \boldsymbol{\xi}(\mathbf{x}_a) + \boldsymbol{\xi}(\mathbf{x}_b)$. It now directly follows from the monotone invariance property (Result 1) that $T'^{(\mathbf{x}_a)}$ and $T'^{(\mathbf{x}_b)}$ have the same set of splitting variables and that $\mathbf{c}_{T'^{(\mathbf{x}_b)}} = \mathbf{c}_{T'^{(\mathbf{x}_a)}} - \boldsymbol{\xi}(\mathbf{x}_a) + \boldsymbol{\xi}(\mathbf{x}_b)$. □

E. COVACT Properties under Non-Identifiable Parameters

Suppose we obtain two LS estimates of $\boldsymbol{\Theta}$, namely, $\hat{\boldsymbol{\Theta}}_a$ and $\hat{\boldsymbol{\Theta}}_b$ ($\hat{\boldsymbol{\Theta}}_a \neq \hat{\boldsymbol{\Theta}}_b$), such that $\boldsymbol{\xi}(\mathbf{x}_{ij}; \hat{\boldsymbol{\Theta}}_a)$ $= (\xi_1(\mathbf{x}_{ij}|\hat{\boldsymbol{\theta}}_{1,a}), \ldots, \xi_P(\mathbf{x}_{ij}|\hat{\boldsymbol{\theta}}_{P,a}))'$ and $\boldsymbol{\xi}(\mathbf{x}_{ij}; \hat{\boldsymbol{\Theta}}_b) = (\xi_1(\mathbf{x}_{ij}|\hat{\boldsymbol{\theta}}_{1,b}), \ldots, \xi_P(\mathbf{x}_{ij}|\hat{\boldsymbol{\theta}}_{P,b}))'$. Consider the covariate adjusted trees $T_a'^{\text{adj}(\mathbf{x})}$ and $T_b'^{\text{adj}(\mathbf{x})}$ constructed using the two adjusted data sets $\hat{\tilde{\mathbf{y}}}_{ij,a} = \mathbf{y}_{ij} - \boldsymbol{\xi}(\mathbf{x}_{ij}; \hat{\boldsymbol{\Theta}}_a)$ and $\hat{\tilde{\mathbf{y}}}_{ij,b} = \mathbf{y}_{ij} - \boldsymbol{\xi}(\mathbf{x}_{ij}; \hat{\boldsymbol{\Theta}}_b)$, respectively, where $\hat{\tilde{\mathbf{y}}}_{ij,b} = \hat{\tilde{\mathbf{y}}}_{ij,a} + \boldsymbol{\xi}(\mathbf{x}_{ij}; \hat{\boldsymbol{\Theta}}_a) - \boldsymbol{\xi}(\mathbf{x}_{ij}; \hat{\boldsymbol{\Theta}}_b)$. Based on the linear invariance property, the following facts hold: (1) the same set of features is chosen for $T_a'^{\text{adj}(\mathbf{x})}$ and $T_b'^{\text{adj}(\mathbf{x})}$; (2) the split $\tilde{Y}_{\nu,a} \leqslant \tilde{c}_{\nu,a}$ in $T_a'^{\text{adj}(\mathbf{x})}$ ($\nu \in (1, 2, \ldots, P)$) is equivalent to the split $\tilde{Y}_{\nu,b} \leqslant \tilde{c}_{\nu,a} + \xi_\nu(\mathbf{x}|\hat{\boldsymbol{\theta}}_{\nu,a}) - \xi_\nu(\mathbf{x}|\hat{\boldsymbol{\theta}}_{\nu,b})$ in $T_b'^{\text{adj}(\mathbf{x})}$ for any given covariate value $\mathbf{x}$. In other words, the observations that fall in the left descendant node of the split $\tilde{Y}_{\nu,a} \leqslant \tilde{c}_{\nu,a}$ are identical to those that fall in the left descendant node of the split $\tilde{Y}_{\nu,b} \leqslant \tilde{c}_{\nu,a} + \xi_\nu(\mathbf{x}|\boldsymbol{\theta}_{\nu,a}) - \xi_\nu(\mathbf{x}|\boldsymbol{\theta}_{\nu,b})$ and likewise for the right descendant nodes, so that $T_a'^{\text{adj}(\mathbf{x})}$ and $T_b'^{\text{adj}(\mathbf{x})}$ yield the same

classification results.

F. Application to Post-Mortem Tissue Data

In our application of COVACT to the biomarker data examined by Sweet *and others* (2003, 2004, 2007, 2008), we first use SAS PROC REG (SAS Institute Inc., 2014) to adjust each biomarker for the effects of age at death, gender, PMI, and brain tissue storage time, all of which are assumed to not depend on diagnostic group. For each subject, storage time varied across the four studies and, thus, storage time values are averaged across the four studies to obtain a single storage time for each subject. As a preliminary analysis, we verified the joint normality of the adjusted biomarker data using the approach of Kankainen *and others* (2007). We also fit an analysis of covariance model to each biomarker, controlling for the effects of group, the included covariates, and the interactions between group and each covariate. For each biomarker, none of the interactions involving group are significant at the 0.05 level. We then apply COVACT by applying the traditional approach by Breiman, Friedman, Olshen, and Stone (Breiman *and others*, 1984) (denoted as BFOS) to the covariate adjusted biomarkers. For comparative purposes, we also apply (1) LDA to the covariate adjusted biomarkers and (2) the traditional BFOS approach to the unadjusted biomarkers. While LDA is implemented using SAS PROC DISCRIM, COVACT and the unadjusted classification trees are constructed using Salford Systems CART® software, based on the Gini index and assuming equal priors. When constructing each tree, 15-fold cross validation (CV) is used to implement the minimal cost complexity pruning procedure (see BFOS for more details), which determines the set of feature variables, splits, and misclassification rate for the final tree. We also used 15-fold CV to compute the misclassification rate for LDA to help ensure that it could be compared with the misclassification rates for our classification trees.

G. Simulation Study

G.1 *Description*

To more extensively compare the classification performances of COVACT and the traditional BFOS approach, we conduct a simulation study based on data settings that mimic those examined by Sweet *and others*. Specifically, we generate feature values using the six biomarkers measured across these studies and covariate values using the values for subject's age at death, PMI, and brain tissue storage time (averaged across all four studies). In conducting our study, we are interested in evaluating how the degree of correlation between features and covariates impacts classification accuracy for COVACT relative to traditional classification trees and, thus, only consider the continuous covariates of subject's age at death, PMI, and brain tissue storage time.

We conduct our study by implementing seven different simulation scenarios. In scenario 1, we first generate J $\mathbf{x}_{ind}$ covariate values from a normal distribution using the sample mean and pooled sample variance (denoted by $\hat{\mathbf{\Sigma}}_{XX}$) for subject's age at death, where we consider $J = 30, 60, 240$. Next, we compute the vector of sample means for $P = 6$ biomarkers in each of the control and schizophrenia diagnostic groups (denoted by $\hat{\boldsymbol{\mu}}_{Y,1}$ and $\hat{\boldsymbol{\mu}}_{Y,2}$), along with the pooled sample covariance matrix $\hat{\mathbf{\Sigma}}_{YY}$ for all six biomarkers and the vector of pooled sample covariances $\hat{\mathbf{\Sigma}}_{YX} = (\hat{\sigma}_{YX,1}, \ldots, \hat{\sigma}_{YX,P})'$ between each biomarker and age at death. We then generate J feature values by simulating each feature observation $\mathbf{y}_{ind} = (y_{ind,1}, \ldots, y_{ind,P})$ from a conditional normal distribution with mean $\hat{\boldsymbol{\mu}}_{Y,i} + \hat{\mathbf{\Sigma}}_{YY}\hat{\mathbf{\Sigma}}_{XX}^{-1}\mathbf{x}_{ind}$ and covariance $\hat{\mathbf{\Sigma}}_{YY} - \hat{\mathbf{\Sigma}}_{YX}\hat{\mathbf{\Sigma}}_{XX}^{-1}\hat{\mathbf{\Sigma}}_{YX}'$ in group i ($i = 1, 2$), where assignment into group i is randomly determined. In addition, we consider subsets of $P = 2, 3, 4$ biomarkers, where the biomarkers in each subset have estimated correlations with age at death that are highest in magnitude. For each $P = 2, 3, 4$, we simulate $J = 30, 60, 240$ feature values as we did for $P = 6$, except that the elements in $\hat{\boldsymbol{\mu}}_{Y,i}$, $\hat{\mathbf{\Sigma}}_{YY}$, and $\hat{\mathbf{\Sigma}}_{YX}$ now correspond to the subset of P biomarkers whose correlations with age at death are highest in magnitude. In scenarios 2 and 3, we use the same simulation scheme as in scenario 1, except that

we replace age at death with PMI in scenario 2 and brain tissue storage time in scenario 3.

For the remaining scenarios, we base our simulations on subsets of the covariate set of age at death, PMI, and brain tissue storage time. In scenario 4, we first generate J $\mathbf{x}_{ind}$ covariate values from a normal distribution using the sample mean vector and pooled sample covariance matrix $\hat{\mathbf{\Sigma}}_{XX}$ for age at death and PMI. We also compute the vectors of sample means $\hat{\boldsymbol{\mu}}_{Y,1}, \hat{\boldsymbol{\mu}}_{Y,2}$ for $P = 6$ biomarkers in each diagnostic group, the pooled sample covariance matrix $\hat{\mathbf{\Sigma}}_{YY}$ for all six biomarkers, and the matrix of pooled sample covariances $\hat{\mathbf{\Sigma}}_{YX}$ between each biomarker and age at death, and between each biomarker and PMI. These mean and covariance estimates are then used to generate J $\mathbf{y}_{ind}$ feature values from a conditional normal distribution as in scenarios 1-3. In addition, we consider subsets of $P = 2, 3, 4$ biomarkers where the biomarkers in each subset have the highest R^2 values when linearly regressed on age at death and PMI. For each $P = 2, 3, 4$, we simulate J feature values as we did for $P = 6$, except that the elements in $\hat{\boldsymbol{\mu}}_{Y,i}$, $\hat{\mathbf{\Sigma}}_{YY}$, and $\hat{\mathbf{\Sigma}}_{YX}$ now correspond to the subset of P biomarkers with the highest R^2 values when linearly regressed on age at death and PMI. In scenarios 5 through 7, we use the same simulation scheme as in scenario 4, except that we consider the covariate subsets of (age at death, brain tissue storage time) in scenario 5, (PMI, brain tissue storage time) in scenario 6, and (age at death, PMI, brain tissue storage time) in scenario 7.

In each scenario, we generate 10,000 data sets, where each contains the J simulated covariate and feature observations $\mathbf{x}_{ind}$ and $\mathbf{y}_{ind} = (y_{ind,1}, \ldots, y_{ind,P})$ $(J = 30, 60, 240; P = 2, 3, 4, 6)$. Based on the generation of each feature using a conditional normal distribution, where all covariate effects are linear and do not differ by group, we have that the generated data satisfies the main assumption of COVACT. Using the Gini index and assuming equal priors, we apply COVACT and the traditional BFOS construction approach to each data set, using the **tree** package (Ripley, 2014) in R software version 3.1.2 (R Core Team, 2014).

G.2 *Conclusions*

We present the estimated means and standard deviations (SDs) for the observed misclassification rates (computed across simulations) for COVACT and the traditional BFOS approach across the seven scenarios and $J = 30, 60, 240$ values for $P = 2, 3$ features in Table 1 and for $P = 4, 6$ features in Table 2. In examining these estimates, we do see that across the different scenarios and numbers of observations and features considered, the mean misclassification rates for COVACT and the traditional BFOS approach are fairly close in magnitude. We note that for the data measured across the four Sweet *and others* studies, which served as the basis for our simulation study, the estimated linear regression coefficients for age at death, PMI, and brain tissue storage time were quite small for each of the six biomarkers. Across the four studies, these coefficients ranged in magnitude from 0.000002 for age with respect to SP-IR puncta density for BA 41 to 0.01 for age with respect to somal volume for BA 42. Therefore, the relatively close magnitudes of the misclassification rates for COVACT and the traditional BFOS approach may be explained by minimal effects of the observed covariates. However, it still remains evident that the means and SDs for these misclassification rates are generally lower for COVACT relative to the traditional BFOS approach, regardless of which scenario or which number of observations and features we consider. We note that this also holds when we examine only control subjects (results in Tables 3 and 4) and schizophrenia subjects (results in Tables 5 and 6) across simulations, thereby indicating that COVACT is generally less conservative than the traditional BFOS approach.

For each approach, we have that misclassification rates, on average, decrease as the number of features P increases. In comparing these rates for scenarios 1-3 (adjusting for the effect of one covariate) with those for scenarios 4-6 (adjusting for the effects of two covariates) and scenario 7 (adjusting for the effects of three covariates), we see, for each J and P, that the decrease in misclassification rates for COVACT relative to the traditional BFOS approach becomes more pronounced as we increase the number of covariates. Therefore, it is evident in this study that the

more covariate effects we account for in applying COVACT, the more of an advantage COVACT displays with respect to classification accuracy compared with the traditional BFOS approach.

Among all of the scenarios and combinations of J and P considered, the results for scenario 7 when there are $J = 30$ observations for $P = 6$ features are most analogous to those for the Sweet *and others* biomarker data application. The fact that the mean observed misclassification rate for this simulation scenario is lower for COVACT helps to lend additional support to the increase in classification accuracy shown by COVACT over the traditional BFOS approach.

References

Breiman, L, Friedman, JH, Olshen, RA and Stone, CJ. (1984). *Classification and Regression Trees*. Belmont, CA: Wadsworth Int. Group.

Kankainen, A, Taskinen, S and Oja, H. (2007). Tests of multinormality based on location vectors and scatter matrices. *Statistical Methods and Applications* **16**, 357379.

R Core Team. (2014). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.

Ripley, Brian. (2014). *tree: Classification and regression trees*. R package version 1.0-35.

SAS Institute Inc. (2014). *SAS 9.4 Help and Documentation*. SAS Institute Inc., Cary, NC.

Sweet, RA, Bergen, SE, Sun, Z, Marcsisin, MJ, Sampson, AR and Lewis, DA. (2007). Anatomical Evidence of Impaired Feedforward Auditory Processing in Schizophrenia. *Biol Psychiatry* **61**, 854–864.

Sweet, RA, Bergen, SE, Sun, Z, Sampson, AR, Pierri, JN and Lewis, DA. (2004). Pyramidal Cell Size Reduction in Schizophrenia: Evidence for Involvement of Auditory Feedforward Circuits. *Biol Psychiatry* **55**, 1128–1137.

Sweet, RA, Henteleff, RA, Zhang, W, Sampson, AR and Lewis, DA. (2008). Reduced Dendritic Spine Density in Auditory Cortex of Subjects with Schizophrenia. *Neuropsychopharmacology* **34**, 374–389.

Sweet, RA, Pierri, JN, Auh, S, Sampson, AR and Lewis, DA. (2003). Reduced Pyramidal Cell Somal Volume in Auditory Association Cortex of Subjects with Schizophrenia. *Neuropsychopharmacology* **28**, 599–609.

Tu, XM, Kowalski, J, Randall, J, Mendoza-Blanco, J, Shear, MK, Monk, TH, Frank, E and Kupfer, DJ. (1997). Generalized Covariance-Adjusted Discriminants: Perspective and Application. *Biometrics* **53**, 900–909.

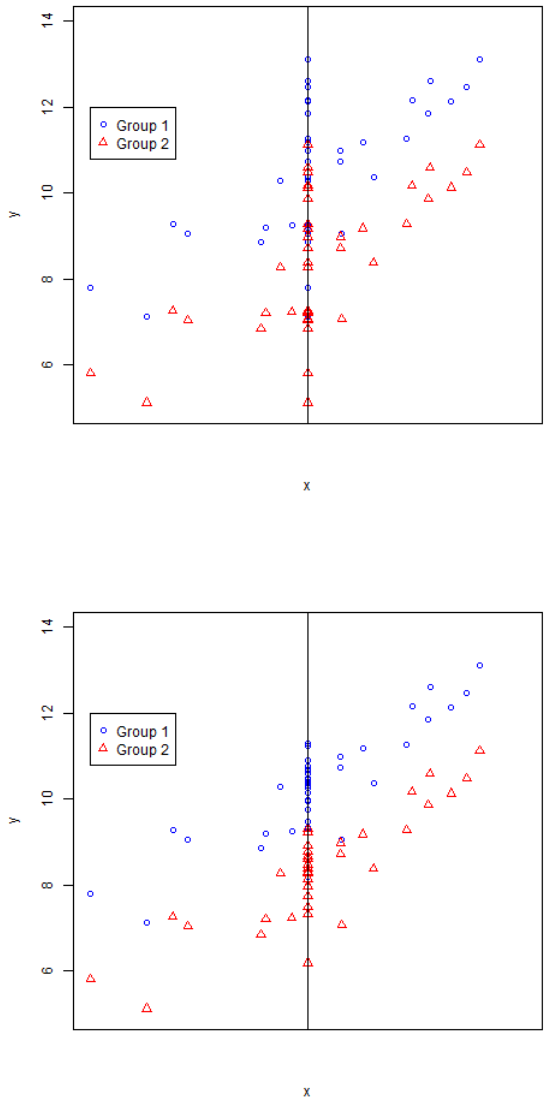

Fig. 1. Diagrams of feature data y and covariate data x, which are linearly related with slope 10/3. The circles and triangles on the center line denote the observed feature values y for groups 1 and 2 in the top diagram (Figure 1(a)) and the adjusted feature values y_{adj} in the bottom diagram (Figure 1(b)). In both diagrams, the circles and triangles not on the center line denote the joint (x, y) values.

Table 1. Means and SDs for Observed Misclassification Rates among All Subjects for $P = 2, 3$

P	Scenario	Approach	J=30		J=120		J=240	
			Mean	SD	Mean	SD	Mean	SD
2	1	COVACT	0.17	0.10	0.19	0.06	0.20	0.06
		Unadjusted	0.17	0.11	0.19	0.07	0.20	0.07
	2	COVACT	0.16	0.10	0.17	0.06	0.18	0.06
		Unadjusted	0.19	0.12	0.21	0.08	0.22	0.07
	3	COVACT	0.13	0.09	0.14	0.05	0.15	0.05
		Unadjusted	0.17	0.10	0.18	0.07	0.20	0.06
	4	COVACT	0.16	0.10	0.17	0.06	0.18	0.06
		Unadjusted	0.19	0.12	0.21	0.08	0.22	0.08
	5	COVACT	0.13	0.08	0.14	0.05	0.15	0.05
		Unadjusted	0.17	0.10	0.19	0.07	0.20	0.06
	6	COVACT	0.15	0.10	0.16	0.06	0.17	0.06
		Unadjusted	0.20	0.12	0.21	0.08	0.22	0.08
	7	COVACT	0.15	0.10	0.16	0.06	0.17	0.06
		Unadjusted	0.20	0.12	0.21	0.08	0.22	0.07
3	1	COVACT	0.16	0.11	0.18	0.07	0.20	0.06
		Unadjusted	0.17	0.11	0.18	0.07	0.20	0.07
	2	COVACT	0.13	0.09	0.14	0.06	0.15	0.06
		Unadjusted	0.16	0.11	0.17	0.07	0.18	0.06
	3	COVACT	0.13	0.09	0.13	0.06	0.14	0.05
		Unadjusted	0.16	0.11	0.18	0.07	0.19	0.07
	4	COVACT	0.13	0.09	0.14	0.06	0.15	0.05
		Unadjusted	0.16	0.11	0.17	0.07	0.18	0.06
	5	COVACT	0.12	0.09	0.13	0.05	0.14	0.05
		Unadjusted	0.16	0.11	0.17	0.07	0.19	0.06
	6	COVACT	0.12	0.09	0.13	0.05	0.14	0.05
		Unadjusted	0.16	0.11	0.17	0.07	0.18	0.06
	7	COVACT	0.12	0.09	0.12	0.05	0.13	0.05
		Unadjusted	0.16	0.11	0.17	0.07	0.19	0.06

Table 2. Means and SDs for Observed Misclassification Rates among All Subjects for $P = 4, 6$

P	Scenario	Approach	J=30		J=120		J=240	
			Mean	SD	Mean	SD	Mean	SD
4	1	COVACT	0.15	0.11	0.16	0.07	0.18	0.06
		Unadjusted	0.15	0.11	0.16	0.07	0.18	0.07
	2	COVACT	0.14	0.10	0.14	0.06	0.15	0.06
		Unadjusted	0.16	0.12	0.17	0.07	0.18	0.06
	3	COVACT	0.13	0.10	0.13	0.06	0.14	0.05
		Unadjusted	0.16	0.12	0.17	0.07	0.19	0.07
	4	COVACT	0.13	0.10	0.14	0.06	0.15	0.06
		Unadjusted	0.16	0.12	0.17	0.07	0.18	0.06
	5	COVACT	0.12	0.09	0.13	0.06	0.14	0.05
		Unadjusted	0.15	0.11	0.17	0.07	0.18	0.07
	6	COVACT	0.11	0.08	0.11	0.05	0.12	0.05
		Unadjusted	0.15	0.11	0.15	0.07	0.17	0.06
	7	COVACT	0.11	0.08	0.11	0.05	0.12	0.05
		Unadjusted	0.14	0.11	0.16	0.07	0.17	0.06
6	1	COVACT	0.13	0.10	0.14	0.07	0.16	0.06
		Unadjusted	0.14	0.11	0.15	0.07	0.16	0.06
	2	COVACT	0.12	0.10	0.13	0.06	0.14	0.06
		Unadjusted	0.14	0.11	0.15	0.07	0.16	0.06
	3	COVACT	0.12	0.09	0.11	0.06	0.12	0.05
		Unadjusted	0.14	0.11	0.15	0.07	0.16	0.06
	4	COVACT	0.12	0.10	0.13	0.06	0.14	0.06
		Unadjusted	0.14	0.11	0.15	0.07	0.16	0.06
	5	COVACT	0.11	0.09	0.11	0.06	0.12	0.05
		Unadjusted	0.14	0.11	0.15	0.07	0.16	0.06
	6	COVACT	0.11	0.09	0.11	0.06	0.12	0.05
		Unadjusted	0.14	0.11	0.15	0.07	0.16	0.06
	7	COVACT	0.11	0.09	0.11	0.06	0.11	0.05
		Unadjusted	0.14	0.11	0.15	0.07	0.16	0.06

Table 3. Means and SDs for Observed Misclassification Rates among Control Subjects for $P = 2, 3$

P	Scenario	Approach	J=30		J=120		J=240	
			Mean	SD	Mean	SD	Mean	SD
2	1	COVACT	0.15	0.13	0.19	0.11	0.20	0.10
		Unadjusted	0.15	0.14	0.19	0.11	0.21	0.10
	2	COVACT	0.15	0.13	0.17	0.10	0.18	0.09
		Unadjusted	0.17	0.15	0.21	0.13	0.22	0.11
	3	COVACT	0.13	0.12	0.14	0.09	0.15	0.07
		Unadjusted	0.15	0.14	0.19	0.11	0.20	0.10
	4	COVACT	0.14	0.13	0.17	0.10	0.18	0.09
		Unadjusted	0.17	0.15	0.21	0.13	0.22	0.11
	5	COVACT	0.13	0.12	0.14	0.08	0.15	0.07
		Unadjusted	0.16	0.14	0.19	0.11	0.20	0.10
	6	COVACT	0.14	0.13	0.16	0.10	0.17	0.09
		Unadjusted	0.17	0.15	0.21	0.13	0.22	0.11
	7	COVACT	0.14	0.12	0.16	0.09	0.17	0.08
		Unadjusted	0.17	0.15	0.21	0.13	0.22	0.11
3	1	COVACT	0.14	0.13	0.18	0.11	0.20	0.10
		Unadjusted	0.15	0.13	0.18	0.11	0.20	0.10
	2	COVACT	0.12	0.12	0.14	0.09	0.15	0.08
		Unadjusted	0.14	0.13	0.17	0.10	0.18	0.09
	3	COVACT	0.12	0.12	0.14	0.09	0.15	0.07
		Unadjusted	0.14	0.13	0.18	0.11	0.19	0.09
	4	COVACT	0.12	0.12	0.14	0.09	0.15	0.08
		Unadjusted	0.14	0.13	0.17	0.10	0.19	0.09
	5	COVACT	0.11	0.11	0.13	0.08	0.14	0.07
		Unadjusted	0.14	0.13	0.17	0.11	0.19	0.09
	6	COVACT	0.12	0.11	0.13	0.08	0.14	0.07
		Unadjusted	0.14	0.13	0.17	0.10	0.18	0.09
	7	COVACT	0.11	0.11	0.12	0.08	0.13	0.07
		Unadjusted	0.14	0.13	0.17	0.11	0.19	0.09

Table 4. Means and SDs for Observed Misclassification Rates among Control Subjects for $P = 4, 6$

P	Scenario	Approach	J=30		J=120		J=240	
			Mean	SD	Mean	SD	Mean	SD
4	1	COVACT	0.13	0.12	0.16	0.10	0.18	0.09
		Unadjusted	0.13	0.13	0.17	0.11	0.18	0.09
	2	COVACT	0.12	0.12	0.14	0.09	0.15	0.08
		Unadjusted	0.14	0.13	0.17	0.11	0.18	0.09
	3	COVACT	0.12	0.12	0.14	0.09	0.15	0.08
		Unadjusted	0.14	0.13	0.17	0.11	0.19	0.10
	4	COVACT	0.12	0.12	0.14	0.09	0.15	0.08
		Unadjusted	0.14	0.13	0.17	0.11	0.19	0.09
	5	COVACT	0.11	0.11	0.13	0.08	0.14	0.07
		Unadjusted	0.13	0.13	0.17	0.10	0.19	0.09
	6	COVACT	0.10	0.10	0.11	0.07	0.12	0.06
		Unadjusted	0.13	0.12	0.16	0.10	0.17	0.09
	7	COVACT	0.10	0.11	0.11	0.07	0.12	0.06
		Unadjusted	0.13	0.12	0.16	0.10	0.17	0.09
6	1	COVACT	0.12	0.12	0.15	0.10	0.16	0.08
		Unadjusted	0.12	0.12	0.15	0.10	0.17	0.09
	2	COVACT	0.11	0.11	0.13	0.09	0.14	0.08
		Unadjusted	0.12	0.12	0.15	0.10	0.16	0.09
	3	COVACT	0.10	0.11	0.11	0.08	0.12	0.07
		Unadjusted	0.12	0.12	0.15	0.10	0.16	0.09
	4	COVACT	0.11	0.11	0.13	0.09	0.14	0.08
		Unadjusted	0.12	0.12	0.15	0.10	0.16	0.08
	5	COVACT	0.10	0.10	0.11	0.08	0.12	0.07
		Unadjusted	0.12	0.12	0.15	0.10	0.16	0.08
	6	COVACT	0.10	0.11	0.11	0.08	0.12	0.07
		Unadjusted	0.12	0.12	0.15	0.10	0.16	0.09
	7	COVACT	0.10	0.10	0.11	0.08	0.12	0.07
		Unadjusted	0.12	0.12	0.15	0.10	0.16	0.09

Table 5. Means and SDs for Observed Misclassification Rates among Schizophrenia Subjects for $P = 2, 3$

P	Scenario	Approach	J=30		J=120		J=240	
			Mean	SD	Mean	SD	Mean	SD
2	1	COVACT	0.15	0.14	0.19	0.11	0.20	0.10
		Unadjusted	0.16	0.14	0.19	0.11	0.21	0.10
	2	COVACT	0.14	0.13	0.17	0.10	0.18	0.09
		Unadjusted	0.17	0.14	0.21	0.12	0.23	0.11
	3	COVACT	0.13	0.12	0.14	0.08	0.15	0.07
		Unadjusted	0.15	0.14	0.19	0.11	0.20	0.10
	4	COVACT	0.14	0.13	0.17	0.10	0.18	0.09
		Unadjusted	0.17	0.15	0.21	0.12	0.22	0.11
	5	COVACT	0.12	0.11	0.14	0.09	0.15	0.07
		Unadjusted	0.15	0.14	0.19	0.11	0.20	0.10
	6	COVACT	0.14	0.12	0.17	0.10	0.17	0.09
		Unadjusted	0.17	0.15	0.21	0.12	0.22	0.12
	7	COVACT	0.13	0.12	0.16	0.09	0.17	0.08
		Unadjusted	0.17	0.14	0.21	0.13	0.22	0.11
3	1	COVACT	0.14	0.13	0.18	0.11	0.20	0.10
		Unadjusted	0.15	0.13	0.18	0.11	0.20	0.10
	2	COVACT	0.12	0.12	0.14	0.09	0.15	0.08
		Unadjusted	0.14	0.13	0.17	0.10	0.18	0.09
	3	COVACT	0.12	0.12	0.14	0.08	0.15	0.07
		Unadjusted	0.15	0.13	0.18	0.11	0.19	0.10
	4	COVACT	0.12	0.12	0.14	0.09	0.15	0.08
		Unadjusted	0.14	0.13	0.17	0.10	0.18	0.09
	5	COVACT	0.11	0.11	0.13	0.08	0.14	0.07
		Unadjusted	0.14	0.13	0.17	0.11	0.19	0.09
	6	COVACT	0.12	0.11	0.13	0.08	0.14	0.07
		Unadjusted	0.14	0.13	0.17	0.10	0.19	0.09
	7	COVACT	0.11	0.11	0.13	0.08	0.13	0.07
		Unadjusted	0.14	0.13	0.17	0.11	0.19	0.09

Table 6. Means and SDs for Observed Misclassification Rates among Schizophrenia Subjects for $P = 4, 6$

P	Scenario	Approach	J=30		J=120		J=240	
			Mean	SD	Mean	SD	Mean	SD
4	1	COVACT	0.13	0.12	0.16	0.10	0.18	0.09
		Unadjusted	0.13	0.13	0.16	0.10	0.18	0.09
	2	COVACT	0.12	0.12	0.14	0.09	0.16	0.08
		Unadjusted	0.14	0.13	0.17	0.11	0.19	0.09
	3	COVACT	0.12	0.12	0.13	0.09	0.15	0.08
		Unadjusted	0.14	0.13	0.17	0.11	0.19	0.10
	4	COVACT	0.12	0.12	0.14	0.09	0.15	0.08
		Unadjusted	0.14	0.13	0.17	0.10	0.19	0.09
	5	COVACT	0.11	0.11	0.13	0.08	0.14	0.07
		Unadjusted	0.13	0.13	0.17	0.10	0.19	0.09
	6	COVACT	0.10	0.10	0.11	0.08	0.12	0.06
		Unadjusted	0.13	0.12	0.16	0.10	0.17	0.09
	7	COVACT	0.10	0.10	0.11	0.08	0.12	0.06
		Unadjusted	0.13	0.12	0.16	0.10	0.17	0.09
6	1	COVACT	0.12	0.12	0.15	0.10	0.16	0.09
		Unadjusted	0.12	0.12	0.15	0.10	0.16	0.09
	2	COVACT	0.11	0.11	0.13	0.09	0.14	0.08
		Unadjusted	0.12	0.12	0.15	0.10	0.16	0.09
	3	COVACT	0.11	0.11	0.11	0.08	0.12	0.07
		Unadjusted	0.12	0.12	0.15	0.10	0.16	0.09
	4	COVACT	0.11	0.11	0.13	0.09	0.14	0.08
		Unadjusted	0.12	0.12	0.15	0.10	0.17	0.09
	5	COVACT	0.10	0.10	0.11	0.08	0.12	0.07
		Unadjusted	0.12	0.12	0.15	0.10	0.16	0.09
	6	COVACT	0.10	0.10	0.11	0.08	0.12	0.07
		Unadjusted	0.12	0.12	0.15	0.10	0.16	0.09
	7	COVACT	0.10	0.10	0.11	0.08	0.11	0.06
		Unadjusted	0.12	0.12	0.15	0.10	0.17	0.09