Kernel machine score test for pathway analysis in the presence of semi-competing risks

Matey Neykov, Boris P. Hejblum, Jennifer Sinnott

Kernel Machine Score Test for Pathway Analysis in the Presence of Semi-Competing Risks

Journal Title
XX(X):2–26

DOI: 10.1177/ToBeAssigned
www.sagepub.com/
SAGE

Matey Neykov[1], Boris P. Hejblum[2] and Jennifer A. Sinnott[3]

Abstract

In cancer studies, patients often experience two different types of events: a non-terminal event such as recurrence or metastasis, and a terminal event such as cancer-specific death. Identifying pathways and networks of genes associated with one or both of these events is an important step in understanding disease development and targeting new biological processes for potential intervention. These correlated outcomes are commonly dealt with by modeling progression-free survival, where the event time is the minimum between the times of recurrence and death. However, identifying pathways only associated with progression-free survival may miss out on pathways that affect time to recurrence but not death, or vice versa. We propose a combined testing procedure for a pathway's association with both the cause-specific hazard of recurrence and the marginal hazard of death. The dependency between the two outcomes is accounted for through perturbation resampling to approximate the test's null distribution, without any further assumption on the nature of the dependency. Even complex nonlinear relationships between pathways and disease progression or death can be uncovered thanks to a flexible kernel machine framework. The superior statistical power of our approach is demonstrated in numerical studies and in a gene expression study of breast cancer.

Keywords

Kernel Machines, Pathway Analysis, Resampling, Score Test, Semi-Competing Risks

Prepared using sagej.cls *[Version: 2015/06/09 v1.01]*

1 Introduction

Identifying genomic disruptions associated with cancer prognosis is an important first step in improving scientific understanding of the progression of the disease and identifying important biological processes to potentially target for treatment development. While we could pursue this goal gene-by-gene, interrogating data at the pathway level may be a more appealing approach. Examining pathways leverages the extensive and ever-expanding research of biologists working to improve our knowledge of how genes work together in groups. For example, the Molecular Signatures Database[1] now contains eight major collections of pathways, including a collection of so-called hallmark gene sets which culls information on fifty important biological states and processes from founder gene sets proposed by many different research groups. Moreover, there has been evidence in certain cancers that while many individuals will experience dysregulation in a few key pathways, the actual component of the pathway that is altered may differ across individuals.[2;3] Thus, focusing on commonly dysregulated pathways instead of commonly dysregulated genes may be a more effective approach for identifying key processes in disease progression.

A number of statistical methods have emerged in recent years to assess the significance of the association between a set of genes and an outcome of interest[4–13]. Methods based on kernel machines (KMs) are particularly attractive because they allow a lot of modeling flexibility while relying on specification of "similarity" between two individuals' pathway expression, which can be easier to conceptualize and discuss with disease researchers. KM methods can be implemented with a linear kernel to induce a model with linear effects, or with a variety of nonlinear kernels to allow increasingly complex models with nonlinear effects without explicitly specifying the form of those effects. Furthermore, when the sampling distributions of KM tests are estimated with perturbation resampling, it is easy to combine information across tests using different kernels and control for multiple testing when many pathways are under consideration.[8;11;12].

When testing for a pathway's association with survival, methods have been developed for the Cox proportional hazards model and the accelerated failure time model.[8;12] However, in studies of disease survival, it is common for patients to experience both non-terminal and terminal disease related events.

[1]Department of Operations Research and Financial Engineering, Princeton University, Princeton, NJ, USA
[2]Department of Biostatistics, Harvard University, Boston, MA, USA
[3]Department of Statistics, Ohio State University, Columbus, OH, USA

Corresponding author:
Jennifer A. Sinnott, Department of Statistics, 1958 Neil Ave, Columbus, OH 43210, USA
Email: jsinnott@stat.osu.edu

Prepared using sagej.cls

For example, in studies of survival following breast cancer diagnosis, patients may experience both recurrence of the disease (a non-terminal event) and death (a terminal event). One common approach for dealing with this kind of data is simply to ignore information on the non-terminal event and focus on overall survival (OS); this approach may be appealing because OS is the most clinically relevant and accurately measured endpoint.[14;15] A second common approach is to model so-called progression-free survival (PFS), in which the failure time is the time to either recurrence or death – whichever comes first.[16–18]. Using PFS has some benefits, especially in clinical trials. In clinical trials, focusing on PFS allows assessment of treatment efficacy sooner, and thus is more efficient when progression is a surrogate marker for death.[19] A drug's impact on PFS may also be a more accurate assessment of its value because patients' treatment plans are likely reassessed at the time of recurrence without regard to randomization.[16;17]

However, in a study relating tumor gene expression to survival, methods that use only PFS may lose substantial information by eliminating follow-up time contributed by patients after they experience recurrence. Moreover, it is plausible that a biological process hastens time to recurrence but not time to death, or vice versa; or that dysregulation in some subnetwork of a pathway brings about recurrence while dysregulation in another accelerates the time to death. Identifying such biological processes and pathways may be important for both prediction and biological understanding, but may be difficult if analysis focuses only on a composite endpoint combining recurrence and death. Indeed, our simulations demonstrate that methods based on PFS can have poor performance in exactly these situations.

To more fully capture disease progression integrating all available information on recurrence and death, numerous methods are emerging to address this so-called Semi-Competing Risks (SCR) setting. The main statistical challenge for SCR data is that the marginal survival function for the non-terminal event is not nonparametrically identifiable.[20] Specifically, let T^R be the time to the non-terminal event and T^D be the time to the terminal event (R and D here stand for "recurrence" and "death"). Then we can define the joint survival function $S(t^R, t^D) = P(T^R > t^R, T^D > t^D)$, but we can only learn about this function on the "wedge" region $0 < t^R < t^D$, because recurrence times after death are never observable. Because of this, the marginal survival function of T^R is not estimable without further modeling assumptions. Some methods to deal with SCR model the two marginal distributions of T^R and T^D, and either leave their joint distribution arbitrary or model it using a copula.[20–29] Other methods model the two cause-specific hazards for recurrence and death, as well as a third conditional hazard for the time to death following recurrence.[30–37] To our knowledge, previous methods for the SCR setting (as well as the competing risks [CR] setting) have focused on modeling linear effects of the covariates on the log-hazards, but have not been designed to effectively capture nonlinear effects of covariates on the log-hazards.

To test whether a pathway is associated with recurrence and/or death, we propose a novel approach, in which we separately test whether the pathway is associated with the cause-specific hazard for recurrence

and the marginal hazard for death. This approach is more appealing than the first set of approaches that model the marginal distributions of T^R and T^D because it only imposes modeling assumptions on observable quantities. It also provides more flexibility than the second set of approaches that model cause-specific and conditional hazards because it only imposes structure on two hazards – the cause-specific hazard of recurrence and the overall hazard of death. The two component pathway tests are KM-based tests, and thus allow flexible, data-driven detection of any nonlinear effects that the pathway genes may have on either the non-terminal or terminal events. In order to combine information from the two tests, we add the test statistics together, which is mathematically equivalent to making a working independence assumption about T^R and T^D. In actuality, the two event time are likely to be correlated; therefore, we estimate the null distribution of this combined test statistic using a perturbation resampling approach that preserves the correlation between the two event times, but does not explicitly specify the nature of the relationship.

The paper proceeds as follows. In section 2 we describe our KM SCR framework: in particular, in subsection 2.3 we develop a score test extending the work of Cai *et al.*[8] for the Cox KM model; full details of the derivation as well as the asymptotic behavior of the test are detailed in the Appendix. In subsection 2.4 we propose a perturbation resampling procedure to approximate the null distribution of the score test. In section 3, we present numerical studies, in which we compare our proposed approach to pathway tests for three alternative models commonly used in practice – one modeling PFS, one modeling OS alone, and one modeling the two event types as CR. Our proposed approach is illustrated in an application relating breast tumor gene expression to recurrence and death in section 4, and we demonstrate how our SCR model improves over the more common PFS model in real data. Some final comments are made in section 5. The approach is implemented in the R package `kernscr` available on CRAN at https://cran.r-project.org/web/packages/kernscr.

2 Methods

2.1 Problem Description and Notation

Let T^R be the time to the non-terminal event and T^D be the time to the terminal event. In our setting, T^R is the time to cancer recurrence, and T^D is the time to cancer-specific death. There are four possible observation patterns – we may observe:

1. T^R followed by T^D: patients who experience the non-terminal event followed by the terminal event
2. T^R only: patients who experience the non-terminal event but do not experience the terminal event during follow-up

3. T^D only: patients who experience the terminal event without the non-terminal event occurring
4. neither T^R nor T^D: patients who do not experience either kind of event during follow-up

We let C be the time of censoring for each subject, which is the time of last follow-up or the time of death due to another cause; C is observed for subjects with patterns 2 and 4. The goal of the study is to evaluate whether a set of genes (or SNPs, or other biomarkers) is associated with recurrence and/or death. The genes in the gene set will be denoted by $\mathbf{Z}_{p\times 1}$, a vector of length p. To have the best chance of identifying a gene set that is relevant to progression and can add value to existing clinical predictors, we may want to include information on known clinical variables U such as tumor grade and stage. We will assume that the variable U is discrete and is coded to take values $\{1, 2, \ldots, K\}$ – conceptually, these represent different risk strata, and will be used as strata in our model. For example, in breast cancer, the levels of U could correspond to different risk strata defined by combinations of factors in the Gail model.[38]

Because T^D may censor T^R, and C may censor both times, we define $X^D = T^D \wedge C$, $\Delta^D = \mathbf{I}(T^D \leq C)$, $X^R = T^R \wedge X^D$, and $\Delta^R = \mathbf{I}(T^R \leq X^D)$. Here $a \wedge b$ denotes the minimum of a and b, and $\mathbf{I}(A)$ is the indicator function taking the value 1 if A occurs and 0 otherwise. For convenience of notation, we also define $\mathbf{W} = (\mathbf{Z}^\intercal, U)^\intercal$. The observed data for n iid individuals is then:

$$\mathcal{O} = \left\{ \left(X_i^R, X_i^D, \Delta_i^R, \Delta_i^D, \mathbf{W}_i^\intercal \right) : i = 1, \ldots, n \right\}.$$

We assume that censoring is non-informative and that censoring is independent of the event times conditional on the covariates: $C \perp (T^R, T^D) \mid \mathbf{W}$. The time to death T^D may also censor T^R, but these two times are likely to be correlated, so that this censoring is informative. We will describe several approaches that could be used to model this data. These approaches will impose structure on either the *cause-specific hazard* of a certain event $E \in \{R, D\}$ – which we will denote by λ_{c} – or the *marginal hazard* of an event $E \in \{R, D\}$ – which we will denote by λ_{m}. That is, we will describe different approaches that model either the instantaneous rate of occurrence of event E given that the individual is still at risk of experiencing *both* R and D – the cause-specific hazard:

$$\lambda_{\mathrm{c}}^E(t; \mathbf{W}) = \lim_{\varepsilon \to 0} P(t \leq T^E < t + \varepsilon \mid T^R \geq t, T^D \geq t, \mathbf{W})/\varepsilon, E \in \{R, D\}$$

or the instantaneous rate of occurrence of event E given that the individual is still at risk of E, without regard to the other event type – the marginal hazard:

$$\lambda_{\mathrm{m}}^E(t; \mathbf{W}) = \lim_{\varepsilon \to 0} P(t \leq T^E < t + \varepsilon \mid T^E \geq t, \mathbf{W})/\varepsilon, E \in \{R, D\}$$

2.2 The Kernel Machine Semi-Competing Risks Model

Our proposed testing method will be abbreviated "SCR" to indicate Semi-Competing Risks. We model the *cause-specific hazard* for the non-terminal event and the *marginal hazard* for the terminal event:

$$\begin{aligned}\lambda^R_{\mathfrak{c}}(t;\mathbf{W}) &= \lambda^R_{\mathfrak{c}U}(t)\exp\{h_R(\mathbf{Z})\}\\ \lambda^D_{\mathfrak{m}}(t;\mathbf{W}) &= \lambda^D_{\mathfrak{m}U}(t)\exp\{h_D(\mathbf{Z})\}\end{aligned} \tag{1}$$

Here, the cause-specific hazard $\lambda^R_{\mathfrak{c}}(t;\mathbf{W})$ that a person with covariates $\mathbf{W}$ experiences recurrence is the product of an unspecified cause-specific baseline hazard $\lambda^R_{\mathfrak{c}U}(t)$ associated with the particular clinical risk stratum defined by U and the exponential of some function h_R of the genomic covariates. The marginal hazard of death for the same person is modeled similarly. The functions h_R and h_D are unknown, centered, smooth functions of the genomic covariates, and we will ultimately be interested in testing $H_0 : h_R(\cdot) = 0, h_D(\cdot) = 0$, within risk strata defined by U. This null suggests that none of the genes in the pathway is associated with either recurrence or death. Note that under the null, the model makes no proportional hazards assumptions, assuming only strata-specific unspecified hazard functions; we discuss this further in the Discussion (Section 5).

We propose to specify the functions h_R and h_D using the KM framework. Specifically, we assume that both h_R and h_D belong to $\mathcal{H}_K$, the Hilbert space generated by a chosen positive definite kernel $K(\cdot,\cdot;\rho)$. The kernel K is a measure of similarity between two vectors of genomic measurements, and may depend on a possibly unknown scaling parameter ρ. Different choices of kernel K will yield different collections of possible functions $h(\cdot)$. For example, specifying the *linear kernel* $K(\mathbf{z}_1,\mathbf{z}_2) = \mathbf{z}_1^\intercal\mathbf{z}_2$ forces $h(\mathbf{z}) = \beta^\intercal\mathbf{z}$, a linear function of the covariates. The *quadratic kernel* $K(\mathbf{z}_1,\mathbf{z}_2;\rho) = (\rho + \mathbf{z}_1^\intercal\mathbf{z}_2)^2$ yields a Hilbert space $\mathcal{H}_K$ spanned by basis functions $\{z_j, z_jz_{j'} : j, j' = 1,\ldots,p\}$, which incorporates main effects, quadratic effects, and 2-way interactions. To allow for more complex non-linear effects, one may consider the *Gaussian kernel*, defined by $K(\mathbf{z}_1,\mathbf{z}_2;\rho) = \exp\{-\|\mathbf{z}_1-\mathbf{z}_2\|^2/\rho\}$. The resulting function space $\mathcal{H}_K$ is generated by the radial basis functions. Many other kernels are available for use depending on the data available and the assumptions researchers wish to make: these include the Bessel kernel, the B-Spline kernel, and the sigmoid kernel. Adaptations such as kernel principal components analysis (PCA) can be implemented to reduce the complexity of the feature space associated with the kernel function and thus further improve the performance of KM methods in data sets of moderate sample size.[8;39]

When we do not assume the effects of the pathway variables $\mathbf{Z}$ on recurrence and death are identically 0, we may obtain estimators for each h_E (for $E \in \{R, D\}$) by maximizing the associated partial

likelihood functions each penalized by the norm of h_E in the Hilbert space, $\|h_E\|_{\mathcal{H}_K}$:

$$\log(L_P^E(h_E)) = \sum_{i=1}^{n}\sum_{k=1}^{K}\mathbf{I}(U_i=k)\int \log\left[\frac{\exp\{h_E(\mathbf{Z}_i)\}}{\sum_{l=1}^{n}\mathbf{I}(U_l=k)Y_l^E(s)\exp\{h_E(\mathbf{Z}_l)\}}\right] dN_i^E(s) - \frac{\mathfrak{d}_E}{2}\|h_E\|^2_{\mathcal{H}_K}.$$

where $N_i^E(t) = \Delta_i^E\mathbf{I}(X_i^E \le t)$ for both event types, but the at-risk indicators differ: $Y_i^R(t) = \mathbf{I}(X_i^R \ge t)\mathbf{I}(X_i^D \ge t)$ and $Y_i^D(t) = \mathbf{I}(X_i^D \ge t)$. The $\mathfrak{d}_E$ are penalty parameters that control the smoothness of the functions h_E. By the representer theorem,[40;41] the maximizer of $\log(L_P^E(h_E))$ for h_E must take a dual form, $\widehat{h}_E(\mathbf{z}) = \sum_{l=1}^{n}\alpha_{El}K(\mathbf{Z}_l,\mathbf{z};\rho)$, where the α_{El} are unknown parameters. Under this dual representation, we maximize (with a slight abuse of notation):

$$\log(L_P^E(\boldsymbol{\alpha}_E)) = \sum_{i=1}^{n}\sum_{k=1}^{K}\mathbf{I}(U_i=k)\int \log\left[\frac{\exp(\boldsymbol{\alpha}_E^\intercal\mathbf{K}_i(\rho))}{\sum_{l=1}^{n}\mathbf{I}(U_l=k)Y_l^E(s)\exp(\boldsymbol{\alpha}_E^\intercal\mathbf{K}_l(\rho))}\right] dN_i^E(s) - \frac{\mathfrak{d}_E}{2}\boldsymbol{\alpha}_E^\intercal\mathbb{K}(\rho)\boldsymbol{\alpha}_E \tag{2}$$

where $\mathbf{K}_i(\rho) = (\mathbb{K}_{i1}(\rho), \mathbb{K}_{i2}(\rho), \dots \mathbb{K}_{in}(\rho))^\intercal$ is the i^{th} row of the kernel matrix $\mathbb{K}(\rho)$. The penalized partial likelihood functions (2) may be viewed as arising from the random effects model in which we assume:

$$\begin{aligned}\lambda_{\mathfrak{c}i}^R(t) &= \lambda_{\mathfrak{c}U_i}^R(t)\exp\{\boldsymbol{\alpha}_R^\intercal\mathbb{K}_i(\rho)\}\\ \lambda_{\mathfrak{m}i}^D(t) &= \lambda_{\mathfrak{m}U_i}^D(t)\exp\{\boldsymbol{\alpha}_D^\intercal\mathbb{K}_i(\rho)\}\end{aligned} \tag{3}$$

for $\boldsymbol{\alpha}_E = \tau_E\boldsymbol{\varepsilon}_E$, where $\boldsymbol{\varepsilon}_E$ is multivariate normal with $E[\boldsymbol{\varepsilon}_E] = 0$ and $\mathrm{Var}(\boldsymbol{\varepsilon}_E) = \mathbb{K}(\rho)^-$, and $\mathfrak{d}_E^{-1} = \tau_E^2$. Here $\mathbb{K}(\rho)^-$ is the Moore-Penrose generalized inverse of $\mathbb{K}(\rho)$. Analogous connections between KM models and the mixed model framework have successfully been used to fit other KM regression models.[6–8;12] Conditional on the random effects $\boldsymbol{\varepsilon}_E$, the partial likelihoods may be written:

$$\log(L_{\boldsymbol{\varepsilon}_E}^E(\tau_E;\rho)) = \sum_{i=1}^{n}\sum_{k=1}^{K}\mathbf{I}(U_i=k)\int \log\left[\frac{\exp\{\tau_E\boldsymbol{\varepsilon}_E^\intercal\mathbf{K}_i(\rho)\}}{\widehat{U}_{Ek}^{(0)}(s;\rho,\tau_E)}\right] dN_i^E(s) \tag{4}$$

where we define $\widehat{U}_{Ek}^{(m)}(s;\rho,\tau_E) = \sum_{i=1}^{n}\mathbf{I}(U_i=k)Y_i^E(s)\exp\{\tau_E\boldsymbol{\varepsilon}_E^\intercal\mathbf{K}_i(\rho)\}\cdot\{\boldsymbol{\varepsilon}_E^\intercal\mathbf{K}_i(\rho)\}^m$ for $m = 0,1,2$.

2.3 Score test

With the model formulated as in (3), the null hypothesis becomes $H_0 : \tau_R = 0, \tau_D = 0$. Because these values are on the boundary of the parameter space, the likelihood ratio test has a nonstandard distribution; however, we can derive a score test similar to that derived in Cai et al.[8] for the Cox KM model with a

single event type of interest. To proceed in our setting, we will derive a test statistic under an assumption of working independence for T^R and T^D. This gives us a guide for how to combine information about the relationship between the covariates and each event time, but we would not expect these two times to be independent in practice. Thus, although the working independence assumption is used to justify the choice of the test statistic, both the asymptotic distribution we derive in the Appendix (Section 5) and the resampling approach used to estimate the null distribution in finite sample (Subsection 2.4) are valid when the two times are correlated, as we expect to occur in practice.

If we assume working independence for T^R and T^D; independence of the random effects ε_R and ε_D; and that we are testing H_0 against an alternative of the form $H_A : \tau_R = \frac{\tau_D}{\eta} = \tau$ for some known constant η; then the joint likelihood is:

$$\mathcal{M}(\tau;\rho) = \log\left\{L^R_{\boldsymbol{\varepsilon}_R}(\tau;\rho)\right\} + \log\left\{L^D_{\boldsymbol{\varepsilon}_D}(\eta\tau;\rho)\right\}.$$

Then, following Cai et al.[8] and Commenges and Andersen,[42] a score test may be based on the statistic:

$$\widehat{Q}(\rho,\eta) = E_{\varepsilon_R,\varepsilon_D}\left[\left\{\left.\frac{\partial\mathcal{M}(\tau;\rho)}{\partial\tau}\right|_{\tau=0}\right\}^2\right] + E_{\varepsilon_R,\varepsilon_D}\left[\left.\frac{\partial^2\mathcal{M}(\tau;\rho)}{\partial\tau^2}\right|_{\tau=0}\right].$$

Derivations are in the Appendix. In the case where $\eta = 1$ the form of the score statistic $\widehat{Q}(\rho) = \widehat{Q}(\rho,1)$ is:

$$\widehat{Q}(\rho) = \sum_{E\in\{R,D\}}\widehat{Q}_E(\rho) \quad \text{where } \widehat{Q}_E(\rho) = \widehat{\mathbf{M}}^E(\infty)^\intercal\mathbb{K}(\rho)\widehat{\mathbf{M}}^E(\infty) - n\widehat{q}_E(\rho) \tag{5}$$

where $\widehat{\mathbf{M}}^E(s) = (\widehat{M}^E_1(s), \widehat{M}^E_2(s), \ldots, \widehat{M}^E_n(s))^\intercal$ and each $\widehat{M}^E_i(s) = N^E_i(s) - \int_0^s Y^E_i(u)d\widehat{\Lambda}^E_{U_i}(u)$, $\widehat{\Lambda}^E_k(s) = \sum_{j=1}^n \mathbf{I}(U_j = k)\int_0^s \frac{dN^E_j(u)}{\widehat{S}^{(0)}_{Ek}(u)}$, $\widehat{S}^{(0)}_{Ek}(s) = \sum_{i=1}^n \mathbf{I}(U_i = k)Y^E_i(s)$, and $\widehat{q}_E(\rho) = \frac{1}{n}\sum_{k=1}^K \int\sum_{j=1}^n \mathbf{I}(U_j = k)Y^E_j(s)K_{jj}(\rho)d\widehat{\Lambda}^E_k(s) - \frac{1}{n}\sum_{k=1}^K\int\frac{\sum_{j=1}^n\sum_{l=1}^n \mathbf{I}(U_j=k)\mathbf{I}(U_l=k)Y^E_j(s)Y^E_l(s)K_{jl}(\rho)}{\widehat{S}^{(0)}_{Ek}(s)}d\widehat{\Lambda}^E_k(s)$.

2.4 A resampling procedure for approximating the null distribution

In the Appendix, we show that $n^{-1}\widehat{Q}(\rho) + \widehat{q}(t)$ is asymptotically equivalent to the process:

$$\widehat{\mathcal{W}}(\rho) = \sum_{E\in\{R,D\}} \left[\int\int K(\mathbf{z}_1,\mathbf{z}_2;\rho) d\widehat{\mathbb{W}}^E_M(\mathbf{z}_1) d\widehat{\mathbb{W}}^E_M(\mathbf{z}_2) + \right.$$
$$- 2\int \left\{ \int \sum_{k=1}^K p_k \omega^E_k(\mathbf{z},t,\rho) d\widehat{\mathbb{W}}^E_{\Lambda^E_k}(t) \right\} d\widehat{\mathbb{W}}^E_M(\mathbf{z})$$
$$\left. + \int\int \sum_{k=1}^K \sum_{k'=1}^K p_k p_{k'} \omega^E_{kk'}(t,s,\rho) d\widehat{\mathbb{W}}^E_{\Lambda^E_k}(t) d\widehat{\mathbb{W}}^E_{\Lambda^E_{k'}}(s) \right]$$

where $\widehat{\mathbb{W}}^E_M(\mathbf{z}) = n^{-\frac{1}{2}} \sum_{i=1}^n M^E_i(\infty)\mathbf{I}(\mathbf{Z}_i \le \mathbf{z})$, $M^E_i(t) = N^E_i(t) - \int_0^t Y^E_i(s) d\Lambda^E_{U_i}(s)$, $p_k = P(U = k)$, $\omega^E_k(\mathbf{z},t,\rho) = E[K(\mathbf{Z},\mathbf{z};\rho)Y^E(t) \mid U = k]$, $\omega^E_{kk'}(t,s,\rho) = E[K(\mathbf{Z}_1,\mathbf{Z}_2;\rho)Y^E_1(t)Y^E_2(s) \mid U_1 = k, U_2 = k']$. This form is very close to the form given in Cai et al.;[8] the main differences are that we are summing the test statistics over the two event types, and that the covariates U are modeled as strata. Also note that $\Lambda^R_k(t) = \Lambda^R_{\mathfrak{c}k}(t)$ and $\Lambda^D_k(t) = \Lambda^D_{\mathfrak{m}k}(t)$, but we drop the subscripts distinguishing the types of hazards being modeled for notational simplicity. As discussed in the Appendix, the arguments in Cai et al.[8] may be easily generalized to our setting, in order to establish that $\widehat{\mathcal{W}}(\rho)$ converges weakly to a zero-mean Gaussian process $\mathcal{W}(\rho)$.

In finite samples, we may approximate the distribution of $\widehat{Q}(\rho)$ by perturbation resampling. Specifically, suppose we generate a vector of iid standard normal variates $\mathcal{V} = (\mathcal{V}_1,\ldots,\mathcal{V}_n)^\intercal$, and calculate:

$$\widehat{\mathcal{W}}^*(\rho) = \sum_{E\in\{R,D\}} \left[\int\int K(\mathbf{z}_1,\mathbf{z}_2;\rho) d\widehat{\mathbb{W}}^{E*}_M(\mathbf{z}_1) d\widehat{\mathbb{W}}^{E*}_M(\mathbf{z}_2) + \right.$$
$$- 2\int \left\{ \int \sum_{k=1}^K \widehat{p}_k \widehat{\omega}^E_k(\mathbf{z},t,\rho) d\widehat{\mathbb{W}}^{E*}_{\Lambda^E_k} \right\} d\widehat{\mathbb{W}}^{E*}_M(\mathbf{z})$$
$$\left. + \int\int \sum_{k=1}^K \sum_{k'=1}^K \widehat{p}_k \widehat{p}_{k'} \widehat{\omega}^E_{kk'}(t,s,\rho) d\widehat{\mathbb{W}}^{E*}_{\Lambda^E_k}(t) d\widehat{\mathbb{W}}^{E*}_{\Lambda^E_{k'}}(s) \right]$$

where $\widehat{\mathbb{W}}^{E*}_M(\mathbf{z}) = n^{-\frac{1}{2}} \sum_{i=1}^n \widehat{M}^E_i(\infty)\mathbf{I}(\mathbf{Z}_i \le \mathbf{z})\mathcal{V}_i$, $\widehat{\mathbb{W}}^{E*}_{\Lambda^E_k}(t) = n^{-\frac{1}{2}} \sum_{i=1}^n \widehat{W}^E_{ki}(t)\mathcal{V}_i$, $\widehat{\omega}^E_k(\mathbf{z},t,\rho) = \frac{1}{n}\sum_{i=1}^n \mathbf{I}(U_i = k)K(\mathbf{Z}_i,\mathbf{z};\rho)Y^E_i(t)$, $\widehat{\omega}^E_{kk'}(t,s,\rho) = \frac{1}{n^2}\sum_{i=1}^n\sum_{j=1}^n \mathbf{I}(U_i = k)\mathbf{I}(U_j = k')K(\mathbf{z}_i,\mathbf{z}_j;\rho)Y^E_i(t)Y^E_j(s)$, $\widehat{p}_k = \frac{1}{n}\sum_{i=1}^n \mathbf{I}(U_i = k)$, and where $\widehat{W}^E_{ki}(t) = \int_0^t \frac{\mathbf{I}(U_i=k)d\widehat{M}^E_i(s)}{\widehat{S}^{(0)}_{Ek}(s)}$ is the empirical counterpart of $W^E_{ki}(t)$, defined in the Appendix. By assigning each individual a perturbation variate $\mathcal{V}_i$ and associating this variate with the individual in each process

$\widehat{\mathbb{W}}^{R*}_M(\mathbf{z}), \widehat{\mathbb{W}}^{R*}_{\Lambda^R_k}(t), \widehat{\mathbb{W}}^{D*}_M(\mathbf{z}), \widehat{\mathbb{W}}^{D*}_{\Lambda^R_k}(t)$, this resampling procedure preserves the correlation between the statistics associated with R and D. Moreover, conditional on the data, $\widehat{\mathcal{W}}^*(\rho)$ has the same limiting distribution as $\widehat{\mathcal{W}}(\rho)$. Thus, by obtaining B realizations of $\mathcal{V}$, and for each of these realizations $b = 1, \ldots, B$ calculating a perturbed statistic $\widehat{\mathcal{W}}^{*(b)}(\rho)$, we can approximate the null distribution of $\widehat{\mathcal{W}}(\rho)$ by $\vec{\mathcal{W}}^*(\rho) = \{\widehat{\mathcal{W}}^*_{(b)}(\rho), b = 1, \ldots, B\}$. Then we may approximate the null distribution of $n^{-1}\widehat{Q}(\rho)$ by $\widehat{\mathcal{Q}}^{*(b)}(\rho) = \widehat{\mathcal{W}}^*_{(b)}(\rho) - \bar{q}^*(\rho)$, where $\bar{q}^*(\rho)$ is the empirical mean of $\vec{\mathcal{W}}^*(\rho)$. For example, for any fixed ρ, an approximate p-value for the test may be calculated by $\frac{1}{B}\sum_{b=1}^{B} \mathbf{I}\left\{\widehat{\mathcal{Q}}^{*(b)}(\rho) > n^{-1}\widehat{Q}(\rho)\right\}$.

2.5 Kernels with tuning parameters

The linear kernel does not rely on a tuning parameter ρ, but many kernels, including the quadratic and Gaussian kernels used here, rely on a tuning parameter that controls the complexity of the associated functions h. Under H_0, the matrix $\mathbb{K}(\rho)$ disappears, so that ρ is not estimable; however, we may follow the approach in Cai et al.[8] of using the statistic:

$$\widehat{S} = \sup_{\rho \in [\rho_L, \rho_U]} \left\{ n^{-1}\widehat{Q}(\rho)/\widehat{\sigma}(\rho) \right\}$$

where $\widehat{\sigma}^2(\rho)$ is an estimate of the variance $\sigma^2(\rho)$ of $\mathcal{W}(\rho)$. To get such an estimate, we can calculate the variance of the perturbations: $\widehat{\sigma}^2(\rho) = \text{Var}\left\{\vec{\mathcal{W}}^*(\rho)\right\}$. We then may calculate an approximate p-value by calculating $\frac{1}{B}\sum_{b=1}^{B} \mathbf{I}(\widehat{S}^{*(b)} > \widehat{S})/B$, where $\widehat{S}^{*(b)} = \sup_{\rho \in [\rho_L, \rho_U]} \left\{ \widehat{\mathcal{Q}}^{*(b)}(\rho)/\widehat{\sigma}(\rho) \right\}, b = 1, \ldots, B$. For the Gaussian and quadratic kernels, we determine the range $[\rho_L, \rho_U]$ of ρ by requiring that the associated kernel matrices $\mathbb{K}(\rho)$ have eigenvalues $\widehat{\lambda}_i$ that decay at a polynomial rate $O(i^{-\alpha})$ for some range of $\alpha > 1$. This approach is motivated by work by Braun,[43] who bounds the error due to projecting the feature space onto the first r principal components using terms whose behavior depends on the decay properties of the eigenvalues. By experimentation we found that $\alpha \in [1.5, 4]$ yielded a reasonable range of feature space complexity. To implement this, for each ρ, we regressed the logarithms of the top 90% of eigenvalues on their indices and used the regression coefficient as an estimate of α; we considered values ρ whose associated α fell in the specified range.

2.6 Alternative Models

As discussed in the introduction, a popular modeling alternative for this setting is to focus on progression-free survival (PFS). To do this, let RD stand for recurrence-or-death, and define $T^{RD} = T^R \wedge T^D$, $X^{RD} = T^{RD} \wedge C$, $\Delta^{RD} = \mathbf{I}(T^{RD} \leq C)$. Then we model the (necessarily marginal) hazard

of recurrence-or-death by:

$$\lambda^{RD}_{\mathtt{m}}(t) = \lambda^{RD}_{\mathtt{m}U}(t) \exp\left\{h_{RD}(\mathbf{Z})\right\},$$

where $h_{RD}(\cdot) \in \mathcal{H}_K$ for some kernel K. The previous method developed in Cai et al.[8] applies directly.

Another popular modeling approach is to disregard information on metastases, and focus entirely on the marginal model for death – we will call this model OS for overall survival:

$$\lambda^{D}_{\mathtt{m}}(t; \mathbf{W}) = \lambda^{D}_{\mathtt{m}U}(t) \exp\left\{h_{D}(\mathbf{Z})\right\}.$$

Once again, the previous method developed in Cai et al.[8] applies directly.

Another alternative model which we will compare is the competing risks model (CR), in which we model both cause-specific hazards of the times to recurrence and death and disregard follow-up time after recurrence occurs:

$$\begin{aligned} \lambda^{R}_{\mathtt{c}}(t; \mathbf{W}) &= \lambda^{R}_{\mathtt{c}U}(t) \exp\left\{h_{R}(\mathbf{Z})\right\} \\ \lambda^{D}_{\mathtt{c}}(t; \mathbf{W}) &= \lambda^{D}_{\mathtt{c}U}(t) \exp\left\{h_{D}(\mathbf{Z})\right\} \end{aligned}$$

A score test for this model may be derived in the same manner as outlined above for the SCR model, under an assumption of working independence for the two event times. By comparing the power of this model to the power of the SCR model, we can quantify the added benefit seen by not just separating recurrence and death, but also by including information on death after recurrence.

3 Simulation study

We present the results of simulation studies comparing KM-based testing under our proposed SCR model and the three alternative models described in Subsection 2.6 – the PFS model, the CR model, and the OS model. We used three kernels – linear, quadratic, and Gaussian. All tests were run using 1000 perturbations with perturbation variables $\mathcal{V}$ generated from a standard normal distribution. The numbers presented are based on 2000 simulations for empirical size and 1000 simulations for empirical power.

For simplicity, we do not consider any risk stratifying clinical covariate U. For $p = 5$ and 10, we generate genomic covariates $\mathbf{Z}_{p\times 1}$ independently for each individual, where $\mathbf{Z}$ is multivariate normal with mean 0, variance 1, and correlation between each pair of genes given by $\rho_{\mathbf{Z}}$. We consider $\rho_{\mathbf{Z}} = 0.2$ to reflect low correlation between genes and $\rho_{\mathbf{Z}} = 0.5$ to reflect moderate correlation between genes. We consider studies of size $n = 100$ and $n = 200$. The times until the non-terminal event and the terminal

event are generated according to:

$$T^R = \exp\{C_R + h_R(\mathbf{Z}) + E_R\}$$
$$T^D = \exp\{C_D + h_D(\mathbf{Z}) + E_D\}.$$

Here, the error terms E_R and E_D marginally follow extreme value distributions but are generated to be correlated. Specifically, we let $\mathbf{Y} \sim \text{MVN}((0,0)^\intercal, \Sigma)$, where $\Sigma = \left(\begin{smallmatrix} 1 & 0.5 \\ 0.5 & 1 \end{smallmatrix}\right)$; then we transform $\mathbf{E} = \log\{-\log\{\Phi(\mathbf{Y})\}\}$, where Φ is the standard normal cdf, and let the first value be E_R and the second value be E_D. The functions $h_R(\mathbf{Z})$ and $h_D(\mathbf{Z})$ differ across simulation settings, and the constants C_R and C_D are selected to adjust the relative sizes of T^R, T^D, and C, which is the censoring time generated independently from an exponential distribution with mean $\frac{1}{30}$. Thus, correlation between the recurrence time T^R and the death time T^D is induced both through the error terms E_R and E_D, and through h_R and h_D if they are nonzero and involve common or correlated genes.

Table 1 displays the empirical sizes across different data simulation settings, for each of the four models and each of the three kernels. In this setting, approximately 31% recur and then die; 14% only recur during follow-up; 28% die (without recurrence) before the end of follow-up, and 27% are censored before they experience either type of event. In general, the sizes appear to be well-calibrated. The tests using the linear kernel have size near 5%, while the tests using the nonlinear kernels have levels below that. The tests using the Gaussian and quadratic kernels are particularly conservative when the between-gene correlation is low ($\rho_{\mathbf{Z}} = 0.2$) and $n = 100$. On average, the test levels improve with increasing sample size n.

In Table 2, we compare test performance when the true signal is linear. Specifically, we let the signal for recurrence be $h_R(\mathbf{Z}) = Z_1/4$, and the signal for death be $h_D(\mathbf{Z}) = Z_2/4 - Z_3/4 + Z_4/4$. In this setting, approximately 30% experience both recurrence and death; 14% experience recurrence but not death during follow-up; 28% experience death without recurrence; and 28% are censored. Here, not surprisingly, the linear kernel (L) typically performs the best for all the models. In this setting, we expect the SCR model to outperform the PFS model because the PFS model merges the two event types, but the genes that are associated with time to death and time to recurrence differ (though they are correlated). And indeed, under the linear kernel, the SCR model typically outperforms the others – for example, when $n = 200$, $p = 5$ and $\rho_{\mathbf{Z}} = 0.5$, and the linear kernel is used, the power to detect the pathway with the SCR-based test is 77.1%. In comparison, the PFS-based test has power of only 62.7%. The CR-based test has an even lower power of 60.8%, thus demonstrating a gain in power of 16.3% by the SCR test from including information after recurrence for the 44% of subjects who recur during follow-up. Finally, the OS test has power 67.8%, reflecting some decrease in power when information on the recurrence time is not taken into account – though interestingly, in this setting the OS-based test is frequently a strong

Table 1. Empirical sizes (in %) – i.e., the proportion of significant p-values when both $h_R(\cdot) \equiv 0$ and $h_D(\cdot) \equiv 0$ – of the four tests under consideration, at level $\alpha = 5\%$. The tests compared are: SCR (Semi-competing Risks); PFS (Progression-Free Survival); CR (Competing Risks); and OS (Overall Survival). For each of these model formulations, KM (kernel machine) tests are run with Gaussian (G), Linear (L), and Quadratic (Q) kernels. Comparisons are made for sample sizes $n = 100$ and $n = 200$; number of pathway variables $p = 5$ and $p = 10$; and pathway variable pairwise correlation $\rho_{\mathbf{Z}} = 0.2$ and 0.5. In this setting, approximately 31% recur and then die; 14% only recur during follow-up; 28% die (without recurrence) before the end of follow-up, and 27% are censored before they experience either type of event.

			$n = 100$				$n = 200$			
			SCR	PFS	CR	OS	SCR	PFS	CR	OS
$p = 5$	$\rho = 0.2$	G	2.2	2.1	1.6	2.9	3.7	3.8	3.5	4.8
		L	4.2	4.7	4.5	5.1	4.1	5.1	5.1	4.6
		Q	2.4	2.6	2.2	3.4	4.3	4.3	3.4	4.2
	$\rho = 0.5$	G	3.0	2.5	2.3	3.8	4.2	4.0	4.1	4.9
		L	5.4	5.4	5.1	5.8	5.0	5.2	4.8	4.7
		Q	3.6	3.6	3.0	4.2	4.9	4.8	4.2	5.3
$p = 10$	$\rho = 0.2$	G	1.0	1.6	0.5	2.1	2.0	3.0	1.7	3.0
		L	4.2	4.0	4.5	4.6	5.0	4.8	4.2	4.5
		Q	1.7	1.9	1.1	2.8	2.4	3.8	2.2	3.1
	$\rho = 0.5$	G	2.9	3.2	1.8	3.8	3.5	4.5	3.0	4.5
		L	5.3	5.1	4.7	4.9	4.8	4.3	4.9	4.8
		Q	3.4	3.6	2.5	4.0	4.0	5.6	3.9	4.5

performer, and is sometimes more powerful when nonlinear kernels are used, perhaps due to the strength of the signal $h_D(\mathbf{Z})$ and the involvement of multiple genes.

We also consider a nonlinear setting with $h_R(\mathbf{Z}) = 1.5\sin(Z_1 + Z_2^2), h_D(\mathbf{Z}) = (Z_4 - Z_5)^2/4$ (Table 3). Here, approximately 53% of subjects experience both recurrence and death; 20% experience recurrence without death; 18% die before recurring; and 9% are censored. Once more the SCR model typically outperforms the others, and the Gaussian kernel typically outperforms the linear and even the quadratic kernels. The OS model is particularly weak in this setting because it ignores the additional information provided by the 73% of subjects who experience recurrence. For example, when $n = 200$, $p = 5$, and $\rho_{\mathbf{Z}} = 0.5$, and the Gaussian kernel is used, the power to detect the pathway is 95.1% with the SCR model; 88.2% when the PFS model is used; 94.5% when the CR model is used; and only 9.2% when OS is used. When $n = 100$, $p = 10$, and $\rho_{\mathbf{Z}} = 0.2$, the Gaussian kernel has slightly lower power than the linear kernel. This is likely because of the difficulty in finding the best tuning parameter for the nonlinear kernel when there is not much data, the covariates are only weakly correlated, and the signal in the pathway is sparser. We see this phenomenon diminish when we either increase the sample size or increase the correlation among the genomic covariates.

Table 2. Power (in %) of detecting a linear signal, specifically letting $h_R(\mathbf{Z}) = \mathbf{Z}_1/4$ for recurrence and $h_D(\mathbf{Z}) = \mathbf{Z}_2/4 - \mathbf{Z}_3/4 + \mathbf{Z}_4/4$ for death. The tests are run at $\alpha = 5\%$, and the four compared are as before: SCR (Semi-competing Risks); PFS (Progression-Free Survival); CR (Competing Risks); and OS (Overall Survival). For each of these model formulations, KM (kernel machine) tests are run with Gaussian (G), Linear (L), and Quadratic (Q) kernels. Comparisons are made for sample sizes $n = 100$ and $n = 200$; number of pathway variables $p = 5$ and $p = 10$; and pathway variable pairwise correlation $\rho_{\mathbf{Z}} = 0.2$ and 0.5. In this setting, approximately 30% recur and then die; 14% only recur during follow-up; 28% die (without recurrence) before the end of follow-up, and 28% are censored before they experience either type of event.

			$n = 100$				$n = 200$			
			SCR	PFS	CR	OS	SCR	PFS	CR	OS
$p = 5$	$\rho = 0.2$	G	34.8	16.7	25.1	39.0	80.8	39.9	71.4	78.3
		L	48.1	26.1	41.5	48.6	89.1	49.5	80.3	84.5
		Q	39.4	20.3	25.9	43.2	84.5	45.9	73.3	80.7
	$\rho = 0.5$	G	27.2	24.9	17.7	29.7	65.6	48.6	51.9	61.8
		L	38.5	34.4	30.8	34.7	77.1	62.7	60.8	67.8
		Q	32.6	29.4	20.6	32.0	68.5	56.9	55.1	62.6
$p = 10$	$\rho = 0.2$	G	14.9	10.9	5.9	18.5	63.6	35.0	46.4	63.7
		L	33.9	16.6	25.3	34.0	70.8	40.3	58.7	68.9
		Q	19.6	12.6	10.7	21.2	64.7	37.0	48.5	63.2
	$\rho = 0.5$	G	25.0	24.1	14.0	25.2	61.4	55.0	46.5	53.6
		L	33.6	30.4	23.9	28.2	66.3	56.0	51.1	56.3
		Q	25.3	26.4	15.7	25.4	61.8	56.7	47.1	54.0

4 Application to Breast Cancer Gene Expression

Genomics has proven to be of great importance in breast cancer. Besides high risk mutations identified on *BRCA1* and *BRCA2* genes, numerous genomic biomarkers have been developed for clinical use in this disease. Such discoveries have translated into better prevention, better diagnoses and better treatment for patients. However, further understanding of the molecular pathways of pathogenesis in breast cancer remains critical for the development of better targeted therapies.[44] Here, following Cai *et al.*[8], we focus on 70 genetic pathways (with numbers of genes ranging from 2 to 235 [median=22]) from the Molecular Signatures Database[1] and their potential association with breast cancer survival. Cai *et al.* identified pathways associated with OS regardless of the development of metastases. Here, we compare using the PFS model, the most commonly chosen model for integrating the two types of events, to the proposed SCR model. We apply both the SCR and PFS models using a Gaussian kernel, with kernel PCA[39] including principal components accounting for 90% of the total variability, to assess the overall effect of each of these pathways on breast cancer progression and survival, using data from a study by van de Vijver *et al.*[45] Gene expression was measured for 260 patients with primary breast carcinomas from the Netherlands Cancer Institute. The median follow-up time was 8.8 years; 23% died

Table 3. Power (in %) of detecting a nonlinear signal, specifically letting $h_R(\mathbf{Z}) = 1.5\sin(\mathbf{Z}_1 + \mathbf{Z}_2^2)$ for recurrence and $h_D(\mathbf{Z}) = (\mathbf{Z}_4 - \mathbf{Z}_5)^2/4$ for death. The tests are run at $\alpha = 5\%$, and the four compared are as before: SCR (Semi-competing Risks); PFS (Progression-Free Survival); CR (Competing Risks); and OS (Overall Survival). For each of these model formulations, KM (kernel machine) tests are run with Gaussian (G), Linear (L), and Quadratic (Q) kernels. Comparisons are made for sample sizes $n = 100$ and $n = 200$; number of pathway variables $p = 5$ and $p = 10$; and pathway variable pairwise correlation $\rho_{\mathbf{Z}} = 0.2$ and 0.5. In this setting, approximately 53% recur and then die; 20% only recur during follow-up; 18% die (without recurrence) before the end of follow-up, and 9% are censored before they experience either type of event.

			$n = 100$				$n = 200$			
			SCR	PFS	CR	OS	SCR	PFS	CR	OS
$p = 5$	$\rho = 0.2$	G	64.4	49.8	55.3	8.3	98.3	91.0	95.2	24.3
		L	46.9	45.6	47.0	6.4	84.7	82.3	84.6	4.5
		Q	41.5	39.3	39.1	4.2	83.9	77.5	81.4	6.7
	$\rho = 0.5$	G	55.3	46.2	53.5	4.0	95.1	88.2	94.5	9.2
		L	34.1	33.2	35.0	6.4	65.2	62.2	68.7	5.0
		Q	29.1	29.0	25.8	3.9	65.6	61.1	62.9	7.2
$p = 10$	$\rho = 0.2$	G	23.4	25.0	21.3	2.5	70.1	66.7	68.2	3.9
		L	32.5	32.7	31.9	5.5	70.0	67.3	70.9	4.9
		Q	24.5	24.4	22.0	2.8	64.0	63.1	62.7	3.5
	$\rho = 0.5$	G	29.8	29.4	28.0	3.4	65.3	61.5	63.6	3.7
		L	25.5	25.8	27.2	6.1	52.4	48.0	52.7	4.3
		Q	22.2	23.5	21.5	3.9	50.5	49.0	50.1	4.0

after presenting metastases, 9% were censored with metastases, 2% died without metastases, and 66% were censored before experiencing either event. Because we are testing 70 pathways simultaneously, we use the multiple testing adjustment method using perturbation described in Cai *et al.*[8] In the SCR model, incidence of metastases was the non-terminal event, and death was the terminal event. PFS was defined as time to either death or metastases, whichever occurred first. Tests were run with 10,000 perturbations. Note that the number of perturbations determines that smallest size p-value that one can reliably estimate; with 10,000 perturbations, the smallest nonzero nominal p-value is 0.0001, and any more significant associations report a p-value of 0. The smallest nonzero adjusted p-value one can reliably estimate is approximately $0.0001 \times 70 = 0.007$. Smaller p-values could be estimated by running more perturbations.

Figure 1 displays the raw p-values and the adjusted p-values for the proposed score test from both models. As expected, the SCR model shows greater power. Indeed p-values from the SCR model are typically lower than those from the PFS model (51 out of 70 unadjusted; 58 out of 70 adjusted). With family-wise error rate of 0.05, 24 pathways are significant using the SCR model while 21 are significant using PFS. In this data set, the difference between these two methods is not dramatic, perhaps because

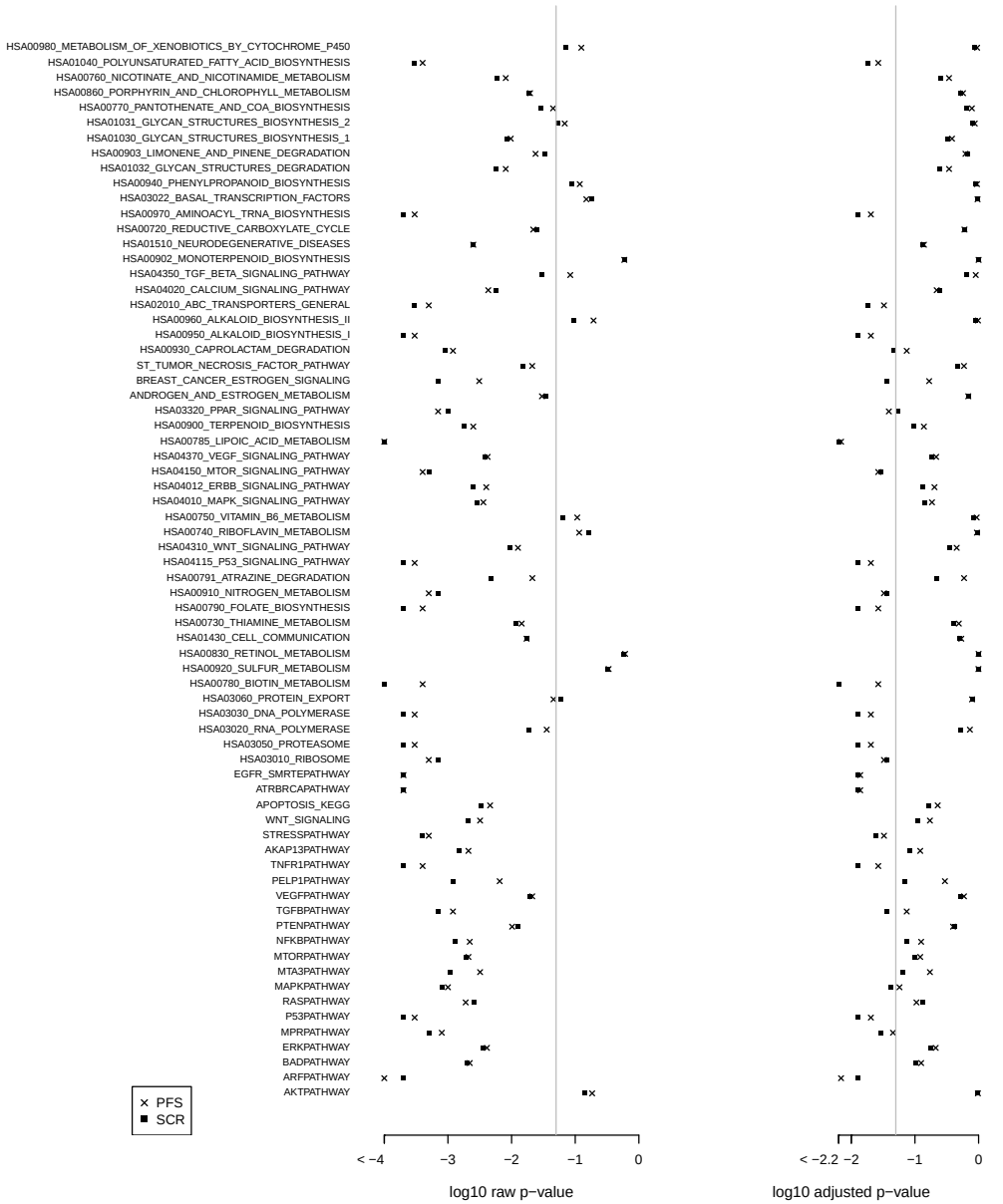

Figure 1. log10 p-values for testing the overall effect of 70 genetic pathways on breast cancer survival based on the KM (kernel machine) score test with Gaussian kernel using kernel PCA (principal components analysis; with a 90% of total variability threshold). The crosses and the squares represent PFS and SCR models, respectively. These results are based on B = 20,000 perturbations.

the total proportion of either type of event is not very large. We would expect the PFS- and SCR-based approaches to produce similar results when, for example, a few key genes in a pathway hasten both time to progression and time to death with similar effects. The SCR-based model will be better than the PFS-based model at detecting pathways in which progression and death are associated with different genes' dysregulation within the pathways. And indeed, in this example, the SCR model identifies three more pathways than the PFS model at the family-wise error rate of 0.05. These pathways include the *TGF*β pathway and the *MAPK* pathway, which have been implicated in breast cancer. For example, hyperactivation of the *MAPK* pathway is directly associated with the appearance of certain subtypes of breast cancer.[46] Similarly, the *TGF*β pathway has recently been identified as a good prognostic indicator for certain subtypes of breast cancer,[47] while its implication in breast cancer has been known and studied for more than two decades.[48;49] Thus, the increased power of the SCR KM model over the PFS KM model may directly translate into identifying more biologically meaningful associations.

5 Discussion

In this paper we present a flexible test procedure for assessing whether a pathway is associated with recurrence and/or death. The proposed approach is appealing because it tests marginally for the relationship between the pathway and the most clinically relevant and precisely measured outcome (disease-specific death), and augments the test by testing for a potential relationship between the pathway and the non-terminal event. Using information on both events is preferable even when overall survival is the most clinically relevant endpoint, because integrating information on recurrence improves our ability to identify important pathways if the recurrence endpoint is a meaningful landmark between diagnosis and death, and is especially helpful if the survival time after recurrence is variable and governed by patient response to a variety of post-recurrence treatments.[16] The proposed approach is implemented in the R package `kernscr` available on CRAN at https://cran.r-project.org/web/packages/kernscr.

As mentioned, one reason that PFS is sometimes preferred, especially in the clinical trial setting, is that it is not affected by changes in treatment that occur after a patient presents with a recurrence. A method such as ours that combines information from recurrence and eventual death may elide over this crucial source of heterogeneity. Although we cannot fully investigate this issue in the breast cancer data because we lack information about treatment, our proposed method is still informative in this setting for identifying potentially important pathways – we just need to be careful about the interpretation. Specifically, a pathway may be associated with death either independently of treatment or in a manner mediated by treatment, if the gene expression impacts either treatment selection or treatment response. Such a pathway is likely worth investigating further regardless. Furthermore, other work with nonlinear KM regression has demonstrated that nonlinear kernels can successfully detect signals that are based

on mixture models – for example, latent classes with differential covariate effects.[12] Thus, nonlinear KM methods may in fact be well suited to detecting pathways exhibiting heterogeneous patient effects – though more research is needed to investigate the impacts of different sources of patient heterogeneity on the efficacy of KM regression methods.

Some previous work in the SCR setting has focused on modeling marginal associations with the two event types and modeling their joint relationship using a copula model; by comparison, our approach only makes modeling assumptions on observable quantities, and the relationship between the non-terminal and terminal event is left unspecified. Extending KM methods to copula models warrants further research. Other previous work models cause-specific hazards for the two events as well as a conditional hazard for the terminal event given that the non-terminal event has happened; by comparison, our approach makes fewer modeling assumptions by modeling only the cause-specific hazard for the non-terminal event and the marginal hazard for the terminal event. It would, however, be interesting to extend KM methods to this model as well, particularly if further information, such as treatment changes made at time of recurrence, were available and could be integrated into the model. Previous work in the SCR setting has not been able to capture covariates' nonlinear effects on the two events. By using a flexible KM framework, our approach is able to capture nonlinear effects through the specification of a nonlinear kernel.

A potential issue with our approach is that it is unclear *a priori* how to combine the tests for the two events because the events are likely to be correlated and the correlation structure is unknown. To resolve this issue, we add the two test statistics together, and model their joint null distribution using a perturbation resampling method that accounts for their correlation without explicitly specifying it. Adding together the test statistics may not be the optimal way of combining information for the two event types, and future work could consider other possible approaches, such as taking the minimum p-value of the two tests.[11;12] Further extensions and adaptations of this model are possible. For example, rather than choosing a particular kernel, we could test several candidate kernels (such as a linear and a nonlinear kernel) and combine the information in an omnibus test using the perturbation resampling structure[12] – this could be useful in practical settings if researchers are unsure of what similarity metric may be most appropriate for their data and what type of model complexity they wish to consider.

As outlined in the text, our models currently assume that the non-genotypic covariates are discrete and identify "risk strata" which we use as stratifying variables in the proportional hazards models. We made this choice in order to better guarantee test validity in general situations; specifically, we wanted to avoid having the null distribution of the test statistic rely on the assumption that the cause-specific hazard of recurrence and the marginal hazard of death are both proportional in the clinical covariates. We feel that modeling clinical covariates as strata is sufficient in many applications, where combinations of covariates classify individuals into certain risk strata. However, if it is of interest to include non-genotypic covariates

as linear effects, and the two proportional hazards models specified under the null seem reasonable, then the proposed method is easily extendible to this case, with those covariates included in a similar way as in Cai et al. (2011)[8].

Our proposed joint test exhibits systematically higher power when compared with other testing approaches when different components of the pathway are associated with recurrence and death, as demonstrated by our simulation studies. In an analysis of a real gene expression study of breast cancer, it also identifies more pathways than the analogous test based only on PFS. The SCR KM method proposed here is a powerful tool that can allow disease researchers to better leverage the rich information in their data and to better identify biologically significant pathways to investigate more in depth in targeted studies.

Acknowledgements

BPH was supported by the National Institutes of Health (NIH) grant U54 HG007963. JAS was supported by the National Institutes of Health (NIH) grant T32 CA09001 and the A. David Mazzone Career Development Award. The authors express their deepest gratitude to Professor Tianxi Cai, Harvard University, for her help in preparing this manuscript.

References

1. Subramanian A, Tamayo P, Mootha VK et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proceedings of the National Academy of Sciences of the United States of America* 2005; 102(43): 15545–15550.
2. McLendon R, Friedman A, Bigner D et al. Comprehensive genomic characterization defines human glioblastoma genes and core pathways. *Nature* 2008; 455(7216): 1061–1068.
3. Jones S, Zhang X, Parsons DW et al. Core signaling pathways in human pancreatic cancers revealed by global genomic analyses. *Science* 2008; 321(5897): 1801–1806.
4. Goeman JJ, Van De Geer SA, De Kort F et al. A global test for groups of genes: testing association with a clinical outcome. *Bioinformatics* 2004; 20(1): 93–99.
5. Goeman JJ, Oosting J, Cleton-Jansen AM et al. Testing association of a pathway with survival using gene expression data. *Bioinformatics* 2005; 21(9): 1950–1957.
6. Liu D, Lin X and Ghosh D. Semiparametric regression of multidimensional genetic pathway data: Least-squares kernel machines and linear mixed models. *Biometrics* 2007; 63(4): 1079–1088.
7. Liu D, Ghosh D and Lin X. Estimation and testing for the effect of a genetic pathway on a disease outcome using logistic kernel machine regression via logistic mixed models. *BMC bioinformatics* 2008; 9(1): 292.

8. Cai T, Tonini G and Lin X. Kernel Machine Approach to Testing the Significance of Multiple Genetic Markers for Risk Prediction. *Biometrics* 2011; 67(3): 975–986.
9. Lin X, Cai T, Wu MC et al. Kernel machine SNP-set analysis for censored survival outcomes in genome-wide association studies. *Genetic Epidemiology* 2011; 35(7): 620–631.
10. Wu MC, Lee S, Cai T et al. Rare-variant association testing for sequencing data with the sequence kernel association test. *The American Journal of Human Genetics* 2011; 89(1): 82–93.
11. Wu MC, Maity A, Lee S et al. Kernel machine snp-set testing under multiple candidate kernels. *Genetic epidemiology* 2013; 37(3): 267–275.
12. Sinnott JA and Cai T. Omnibus risk assessment via accelerated failure time kernel machine modeling. *Biometrics* 2013; 69(4): 861–873.
13. Hejblum BP, Skinner J and Thiébaut R. Time-Course Gene Set Analysis for Longitudinal Gene Expression Data. *PLOS Computational Biology* 2015; 11(6): e1004310. DOI:10.1371/journal.pcbi.1004310.
14. Panageas KS, Ben-Porat L, Dickler MN et al. When you look matters: the effect of assessment schedule on progression-free survival. *Journal of the National Cancer Institute* 2007; 99(6): 428–432.
15. Burzykowski T, Buyse M, Piccart-Gebhart MJ et al. Evaluation of tumor response, disease control, progression-free survival, and time to progression as potential surrogate end points in metastatic breast cancer. *Journal of Clinical Oncology* 2008; 26(12): 1987–1992.
16. Broglio KR and Berry DA. Detecting an overall survival benefit that is derived from progression-free survival. *Journal of the National Cancer Institute* 2009; 101(23): 1642.
17. Buyse M, Sargent DJ and Saad ED. Survival is not a good outcome for randomized trials with effective subsequent therapies. *Journal of Clinical Oncology* 2011; 29(35): 4719–4720.
18. Amir E, Seruga B, Kwong R et al. Poor correlation between progression-free and overall survival in modern clinical trials: are composite endpoints the answer? *European Journal of Cancer* 2012; 48(3): 385–388.
19. Sherrill B, Kaye JA, Sandin R et al. Review of meta-analyses evaluating surrogate endpoints for overall survival in oncology. *OncoTargets and therapy* 2012; 5: 287.
20. Fine JP, Jiang H and Chappell R. On semi-competing risks data. *Biometrika* 2001; 88(4): 907–919.
21. Cook RJ and Lawless JF. Marginal analysis of recurrent events and a terminating event. *Statistics in medicine* 1997; 16(8): 911–924.
22. Ghosh D and Lin D. Nonparametric analysis of recurrent events and death. *Biometrics* 2000; 56(2): 554–562.
23. Ghosh D and Lin D. Marginal regression models for recurrent and terminal events. *Statistica Sinica* 2002; 12(3): 663–688.
24. Jiang H, Fine JP and Chappell R. Semiparametric analysis of survival data with left truncation and dependent right censoring. *Biometrics* 2005; 61(2): 567–575.
25. Ghosh D. Semiparametric inferences for association with semi-competing risks data. *Statistics in medicine* 2006; 25(12): 2059–2070.

26. Peng L and Fine JP. Regression modeling of semicompeting risks data. *Biometrics* 2007; 63(1): 96–108.
27. Lakhal L, Rivest LP and Abdous B. Estimating survival and association in a semicompeting risks model. *Biometrics* 2008; 64(1): 180–188.
28. Hsieh JJ, Wang W and Adam Ding A. Regression analysis based on semicompeting risks data. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 2008; 70(1): 3–20.
29. Fu H, Wang Y, Liu J et al. Joint modeling of progression-free survival and overall survival by a bayesian normal induced copula estimation model. *Statistics in medicine* 2013; 32(2): 240–254.
30. Liu L, Wolfe RA and Huang X. Shared frailty models for recurrent events and a terminal event. *Biometrics* 2004; 60(3): 747–756.
31. Ye Y, Kalbfleisch JD and Schaubel DE. Semiparametric analysis of correlated recurrent and terminal events. *Biometrics* 2007; 63(1): 78–87.
32. Zeng D and Lin D. Semiparametric transformation models with random effects for joint analysis of recurrent and terminal events. *Biometrics* 2009; 65(3): 746–752.
33. Xu J, Kalbfleisch JD and Tai B. Statistical analysis of illness–death processes and semicompeting risks data. *Biometrics* 2010; 66(3): 716–725.
34. Zeng D, Chen Q, Chen MH et al. Estimating treatment effects with treatment switching via semicompeting risks models: an application to a colorectal cancer study. *Biometrika* 2012; 99(1): 167–184.
35. Zhang Y, Chen MH, Ibrahim JG et al. Bayesian gamma frailty models for survival data with semi-competing risks and treatment switching. *Lifetime data analysis* 2014; 20(1): 76–105.
36. Hu C and Tsodikov A. Joint modeling approach for semicompeting risks data with missing nonterminal event status. *Lifetime data analysis* 2014; 20(4): 563–583.
37. Lee KH, Haneuse S, Schrag D et al. Bayesian semiparametric analysis of semicompeting risks data: investigating hospital readmission after a pancreatic cancer diagnosis. *Journal of the Royal Statistical Society: Series C (Applied Statistics)* 2015; 64(2): 253–273.
38. Gail MH, Brinton LA, Byar DP et al. Projecting individualized probabilities of developing breast cancer for white females who are being examined annually. *Journal of the National Cancer Institute* 1989; 81(24): 1879–1886.
39. Schölkopf B, Smola A and Müller KR. Nonlinear component analysis as a kernel eigenvalue problem. *Neural computation* 1998; 10(5): 1299–1319.
40. Kimeldorf G and Wahba G. A correspondence between bayesian estimation on stochastic processes and smoothing by splines. *The Annals of Mathematical Statistics* 1970; 41(2): 495–502.
41. Scholkopf B and Smola A. *Learning with kernels*. MIT Press Cambridge, Mass, 2002.
42. Commenges D and Andersen PK. Score Test of Homogeneity for Survival Data. *Lifetime data analysis* 1995; 1: 145–156.

43. Braun M. *Spectral properties of the kernel matrix and their application to kernel methods in machine learning*. PhD Thesis, PhD thesis, University of Bonn, 2005.
44. Olopade OI, Grushko Ta, Nanda R et al. Advances in breast cancer: Pathways to personalized medicine. *Clinical Cancer Research* 2008; 14(24): 7988–7999.
45. van de Vijver MJ, He YD, van't Veer LJ et al. A gene-expression signature as a predictor of survival in breast cancer. *The New England journal of medicine* 2002; 347(25): 1999–2009.
46. Oh AS, Lorant LA, Holloway JN et al. Hyperactivation of MAPK induces loss of ERalpha expression in breast cancer cells. *Molecular endocrinology (Baltimore, Md)* 2001; 15(8): 1344–1359.
47. Teschendorff AE, Gomez S, Arenas A et al. Improved prognostic classification of breast cancer defined by antagonistic activation patterns of immune response pathway modules. *BMC cancer* 2010; 10(1): 604. DOI: 10.1186/1471-2407-10-604.
48. Buck MB and Knabbe C. TGF-Beta Signaling in Breast Cancer. *Annals of the New York Academy of Sciences* 2006; 1089(1): 119–126. DOI:10.1196/annals.1386.024.
49. Barcellos-Hoff M and Akhurst RJ. Transforming growth factor-β in breast cancer: too much, too late. *Breast Cancer Research* 2009; 11(1): 202. DOI:10.1186/bcr2224.
50. Fleming T and Harrington D. *Counting Processes and Survival Analysis*, volume 169. John Wiley & Sons, 1991.

Supplemental material

Derivation of the Score Test

Using the notation in section 2.3, we have our objective function $\mathcal{M}(\tau;\rho) = \log(L^R_{\boldsymbol{\varepsilon}_R}(\tau;\rho)) + \log(L^D_{\boldsymbol{\varepsilon}_D}(\eta\tau;\rho))$ and we need to calculate $\frac{\partial \mathcal{M}(\tau;\rho)}{\partial \tau}\Big|_{\tau=0}$ and $\frac{\partial^2 \mathcal{M}(\tau;\rho)}{\partial \tau^2}\Big|_{\tau=0}$. First, for $E \in \{R, D\}$

$$\begin{aligned}
\frac{\partial}{\partial \tau} \log L^E_{\boldsymbol{\varepsilon}_E}(\tau;\rho)\Big|_{\tau=0} &= \sum_{i=1}^{n}\sum_{k=1}^{K} \mathbf{I}(U_i = k) \int \left\{ K_i(\rho)^\intercal \varepsilon_E - \frac{\widehat{S}^{(1)}_{Ek}(s;\rho)}{\widehat{S}^{(0)}_{Ek}(s)} \right\} dN^E_i(s) \\
&= \sum_{i=1}^{n} \left[\sum_{k=1}^{K} \mathbf{I}(U_i = k) \int \left\{ dN^E_i(s) - Y^E_i(s) d\widehat{\Lambda}^E_k(s) \right\} \right] K_i(\rho)^\intercal \varepsilon_E \\
&= \sum_{i=1}^{n} \widehat{M}^E_i(\infty) K_i(\rho)^\intercal \varepsilon_E
\end{aligned}$$

where $\widehat{S}^{(m)}_{Ek}(s;\rho) = \widehat{U}^{(m)}_{Ek}(s;\rho,0) = \sum_{i=1}^n \mathbf{I}(U_i = k) Y_i^E(s)\{K_i(\rho)^\intercal \boldsymbol{\varepsilon}_E\}^m$ – though we may write $\widehat{S}^{(m)}_{Ek}(s;\rho) = \widehat{S}^{(m)}_{Ek}(s)$ since it does not depend on ρ; $\widehat{\Lambda}^E_k(s) = \sum_{j=1}^n \mathbf{I}(U_j = k) \int_0^s \frac{dN_j^E(u)}{\widehat{S}^{(0)}_{Ek}(u)}$; and

$$\widehat{M}_i^E(s) = \sum_{k=1}^K \mathbf{I}(U_i = k) \int_0^s \left\{ dN_i^E(u) - Y_i^E(u) d\widehat{\Lambda}^E_k(u) \right\} = N_i^E(s) - \int_0^s Y_i^E(u) d\widehat{\Lambda}^E_{U_i}(u).$$

Next, we may calculate:

$$E_{\varepsilon_E}\left(\left[\frac{\partial \log\{L^E_{\boldsymbol{\varepsilon}_E}(\tau;\rho)\}}{\partial \tau}\bigg|_{\tau=0} \right]^2 \right) = \sum_{i=1}^n \sum_{j=1}^n \widehat{M}_i^E(\infty)\widehat{M}_j^E(\infty) K_{ij}(\rho) = \widehat{\mathbf{M}}^E(\infty)^\intercal \mathbb{K}(\rho) \widehat{\mathbf{M}}^E(\infty)$$

since $E_{\varepsilon_E}[\varepsilon_E \varepsilon_E^\intercal] = \mathbb{K}(\rho)^-$ and $K_i(\rho)^\intercal \mathbb{K}(\rho)^- K_j(\rho) = K_{ij}(\rho)$. Thus, because we assume ε_R and ε_D are independent and both have mean 0,

$$E_{\varepsilon_R,\varepsilon_D}\left[\left\{ \frac{\partial \mathcal{M}(\tau;\rho)}{\partial \tau}\bigg|_{\tau=0} \right\}^2 \right] = \widehat{\mathbf{M}}^R(\infty)^\intercal \mathbb{K}(\rho) \widehat{\mathbf{M}}^R(\infty) + \eta^2 \widehat{\mathbf{M}}^D(\infty)^\intercal \mathbb{K}(\rho) \widehat{\mathbf{M}}^D(\infty)$$

Next, for $E \in \{R, D\}$

$$\begin{aligned}
\frac{\partial^2}{\partial \tau^2} L^E_{\varepsilon_E}(\tau,\rho)\bigg|_{\tau=0} &= \sum_{i=1}^n \sum_{k=1}^K \mathbf{I}(U_i = k) \int \left\{ \frac{\widehat{S}^{(1)}_{Ek}(s;\rho)^2}{\widehat{S}^{(0)}_{Ek}(s)^2} - \frac{\widehat{S}^{(2)}_{Ek}(s;\rho)}{\widehat{S}^{(0)}_{Ek}(s)} \right\} dN_i^E(s) \\
&\quad - \sum_{i=1}^n \sum_{k=1}^K \mathbf{I}(U_i = k) \int \frac{\sum_{j=1}^n \mathbf{I}(U_j = k) Y_j^E(s) \left\{ \mathbf{K}_j(\rho)^\intercal \varepsilon_E \right\}^2}{\widehat{S}^{(0)}_{Ek}(s)} dN_i^E(s) \\
&= \sum_{k=1}^K \int \frac{\sum_{j=1}^n \sum_{l=1}^n \mathbf{I}(U_j = k)\mathbf{I}(U_l = k) Y_j^E(s) Y_l^E(s) \left\{ \mathbf{K}_j(\rho)^\intercal \varepsilon_E \varepsilon_E^\intercal \mathbf{K}_l(\rho) \right\}}{\widehat{S}^{(0)}_{Ek}(s)} d\widehat{\Lambda}^E_k(s) \\
&\quad - \sum_{k=1}^K \int \sum_{j=1}^n \mathbf{I}(U_j = k) Y_j^E(s) \left\{ \mathbf{K}_j(\rho)^\intercal \varepsilon_E \varepsilon_E^\intercal \mathbf{K}_j(\rho) \right\} d\widehat{\Lambda}^E_k(s)
\end{aligned}$$

So, if we define

$$\begin{aligned}\widehat{q}_E(\rho) =& -\frac{1}{n}E_{\varepsilon_E}\left[\left.\frac{\partial^2 \log(L^E_{\boldsymbol{\varepsilon}_E}(\tau;\rho))}{\partial\tau^2}\right|_{\tau=0}\right]\\ =&\frac{1}{n}\sum_{k=1}^{K}\int\sum_{j=1}^{n}\mathbf{I}(U_j=k)Y_j^E(s)K_{jj}(\rho)d\widehat{\Lambda}_k^E(s)\\ &-\frac{1}{n}\sum_{k=1}^{K}\int\frac{\sum_{j=1}^{n}\sum_{l=1}^{n}\mathbf{I}(U_j=k)\mathbf{I}(U_l=k)Y_j^E(s)Y_l^E(s)K_{jl}(\rho)}{\widehat{S}^{(0)}_{Ek}(s)}d\widehat{\Lambda}_k^E(s)\end{aligned}$$

then

$$E_{\varepsilon_R,\varepsilon_D}\left[\left.\frac{\partial^2\mathcal{M}(\tau;\rho)}{\partial\tau^2}\right|_{\tau=0}\right] = -n\widehat{q}_R(\rho) - n\eta^2\widehat{q}_D(\rho)$$

Then the final expression for the score test statistic is:

$$\widehat{Q}(\rho;\eta) = \widehat{\mathbf{M}}^R(\infty)^\intercal\mathbb{K}(\rho)\widehat{\mathbf{M}}^R(\infty) - n\widehat{q}_R(\rho) + \eta^2\left\{\widehat{\mathbf{M}}^D(\infty)^\intercal\mathbb{K}(\rho)\widehat{\mathbf{M}}^D(\infty) - n\widehat{q}_D(\rho)\right\}.$$

Asymptotic Distribution of the Test Statistic

Our asymptotic results are outlined below, but follow from those presented in Cai et al.[8]. Our test statistic is the sum of two statistics of the form presented in Cai et al.[8]; one arising from T^R and one arising from T^D. The main point to jusitfy is that $\left\{\widehat{\mathbb{W}}^R_M(\mathbf{z}),\widehat{\mathbb{W}}^D_M(\mathbf{z}),\widehat{\mathbb{W}}^R_{\Lambda^E_1}(t),\ldots,\widehat{\mathbb{W}}^R_{\Lambda^E_K}(t),\widehat{\mathbb{W}}^D_{\Lambda^E_1}(t),\ldots,\widehat{\mathbb{W}}^D_{\Lambda^E_K}(t)\right\}$ converge jointly to zero-mean Gaussian processes $\left\{\mathbb{W}^R_M(\mathbf{z}),\mathbb{W}^D_M(\mathbf{z}),\mathbb{W}^R_{\Lambda^E_1}(t),\ldots,\mathbb{W}^R_{\Lambda^E_K}(t),\mathbb{W}^D_{\Lambda^E_1}(t),\ldots,\mathbb{W}^D_{\Lambda^E_K}(t)\right\}$. This holds because the functions $\mathbf{I}(\mathbf{Z}\leq\mathbf{z})$, $M_i^R(t)$, and $M_i^D(t)$ have finite pseudodimensions. Because of this joint convergence, the asymptotic results from Cai et al.[8] generalize to our setting.

To write out the details more explicitly, we assume that the kernel tuning parameter ρ is in some interval $[\rho_L,\rho_U]$ and that the genomic variables $\mathbf{Z}$ have bounded support. We assume that X_R and X_D have a finite support $[0,\tau]$, that the density of C given covariates $\mathbf{Z}$ is continuous and bounded, and that the marginal density of C is bounded away from 0 on $[0,\tau]$. We assume that the kernel K is continuously differentiable with respect to all its arguments. To find the asymptotic distribution of $n^{-1}\widehat{Q}(\rho)$, we first note by Fleming and Harrington[50],

$$\sup_{t\in[0,\tau]}\left\{\left|\widehat{\Lambda}_k^R(t)-\widehat{\Lambda}_k^R(t)\right|+\left|\widehat{\Lambda}_k^D(t)-\widehat{\Lambda}_k^D(t)\right|\right\} = o_P(1)$$

and that

$$\widehat{\mathbb{W}}^E_{\Lambda^E_k}(t) = n^{\frac{1}{2}}\left\{\widehat{\Lambda}^E_k(t) - \Lambda^E_k(t)\right\} = n^{-\frac{1}{2}}\sum_{i=1}^{n} W^E_{ki}(t) + O_P(n^{-\frac{1}{2}}),$$

where $W^E_{ki}(t) = \int_0^t \frac{\mathbf{I}(U_i=k)dM^E_i(s)}{\mathcal{G}_{Ek}(s)}$, for $\mathcal{G}_{Ek}(s) = E\left[Y^E(t) \mid U = k\right]$. We may see that $\widehat{M}^E_i(s) = M^E_i(s) - n^{-\frac{1}{2}}\int Y^E_i(t)d\widehat{\mathbb{W}}^E_{\Lambda^E_{U_i}(t)} + O_p(n^{-1})$, so that by mimicking the arguments in Cai et al.[8], $n^{-1}\widehat{Q}(\rho) + q(\rho)$ is asymptotically equivalent to the process:

$$\begin{aligned}\widehat{\mathcal{W}}(\rho) = \sum_{E\in\{R,D\}}\Bigg[&\int\int K(\mathbf{z}_1,\mathbf{z}_2;\rho)d\widehat{\mathbb{W}}^E_M(\mathbf{z}_1)d\widehat{\mathbb{W}}^E_M(\mathbf{z}_2) + \\ &-2\int\left\{\int\sum_{k=1}^{K} p_k\omega^E_k(\mathbf{z},t,\rho)d\widehat{\mathbb{W}}^E_{\Lambda^E_k}(t)\right\}d\widehat{\mathbb{W}}^E_M(\mathbf{z}) \\ &+\int\int\sum_{k=1}^{K}\sum_{k'=1}^{K} p_k p_{k'}\omega^E_{kk'}(t,s,\rho)d\widehat{\mathbb{W}}^E_{\Lambda^E_k}(t)d\widehat{\mathbb{W}}^E_{\Lambda^E_{k'}}(s)\Bigg]\end{aligned}$$

which is a zero-mean stochastic process, where:

$$\begin{aligned}\widehat{\mathbb{W}}^E_M(\mathbf{z}) =& n^{-\frac{1}{2}}\sum_{i=1}^{n} M^E_i(\infty)\mathbf{I}(\mathbf{Z}_i \le \mathbf{z}) \\ M^E_i(t) =& N^E_i(t) - \int_0^t Y^E_i(s)d\Lambda^E_{U_i}(s) \\ \omega^E_k(\mathbf{z},t,\rho) =& E\left[K(\mathbf{Z},\mathbf{z};\rho)Y^E(t) \mid U = k\right] \\ \omega^E_{kk'}(t,s,\rho) =& E\left[K(\mathbf{Z}_1,\mathbf{Z}_2;\rho)Y^E_1(t)Y^E_2(s) \mid U_1 = k, U_2 = k'\right] \\ p_k =& P(U = k) \\ q(\rho) =& \sum_{E\in\{R,D\}}\sum_{k=1}^{K} p_k\Bigg[\int E[K(\mathbf{Z},\mathbf{Z},\rho)Y^E(t)|U=k] \\ &-\omega^E_{kk}(t,t,\rho)/E[Y^E(t)|U=k]d\Lambda^E_k(t)\Bigg]\end{aligned}$$

Finally, because as we noted $\left\{\widehat{\mathbb{W}}^R_M(\mathbf{z}), \widehat{\mathbb{W}}^D_M(\mathbf{z}), \widehat{\mathbb{W}}^R_{\Lambda^E_1}(t), \ldots, \widehat{\mathbb{W}}^R_{\Lambda^E_K}(t), \widehat{\mathbb{W}}^D_{\Lambda^E_1}(t), \ldots, \widehat{\mathbb{W}}^D_{\Lambda^E_K}(t)\right\}$ converge jointly to zero-mean Gaussian processes $\left\{\mathbb{W}^R_M(\mathbf{z}), \mathbb{W}^D_M(\mathbf{z}), \mathbb{W}^R_{\Lambda^E_1}(t), \ldots, \mathbb{W}^R_{\Lambda^E_K}(t), \mathbb{W}^D_{\Lambda^E_1}(t), \ldots, \mathbb{W}^D_{\Lambda^E_K}(t)\right\}$, it follows

that $\widehat{\mathcal{W}}(\rho)$ converges weakly to the process:

$$\begin{aligned}\mathcal{W}(\rho) = \sum_{E\in\{R,D\}} \Bigg[\int\int K(\mathbf{z}_1,\mathbf{z}_2;\rho) d\mathbb{W}_M^E(\mathbf{z}_1) d\mathbb{W}_M^E(\mathbf{z}_2) + \\ - 2\int \left\{ \int \sum_{k=1}^{K} p_k \omega_k^E(\mathbf{z},t,\rho) d\mathbb{W}_{\Lambda_k^E}^E(t) \right\} d\mathbb{W}_M^E(\mathbf{z}) \\ + \int\int \sum_{k=1}^{K}\sum_{k'=1}^{K} p_k p_{k'} \omega_{kk'}^E(t,s,\rho) d\mathbb{W}_{\Lambda_k^E}^E(t) d\mathbb{W}_{\Lambda_{k'}^E}^E(s) \Bigg] .\end{aligned}$$