Biostatistics (2020) **00**, 00, *pp.* 1–16
doi:10.1093/biostatistics/kxaa049

An optimal kernel-based multivariate U-statistic to test for associations with multiple phenotypes

YALU WEN*

Department of Statistics, University of Auckland, Auckland, New Zealand
y.wen@auckland.ac.nz

QING LU

Department of Biostatistics, College of Public Health, University of Florida, Gainesville, FL, USA

Summary

Set-based analysis that jointly considers multiple predictors in a group has been broadly conducted for association tests. However, their power can be sensitive to the distribution of phenotypes, and the underlying relationships between predictors and outcomes. Moreover, most of the set-based methods are designed for single-trait analysis, making it hard to explore the pleiotropic effect and borrow information when multiple phenotypes are available. Here, we propose a kernel-based multivariate U-statistics (KMU) that is robust and powerful in testing the association between a set of predictors and multiple outcomes. We employed a rank-based kernel function for the outcomes, which makes our method robust to various outcome distributions. Rather than selecting a single kernel, our test statistics is built based on multiple kernels selected in a data-driven manner, and thus is capable of capturing various complex relationships between predictors and outcomes. The asymptotic properties of our test statistics have been developed. Through simulations, we have demonstrated that KMU has controlled type I error and higher power than its counterparts. We further showed its practical utility by analyzing a whole genome sequencing data from Alzheimer's Disease Neuroimaging Initiative study, where novel genes have been detected to be associated with imaging phenotypes.

Keywords: Gene-set association analysis; Multiple phenotypes; Multivariate U-statistics; Non-additive effects; Optimal kernel functions.

1. Introduction

The advances in high-throughput biotechnologies allow researchers to systematically investigate the role of a deep catalog of predictors on the development of complex human traits (Welter *and others*, 2014). To date, the traditional genome-wide association studies that assess the effects of predictors one at a time have successfully detected millions of markers. However, their power can be limited partially due to weak marginal associations, multiple comparison issues, and the lack of considering potential interactions among predictors.

*To whom correspondence should be addressed.

The common strategy to address these limitations is to perform a joint association test, where a set-based analysis is conducted to assess the cumulative effects of all predictors within the set (e.g., a gene or pathway) (He *and others*, 2017; Wei and Lu, 2017; He *and others*, 2019). The subtle associations embedded in each predictor are aggregated and the number of tests is greatly reduced. Among the existing set-based approaches, the kernel machine regression (KMR) models that assess the association through measuring and comparing predictors' and phenotypic similarities have been widely used (Liu *and others*, 2007; Wu *and others*, 2011, 2013; He *and others*, 2017; Larson *and others*, 2017). Despite their popularity, the performance of KMR greatly depends on the kernel functions used in measuring similarities. They achieve the best performance when the chosen kernel mimics the underlying mechanisms, but their power can be substantially reduced if otherwise (Wu *and others*, 2013; He *and others*, 2019). Indeed, the kernels reflect the prior beliefs about the disease mechanisms. For example, a linear kernel implicitly indicates there is an additive relationship between the predictors and outcome. In practice, the true disease model is usually unknown in advance, making it hard to pre-select kernels. A widely employed method for kernel selection is through a perturbation procedure, where a single kernel that leads to the minimum p-value is selected and its distribution is derived empirically (Wu *and others*, 2013; Larson *and others*, 2017). However, such a procedure can be computationally intensive, especially for large-scale genomic data. Moreover, existing methods only select one kernel from all the candidates, which may fail to capture complex relationships (Wu *and others*, 2013; Larson *and others*, 2017; He *and others*, 2019). For example, if there are both additive and pairwise interaction effects, using only one kernel function (e.g., the linear or quadratic kernels) is unlikely to fully capture the information.

Recent studies have indicated that many genetic variants are associated with multiple outcomes (Solovieff *and others*, 2013). Joint analysis of multiple traits can substantially boost the power in detecting biomarkers as compared to single-outcome analysis (Aschard *and others*, 2014; Zhan *and others*, 2017; Dutta *and others*, 2019). It can borrow information across multiple traits and thus amplify the marginal association signals, making them easy to detect (Aschard *and others*, 2014). It also allows for the exploration of pleiotropic effects, and greatly facilitates the investigation of the underlying biology (Zhan *and others*, 2017). It enables the investigation of the same disease from different perspectives, as many traits are inherently multi-phenotypic with each reflecting a different aspect of the disease (Alberti *and others*, 2005; Wei and Lu, 2017).

Many methods have been developed for multiple phenotype analyses. These include (1) integrating results from univariate analysis for each trait (van der Sluis *and others*, 2013); (2) collapsing multiple phenotypes into a single score through dimension reduction methods (Klei *and others*, 2008); and (3) multivariate analysis techniques (Wei and Lu, 2017; Zhan *and others*, 2017; Dutta *and others*, 2019). While the first two categories are easy to implement, they either fail to consider the correlations among multiple phenotypes or the weights for combining them are not optimal, leading to the loss of power (Zhan *and others*, 2017). Among the multivariate methods, KMRs are one of the most popular models. For example, Wu and Pankow (2016) developed a score based test statistics for the association analysis with multiple traits. Broadaway *and others* (2016) proposed a dual-kernel approach, where hypothesis test is conducted through comparing the genetic and the multi-trait similarities. Zhan *and others* (2017) further developed the dual kernel-based association test where the dimension of phenotypes can be larger than the sample size. Dutta *and others* (2019) developed a general framework for testing the pleiotropic effects on continuous phenotypes using a multivariate KMR. Despite these advances, KMRs have several key limitations. First, they are semi-parametric models that rely on the distributional assumptions, and thus cannot handle multivariate traits with any arbitrary distributions. Second, KMRs usually concatenate all the phenotypes to form an outcome vector, and can be computationally demanding, especially when a large number of phenotypes are considered. Third, for modern biomedical data that includes both traditional variables and various new data types (e.g., shapes and images) (Wei and Lu, 2017), it can be challenging to integrate these complex data types into the KMR framework, which usually works in a vector space (e.g., outcome is continuous or binary).

Motivated by KMRs and the recent development in U-statistics (Wei and Lu, 2017; He *and others*, 2019), we propose a kernel-based multivariate U-statistic (referred to as KMU) to assess the association between a group of predictors and multiple phenotypes, where the similarities in predictors are compared to those in outcomes. Compared to existing methods, our KMU has several key advantages. First, it makes no distributional assumptions about the outcomes of interest, and thus is robust and powerful against various distributions. Second, rather than selecting one kernel function from a candidate set (He *and others*, 2019; Wu *and others*, 2013), we consider all possible combinations of kernel functions and choose the one that best models the data. The asymptotic distribution of the test statistics in KMU is developed and no computationally intensive perturbation procedure is required. Third, KMU embeds both the predictors and outcomes in the Reproducible Kernel Hilbert Space (RKHS), and thus can handle high-dimensional predictors and outcomes that include complex-objects. In Section 2, we studied the theoretical properties of KMU in a general setting, where asymptotic properties are derived. Results from simulation studies and a whole genome sequencing (WGS) data application were presented in Sections 3 and 4, respectively.

2. Statistical methods

2.1. *General setting and rationale*

We assume a study of n independent individuals. Let p_y and p_g represent the number of outcomes and genetic variants, respectively. For the ith subject ($i = 1, \ldots, n$), let $\boldsymbol{X_i} = (x_{i1}, \ldots, x_{im})'$ be a vector of m covariates (e.g., age and gender), $\boldsymbol{G_i} = (G_{i1}, G_{i2}, \ldots, G_{ip_g})'$ be a vector of random variable for the predictors within a set, which can be defined using existing criteria (e.g., a gene or pathway). Let $\boldsymbol{g_i} = (g_{i1}, g_{i2}, \cdots, g_{ip_g})'$ be the observed value of $\boldsymbol{G}_i$. For this study, we assume the number of predictors p_g is large and can be greater than n. We use $\boldsymbol{Y_i}$ and $\boldsymbol{y_i}$ to respectively denote the random outcomes and its observed values for subject i, where the random outcomes can be a random variable (i.e., $Y_i \in \mathbb{R}$), a random vector ($\boldsymbol{Y_i} \in \mathbb{R}^{p_y}$), a random matrix ($\boldsymbol{Y_i} \in \mathbb{R}^{p_y \times p_y}$), and other data types (e.g., shapes or graphs). Note that for both random outcomes and predictors, we let them take values on metric spaces (i.e., $\boldsymbol{Y} \in \Omega_Y$ and $\boldsymbol{G} \in \Omega_G$) without any assumptions for their distributions.Given the sample, our research interest lies in whether the outcomes $\boldsymbol{Y}$ and predictors $\boldsymbol{G}$ are associated. Since no distributional assumptions were employed for both $\boldsymbol{Y}$ and $\boldsymbol{G}$, they can be of various forms (e.g., $\boldsymbol{Y}$ can be shape data). Therefore, it is not straightforward to use regressions (e.g., KMRs) for the hypothesis testing. However, as shown in Wei and Lu, 2017, it is relatively easy to construct kernel functions to obtain the pairwise similarities for both outcomes and predictors. Let $k(\cdot, \cdot) : \Omega_Y \times \Omega_Y \to \mathbb{R}$ and $s(\cdot, \cdot) : \Omega_G \times \Omega_G \to \mathbb{R}$ be kernel functions to measure the similarities between outcome $\boldsymbol{Y}$ and predictor $\boldsymbol{G}$. Intuitively, if $\boldsymbol{Y}$ and $\boldsymbol{G}$ are associated, then the high similarities between $\boldsymbol{G}_i$ and $\boldsymbol{G}_j$ (i.e.,$s(\boldsymbol{G}_i, \boldsymbol{G}_j)$) lead to high similarities between outcomes $\boldsymbol{Y_i}$ and $\boldsymbol{Y_j}$ (i.e., $k(\boldsymbol{Y}_i, \boldsymbol{Y}_j)$) (Wei and Lu, 2017).

We use $\boldsymbol{K}$ and $\boldsymbol{S}$ to denote the kernel matrices for $k(\cdot, \cdot)$ and $s(\cdot, \cdot)$, respectively. Given the predictors and outcomes for subjects i and j, their outcome and predictor similarities can be written as $K_{ij} = k(\boldsymbol{y_i}, \boldsymbol{y_j})$ and $S_{ij} = s(\boldsymbol{g_i}, \boldsymbol{g_j})$. We define the centered outcome-similarity as $\tilde{K}_{ij} = \tilde{k}(\boldsymbol{y_i}, \boldsymbol{y_j}) = k(\boldsymbol{y_i}, \boldsymbol{y_j}) - E(k(\boldsymbol{y_i}, \boldsymbol{Y_j})) - E(k(\boldsymbol{Y_i}, \boldsymbol{y_j})) + E(k(\boldsymbol{Y_i}, \boldsymbol{Y_j}))$ and define the centered predictor-similarity $\tilde{S}_{ij} = \tilde{s}(\boldsymbol{g_i}, \boldsymbol{g_j})$ in a similar manner. It is straightforward to show that $E(\tilde{k}(\boldsymbol{Y_i}, \boldsymbol{Y_j})) = 0$ and $E(\tilde{s}(\boldsymbol{G_i}, \boldsymbol{G_j})) = 0$. Given n samples, the centralized kernel matrices (i.e., $\tilde{\boldsymbol{K}}$ and $\tilde{\boldsymbol{S}}$) can be estimated empirically (He *and others*, 2019). The relationships between two similarities can be assessed via testing $\rho = 0$ with $\hat{\rho} = \sum_{i<j} \tilde{K}_{ij}\tilde{S}_{ij} / (\sum_{i<j} \tilde{K}_{ij}^2 \sum_{i<j} \tilde{S}_{ij}^2)^{1/2}$, which is equivalent to test whether the numerator $\sum_{i<j} \tilde{K}_{ij}\tilde{S}_{ij} \neq 0$ (Tzeng *and others*, 2009). As presented in the following sections, $\sum_{i<j} \tilde{K}_{ij}\tilde{S}_{ij}$ is of the same form as our proposed test statistic and those used in Wei and Lu, 2017 and He *and others*, 2019.

2.2. *Hypothesis test based on a single kernel for predictors and outcomes*

Let $k \otimes s : (\Omega_Y \times \Omega_G) \times (\Omega_Y \times \Omega_G) \rightarrow \mathbb{R}$ be a kernel function such that $(k \otimes s)\left((\boldsymbol{g}_i, \boldsymbol{y}_i), (\boldsymbol{g}_j, \boldsymbol{y}_j)\right) = \tilde{s}(\boldsymbol{g}_i, \boldsymbol{g}_j) \times \tilde{k}(\boldsymbol{y}_i, \boldsymbol{y}_j)$. We define our test statistics as,

$$U_s = \sum_{1 \le i < j \le n} \tilde{K}_{ij}^n \tilde{S}_{ij}^n, \tag{2.1}$$

where $\tilde{K}_{ij}^n$ and $\tilde{S}_{ij}^n$ are the ith row and jth column of the empirical kernel matrices. For notation simplicity, we dropped the superscript n. Under the null hypothesis, we can show that $E(k \otimes s(\boldsymbol{D}_i, \boldsymbol{D}_j)|\boldsymbol{D}_i) \equiv 0$ and $U_s = 0$ (Appendix A1 of the Supplementary material available at *Biostatistics* online), where $\boldsymbol{D}_i = (\boldsymbol{G}_i, \boldsymbol{Y}_i)$. Therefore, U_s is a degenerate U-statistics with the kernel $k \otimes s$. Limit theory for degenerate U-statistics with fixed kernel functions has been well studies by Weber (1981), Shieh (1997), and used by Wei and Lu (2017), Wu *and others* (2011, 2013). In that case, the limit distribution is a mixture of independent χ_1^2. However, the kernel used in U_s implicitly depends on the space $\Omega_Y \times \Omega_G$, and thus it depends on the dimension of outcomes (p_y) and predictors (p_g). Using the spectrum decomposition theorem, each kernel function involved in U_s can be decomposed as

$$\tilde{s}(\boldsymbol{g}_1, \boldsymbol{g}_2) = \sum_{i=1}^{\infty} \theta_{gi} \phi_{gi}(\boldsymbol{g}_1) \phi_{gi}(\boldsymbol{g}_2), \qquad \tilde{k}(\boldsymbol{y}_1, \boldsymbol{y}_2) = \sum_{i=1}^{\infty} \theta_{yi} \phi_{yi}(\boldsymbol{y}_1) \phi_{yi}(\boldsymbol{y}_2),$$

where θ_{gi} (θ_{yi}) and $\phi_{gi}(\cdot)$ ($\phi_{yi}(\cdot)$) are eigenvalues and orthonormal eigenfunctions for the kernel $\tilde{s}(\boldsymbol{g}_1, \boldsymbol{g}_2)$ ($\tilde{k}(\boldsymbol{y}_1, \boldsymbol{y}_2)$). Let $p = p_g + p_y$, and $V_{s,m} = \sum_{i=1}^{\infty} \theta_{si}^m$ for any positive integer m, where $s \in (g, y)$. Using the Martingale Central Limit theorem (Brown, 1971) and the theorem in Hall (1984), we have the following result for U_s.

Theorem 2.1 Given $(k \otimes s)$ is symmetric and $E(k \otimes s(\boldsymbol{D}_1, \boldsymbol{D}_2)|\boldsymbol{D}_1) \equiv 0$. Assume $E\left[(k \otimes s)^2(\boldsymbol{D}_1, \boldsymbol{D}_2)\right] < \infty$ for each p. If $\dfrac{V_{g,4}}{V_{g,2}^2} \times \dfrac{V_{y,4}}{V_{y,2}^2} \rightarrow 0$ as $p \rightarrow \infty$, then U_s is asymptotically normally distributed with mean zero and variance given by $\frac{1}{2} n^2 E\left[(k \otimes s)^2(\boldsymbol{D}_1, \boldsymbol{D}_2)\right]$.

The details of the proof is shown in Appendix A2 of the Supplementary material available at *Biostatistics* online, and we estimate $E\left[(k \otimes s)^2(\boldsymbol{D}_i, \boldsymbol{D}_j)\right]$ empirically from the sample.

Various kernel functions can be used to measure the predictors' similarities (i.e., $s(\boldsymbol{g}_i, \boldsymbol{g}_j)$). For example, for both categorical genetic data and continuous gene expression data, a linear kernel that corresponds to the linear additive effects or a quadratic kernel that captures interaction effects can be used.

Similarly, various kernels can be used for the outcomes. As outcomes can be measured on different scales and are subjected to outliers, we propose to use a rank-based kernel to improve the robustness of the test and accommodate the situations where outcomes come from various distributions (e.g., Cauchy). Define $\boldsymbol{P}_0 = \boldsymbol{I}_n - \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}$ and $\boldsymbol{e} = (\boldsymbol{\epsilon}_1, \boldsymbol{\epsilon}_2, \ldots, \boldsymbol{\epsilon}_n)'$, where $\boldsymbol{e}$ can be viewed as residuals after adjusting the effects of covariates $\boldsymbol{X} = (\boldsymbol{X_1}, \boldsymbol{X_2}, \ldots, \boldsymbol{X_n})'$. It is apparent that $\boldsymbol{e}$ is residual from a linear regression model, when outcomes are normally distributed. Let $R_{il} = \text{rank}(e_{il})$, where e_{il} is the residual for the lth outcome for subject i. Note that outcomes can be measured on different scales, and thus the rank is evaluated for each outcome. Let $\boldsymbol{R}_i = (R_{i1}, R_{i2}, \ldots, R_{ip_y})$, and we define the rank-based kernel for the outcomes as,

$$k(\boldsymbol{y}_i, \boldsymbol{y}_j) = \frac{\boldsymbol{R}_i' \boldsymbol{R}_j}{p_y}. \tag{2.2}$$

U_s embeds both the predictors and outcomes in RKHS, and can work well when both of them are high-dimensional. In addition, U_s has weaker assumptions as compared to those in He *and others* (2019), where they showed that many widely used kernels (e.g., linear, polynomial, and identity-by-state kernels) satisfy the assumption $\frac{V_{g,4}}{V_{g,2}^2} \to 0$ as $p_g \to \infty$ under mild conditions. Provided $\frac{V_{y,4}}{V_{y,2}^2}$ is bounded, all the kernels that meet their assumptions work for U_s. Indeed, it is easy to see that $\frac{V_{y,4}}{V_{y,2}^2} = 1$ under their settings, and thus their method can be viewed as a special case of our method. In a similar fashion, our method works with a range of kernels under the assumption that $\frac{V_{y,4}}{V_{y,2}^2} \to 0$ as $p_y \to \infty$ for outcomes and $\frac{V_{g,4}}{V_{g,2}^2}$ is bounded for predictors. The rank-based kernel function $\tilde{k}(\boldsymbol{y_1}, \boldsymbol{y_2})$ proposed for measuring outcome-similarity can be viewed as a linear kernel, and thus satisfies the assumption $\frac{V_{y,4}}{V_{y,2}^2} \to 0$ as $p_y \to \infty$ with mild conditions. When outcomes are high-dimensional, our method is flexible in choosing kernel functions for predictors. Instead of requesting a relatively strong condition of $\frac{V_{g,4}}{V_{g,2}^2} \to 0$ as $p_g \to \infty$, we only require $\frac{V_{g,4}}{V_{g,2}^2}$ to be bounded.

2.3. *Hypothesis testing based on multiple candidate kernels and their combinations*

In the previous section, we proposed U_s (equation 2.1) that is robust to various distributions and outliers via defining a rank-based kernel function for the outcomes. To increase the power of the test, we want to choose the optimal kernel function $s(\boldsymbol{g}_i, \boldsymbol{g}_j)$ for predictors, which can be challenging in practice. The form of functions resided in the RKHS generated by the kernel function $s(\cdot, \cdot)$ is characterized by itself, and thus the kernel used for measuring predictors' similarities reflects the assumptions about the functional relationships between the outcomes and predictors. For example, a linear kernel indicates a linear additive relationship, and a quadratic kernel implies the pairwise interaction effects. While kernel functions are flexible in modeling various types of effects, challenges arise given the true effects are unknown in advance. It has been shown that misspecified kernel functions can substantially limit the power of an association test (He *and others*, 2019; Wu *and others*, 2013).

To facilitate the selection of optimal kernels that capture the relationships between predictors and outcomes, we propose a multivariate U-statistics based on U_s, where a set of M candidate kernels $(s_1(\cdot, \cdot), s_2(\cdot, \cdot), \ldots, s_M(\cdot, \cdot))$ and their combinations are considered. Let $\boldsymbol{\psi}_l = (\psi_{l1}, \psi_{l2}, \ldots, \psi_{lM})$ be a weight vector associated with the lth combination of the M kernel functions, where $|\psi_{lj}| \in (0, 1)$ for all $j \in (1, \ldots, M)$. Correspondingly, the kernel function and its kernel matrix under the lth combination is $s^l(\cdot, \cdot) = \sum_m \psi_{lm} s_m(\cdot, \cdot)$ and $\tilde{\boldsymbol{S}}^l = \sum_m^M \psi_{lm} \tilde{\boldsymbol{S}}_m$. To test whether the predictors and outcomes are associated given $\tilde{\boldsymbol{S}}^l$, $U_l = \sum_{i<j} \tilde{K}_{ij} \tilde{S}^l_{ij}$ statistics is constructed and Theorem 2.1 is used to derive its p-value (denoted by p_l). Different from existing literature that only selects one kernel to capture a specific type of effect (He *and others*, 2019; Wu *and others*, 2013), we use the idea of multi-kernel learning algorithms (Dereli *and others*, 2019), where multiple kernels are selected to capture several types of effects. For example, our method allows for the selection of linear and quadratic kernels simultaneously so that both the linear and pairwise interaction effects can be captured.

Given a total of N_M combinations of the M kernels, a vector of $\boldsymbol{U} = (\boldsymbol{U}_O, \boldsymbol{U}_C)' = (U_1, U_2, \ldots, U_{N_M})'$ can be constructed and their p-values $\boldsymbol{p} = (\boldsymbol{p}_O, \boldsymbol{p}_C)' = (p_1, p_2, \ldots, p_{N_M})$ can be obtained, where $\boldsymbol{U}_O$ is a $M \times 1$ vector of test statistics associated with the M kernel itself (i.e., $\psi_{lm} = 1, \psi_{lk} = 0, \forall k \neq m$) and $\boldsymbol{U}_C$ is a $(N_M - M) \times 1$ vector of test statistics associated with kernel combinations (i.e., at least two kernels are selected). We defined our test statistics (U_{kmu}) as the minimum of p-values from all combinations of

M kernels:

$$U_{\text{kmu}} = \underset{l\in(1,\cdots,N_M)}{\text{argmin}}\; p_l. \tag{2.3}$$

The distribution of U_{kmu} is usually obtained through a perturbation procedure (Wu *and others*, 2013). However, it is computationally expensive, especially when the combinations of kernels are also considered. Using Theorem 2.1 that U_l follows a normal distribution under the null hypothesis, the analytical form of the distribution of U_{kmu} can be derived, which relies on the well-developed results for multivariate normal distribution. Let $\boldsymbol{\Psi} = (\boldsymbol{\psi}_1, \boldsymbol{\psi}_2, \ldots, \boldsymbol{\psi}_{N_M})'$ be the weight matrix associated with all the combination of kernels. The $\boldsymbol{U}$ vector can be written as $\boldsymbol{U} = \boldsymbol{\Psi}\boldsymbol{U}_O$, and we have the following result (proofs are in Appendix A3 of the Supplementary material available at *Biostatistics* online).

THEOREM 2.2 Assume condition $\dfrac{V_{gm,4}}{V_{gm,2}^2} \times \dfrac{V_{y,4}}{V_{y,2}^2} \to 0$ as $p \to \infty$ is satisfied for all candidate kernels $S_m, \forall m \in (1, \ldots, M)$.

1. For $\boldsymbol{U_O}$ that is derived based on a single kernel from the candidate set, we have $\boldsymbol{U_O} \xrightarrow{d} \boldsymbol{Z}_O$, where $\boldsymbol{Z_O} = (Z_1, Z_2, \ldots, Z_M)'$ follows a multivariate normal distribution with mean $\mathbf{0}_M$ and covariance $\boldsymbol{\Sigma}_{M\times M}$.
2. For $\boldsymbol{U}$ that is derived based on a single kernel and multiple candidate kernels, we have $\boldsymbol{U} \xrightarrow{d} \boldsymbol{Z}$, where $\boldsymbol{Z} = (Z_1, Z_2, \cdots, Z_{N_M})$ follows a degenerate multivariate normal distribution with mean $\mathbf{0}_{N_M}$ and covariance $\boldsymbol{\Sigma} = \boldsymbol{\Psi}\boldsymbol{\Sigma}_{M\times M}\boldsymbol{\Psi}'$

If only a single kernel is selected from the candidate set as existing literature (He *and others*, 2019; Wu *and others*, 2013), we can simplify the overall test statistics as

$$U^0_{\text{kmu}} = \underset{l\in(1,\ldots,M)}{\text{argmin}}\, p_l, \tag{2.4}$$

where $p_l \in \boldsymbol{p}_O$. While the perturbation procedure can be employed to obtain its distribution, we directly derive its asymptotic results based on the first part of Theorem 2.2. For a given value of α, we have

$$\begin{aligned} Pr(U^0_{\text{kmu}} < \alpha) =& 1 - Pr(U^0_{\text{kmu}} \geq \alpha) = 1 - Pr(p_l \geq \alpha, \forall p_l \in \boldsymbol{p}_O) \\ =& 1 - Pr\left[Pr\left(U_l > |u_l|\right) \geq \alpha, \forall l \in (1, \ldots, M)\right] \\ =& \left[1 - Pr\Big(|\boldsymbol{\Sigma}^{-1}_{0,M\times M}\boldsymbol{Z}_O| \leq \Phi^{-1}(1-\alpha/2)\mathbf{1}_M\Big)\right](1 + o(1)), \end{aligned} \tag{2.5}$$

where $\boldsymbol{\Sigma}_{0,M\times M} = \text{diag}(\sigma_1, \ldots, \sigma_M)$ and σ_l^2 is the lth diagonal element of matrix $\boldsymbol{\Sigma}_{M\times M}$. The leading term in equation (2.5) can be efficiently calculated by many existing software (e.g., `mvnorm` package in R).

As evidenced by the multi-kernel learning algorithms, there are many situations where a single kernel is not sufficient in capturing complex relationships. To allow multiple kernels being selected, we define the test statistics using equation 2.3 and derived its distribution based on the second part of Theorem 2.2. For a given value of α, the tail probability is

$$\begin{aligned} Pr(U_{\text{kmu}} < \alpha) =& 1 - Pr(U_{\text{kmu}} \geq \alpha) = 1 - Pr(p_l \geq \alpha, \forall p_l \in \boldsymbol{p}) \\ =& \left[1 - Pr\Big(|\boldsymbol{\Sigma}_0^{-1}\boldsymbol{Z}| \leq \Phi^{-1}(1-\alpha/2)\mathbf{1}_{N_M}\Big)\right](1 + o(1)), \end{aligned} \tag{2.6}$$

where $\mathbf{\Sigma}_0 = \text{diag}(\sigma_1, \ldots, \sigma_{N_M})$ and σ_l^2 is the lth diagonal element of the covariance matrix $\mathbf{\Sigma}$. While different from the case where a single kernel is selected, $\boldsymbol{Z}$ follows a degenerate multivariate normal distribution and its dimension can be large (i.e., grows exponentially with the number of kernels). Therefore, the standard software is not directly applicable. To facilitate the computation of the tail probability, we used the fact that all elements in $\boldsymbol{Z}$ are linear combinations of the elements in $\boldsymbol{Z_O}$. Let $\mathbf{\Lambda} = \text{diag}(\lambda_1, \lambda_2, \ldots, \lambda_M)'$ and $\boldsymbol{A}$ be the matrices of eigenvalues and the corresponding eigenvectors for the covariance matrix $\mathbf{\Sigma}_{M\times M}$. $\boldsymbol{Z}$ can be written as $\boldsymbol{Z} = \mathbf{\Psi}\boldsymbol{Z_O} = \mathbf{\Psi}\boldsymbol{A}\mathbf{\Lambda}^{1/2}\boldsymbol{Z}_0$, where $\boldsymbol{Z}_0$ is a $M \times 1$ vector with each element being independent standard normal variables. Instead of treating $\boldsymbol{Z}$ as a random variable, we consider the low-dimensional independently normally distributed $\boldsymbol{Z}_0$. Solving equation (2.6) is equivalent to the computation of a normal distribution with integration region specified through a set of linear inequalities. This can be efficiently calculated through a stochastic multiple integration algorithm designed by Genz (1992) and Genz and Bretz (1999).

While we mainly focus on the kernel selection for the predictors in this article, the Theorem 2.2 and equation (2.6) also apply for outcome kernel selections. Therefore, our proposed method can be applied to select optimal kernels for both predictors and outcomes.

3. Simulation

We conducted simulation studies to evaluate the performance of our method under different outcome distributions and disease models. For biological studies, the predictors can be both continuous (e.g., gene expression data) and categorical (e.g., genetic data). Therefore, we considered both cases. For continuous predictors, we set the number of variables to be 30 and 100, and each predictor was simulated from a standard normal distribution. For categorical predictors, we randomly selected a gene from chromosome 13 and used the HAPGEN2 to simulate the genotypes (Su *and others*, 2011), where both common and rare variants were included.

To form the kernel set for the proposed method, a linear kernel for the additive effects and a quadratic kernel for pairwise interactions were considered for both continuous and categorical predictors. In addition to these two kernels, we included a Gaussian kernel to capture other non-additive effects for continuous predictors and a weighted linear kernel with beta weights to account for the effects from rare variants. We compared our methods to the widely used SKAT-based methods with the default settings (Sequence kernel association test (SKAT)). These include Lee *and others* (2012) that assumes the outcomes have homogeneous causes (denoted by SKATh) and Aschard *and others* (2014) that allows for heterogeneous effects (denoted by SKATp). We evaluated the performance of all methods under various sample sizes (i.e., 500, 1000, and 2000). 5000 and 1000 Monte Carlo simulations were conducted under each setting to evaluate the type I error and power, respectively.

3.1. *Simulation I: The impact of outcome distributions*

We first evaluated the impact of outcome distributions on the performance of our method. We simulated two outcomes for each subject ($\boldsymbol{Y}_i = (Y_{i1}, Y_{i2})'$):

$$Y_i = \alpha_0 z_{i0} + \alpha_1 z_{i1} + \left[\sum_j^2 \sum_k \left(\beta_k I_{k\in S_m} + \beta_{jk} I_{k\in S_{jm}}\right) x_{ik}\right] + \boldsymbol{\epsilon}_i, \tag{3.7}$$

where $z_{i0} \sim N(0, 1)$ and $z_{i1} \sim Ber(0.3)$ represent the continuous and categorical demographic variables (e.g., age and gender), and α_is are their corresponding effects. x_{ik} represents the kth predictor for the i subject. S_m and S_{jm} denote the set of variables that are associated with both outcomes and only the jth outcome ($j \in \{1, 2\}$), respectively. β_k (β_{jk}) denote the shared (outcome-specific) effects for the kth

Table 1. *Type I errors under different distributions for correlated outcomes*

		Continuous						**Categorical**		
		***p* = 30**			***p* = 100**			***p* varies**		
Distribution	***N***	**Ukmu**	**SKATh**	**SKATp**	**Ukmu**	**SKATh**	**SKATp**	**Ukmu**	**SKATh**	**SKATp**
Gaussian	500	0.0480	0.0431	0.0438	0.0512	0.0341	0.0344	0.0529	0.0485	0.0490
	1000	0.0493	0.0476	0.0481	0.0513	0.0403	0.0433	0.0532	0.0462	0.0455
	2000	0.0467	0.0456	0.0476	0.0504	0.0455	0.0469	0.0496	0.0486	0.0480
T-dist(df = 2)	500	0.0504	0.0445	0.0425	0.0469	0.0367	0.0350	0.0513	0.0465	0.0475
	1000	0.0537	0.0489	0.0494	0.0504	0.0425	0.0428	0.0504	0.0491	0.0432
	2000	0.0531	0.0510	0.0496	0.0521	0.0456	0.0425	0.0519	0.0500	0.0443
Mixture	500	0.0475	0.0440	0.0441	0.0477	0.0330	0.0341	0.0527	0.0481	0.0440
	1000	0.0504	0.0493	0.0478	0.0515	0.0453	0.0447	0.0517	0.0487	0.0470
	2000	0.0490	0.0493	0.0500	0.0488	0.0449	0.0460	0.0503	0.0498	0.0461

predictor, and is sampled from a uniform distribution. $I_{k\in S} = 1$ if k is in set S, and 0 otherwise. Among all the predictors, 20% of them are causal, of which 70% are shared by both outcomes. We considered three distributions for $\boldsymbol{\epsilon}_i$, including a multivariate normal distribution, a multivariate t-distribution with 2 degrees of freedom and a mixture of 90% normal and 10% Cauchy distribution. We simulated two scenarios for each distribution, including $\boldsymbol{\epsilon}_i$ are independent and correlated. The details of simulation settings and effect sizes were summarized in Table S1 of the Supplementary material available at *Biostatistics* online.

All methods have reasonably controlled the type I errors for both continuous and categorical predictors (Table 1 and Table S2 of the Supplementary material available at *Biostatistics* online). The empirical power for the correlated and independent outcomes are shown in Figure 1 and Figure S1 of the Supplementary material available at *Biostatistics* online, respectively. For all methods, the power increases as sample size increases. When the outcomes follow normal distributions, our method performs similarly to SKAT-based methods, regardless of the correlations between outcomes. However, when the outcomes are from a non-exponential family, our method performs substantially better than the others. This is mainly because we employ a rank-based kernel to measure outcome-similarity, which is robust under various distributions.

3.2. *Simulation II: The impact of underlying disease models*

In this set of simulations, we evaluated the impact of the underlying disease models on the performance of our method. For continuous predictors, we simulated the outcomes as

$$\begin{aligned}\boldsymbol{Y}_i =& \alpha_0 z_{i0} + \alpha_1 z_{i1} + \alpha^m \times \left[\sum_j^2 \sum_k \left(\beta_k^m I_{k\in S_m} + \beta_{jk}^m I_{k\in S_{jm}}\right) x_{ik}\right] + \\ &\alpha^o \times \left[\sum_j^2 \sum_k \left(\beta_k^o I_{k\in S_o} + \beta_{jk}^o I_{k\in S_{jo}}\right)\left(e^{-x_{ik}^2/100} + \cos x_{ik}^2\right)\right] + \\ &\alpha^p \times \left[\sum_j^2 \sum_{k,k'} \left(\beta_{kk'}^p I_{k,k'\in S_p} + \beta_{jkk'}^p I_{k,k'\in S_{jp}}\right) x_{ik} x_{ik'}\right] + \boldsymbol{\epsilon}_i,\end{aligned}$$

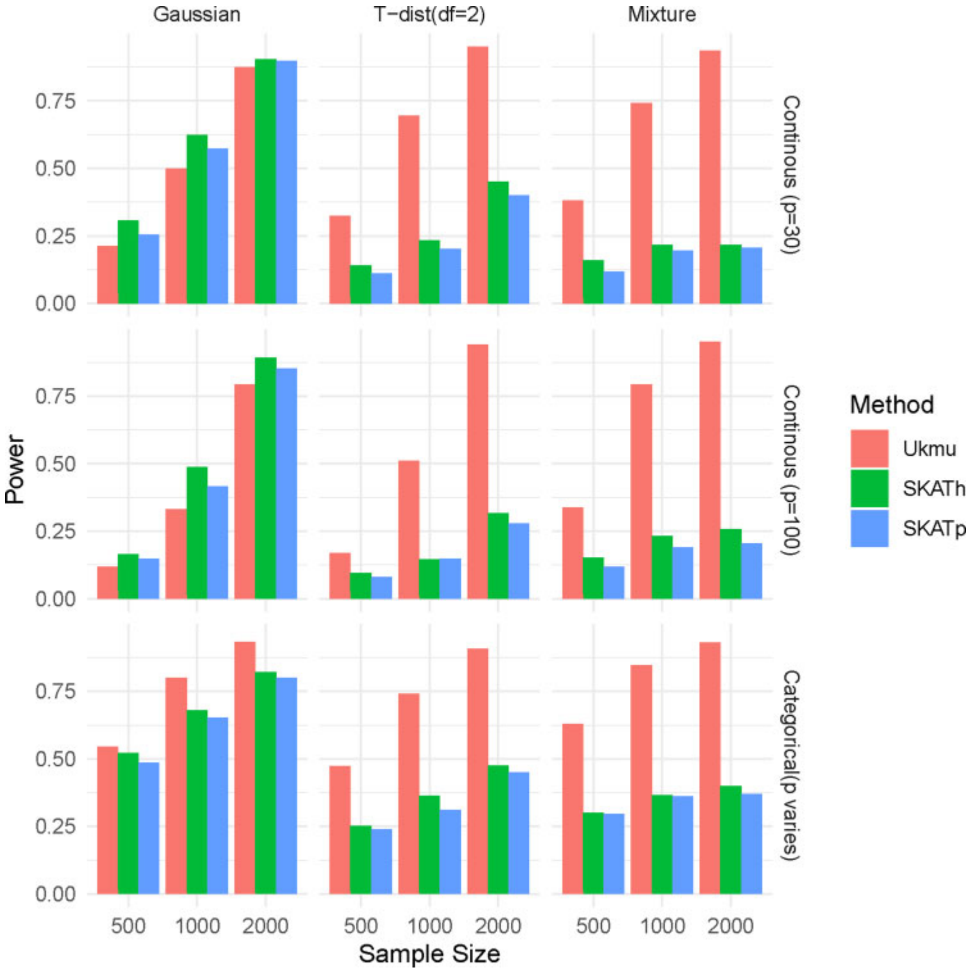

Fig. 1. The impact of outcome distributions when outcomes are correlated

where $\alpha^m = 1$, $\alpha^p = 1$, and $\alpha^o = 1$ if linear, pairwise interaction and other non-additive effects are present, and 0 otherwise. S_i and S_{ji} $(i \in \{m, p, o\}, j \in \{1, 2\})$ respectively denote the set of variables that are associated with both outcomes and only the jth outcome for the ith effect. β_k^i and β_{jk}^i $(i \in \{m, p, o\})$ are the shared effect and outcome-specific effect for the kth predictor, and are sampled from uniform distributions. Similar to simulation 1, 20% of the predictors are set causal, of which 70% are shared by both outcomes for each effect. In total, we considered 6 non-linear models, including the pairwise interaction only model (i.e., $\alpha^p = 1, \alpha^m = \alpha^o = 0$), other non-additive effect only model (i.e., $\alpha^o = 1, \alpha^m = \alpha^p = 0$), and various combinations (Table S3 of the Supplementary material available at *Biostatistics* online). For categorical variables, we simulated the outcomes as,

$$\boldsymbol{Y_i} = \alpha_0 z_{i0} + \alpha_1 z_{i1} + \alpha^m \times \left[\sum_j^2 \sum_k \left(\beta_k^m I_{k \in S_m} + \beta_{jk}^m I_{k \in S_{jm}} \right) x_{ik} \right] +$$

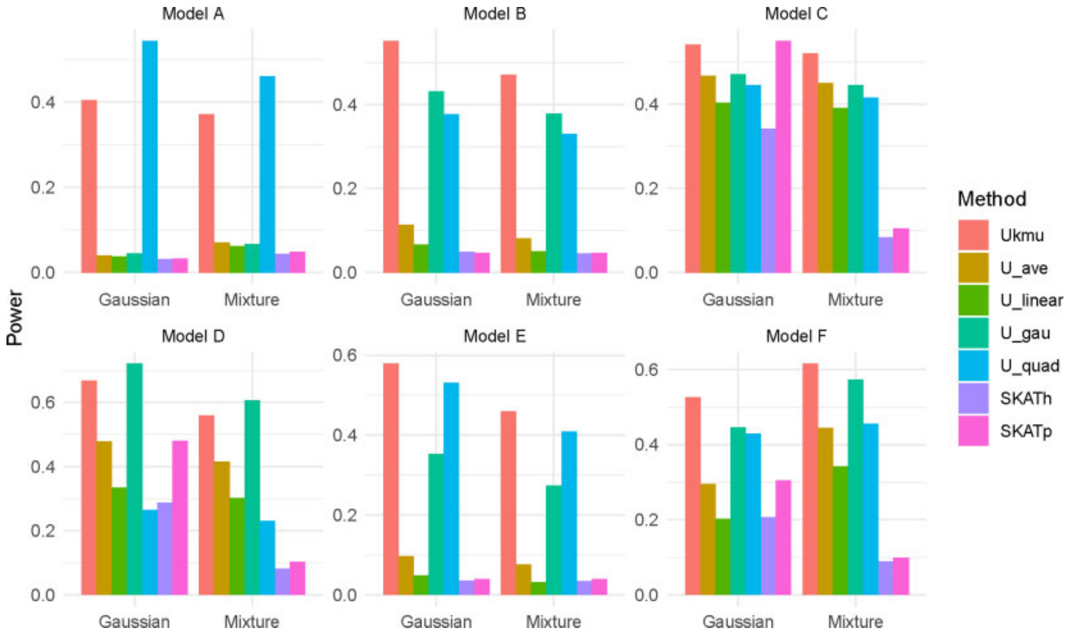

Fig. 2. The impact of disease model when predictors are continuous ($N = 1000$, $p = 100$). Model A: Pairwise interaction effects. Model B: Other non-additive effects. Model C: Linear and pairwise interaction effects. Model D: Linear and other non-additive effects. Model E: Pairwise interaction and other non-additive effects. Model F: Linear, pairwise interaction and other non-additive effects.

$$\alpha^p \times \left[\sum_{j}^{2} \sum_{k,k'} \left(\beta^p_{kk'} I_{k,k' \in S_p} + \beta^p_{jkk'} I_{k,k' \in S_{jp}} \right) x_{ik} x_{ik'} \right] + \boldsymbol{\epsilon}_i.$$

We considered three disease models for categorical predictors, including (1) a linear additive model caused by rare variants; (2) a linear additive model caused by common variants; and (3) a linear additive and pairwise interaction model. For both continuous and categorical predictors,we considered two types of distributions for $\boldsymbol{\epsilon}_i$. These included a multivariate normal as well as a mixture of 90% normal and 10% Cauchy distributions, where variance for $\boldsymbol{\epsilon}_i$ is compound symmetric with correlation equal 0.2. The details of simulation settings and effect sizes were summarized in Table S3 of the Supplementary material available at *Biostatistics* online. For comparison purposes, we built a single-kernel-based U_s with $s \in$ {linear, quad, gau, weight}, and further built a single-kernel-based U_{ave}, where all kernels in the candidate set were averaged.

The power for continuous predictors is shown in Figure 2 and Appendix Figures S2 and S3 of the Supplementary material available at *Biostatistics* online. U_s, whose kernel reflects the underlying disease model, achieves the best performance. U_{kmu} that considers all kernels and their combinations has close-to-the-best performance. For example, when the disease is caused by pairwise interactions (i.e., Model A), U_s with quadratic kernel performs the best, followed by U_{kmu}. However, when the relationship between predictors and outcomes is complex, U_{kmu} that considers multiple kernels achieves the best performance. For example, when $\boldsymbol{Y}_i = \left[\sum_j^2 \sum_k \left(\beta^o_k I_{k \in S_o} + \beta^o_{jk} I_{k \in S_{jo}}\right)\left(e^{-x^2_{ik}/100} + \cos x^2_{ik}\right)\right]$ (i.e., Model B), U_{kmu} performs much better than any of those single-kernel-based U_s. This clearly indicates the benefits of a multi-kernel-based test when the underlying disease model is complicated. In addition, U_{kmu} performs

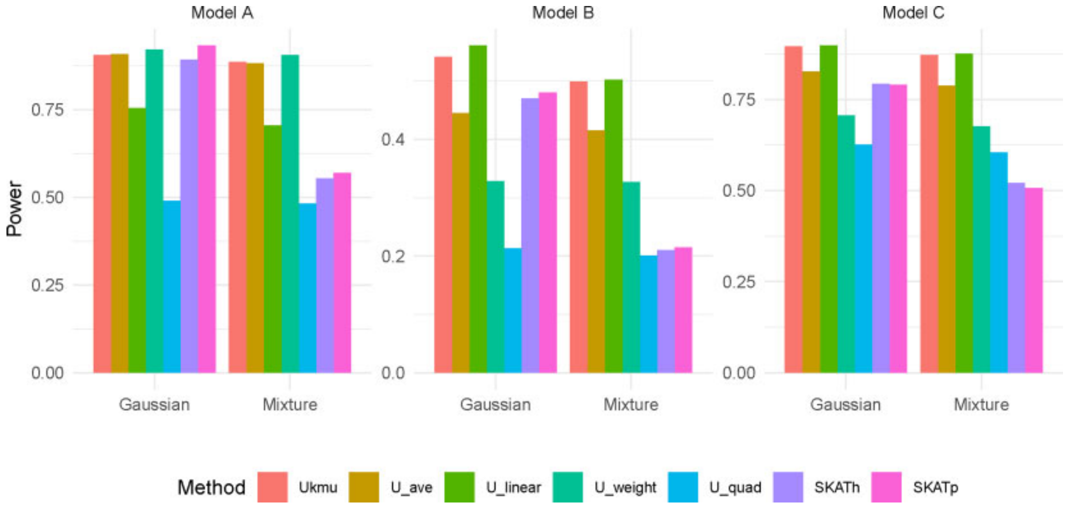

Fig. 3. The impact of disease model when predictors are categorical ($N = 1000$). Model A: Rare variants have effects. Model B: Common variants have effects. Model C: Linear and pairwise interaction effects.

better than the SKAT-based methods, and the advantages are much larger when the effects of predictors on the outcomes are not linear and/or the outcomes follow a non-exponential family distribution.

The power for categorical predictors when $N = 1000$ is shown in Figure 3 and the rest are in Figure S4 of the Supplementary material available at *Biostatistics* online. Similar to continuous predictors, U_s that uses the optimal kernel has the highest power, and U_{kmu} has close-to-the-best performance. When the outcomes are not simply caused by rare variants, U_{kmu} outperforms SKAT-based methods, especially when outcomes are not normally distributed.

4. Real data application

We analyzed the WGS data from the Alzheimer's Disease Neuroimaging Initiative (ADNI) study. ADNI is a longitudinal study that assesses the effects of genetic variants on Alzheimer's Disease (AD) and various AD-related outcomes, including 3D brain imagining and cognitive measurements (Saykin *and others*, 2010). DNA samples from study subjects were obtained and analyzed using the Illumina's non-CLIA (Clinical Laboratory Improvement Amendments) WGS. A total of 808 subjects was included in the data, with 280 controls, 234 and 246 early and late cognitive impaired subjects, and 46 AD patients at their baseline assessment.

We are interested in the association between genetic variants and brain imaging outcomes, including 18F-fluoro-2-deoxyglucose, Hippocampus, Entorhinal, 8F-florbetapir (AV45), Fusiform, and Ventricles measurements. These outcomes were treated as a multivariate outcome and their distributions were shown in Figure S5 of the Supplementary material available at *Biostatistics* online. The rank-based kernel (equation 2.2) was used to measure the outcome similarity. For genetic data, we excluded genetic variants with more than 20% missing, and grouped them based on gene range listed in the GRch37 assembly. After quality control and the grouping process, a total of 22 494 autosomal genes harboring 1.8 million genetic variants were included in our study. To form the candidate kernel set, we considered the linear, weighted linear with beta weight, and the quadratic kernels to capture the effects from common variants, rare variants and the pairwise interactions, respectively. For all the analyses, we included age, gender,

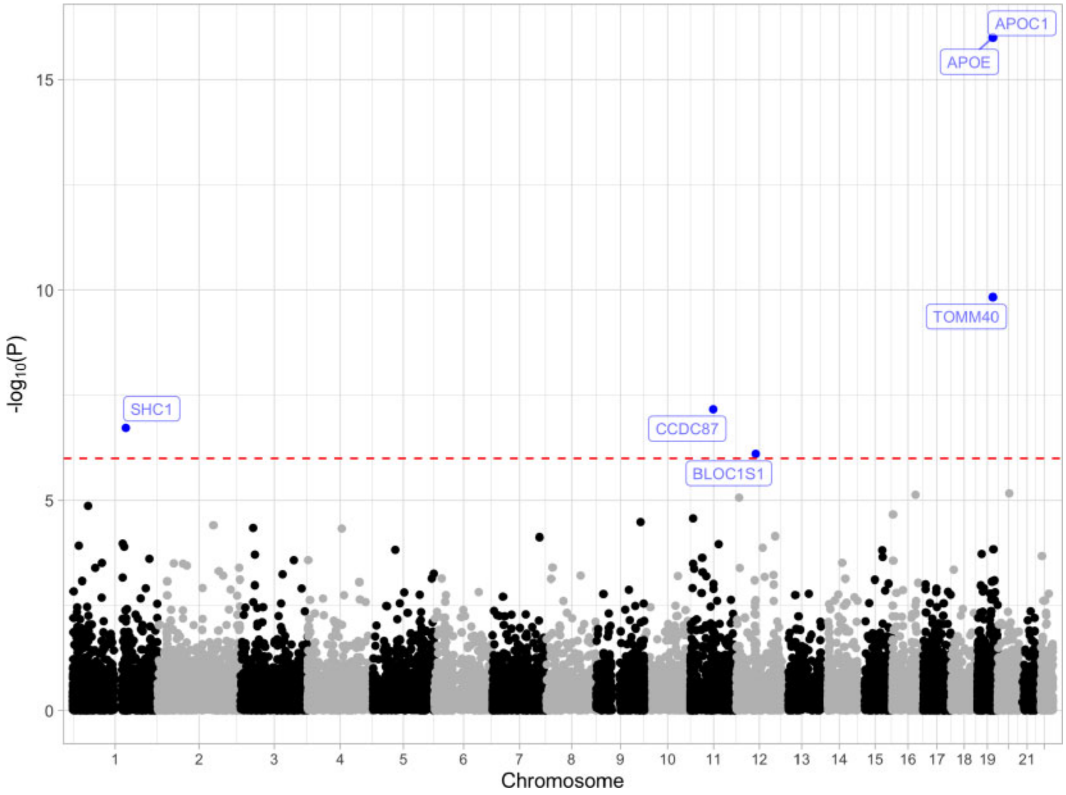

Fig. 4. Manhattan plot for whole genome gene-based analysis using U_{kmu}.

education, and the top 20 genome principle components as the covariates to adjust for the potential confounding effects. For comparison purposes, we also analyzed the dataset with SKAT-based methods with their default settings, where the same set of covariates were used.

QQ plot shows no evidence of systematic bias of the methods (Figure S6 of the Supplementary material available at *Biostatistics* online). The Manhattan plots for U_{kmu} and the others are shown in Figure 4 and Figure S7 of the Supplementary material available at *Biostatistics* online, respectively. Our proposed multivariate U_{kmu} method has detected 6 genes that achieve the genome-wide significance level (i.e., $p \leq 10^{-6}$). In addition to the *APOE* and *APOC* genes that have also been detected by the single-kernel-based U_s methods (Figure S7 of the Supplementary material available at *Biostatistics* online), U_{kmu} has detected 4 additional genes (i.e., *TOMM40*, *SHC1*, *CCDC87*, and *BLOC1S1*). This indicates jointly considering multiple kernels has the potential to capture more complicated relationships between predictors and outcomes, and thus may improve the power of the test. While U_s has detected *APOE* and *APOC* genes that have well-known association with AD, the SKAT-based methods failed to identify any significant genes.

All of the six significant genes detected by U_{kmu} have reported evidence suggesting their association with AD and its related phenotypes. For example, mounting evidence have suggested that *APOE* is related to AD and its associated outcomes (Hoffmann *and others*, 2016). *APOC1* gene is reported to be a genetic risk factor for dementia and cognitive impairment in the elderly, and it has significant impact on hippocampal volumes (Serra-Grabulosa *and others*, 2003). For the *TOMM40* gene, some suggest their effects on AD-related phenotypes are through its linkage disequilibrium with *APOE* (Maruszak *and others*, 2012), while others found evidence of an independent association between *TOMM40* and AD-related phenotypes, such as hippocampal thinning (Burggren *and others*, 2017) and gray matter volume (Johnson *and others*, 2011). Using a weighted burden test, Curtis *and others* (2020) found that *CCDC87* could affect the susceptibility to the late onset AD. Liang *and others* (2012) found that the expression pattern of *SHC1* is associated with AD progression, and Montibeller and de Belleroche (2018) found that the *BLOC1S1* is profoundly up-regulated in the frontal cortex of AD patients.

5. Discussion

In this study, we developed a flexible framework to test the association between a set of predictors and multiple outcomes, while adjusting for the confounding effects. We first proposed a robust U_s statistic, where a rank-based kernel is developed for outcomes. Based on U_s, we further developed U_{kmu} and its asymptotic distribution, where a set of candidate kernels were combined to optimally model the relationship between outcomes and predictors. Through simulations and real data analysis, we showed that U_{kmu} (1) is robust against the distributions of the outcomes; (2) performs as good as the kernel that reflects the underlying disease models; and (3) outperforms single kernel-based methods when the relationships between predictors and outcomes are complex.

Many published work can be viewed as special cases of U_{kmu} with corresponding values of $\boldsymbol{K}$ and $\boldsymbol{S}$, which reflect the assumed disease models. Fundamentally different from these methods that employ only one kernel function to capture a specific type of relation (e.g., linear or pairwise interaction), U_{kmu} works in a similar fashion as multi-kernel learning algorithms, where multiple kernels are combined in a data-driven manner to capture the relationships between predictors and outcomes. Combining multiple kernels in a data-driven manner offers greater flexibility and provides the capacity in modeling complex relationships (e.g., both linear and non-linear effects are present). The proposed U_{kmu} is shown to have controlled type I error and improved power over the weaker kernels in the set, without prior knowledge of the underlying disease mechanisms. While we used equal weights for kernel combinations, other weighting schemes (e.g., using prior information regarding the relationships between predictors and outcomes) can also be employed.

It is well accepted that power for a kernel-based method can be substantially reduced if the selected kernel does not reflect the underlying biology. In the existing literature, efforts were made for selecting a single kernel from a candidate set. These methods are mainly based on a perturbation procedure developed by Wu *and others* (2013), which can be computationally expensive. While we use a data-driven manner to select/combine multiple kernels from a candidate set, our method is not computationally demanding, as the significance is derived based on the asymptotic property of the test statistics. While we mainly focus on the kernel selection for predictors, the asymptotic results also apply for the selection of kernels for both predictors and outcomes. This makes our method efficient and scalable to genome-wide studies with various types of underlying mechanisms.

Different from the widely used KMRs, our method adopted a rank-based kernel function to measure the phenotypic similarity, making it robust against the outcome distributions. As shown in simulation 1, our method has similar performance to that from SKAT-based methods when outcomes come from the exponential family, but significantly outperforms them for other distributions. While we demonstrated the methods with multiple outcomes, it is trivial to see that the proposed rank-based kernel also works for single outcome. Regardless of the number of outcomes, our method can provide robustness against distributional assumptions. We consider this important, as the outcomes can come from any distributions in practice.

In this study, we focused on using linear and weighted linear kernels to capture the effects from common and rare variants in genetic data without considering any other additional information, such as the association signals from prior studies. Our method can be further improved by adopting kernel functions to incorporate prior biological knowledge, and this can be a future revenue of our research.

Nevertheless, we have developed a powerful and flexible framework for testing the association between a set of predictors and outcome(s). The proposed method has robust power regardless of the underlying biology and outcome distributions. It can be easily scaled to genome-wide study and will contribute to detecting biomarkers with pleiotropic effects.

6. Software

Software in the form of R package, together with a sample input data set and complete documentation is at https://github.com/YaluWen/KMU.

Supplementary material

Supplementary material is available online at http://biostatistics.oxfordjournals.org.

Acknowledgments

The authors wish to acknowledge the use of New Zealand eScience Infrastructure (NeSI) high performance computing facilities, consulting support and/or training services as part of this research. New Zealand's national facilities are provided by NeSI and funded jointly by NeSI's collaborator institutions and through the Ministry of Business, Innovation and Employment's Research Infrastructure programme (https://www.nesi.org.nz.).
Conflict of Interest: None declared.

Funding

This project is funded by the Faculty Research Development Fund from the University of Auckland, the Marsden Fund from Royal Society of New Zealand (Project No. 19-UOA-209), the National Library of Medicine (Award No. 1R01LM012848-01), and the National Institute on Drug Abuse (Award No. R01DA043501).

References

Alberti, K. G., Zimmet, P., Shaw, J. and IDF Epidemiology Task Force Consensus Group. (2005). The metabolic syndrome—a new worldwide definition. *Lancet* **366**, 1059–1062.

Aschard, H., Vilhjalmsson, B. J., Greliche, N., Morange, P. E., Tregouet, D. A. and Kraft, P. (2014). Maximizing the power of principal-component analysis of correlated phenotypes in genome-wide association studies. *American Journal of Human Genetics* **94**, 662–676.

Broadaway, K. A., Cutler, D. J., Duncan, R., Moore, J. L., Ware, E. B., Jhun, M. A., Bielak, L. F., Zhao, W., Smith, J. A., Peyser, P. A., *and others*. (2016). A statistical approach for testing cross-phenotype effects of rare variants. *American Journal of Human Genetics* **98**, 525–540.

Brown, B. M. (1971). Martingale central limit theorems. *Annals of Mathematical Statistics* **42**, 59–66.

Burggren, A. C., Mahmood, Z., Harrison, T. M., Siddarth, P., Miller, K. J., Small, G. W., Merrill, D. A. and Bookheimer, S. Y. (2017). Hippocampal thinning linked to longer TOMM40 poly-T variant lengths in the absence of the APOE epsilon4 variant. *Alzheimers & Dementia* **13**, 739–748.

Curtis, D., Bakaya, K., Sharma, L. and Bandyopadhyay, S. (2020). Weighted burden analysis of exome-sequenced late-onset Alzheimer's cases and controls provides further evidence for a role for PSEN1 and suggests involvement of the PI3K/Akt/GSK-3beta and WNT signalling pathways. *Annals of Human Genetics* **84**, 291–302.

Dereli, O., Ouz, C. and Gnen, M. (2019). Path2Surv: pathway/gene set-based survival analysis using multiple kernel learning. *Bioinformatics* **35**, 5137–5145.

Dutta, D., Scott, L., Boehnke, M. and Lee, S. (2019). Multi-SKAT: general framework to test for rare-variant association with multiple phenotypes. *Genetic Epidemiology* **43**, 4–23.

Genz, A. (1992). Numerical computation of multivariate normal probabilities. *Journal of Computational and Graphical Statistics* **1**, 141–149.

Genz, A. and Bretz, F. (1999). Numerical computation of multivariate t-probabilities with application to power calculation of multiple contrasts. *Journal of Statistical Computation and Simulation* **63**, 103–117.

Hall, P. (1984). Central limit theorem for integrated square error of multivariate nonparametric density estimators. *Journal of Multivariate Analysis* **14**, 1–16.

He, T., Li, S., Zhong, P. S. and Cui, Y. (2019). An optimal kernel-based U-statistic method for quantitative gene-set association analysis. *Genetic Epidemiology* **43**, 137–149.

He, Z., Zhang, M., Lee, S., Smith, J. A., Kardia, S. L. R., Diez Roux, A. V. and Mukherjee, B. (2017). Set-based tests for the gene-environment interaction in longitudinal studies. *Journal of American Statistical Association* **112**, 966–978.

Hoffmann, K., Sobol, N. A., Frederiksen, K. S., Beyer, N., Vogel, A., Vestergaard, K., Braendgaard, H., Gottrup, H., Lolk, A., Wermuth, L., *and others*. (2016). Moderate-to-high intensity physical exercise in patients with Alzheimer's disease: a randomized controlled trial. *Journal of Alzheimers Disease* **50**, 443–453.

Johnson, S. C., La Rue, A., Hermann, B. P., Xu, G., Koscik, R. L., Jonaitis, E. M., Bendlin, B. B., Hogan, K. J., Roses, A. D., Saunders, A. M., *and others*. (2011). The effect of TOMM40 poly-T length on gray matter volume and cognition in middle-aged persons with APOE epsilon3/epsilon3 genotype. *Alzheimers & Dementia* **7**, 456–465.

Klei, L., Luca, D., Devlin, B. and Roeder, K. (2008). Pleiotropy and principal components of heritability combine to increase power for association analysis. *Genetic Epidemiology* **32**, 9–19.

Larson, N. B., McDonnell, S., Cannon Albright, L., Teerlink, C., Stanford, J., Ostrander, E. A., Isaacs, W. B., Xu, J., Cooney, K. A., Lange, E., *and others*. (2017). gsSKAT: rapid gene set analysis and multiple testing correction for rare-variant association studies using weighted linear kernels. *Genetic Epidemiology* **41**, 297–308.

Lee, S., Emond, M. J., Bamshad, M. J., Barnes, K. C., Rieder, M. J., Nickerson, D. A., Team, Nhlbi Go Exome Sequencing Project-ESP Lung Project, Christiani, D. C., Wurfel, M. M. and Lin, X. (2012). Optimal unified approach for rare-variant association testing with application to small-sample case-control whole-exome sequencing studies. *American Journal of Human Genetics* **91**, 224–237.

Liang, D., Han, G., Feng, X., Sun, J., Duan, Y. and Lei, H. (2012). Concerted perturbation observed in a hub network in Alzheimer's disease. *PLoS One* **7**, 1–17.

Liu, D., Lin, X. and Ghosh, D. (2007). Semiparametric regression of multidimensional genetic pathway data: least-squares kernel machines and linear mixed models. *Biometrics* **63**, 1079–1088.

Maruszak, A., Peplonska, B., Safranow, K., Chodakowska-Zebrowska, M., Barcikowska, M. and Zekanowski, C. (2012). TOMM40 rs10524523 polymorphism's role in late-onset Alzheimer's disease and in longevity. *Journal of Alzheimers Disease* **28**, 309–322.

Montibeller, L. and de Belleroche, J. (2018). Amyotrophic lateral sclerosis (ALS) and Alzheimer's disease (AD) are characterised by differential activation of ER stress pathways: focus on UPR target genes. *Cell Stress & Chaperones* **23**, 897–912.

Saykin, A. J., Shen, L., Foroud, T. M., Potkin, S. G., Swaminathan, S., Kim, S., Risacher, S. L., Nho, K., Huentelman, M. J., Craig, D. W., *and others*. (2010). Alzheimer's disease neuroimaging initiative biomarkers as quantitative phenotypes: genetics core aims, progress, and plans. *Alzheimers & Dementia* **6**, 265–273.

Serra-Grabulosa, J. M., Salgado-Pineda, P., Junque, C., Sole-Padulles, C., Moral, P., Lopez-Alomar, A., Lopez, T., Lopez-Guillen, A., Bargallo, N., Mercader, J. M., *and others*. (2003). Apolipoproteins E and C1 and brain morphology in memory impaired elders. *Neurogenetics* **4**, 141–146.

Shieh, G. S. (1997). Weighted degenerate U- and V-statistics with estimated parameters. *Statistica Sinica* **7**, 1021–1038.

Solovieff, N., Cotsapas, C., Lee, P. H., Purcell, S. M. and Smoller, J. W. (2013). Pleiotropy in complex traits: challenges and strategies. *Nature Reviews Genetics* **14**, 483–495.

Su, Z., Marchini, J. and Donnelly, P. (2011). HAPGEN2: simulation of multiple disease SNPs. *Bioinformatics* **27**, 2304–2305.

Tzeng, J. Y., Zhang, D., Chang, S. M., Thomas, D. C. and Davidian, M. (2009). Gene-trait similarity regression for multimarker-based association analysis. *Biometrics* **65**, 822–832.

van der Sluis, S., Posthuma, D. and Dolan, C. V. (2013). TATES: efficient multivariate genotype-phenotype analysis for genome-wide association studies. *PLoS Genetics* **9**, e1003235.

Weber, N. C. (1981). Incomplete degenerate U-statistics. *Scandinavian Journal of Statistics* **8**, 120–123.

Wei, C. and Lu, Q. (2017). A generalized association test based on U statistics. *Bioinformatics* **33**, 1963–1971.

Welter, D., MacArthur, J., Morales, J., Burdett, T., Hall, P., Junkins, H., Klemm, A., Flicek, P., Manolio, T., Hindorff, L. *and others*. (2014). The NHGRI GWAS Catalog, a curated resource of SNP-trait associations. *Nucleic Acids Res* **42**(Database issue), D1001–D1006.

Wu, B. and Pankow, J. S. (2016). Sequence kernel association test of multiple continuous phenotypes. *Genetic Epidemiology* **40**, 91–100.

Wu, M. C., Lee, S., Cai, T., Li, Y., Boehnke, M. and Lin, X. (2011). Rare-variant association testing for sequencing data with the sequence kernel association test. *American Journal of Human Genetics* **89**, 82–93.

Wu, M. C., Maity, A., Lee, S., Simmons, E. M., Harmon, Q. E., Lin, X., Engel, S. M., Molldrem, J. J. and Armistead, P. M. (2013). Kernel machine SNP-set testing under multiple candidate kernels. *Genetic Epidemiology* **37**, 267–275.

Zhan, X., Zhao, N., Plantinga, A., Thornton, T. A., Conneely, K. N., Epstein, M. P. and Wu, M. C. (2017). Powerful genetic association analysis for common or rare variants with high-dimensional structured traits. *Genetics* **206**, 1779–1790.

[*Received July 10, 2020; revised September 24, 2020; accepted for publication October 3, 2020*]