<table>
<thead>
<tr>
<th>Statistica Sinica Preprint No: SS-2021-0087</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Title</strong></td>
</tr>
<tr>
<td><strong>Manuscript ID</strong></td>
</tr>
<tr>
<td><strong>URL</strong></td>
</tr>
<tr>
<td><strong>DOI</strong></td>
</tr>
<tr>
<td><strong>Complete List of Authors</strong></td>
</tr>
<tr>
<td><strong>Corresponding Author</strong></td>
</tr>
<tr>
<td><strong>E-mail</strong></td>
</tr>
</tbody>
</table>

Notice: Accepted version subject to English editing.
Principal Components Analysis for Right Censored Data

Benjamin W. Langworthy\(^{1}\star\), Jianwen Cai\(^{1}\), Robert W. Corty\(^{2}\),

Michael R. Kosorok\(^{1}\), and Jason P. Fine\(^{1}\)

\(^{1}\)Department of Biostatistics, University of North Carolina Chapel at Hill

135 Dauer Drive, 3101 McGavran Greenberg Hall, Chapel Hill, NC 27599

\(^{2}\)School of Medicine, University of North Carolina at Chapel Hill

\*Corresponding Author

Abstract: Summary

Principal Components Analysis (PCA) is a common dimension reduction tool that transforms a set of variables into a linearly uncorrelated set of variables. Standard estimators of PCA involve either the eigendecomposition of the estimated covariance matrix or singular value decomposition of the centered data. For right censored failure time data it is not straightforward to estimate the principal components in this way because not all failure times are observed. Standard estimators for the covariance or correlation matrix should not be used in this case as they would require strong assumptions on the form of the joint distribution as well as the marginal distributions beyond the final observation time.

We present a novel, non-parametric estimator for the covariance of multivariate right censored failure time data. This estimator is based on the counting processes and corresponding martingales defined by the failure times. We prove that these estimators are consistent and converge to a gaussian process when properly standardized. We show that these covariance estimates can be used to estimate PCA for the martingales and counting processes for the different failure times. The corresponding estimates of the principal directions are consistent and asymptotically normal. We apply this method to data from a clinical trial of patients with pancreatic cancer and recover a medically-valid low-dimensional representation of adverse events.
1. Introduction

Principal Components Analysis (PCA) first introduced by Pearson (1901), transforms a set of potentially linearly correlated variables into a set of linearly uncorrelated variables called the principal components. All the transformations are linear combinations of the data, and the first principal component is defined to be the linear combination of the data that explains the largest possible variance within the data. All subsequent directions are the linear combinations of the data with maximal variance, subject to the constraint that they are orthogonal to all previous principal components. PCA is frequently used for dimension reduction and removing collinearity within a set of variables. PCA is generally estimated using eigendecomposition of the estimated covariance matrix or singular value decomposition (SVD) of the mean centered data.

PCA is used across many different settings, but to our knowledge one area where PCA has not been used is multivariate time to event data in the presence of right censoring. Such a set up can occur in longitudinal studies where patients are followed for many different event types. We will consider data from a clinical trial for patients with pancreatic cancer, but the techniques can be used for any number of longitudinal studies where individuals are followed over time and data on many different event types are collected. As we will show in such set ups, if there is right censoring for the different event types either due to loss to follow up or administrative end of study, standard PCA estimation techniques cannot be used for the time to event data. To show why this is we consider the multivariate survival setting with \( p \) different event times. For each subject we define \( T^{(j)} \) to be the failure time for
the \( jth \) event time. If we assume that each subject also has an independent censoring time, \( C \), the observed data for the \( jth \) event type consists of \( Y^{(j)} = T^{(j)} \wedge C \) as well as an indicator for whether the observed time is a censoring time or failure time. The full covariance matrix for \( T = [T^{(1)}, \ldots, T^{(p)}] \) cannot be estimated without strong parametric assumptions. This means that PCA cannot be estimated for \( T \). In order to overcome this issue we propose two versions of PCA for multivariate survival data and show how both can be estimated. One version uses the counting processes for each event type, defined as \( N^{(j)}(t) = I(T^{(j)} \leq t) \). The other version uses the martingales based on the decomposition of these counting processes. Even though these counting processes are not observed for all time points if patients are censored, we show that the covariance between the counting processes and martingales for two different event types can both be estimated non-parametrically. This allows for the principal component directions for the counting processes and martingales to be estimated consistently even in the presence of independent censoring. The corresponding component scores can be estimated for each subject up until the time they are censored. In order to make these methods more flexible we also allow for semi-competing risks, and show how to extend both survival PCA methods to this setting.

The estimation of the covariance between counting processes and martingales uses existing methods for estimation of bivariate and univariate survival functions, and univariate hazard functions. [Prentice and Cai (1992)](prentice1992) shows that at a fixed timepoint the covariance for the counting process martingales for two different event types can be written as a function of bivariate and univariate survival functions, as well as univariate hazard functions. Similarly, the covariance between the counting processes for two different event types can be written as a function of the bivariate and univariate survival functions. In this paper we use the
Kaplan-Meier and Nelson-Aalen estimators for the univariate survival and cumulative hazard functions. Estimators for the bivariate survival function include those by Dabrowska et al. (1988), Prentice and Cai (1992), van der Laan (1993), and Lin and Ying (1993). If there are more than two event types estimates of the full covariance and correlation matrices for the martingales or counting processes are found by estimating all the elements individually using bivariate methods. This allows for estimation of principal component directions for the martingales and counting processes using eigendecomposition of their estimated covariance or correlation matrices.

In the presence of competing risks we use previous methods for estimation of bivariate and univariate cause specific hazard functions and cause specific cumulative incidence functions (CIFs). Prentice et al. (1978) gives an overview of cause specific hazards and incidence functions in the univariate setting, and Kalbfleisch and Prentice (2011) gives details on estimating univariate cause specific CIFs. Details on estimating bivariate cause specific hazard functions and CIFs can be found in Cheng et al. (2007). In order to extend our methods to the competing risk setting we use the cause specific counting processes and the martingales based on their decomposition. We are able to show that the covariance between the cause specific martingales or counting processes for two different events can be written as a function of the bivariate and univariate cause specific CIFs as well as the univariate cause specific hazard functions. Using these results we show that the full covariance matrix for cause specific martingales and counting processes can be estimated in the presence of competing risks, which makes it possible to estimate the corresponding principal component directions using eigendecomposition.

Being able to estimate covariance and principal component directions for martingales
and counting processes in the presence of competing risks is of particular interest in sick populations where death acts as a competing risk for many different adverse events. We present one such example using data from the metastatic pancreatic adenocarcinoma clinical trial (MPACT) study for patients with pancreatic cancer. In this study patients are followed and have many different types of adverse events due to both cancer and treatment. Using PCA of the martingales we are able to define medically relevant groupings of the event types. We also show how the principal component scores can be estimated and used as predictors in a Cox Proportional Hazards (PH) model. This can be used to remove multicollinearity among predictors and is analogous to principal component linear regression.

The rest of this paper is structured as follows. Section 2 studies estimation of covariance and correlation matrices for counting processes and martingales. Section 3 defines the estimators for survival PCA and shows that they are consistent and asymptotically normal. Section 4 provides results of simulation studies for survival PCA methods. Section 5 provides an analysis of adverse events among patients in the MPACT study for pancreatic cancer. Section 6 gives concluding remarks. Proofs for theorems can be found in the appendix and additional simulation results can be found in the supplementary materials.

2. Covariance estimation for bivariate counting processes and counting process martingales

2.1 Estimation of covariance in presence of right censoring

As in Section 1 assume that for each subject in a population there are \( p \) event types of interest, and let \( T^{(j)} \) denote the failure time for the \( jth \) event type. We will assume that \( T^{(j)} \) is a continuous random variable. The full vector of failure times for a subject can be
written as \( T = [T^{(1)}, \ldots, T^{(p)}]^T \). If \( t = [t_1, \ldots, t_p]^T \) is a vector of fixed timepoints, the joint distribution for \( T \) is defined as \( F_T(t) = P(T^{(1)} \leq t_1, \ldots, T^{(p)} \leq t_p) \), and the univariate distribution functions are defined as \( F^{(j)}(t_j) = P(T^{(j)} \leq t_j) \) for \( j = 1, \ldots, p \). We will also assume that there is an independent censoring time, \( C \), with distribution \( F_C(c) = P(C \leq c) \). This censoring time is the same for all event types. This assumption is reasonable in the case where all event types are measured for the same subject, and end of study or loss to follow up will make it impossible to get any additional information from that subject for any of the event types. The observed data for the \( j \)th event type is the observed time, \( Y^{(j)} = T^{(j)} \wedge C \), and the censoring indicator, \( \eta^{(j)} = I(T^{(j)} \leq C) \). Define the counting process associated with the \( j \)th event type as \( N^{(j)}(t) = I(T^{(j)} \leq t) \). Note that the value of \( N^{(j)}(t) \) is not always observed, unlike \( N^{(j^1)}(t) = N^{(j)}(t)\eta^{(j)} \), which we will use in later sections for deriving asymptotic properties of estimates.

We can define the cumulative hazard function for the \( j \)th event type at time \( t \) to be \( \Lambda^{(j)}(t) = \int_0^t \lambda^{(j)}(s)ds \), where \( \lambda^{(j)}(t) = \lim_{\delta \to 0} \frac{1}{\delta} P(t \leq T^{(j)} < t + \delta | T^{(j)} \geq t) \). The martingale defined by the decomposition of \( N^{(j)}(t) \) is \( M^{(j)}(t) = N^{(j)}(t) - \Lambda^{(j)}(t \wedge T^{(j)}) \).

Define the covariance between \( N^{(j)}(t_j) \) and \( N^{(j')}(t_{j'}) \) to be \( CN^{(j,j')}(t_j, t_{j'}) \), and similarly the covariance between \( M^{(j)}(t_j) \) and \( M^{(j')}(t_{j'}) \) to be \( CM^{(j,j')}(t_j, t_{j'}) \). Because \( N^{(j)}(t) \) may not be observed for every subject in the presence of censoring, \( CN^{(j,j')}(t_j, t_{j'}) \) and \( CM^{(j,j')}(t_j, t_{j'}) \) cannot be calculated using standard methods. The following equality can be used to estimate \( CN^{(j,j')}(t_j, t_{j'}) \),

\[
CN^{(j,j')}(t_j, t_{j'}) = S^{(j,j')}(t_j, t_{j'}) - S^{(j)}(t_j)S^{(j')}(t_{j'}), \tag{2.1}
\]

where \( S^{(j,j')}(t_j, t_{j'}) = P(T^{(j)} > t_j, T^{(j')} > t_{j'}) \) is the bivariate survival function and \( S^{(j)}(t_j) = P(T^{(j)} > t_j) \) is the univariate survival function. Similarly, the following equality from [Pren-]
is obtained through Stieltjes integration and repeated integration by parts, and can be used to estimate $CM^{(j,j')}(t_j, t_{j'})$,

$$CM^{(j,j')}(t_j, t_{j'}) = S^{(j,j')}(t_j, t_{j'}) - 1 + \int_0^{t_j} S^{(j,j')}(s_j, t_{j'}) \lambda^{(j)}(s_j) ds + \int_0^{t_{j'}} S^{(j,j')}(t_j, s_{j'}) \lambda^{(j)}(s_{j'}) ds_{j'} + \int_0^{t_j} \int_0^{t_j'} S(s_j, s_{j'}) \lambda^{(j)}(s_j) \lambda^{(j')}(s_{j'}) ds_{j'} ds_{j'}.$$  

(2.2)

This allows for the consistent estimation of $CN^{(j,j')}(t_j, t_{j'})$ and $CM^{(j,j')}(t_j, t_{j'})$ by plugging in consistent estimates of all of the quantities on the right hand side of Equations (2.1) and (2.2) respectively. The variance of $N^{(j)}(t)$ and $M^{(j)}(t)$ can also be written as a function of the univariate survival functions, $CN^{(j,j)}(t_j) = S^{(j)}(t_j)\{1 - S^{(j)}(t_j)\}$, and $CM^{(j,j)}(t_j) = 1 - S^{(j)}(t_j)$. We use the Kaplan-Meier estimator for the univariate survival functions and the Nelson-Aalen estimator for the univariate cumulative hazard function. Potential estimators for the bivariate survival function are discussed in Section 1. We use the estimator from Dabrowska et al. (1988) because it is shown to converge weakly to a Gaussian process (Gill et al., 1995), and has good performance in simulations (Cheng et al., 2007).

In addition to covariance, correlation can be a useful measure for describing the relationship between the martingales or counting processes of two event types. Correlation is standardized by the product of the standard deviations of the martingales or counting processes for the two event types. This gives a useful way to compare the strength of linear association of two martingales or counting processes across time in a way that is not influenced by the changing variance of the martingales and counting processes over time. Define $RN^{(j,j')}(t_j, t_{j'}) = Cor\{N^{(j)}(t_j), N^{(j')}(t_{j'})\}$ and $RM^{(j,j')}(t_j, t_{j'}) = Cor\{M^{(j)}(t_j), M^{(j')}(t_{j'})\}$. 


The following equality can be used to estimate $RN^{(j,j')}(t_j, t_{j'})$,

$$RN^{(j,j')}(t_j, t_{j'}) = \frac{CN^{(j,j')}(t_j, t_{j'})}{\sqrt{S^{(j)}(t_j)\{1 - S^{(j)}(t_j)\}} \sqrt{S^{(j')}(t_{j'})\{1 - S^{(j')}(t_{j'})\}}}.$$  \hspace{1cm} (2.3)

Similarly the following equality can be used to estimate $RM^{(j,j')}(t_j, t_{j'})$,

$$RM^{(j,j')}(t_j, t_{j'}) = \frac{CM^{(j,j')}(t_j, t_{j'})}{\sqrt{1 - S^{(j)}(t_j)} \sqrt{1 - S^{(j')}(t_{j'})}}.$$  \hspace{1cm} (2.4)

Note that Equation (2.3) requires that $0 < S^{(j)}(t_j) < 1$ for all $j$ and Equation (2.4) requires that $S^{(j)}(t_j) < 1$ for all $j$ in order to be well defined. The right hand side of Equations (2.3) and (2.4) can both be consistently estimated using the Kaplan-Meier estimator and estimates for $CN^{(j,j')}(t_j, t_{j'})$ and $CM^{(j,j')}(t_j, t_{j'})$ respectively.

One important thing to note is that $N^{(j)}(t)$ only contains information on if an event has happened before time $t$, and no further information on when it happened. This means $CN^{(j,j')}(t_j, t_{j'})$ and $RN^{(j,j')}(t_j, t_{j'})$ are only useful when $0 < S^{(j)}(t_j), S^{(j')}(t_j) < 1$. For example, when $t_j$ or $t_{j'} = \infty$, $CN^{(j,j')}(t_j, t_{j'}) = 0$, and $RN^{(j,j')}(t_j, t_{j'})$ is undefined because either $N^{(j)}(t_j)$ or $N^{(j')}(t_{j'})$ will be a degenerate random variable. Alternatively $M^{(j)}(t)$ contains information both on if an event has happened by time $t$, as well as when it happened if it happened before time $t$. If we evaluate $CM^{(j,j')}(t_j, t_{j'})$ or $RM^{(j,j')}(t_j, t_{j'})$ at $t_j = t_{j'} = \infty$, they are equivalent to the covariance and correlation between $T^{(j)}$ and $T^{(j')}$ after transforming both variables to have an Exponential(1) distribution using a cumulative hazard transformation. That is $CM^{(j,j')}(\infty, \infty) = \text{Cov}\{\Lambda^{(j)}(T^{(j)}), \Lambda^{(j')}(T^{(j)})\}$, and $RM^{(j,j')}(\infty, \infty) = \text{Cor}\{\Lambda^{(j)}(T^{(j)}), \Lambda^{(j')}(T^{(j)})\}$, respectively. However, this cumulative hazard transformation, and other similar transformations such as the CDF transformation are
not possible for all observations in the presence of right censoring. This is why we focus on martingale and counting processes which have correlation and covariance functions indexed by time that can be estimated in the presence of right censoring.

For a fixed set of timepoints, \( t = [t_1, \ldots, t_p]^T \), we define the full covariance matrix for all \( p \) counting processes as

\[
CN(t) = \begin{bmatrix}
CN^{(1,1)}(t_1, t_1) & \cdots & CN^{(1,p)}(t_1, t_p) \\
\vdots & \ddots & \vdots \\
CN^{(p,1)}(t_p, t_1) & \cdots & CN^{(p,p)}(t_p, t_p)
\end{bmatrix}.
\] (2.5)

\( RN(t), CM(t) \) and \( RM(t) \) can be defined analogously. We will denote the estimators of each of these matrices created by using the estimator for each element of the matrix as \( \hat{CN}(t) \), \( \hat{RN}(t) \), \( \hat{CM}(t) \), and \( \hat{RM}(t) \) respectively. Note that these estimates may not be positive semi-definite. In this case there are a number of methods that can be used to transform these matrices to be positive semi-definite or positive-definite (Rousseeuw and Molenberghs, 1993). We define \( \tilde{CN}(t) \) to be the matrix with the same eigenvectors as \( \hat{CN}(t) \), and all negative eigenvalues set to some small non-negative constant. \( \tilde{RN}(t), \tilde{CM}(t), \tilde{RM}(t) \) are defined analogously. If we assume that the true covariance or correlation matrices are positive-definite then it can be shown that transforming these matrices to be positive-definite will not change the limiting behavior at a fixed timepoint using results from Section 2.3 and Weyl’s inequality.

In the presence of censoring the values of \( t \) for which the martingale or counting process covariance and correlation matrices should be estimated depends on the largest observed censoring or failure time for each event type. For a given data set define \( t^*_j \) to be the largest
observed failure or censoring time for the $j$th event type. The Kaplan-Meier estimator for the survival function for event type $j$ is typically not estimated beyond timepoint $t_j^*$ because doing so can introduce bias \cite{Gillespie}. The Nelson-Aalen estimator and non-parametric bivariate survival function estimators will also have this issue. For this reason we only recommend the estimation of $CM^{(j,j')}(t_j, t_{j'})$, $RM^{(j,j')}(t_j, t_{j'})$, $CN^{(j,j')}(t_j, t_{j'})$ and $RN^{(j,j')}(t_j, t_{j'})$ for $t_j \leq t_j^*$ and $t_{j'} \leq t_{j'}^*$. Further, in the case where the final observed time for the $j$th event type is a failure time then $\hat{S}^{(j)}(t_j) = \hat{S}^{(j,j')}(t_j, t_{j'}) = 0$ for $t_j \geq t_j^*$. This means that $\hat{CN}^{(j,j')}(t_j, t_{j'}) = 0$, and $\hat{RN}^{(j,j')}(t_j, t_{j'})$ is not well defined for $t_j \geq t_{j'}$. A similar issue happens before the first observed failure for event type $j$ when $\hat{S}^{(j)}(t_j) = 1$. In this case both $\hat{RN}^{(j,j')}(t_j, t_{j'})$ and $\hat{RM}^{(j,j')}(t_j, t_{j'})$ are not well defined. In general the martingale and counting process covariance and correlations should only be estimated for timepoints between the first observed failure time for and final observed failure or censoring time for each event type.

### 2.2 Estimation of covariance in presence of right censoring and competing risks

In this section we will allow for the introduction of a single competing risk for all of the non-competing event types of interest. As a motivating example consider a study where subjects are followed over time for a number of non-competing adverse events and death acts as a competing risk for each of the adverse events. As before we will assume that each subject has $p$ non-competing failure times, $T = [T^{(1)}, \ldots, T^{(p)}]^T$. Define $\bar{T}$ to be the failure time for the competing risk. In the case where there is more than one competing risk these can be combined into a single competing event time \cite{Cheng}. In the absence of censoring for the $j$th event type we will observe $\bar{T}^{(j)} = T^{(j)} \wedge \bar{T}$ and $\gamma^{(j)} = 2 - I(\bar{T} > T^{(j)})$. If
we again assume an independent censoring time, $C$, we will observe $\tilde{Y}^{(j)} = \tilde{T}^{(j)} \land C$, failure indicator $\tilde{\eta}^{(j)} = I(\tilde{T}^{(j)} \leq C)$, and cause and censoring type indicator $\tilde{z}^{(j)} = \tilde{\eta}^{(j)} \tilde{\gamma}^{(j)}$. This set up can be thought of as a semi-competing risk setting where $T^{(j)}$ will not be observed if $\hat{T}$ is observed first, but $\hat{T}$ may still be observed after $T^{(j)}$ is observed. Previous work on semi-competing risks includes Fine et al. (2001), Cheng et al. (2007), and Jazić et al. (2016). Importantly, as with Cheng et al. (2007), the event types in $T$ do not compete with each other, and it is the association between these non-competing event types that we will focus on in this section.

The cause specific counting process for the $j$th event type and $l$th cause is defined as
\[ \tilde{N}^{(j)}_l(t) = I(\tilde{T}^{(j)} \leq t, \tilde{\gamma}^{(j)} = l) \] for $j = 1, \ldots, p$ and $l = 1, 2$. Similar to the previous section $\tilde{N}^{(j)}_l$ is not always observed, so we will also define $\tilde{N}^{(j)}_l(t) = \tilde{N}^{(j)}_l(t) \tilde{\eta}^{(j)}$ to be used for estimation purposes. The cumulative cause specific hazard for the $j$th event type and $l$th cause evaluated at time $t$ is $\tilde{\lambda}^{(j)}_l(t) = \int_0^t \tilde{\lambda}^{(j)}_l(s) ds$ where $\tilde{\lambda}^{(j)}_l(t) = \lim_{\delta \to 0} \frac{1}{\delta} P(t \leq \tilde{T}^{(j)} < t + \delta, \tilde{\gamma}^{(j)} = l \mid \tilde{T}^{(j)} \geq t)$. The martingale based on the decomposition of $\tilde{N}^{(j)}_l(t)$ is $\tilde{M}^{(j)}_l(t) = \tilde{N}^{(j)}_l(t) - \tilde{N}^{(j)}_l(t \land \tilde{T}^{(j)})$ [Lin, 1997].

We will focus on the covariance between counting processes and martingales for the non-competing adverse events. That is $\tilde{C}N^{(j,j')}_{l_1,l_2}(t_j, t_{j'}) = Cov\{\tilde{N}^{(j)}_{l_1}(t_j), \tilde{N}^{(j')}_{l_2}(t_{j'})\}$ and $\tilde{C}M^{(j,j')}_{l_1,l_2}(t_j, t_{j'}) = Cov\{\tilde{M}^{(j)}_{l_1}(t_j), \tilde{M}^{(j')}_{l_2}(t_{j'})\}$. These quantities cannot be estimated using standard methods in the presence of censoring. The following equality can be used to estimate $\tilde{C}N^{(j,j')}_{l_1,l_2}(t_j, t_{j'})$,

\[ \tilde{C}N^{(j,j')}_{l_1,l_2}(t_j, t_{j'}) = F^{(j,j')}_{l_1l_2}(t_j, t_{j'}) - F^{(j)}_{l_1}(t_j) F^{(j')}_{l_2}(t_{j'}) \] \hspace{1cm} (2.6)

where $F^{(j,j')}_{kl}(t_j, t_{j'}) = P(\tilde{T}^{(j)} \leq t_j, \tilde{\gamma}^{(j)} = k, \tilde{T}^{(j')} \leq t_{j'}, \tilde{\gamma}^{(j')} = l)$ is the bivariate cause
specific CIF and \( F_k^{(j)}(t_j) = P(\tilde{T}^{(j)} \leq t_j, \tilde{\gamma}^{(j)} = k) \) is the univariate cause specific CIF. Note that in our set up \( F_{22}^{(j,j')} (t,t) \) simplifies to a univariate function because the competing risk time is the same for event type \( j \) and \( j' \). For notational purposes, for a bivariate function, \( G(t_1, t_2) \), define \( G(dt_1, dt_2) = G^{(1,1)}(t_1, t_2)dt_1dt_2 \), \( G(dt_1, t_2) = G^{(1,0)}(t_1, t_2)dt_1 \), and \( G(t_1, dt_2) = G^{(0,1)}(t_1, t_2)dt_2 \), where \( G^{(1,1)}(t_1, t_2) \) is the second partial derivative of \( G(t_1, t_2) \) with respect to \( t_1 \) and \( t_2 \), \( G^{(1,0)}(t_1, t_2) \) is the first partial derivative of \( G(t_1, t_2) \) with respect to \( t_1 \), and \( G^{(0,1)}(t_1, t_2) \) is the first partial derivative of \( G(t_1, t_2) \) with respect to \( t_2 \). Using the fact that \( P(\tilde{T}_j \leq s, \tilde{\gamma}_j = k, \tilde{T}_j' > t) = F_k^{(j)} (s) - F_k^{(j,j')}(s,t) - F_k^{(j,j')}(s,t) \), the following result is obtained and can be used to estimate \( \tilde{C} \tilde{M}^{(j,j')}(t,t) \),

\[
\tilde{C} \tilde{M}^{(j,j')}(t,t) = \tilde{\Lambda}_1^{(j)}(t)\tilde{\Lambda}_1^{(j')}(t)\tilde{S}^{(j,j')}(t,t) + \int_0^t \tilde{\Lambda}_1^{(j)}(s)\tilde{\Lambda}_1^{(j')}(s)F_{22}^{(j,j')}(ds)
+ \int_0^t \int_0^t \{1 - \tilde{\Lambda}_1^{(j)}(s_1)\}\{1 - \tilde{\Lambda}_1^{(j')}(s_2)\}F_{11}^{(j,j')}(ds_1, ds_2)
+ \int_0^t \{1 - \tilde{\Lambda}_1^{(j)}(s)\}\{-\tilde{\Lambda}_1^{(j')}(t)\}\{F_1^{(j)}(ds) - F_{11}^{(j,j')}(ds, t) - F_{12}^{(j,j')}(ds, t)\}
+ \int_0^t \{1 - \tilde{\Lambda}_1^{(j')}(s)\}\{-\tilde{\Lambda}_1^{(j)}(t)\}\{F_1^{(j')}(ds) - F_{11}^{(j,j')}(t, ds) - F_{21}^{(j,j')}(t, ds)\}
+ \int_0^t \int_{s_1}^t \{1 - \tilde{\Lambda}_1^{(j)}(s_1)\}\{-\tilde{\Lambda}_1^{(j')}(s_2)\}F_{12}^{(j,j')}(ds_1, ds_2)
+ \int_0^t \int_{s_1}^t \{1 - \tilde{\Lambda}_1^{(j')}(s_1)\}\{-\tilde{\Lambda}_1^{(j)}(s_2)\}F_{12}^{(j,j')}(ds_1, ds_2),
\]

(2.7)

where \( \tilde{S}^{(j,j')}(t_j, t_{j'}) = P(\tilde{T}^{(j)} > t_j, \tilde{T}^{(j')} > t_{j'}) \). Equation (2.7) is obtained Stieltjes integration similar to how Equation (2.2) is obtained in [Prentice and Cai (1992)]. The bivariate cause specific CIFs can be estimated using methods from [Cheng et al. (2007)](Cheng_2007), while the univariate cause specific CIFs and cumulative hazards can be estimated using methods from [Kalbfleisch and Prentice (2011)](Kalbfleisch_2011). Both the bivariate CIF and \( \tilde{S}^{(j,j')}(t_j, t_{j'}) \) require the estimation of a standard non-competing risk bivariate survival function, and we use the Dabrowska estima-
tor just as in Section 2.1. Therefore estimators for \(C \hat{N}^{(j,j')}(t_j, t_{j'})\) and \(C \hat{M}^{(j,j')}(t_j, t_{j'})\) are obtained by consistently estimating all the elements on the right hand side of Equations (2.6) and (2.7) respectively. The variance of \(\tilde{N}^{(j)}(t_j)\) and \(\tilde{M}^{(j)}(t_j)\) can also be written as functions of the bivariate and univariate cause specific CIFs, \(C \check{N}^{(j,j)}(t_j) = F_1^{(j)}(t_j)\{1 - F_1^{(j)}(t_j)\}\), and \(C \check{M}^{(j,j)}(t_j) = F_1^{(j)}(t_j)\).

As in the previous section we can extend these methods to estimate the correlation between the counting processes or martingales. Define \(R \tilde{N}^{(j,j)}(t_j, t_{j'}) = Cor\{\tilde{N}_1^{(j)}(t_j), \tilde{N}_1^{(j')} (t_{j'})\}\) and

\[R \tilde{M}^{(j,j)}(t_j, t_{j'}) = Cor\{\tilde{M}_1^{(j)}(t_j), \tilde{M}_1^{(j')} (t_{j'})\}\]

The following equality is used to estimate \(R \tilde{N}^{(j,j)}(t_j, t_{j'})\),

\[
R \tilde{N}^{(j,j)}(t_j, t_{j'}) = \frac{C \check{N}^{(j,j')} (t_j, t_{j'})}{\sqrt{F_1^{(j)}(t_j)\{1 - F_1^{(j)}(t_j)\}} \sqrt{F_1^{(j')} (t_{j'})\{1 - F_1^{(j')} (t_{j'})\}}} \quad (2.8)
\]

and

\[
R \tilde{M}^{(j,j)}(t_j) = \frac{C \check{M}^{(j,j')} (t_j)}{\sqrt{F_1^{(j)}(t_j) \sqrt{F_1^{(j')} (t_{j'})}}} \quad (2.9)
\]

can be used to estimate \(R \tilde{M}^{(j,j)}(t_j, t_{j'})\). Equation (2.8) requires \(0 < F_1^{(j)}(t_j) < 1\) for all \(j\) and Equation (2.9) requires \(0 < F_1^{(j)}(t)\) for all \(j\) in order to be well defined. The full covariance and correlation matrices at the vector of timepoints \(t\), \(C \check{N}(t), R \tilde{N}(t), C \check{M}(t)\), and \(R \tilde{M}(t)\) can all be defined analogously to \(CN(t)\), \(RN(t)\), \(CM(t)\), and \(RM(t)\) from the previous section. The estimators of these matrices created by estimating each element of the matrix are defined as \(\widehat{C \check{N}}(t), \widehat{R \tilde{N}}(t), \widehat{C \check{M}}(t)\), and \(\widehat{R \tilde{M}}(t)\). These matrix estimates may not be positive-definite. In this case we can define \(\widehat{C \check{N}}(t), \widehat{R \tilde{N}}(t), \widehat{C \check{M}}(t)\), and \(\widehat{R \tilde{M}}(t)\) using the same techniques as in Section 2.1. Just as in section 2.1 the values of \(t\) for which it is useful
to estimate $CN(t), RN(t), CM(t)$, and $RM(t)$ are meaningful depends on the specific data set. It is not useful to consider values of $t_j$ larger than the largest observed failure, censoring or competing risk time for event type $j$, as this will introduce bias. For values of $t_j$ smaller than the first observed event time for event type $j$, the estimates of the competing risk martingale and counting process correlation matrices are not well defined as the estimate of $F_1^{(j)}(t_j)$ will be 0.

### 2.3 Weak convergence of covariance and correlation estimates

In this section we show weak convergence properties for the estimates of the covariance and correlation matrices for martingales and counting processes. Define $PF$ to be the expectation of a random function, $F$, and $H^{(j)}(t) = I(Y^{(j)} ≥ t)$. Further define $\mathbb{P}_nN^{(j)}(t) = \frac{1}{n} \sum_{i=1}^{n} I(Y^{(j)}_i ≤ t, \eta^{(j)}_i = 1)$ and $\mathbb{P}_nH^{(j)}(t) = \frac{1}{n} \sum_{i=1}^{n} I(Y^{(j)}_i ≥ t)$, and for an arbitrary $q \times r$ matrix $M$ define the function $\text{Vec}(M)$ to be the column vector created by stacking the columns of $M$ on top of each other. In general we will assume that $\Lambda^{(j)}$ is estimated with the Nelson-Aalen estimator and $S^{(j)}$ is estimated with the Kaplan-Meier estimator. Theorem 1 shows the weak convergence of the estimates of the elements of $CM$, $CN$, $RM$ and $RN$:

**Theorem 1** Assume that the estimator of $S^{(j,j')}$ converges weakly such that $\sqrt{n}[\mathbb{P}_nN^{(11)} - PN^{(11)}, \ldots, \mathbb{P}_nN^{(p1)} - PN^{(p1)}, \mathbb{P}_nH^{(1)} - PH^{(1)}, \ldots, \mathbb{P}_nH^{(p)} - PH^{(p)}, \hat{S}^{(1,2)} - S^{(1,2)}, \ldots, \hat{S}^{(p-1,p)} - S^{(p-1,p)}]^T \rightsquigarrow [Z_{N1}, \ldots, Z_{Np}, Z_{H1}, \ldots, Z_{Hp}, Z_{S_{12}}, \ldots, Z_{S_{p-1,p}}]^T$ in $D[0, \tau_1]^2 \times \ldots \times D[0, \tau_p]^2 \times D[0, \tau_{12}] \times \ldots \times D[0, \tau_{p-1,p}]$, where $\tau_{jj'} = (\tau_j, \tau_{j'})$, $(D[0, \tau_j], || \cdot ||_\infty)$ is the space of univariate cadlag functions of bounded variation in $[0, \tau_j]$ equipped with uniform norm, $(D[0, \tau_{jj'}], || \cdot ||_\infty)$ is the space of bivariate cadlag functions of bounded variation in $[0, \tau_{jj'}]$ equipped with uniform norm, and $[Z_{N1}, \ldots, Z_{Np}, Z_{H1}, \ldots, Z_{Hp}, Z_{S_{12}}, \ldots, Z_{S_{p-1,p}}]^T$ is a mean zero tight
Gaussian process. Assume that $P(Y^{(j)} > \tau_j, Y^{(j')} > \tau_{j'}) > 0$ for all $1 \leq j, j' \leq p$ and $\Lambda^{(j)} < \infty$ for $j = 1, \ldots, p$, then for any $[0, t] \subset [0, \tau]$ where $\tau = [\tau_1, \ldots, \tau_p]$,

$$\sqrt{n}[\text{Vec}(\hat{C}M) - \text{Vec}(CM)](t) \leadsto Z_{CM} \quad (2.10)$$

$$\sqrt{n}[\text{Vec}(\hat{C}N) - \text{Vec}(CN)](t) \leadsto Z_{CN}, \quad (2.11)$$

where $Z_{CM}$ and $Z_{CN}$ are $p^2$ dimensional mean zero Gaussian processes. Further if $\omega = [\omega_1, \ldots, \omega_p]$, where $\omega_j < \tau_j$ for $j = 1, \ldots, p$ and $P(Y^{(j)} \leq \omega_j, Y^{(j')} \leq \omega_{j'}) > 0$ for all $1 \leq j, j' \leq p$, then for any $[\omega, t] \subset [\omega, \tau]$

$$\sqrt{n}[\text{Vec}(\hat{R}M) - \text{Vec}(RM)](t) \leadsto Z_{RM} \quad (2.12)$$

$$\sqrt{n}[\text{Vec}(\hat{R}N) - \text{Vec}(RN)](t) \leadsto Z_{RN}, \quad (2.13)$$

where $Z_{RM}$ and $Z_{RN}$ are $p^2$ dimensional mean zero Gaussian processes.

The proof for this theorem is presented in the Appendix, and is done by showing that the right hand sides of Equations (2.11) and (2.10) are Hadamard differentiable mappings and applying the functional delta method (Theorem 2.8 in Kosorok (2008)). The additional assumption required for Equations (2.12) and (2.13) is needed to ensure that $RM^{(j,j)}$ and $RN^{(j,j')}$ are well defined. In addition the assumption of joint weak convergence can be shown for the Nelson-Aalen, Kaplan-Meier and Dabrowska estimators using methods similar to Lemma A.6 in the appendix of Cheng et al. (2007). Next we will show weak convergence properties for the martingale and counting process covariance and correlation matrices in the presence of competing risks. Define $\tilde{H}^{(j)}(t) = I(\hat{Y}^{(j)} \geq t)$. Further let $P_n\tilde{N}^{(j1)}_t$ and $P_n\tilde{H}^{(j)}$ be defined analogously to $P_nN^{(j1)}$ and $P_nH^{(j)}$. We will assume that $\tilde{\Lambda}^{(j)}_1$ is estimated using a Nelson-Aalen style estimator and $F^{(j)}_1$ is estimated using methods from Kalbfleisch and
Assume that the estimator of $\hat{S}^{(j,j')}$ converges weakly such that $\sqrt{n}[\widehat{\mathbf{C}} - \mathbf{C}] \Rightarrow Z_{\mathbf{C}},$ where $Z_{\mathbf{C}}$ is a mean zero tight Gaussian process. Assume that $P(\tilde{Y}^{(j)} > \tau_j, \tilde{Y}^{(j')} > \tau_{j'}) > 0$ for all $1 \leq j, j' \leq p$ and $\tilde{X}_k^{(j)} < \infty$ for $j = 1, \ldots, p$ and $k = 1, 2,$ then for any $[\tilde{\omega}, \tilde{\tau}] \subset [0, \tilde{\tau}]$ where $\tilde{\tau} = [\tilde{\tau}_1, \ldots, \tilde{\tau}_p],$

$$\sqrt{n}[\text{Vec}(\widehat{\mathbf{C}}) - \text{Vec}(\mathbf{C})](t) \Rightarrow Z_{\mathbf{C}},$$  \hspace{1cm} (2.14)  

$$\sqrt{n}[\text{Vec}(\widehat{\mathbf{C}}) - \text{Vec}(\mathbf{C})](t) \Rightarrow Z_{\mathbf{C}},$$  \hspace{1cm} (2.15)

where $Z_{\mathbf{C}}$ and $Z_{\mathbf{C}}$ are $p^2$ dimensional mean zero Gaussian processes. Further if $[\tilde{\omega}_1, \ldots, \tilde{\omega}_p]$, where $\tilde{\omega}_j < \tilde{\tau}_j$ for $j = 1, \ldots, p$ and $P(\tilde{Y}^{(j)} \leq \tilde{\omega}_j, \tilde{Y}'^{(j')} \leq \tilde{\omega}_{j'}) > 0$ for all $1 \leq j, j' \leq p,$ then for any $[\tilde{\omega}, \tilde{\tau}]$

$$\sqrt{n}[\text{Vec}(\widehat{\mathbf{R}}) - \text{Vec}(\mathbf{R})](t) \Rightarrow Z_{\mathbf{R}},$$  \hspace{1cm} (2.16)  

where $Z_{\mathbf{R}}$ and $Z_{\mathbf{R}}$ are $p^2$ dimensional mean zero Gaussian processes.

The proof of Theorem 2 is provided in the Appendix, and uses similar methods as the proof for Theorem 1.
3. PCA methods for right censored data

PCA transforms a set of variables into linearly uncorrelated variables. For a $p \times 1$ dimensional random vector $X$ the first principal component direction, $v_1$, is the $p \times 1$ vector for which $\text{Var}(v_1^T X)$ is maximized subject to the constraint $||v_1||_2 = 1$. The $j$th principal component direction is the $p \times 1$ vector for which $\text{Var}(v_j^T X)$ is maximized subject to the constraints $||v_j||_2 = 1$ and $v_j^T v_{j'} = 0$ for $j' < j$. The principal components can be shown to be the eigenvectors of the covariance matrix of $X$. The solutions will not be unique if there are repeated eigenvalues. The principal components, $v_j^T X$, are linearly uncorrelated. The proportion of the variance of the data explained by the $j$th principal component is equal to $\frac{\lambda_j}{\sum_{i=1}^{p} \lambda_i}$, where $\lambda_i$ is the $i$th eigenvalue of the covariance matrix of $X$. Estimates PCA are found through eigendecomposition of the sample covariance matrix or SVD of the mean centered data.

In the presence of right censoring it is not possible to non-parametrically estimate the principal components for $T$, the $p \times 1$ dimensional vector of failure times. This is because it is not possible to estimate the covariance matrix using standard methods without making strong assumptions on the form of the joint distribution. This is the case with or without the presence of competing risks. When there are no competing risks, instead of estimating the principal components for $T$ we consider the principal components for $N(t) = [N^{(1)}(t_1), \ldots, N^{(p)}(t_p)]^T$ or $M(t) = [M^{(1)}(t_1), \ldots, M^{(p)}(t_p)]^T$. The principal directions of $N(t)$ are the eigenvectors of $CN(t)$ and the principal directions of $M(t)$ are the eigenvectors of $CM(t)$. Similarly in the presence of competing risks we consider $\tilde{N}(t) = [\tilde{N}^{(1)}_1(t_1), \ldots, \tilde{N}^{(p)}_1(t_p)]^T$, which has principal directions equal to the eigenvectors of $C\tilde{N}(t)$ and $\tilde{M}(t) = [\tilde{M}^{(1)}_1(t_1), \ldots, \tilde{M}^{(p)}_1(t_p)]^T$ which has principal directions equal to the eigenvectors of $C\tilde{M}(t)$. In all cases the correlation matrix can be used instead of the covari-
ance matrix to get a scaled version PCA.

We obtain estimates for principal directions for \( N(t), M(t), \tilde{N}(t), \) and \( \tilde{M}(t) \) using the eigenvectors of consistent estimates of the relevant covariance or correlation matrices, which were derived in Section 2. Consistent estimates of the proportion of variance explained are based on the corresponding eigenvalues. If we assume that the eigenvalues of \( CM(t), CN(t), \hat{CM}(t), \) or \( \hat{CN}(t) \) are unique then the estimates of the corresponding principal components will be consistent and asymptotically normal based on the results from Theorem 3. This is also true if we use the correlation matrices instead of the covariance matrices.

**Theorem 3** Assume that \( t_1, \ldots, t_n \) are iid realizations of the \( p \times 1 \) random vector \( T \), with joint distribution \( F_T \). Assume that \( \Sigma \) is a \( p \times p \) positive-definite function of \( F_T \) and that \( \hat{\Sigma}_n \) is a positive-definite function of \( (t_1, \ldots, t_n) \). Define \( v_{\Sigma i} (\hat{v}_{\Sigma in}) \) to be the \( i \)th eigenvector of \( \Sigma (\hat{\Sigma}) \) and \( \xi_{\Sigma i} (\hat{\xi}_{\Sigma in}) \) to be the \( i \)th eigenvalue of \( \Sigma (\hat{\Sigma}) \). Assume that \( \sqrt{n} [\text{Vec}(\hat{\Sigma}_n) - \text{Vec}(\Sigma)] \rightarrow_d N(0, \Psi_{\Sigma}) \), where \( \Psi_{\Sigma} \) is some positive definite matrix. Let \( \Xi = \text{diag}(\xi_{\Sigma 1}, \ldots, \xi_{\Sigma p}) \), \( \hat{\Xi}_n = \text{diag}(\hat{\xi}_{\Sigma 1 n}, \ldots, \hat{\xi}_{\Sigma p n}) \), \( V = [v_{\Sigma 1}, \ldots, v_{\Sigma p}] \), and \( \hat{V} = [\hat{v}_{\Sigma 1 n}, \ldots, \hat{v}_{\Sigma p n}] \). Then \( \sqrt{n} [\text{vec}(\hat{V}) - \text{vec}(V)] \rightarrow_d N(0, \Psi_V) \) and \( \sqrt{n} [\text{vec}(\Xi) - \text{vec}(\hat{\Xi})] \rightarrow_d N(0, \Psi_{\Xi}) \), where \( \Psi_V \) and \( \Psi_{\Xi} \) are positive-semidefinite matrices.

The proof for Theorem 3 can be found in the Appendix and follows similar steps to the proof in [Anderson (2003)] for PCA using standard estimation techniques when data have a multivariate normal distribution. When combined with results from Section 2 this shows that estimates for the principal component vectors based on \( \tilde{CM}(t), \tilde{CN}(t), \) or the corresponding correlation matrix estimates are consistent and asymptotically normal.

The principal component scores can also be estimated for those subjects who have not been censored by time point \( t \). In order for this to be the case it must be that either
$C > t_j$, or $\eta^{(j)} = 1$ for $j = 1, \ldots, p$. In this case $N(t) = [N^{(1)}(t_1), \ldots, N^{(p)}(t_p)]^T$, the entire vector of failure counting processes at time point $t$ will be observed. Define $\hat{v}_{cnj}(t)$ to be the $jth$ eigenvector of $\hat{CN}(t)$. Then the estimate of the $jth$ principal component score is $\hat{v}_{cnj}(t)^T N(t)$. In the presence of a competing risk a similar estimate can be made using $\hat{N}_1(t) = [\hat{N}^{(1)}_1(t_1), \ldots, \hat{N}^{(p)}_1(t_p)]^T$ and the $jth$ eigenvector of $\hat{C\bar{N}}(t)$, $\hat{v}_{cnj}$. In the case where a subject is not censored by time point $t$, the entire vector of martingales, $M(t) = [M^{(1)}(t_1), \ldots, M^{(p)}(t_p)]^T$ can be consistently estimated by plugging in a consistent estimate of $\Lambda^{(j)}(t_j \wedge T^{(j)})$ for $j = 1, \ldots, p$. If $\hat{v}_{cnj}(t)$ is the $jth$ eigenvector of $\hat{CM}(t)$, then the estimate of the $jth$ principal component score is $\hat{v}_{cnj}(t)^T \hat{M}(t)$. Similar calculations can be done in the case of a competing risk. In this case the full vector of martingales, $\bar{M}_1(t) = [\bar{M}^{(1)}_1(t_1), \ldots, \bar{M}^{(p)}_1(t_p)]^T$, can be estimated by plugging in consistent estimates of $\bar{\Lambda}^{(j)}_1(t_j \wedge T^{(j)})$, for $j = 1, \ldots, p$. The $jth$ eigenvector of $\hat{C\bar{M}}(t)$ is defined as $\hat{v}_{\bar{cnj}}(t)$, and the estimate of the $jth$ principal component score is $\hat{v}_{\bar{cnj}}(t)^T \hat{\bar{M}}_1(t)$. All of the estimates for principal component scores based on the scaled counting processes or martingales can be estimated in the same way using the eigenvectors of the correlation matrices. In Section 5 we show an example where the principal component scores estimates can be used as covariates in a Cox PH model. This is possible because the form of the partial likelihood for the Cox PH model only requires the covariate values for those subjects who have not been censored by a given timepoint. This allows for principal component scores for censored time to event variables to be used as covariates in a Cox PH model similar to how principal component scores are used for principal component regression.

The estimation techniques for PCA derived above all allow for right censoring due to both loss to follow up or administrative censoring due to the end of the follow up period.
For certain censoring schemes there may be simpler ways to estimate the principal components for $M(t)$ and $N(t)$, however we believe for almost all types of time to event data it will be necessary to estimate the principal components of $M(t)$ and $N(t)$ rather than $T$ directly. In order to estimate the principal components for $T$ this would require no censoring implying unlimited follow up time and no competing risks, both of which are rare for data sets and studies with time to event data. However one setting in which the estimation of the covariance or correlation of $M(t)$ and $N(t)$ could be simplified would be one in which there is only administrative censoring and no loss to follow up. In this case we would assume that all subjects have the same censoring time, $C = c$. It is important to note that unless all subjects have had all of the event types by time $c$ it is still not possible to estimate the principal components for $T$. However, estimation of the principal components for $M(t)$ or $N(t)$ can be simplified for $t_1, \ldots, t_p < c$. In this case $N(t)$ will be fully observed for each subject because none have been censored before the specified timepoints, so the principal components can be estimated using standard estimation of the covariance matrix for $N(t)$. Similarly $M(t)$ involves an indicator variable which will be observed for all subjects and a set of cumulative hazard functions which can all be estimated up to time $c$. This means we can get an estimate of $M(t)$ for each subject and again use standard covariance and correlation estimation techniques to obtain principal component estimates for $M(t)$. Similarly in the competing risk setting if there is no loss to follow up and only administrative censoring at time $c$, it is still not possible to estimate the principal components of $T$ unless every subject has had each of the events before both the censoring time, $c$, and before they have the competing event. However in this set up the necessary indicator functions $\hat{N}_i^{(j)}(t)$ will be observed for all event types up to time $c$, and the necessary cause specific hazard functions
can also be estimated up to time \( t \). This means that as before we will observe \( \tilde{N}(t) \) and can get estimates of \( \tilde{M}(t) \) for each subject \( t_1, \ldots, t_p < c \). From here standard covariance and correlation estimation techniques can be used to estimate the principal components for \( \tilde{N}(t) \) and \( \tilde{M}(t) \).

A question of interest for researchers is whether PCA based on counting processes or martingales is preferred in practice. While each may have their advantages, we generally believe PCA based on martingales is more useful because they contain more information. This is because, as noted previously, \( M^{(j)}(t) \) and \( \tilde{M}^{(j)}(t) \) contain information on if event type \( j \) has happened by timepoint \( t \) as well as when it happened assuming it happened before timepoint \( t \). \( N^{(j)}(t) \) and \( \tilde{N}^{(j)}(t) \) only contain information on if event type \( j \) has happened by timepoint \( t \).

4. Simulation results

Simulations were conducted to examine the estimation of principal components of \( M(t), N(t), \tilde{M}_1(t), \tilde{N}_1(t) \) focusing on results using correlation matrices. Data sets were simulated with \( p = 8, 16 \) and 32. First we simulate \( W = [W^{(1)}, \ldots, W^{(p)}] \) with a multivariate normal distribution were \( W^{(j)} \) has mean zero and standard deviation one for \( j = 1, \ldots, p \). For each of the dimensions the covariance matrix of \( W \) is a block diagonal matrix with four equal size blocks along the diagonals of the form

\[
\Sigma_W = \begin{bmatrix}
A_1 & 0 & 0 & 0 \\
0 & A_2 & 0 & 0 \\
0 & 0 & A_3 & 0 \\
0 & 0 & 0 & A_4
\end{bmatrix}.
\]
$A_1$ has all ones along the diagonal and 0.7 on all off diagonal elements, $A_2$ has all ones along the diagonals and 0.4 on all off diagonal elements, $A_3$ has all ones along the diagonal and 0.2 on all off diagonal elements, and $A_4$ has all ones along the diagonal and 0.1 along all off diagonal elements. $T = [T^{(1)}, \ldots, T^{(p)}]$ is defined to be a transformation of $W$, such that $T^{(j)} \sim \text{Exponential}(1)$. Specifically $T^{(j)} = -\ln\{1 - \Phi(W^{(j)})\}$, where $\Phi(\cdot)$ is the CDF of a standard normal distribution.

The competing risk setting has a similar set up. $W^T$ has the same distribution as above and $[W^T, \tilde{W}]^T$ has a multivariate normal distribution and the correlation between $\tilde{W}$ and $W^{(j)}$ is 0.1 for $j = 1, \ldots, p$. $T$ is still defined as the same transformation of $W$ and $\tilde{T} = -\ln\{1 - \Phi(\tilde{W})\}$.

In all settings the censoring distribution is $[1/4 \cdot C_1] \sim \text{Beta}(1.5, 6.5)$. In addition simulations with the censoring distribution $C_2 \sim \text{Uniform}(0, 4)$ are shown in the supplementary materials. For all settings a sample size of $n = 200$ and $n = 1,000$ are simulated. In the setting without competing risks the average censoring rate for each event was 53% when using $C_1$ regardless of $p$ and 25% when using $C_2$. In the semi-competing risk setting the censoring rate was 34% when using $C_1$ and 13% when using $C_2$. For both censoring schemes in the competing risk setting the competing and non-competing events were equally likely to be observed. The lower censoring rate in the competing risk setting is due to the fact that the censoring time must come before both the non-competing event time and the competing event time in order for the subject to be censored.

For each data set $\tilde{R}_N(t), \tilde{R}_M(t), \tilde{R}_N(t)$, and $\tilde{R}_M(t)$ were estimated using the Dabrowska estimator for bivariate survival functions, the Kaplan-Meier estimator for univariate survival functions, and the Nelson-Aalen estimator for all cumulative hazard functions.
In order to ensure all estimated covariance and correlation matrices are positive-definite a minimum eigenvalue of 0.001 is used. Matrices are estimated at \( t = [1, 1, 1, 1, \ldots]^T \) and \( t = [2, 2, 2, 2, \ldots]^T \). The true covariance matrices are calculated empirically by simulating a single data set with 500,000 subjects and no censoring which allows for standard correlation and covariance estimation methods to be used. In the competing risk setting the cumulative hazard necessary to calculate the martingales is estimated using the Nelson-Aalen estimator based on the 500,000 simulated subjects. In the setting without competing risks the cumulative hazard is known based on the distribution of \( T^{(j)} \). The true principal component directions are calculated as the eigenvectors of the true correlation and covariance matrices. More information on the true correlation matrices can be found in the supplementary materials.

For the \( i \)th simulated data and \( j \)th principal component direction the angle in radians between the estimated direction, \( \hat{v}_{ij} \) and the true direction, \( v_j \), and is calculated as

\[
\text{Angle}(\hat{v}_{ij}, v_j) = \cos^{-1}\left( \frac{||\hat{v}_{ij}^T v_j||}{||\hat{v}_{ij}||_2 ||v_j||_2} \right)
\]

For each setting the bias and standard deviation for the \( j \)th principal direction is calculated as the empirical mean and standard deviation of the angle between the estimated direction and true direction for all 1,000 simulated data sets. Table 1 reports the average angle and standard deviation for the principal component directions using eigendecomposition of \( \widetilde{RN}(t) \), \( \widetilde{RM}(t) \), \( \widetilde{RN}(t) \), and \( \widetilde{RM}(t) \) for \( C_1 \) and \( t = [1, 1, 1, 1, \ldots]^T \). Results for \( t = [2, 2, 2, 2, \ldots]^T \) and \( C_2 \) are reported in the supplementary materials. In all cases just the first four directions are reported due to the four block structure of the underlying covariance and correlation matrices. In Table 1 we can see that as expected the average angle decreases as sample size increases. Further we can see that the average angle for the first direction is actually lower as the number of dimensions increases.
This is likely due to the increase in the leading eigenvalue for higher dimension set ups. In our higher dimension set ups the leading principal component explains a larger percentage of the overall variance than in lower dimension set ups. Additionally because of the nature of our block matrix set up as the dimension of the set ups increases the difference between the first and second eigenvalues also increases. Both of these facts likely contribute to the first eigenvector being estimated more precisely in our simulations. The true eigenvectors and eigenvalues of the matrices can be found in the supplementary materials. For \( RM(t) \) when \( n = 1,000 \) the average angle between the estimated first direction and true first direction is less than 0.30 radians, which is just over 17 degrees. Even 0.58 radians which is the highest average angle between the true and estimated first principal direction for \( RM(t) \) at \( n = 200 \) is just over 33 degrees. At \( t = [1, 1, 1, 1, \ldots]^T \) the estimates based on \( RM(t) \) outperform the estimates based on \( RN(t) \) across all sample sizes, and dimensions. This is also true when \( t = [2, 2, 2, 2, \ldots]^T \) and \( C_2 \) as shown in the supplementary materials. One key driver for this is the variance of the counting processes is much lower than the variances of the martingales, leading to less precise estimates of the relevant correlations.

In the competing risk setting the angle between the true and estimated principal directions based on \( R\tilde{M}(t) \) are similar, but slightly worse than those based on \( RM(t) \) in the non-competing risk section. However, in the competing risk setting the average angle between the estimated and true principal directions using \( R\tilde{N}(t) \) are lower than those using \( R\tilde{M}(t) \). This again is likely due to the leading eigenvalues for \( R\tilde{N}(t) \) being larger with more separation than \( RN(t) \). This is because \( \tilde{N}_1^{(j)}(t) \) and \( \tilde{N}_1^{(j')} (t) \) will tend to be correlated even when \( T^{(j)} \) and \( T^{(j')} \) are uncorrelated. Consider the fact that in order for \( \tilde{N}_1^{(j)}(t) \) to equal one it must be the case that both \( T^{(j)} \leq t \) and \( T^{(j')} \leq \tilde{T} \). If \( T^{(j)} \) and \( T^{(j')} \) are independent and
Table 1: Average (SD) of the angle in radians between the true and estimated PCA directions based on counting process and martingale correlation using with censoring distribution $C_1$

<table>
<thead>
<tr>
<th></th>
<th>Counting Process</th>
<th>Martingale</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>n=200</td>
<td>n=1000</td>
<td>n=200</td>
</tr>
<tr>
<td>8 Dim</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PC 1</td>
<td>0.70 (0.24)</td>
<td>0.35 (0.12)</td>
<td>1.49</td>
</tr>
<tr>
<td>PC 2</td>
<td>0.98 (0.30)</td>
<td>0.56 (0.22)</td>
<td>1.26</td>
</tr>
<tr>
<td>PC 3</td>
<td>1.12 (0.29)</td>
<td>0.78 (0.32)</td>
<td>1.13</td>
</tr>
<tr>
<td>PC 4</td>
<td>1.15 (0.29)</td>
<td>0.82 (0.35)</td>
<td>1.06</td>
</tr>
<tr>
<td>16 Dim</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PC 1</td>
<td>0.46 (0.14)</td>
<td>0.21 (0.06)</td>
<td>2.47</td>
</tr>
<tr>
<td>PC 2</td>
<td>0.73 (0.25)</td>
<td>0.32 (0.09)</td>
<td>1.77</td>
</tr>
<tr>
<td>PC 3</td>
<td>1.03 (0.28)</td>
<td>0.49 (0.19)</td>
<td>1.37</td>
</tr>
<tr>
<td>PC 4</td>
<td>1.19 (0.25)</td>
<td>0.65 (0.26)</td>
<td>1.19</td>
</tr>
<tr>
<td>32 Dim</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PC 1</td>
<td>0.37 (0.08)</td>
<td>0.17 (0.04)</td>
<td>4.41</td>
</tr>
<tr>
<td>PC 2</td>
<td>0.55 (0.15)</td>
<td>0.24 (0.05)</td>
<td>2.79</td>
</tr>
<tr>
<td>PC 3</td>
<td>0.86 (0.25)</td>
<td>0.35 (0.07)</td>
<td>1.87</td>
</tr>
<tr>
<td>PC 4</td>
<td>1.17 (0.24)</td>
<td>0.50 (0.11)</td>
<td>1.43</td>
</tr>
</tbody>
</table>

With competing risk

<table>
<thead>
<tr>
<th></th>
<th>Counting Process</th>
<th>Martingale</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>n=200</td>
<td>n=1000</td>
<td>n=200</td>
</tr>
<tr>
<td>8 Dim</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PC 1</td>
<td>0.41 (0.28)</td>
<td>0.15 (0.06)</td>
<td>2.20</td>
</tr>
<tr>
<td>PC 2</td>
<td>0.95 (0.38)</td>
<td>0.61 (0.40)</td>
<td>1.19</td>
</tr>
<tr>
<td>PC 3</td>
<td>1.15 (0.29)</td>
<td>0.91 (0.38)</td>
<td>1.02</td>
</tr>
<tr>
<td>PC 4</td>
<td>1.17 (0.28)</td>
<td>0.94 (0.37)</td>
<td>0.93</td>
</tr>
<tr>
<td>16 Dim</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PC 1</td>
<td>0.36 (0.23)</td>
<td>0.14 (0.04)</td>
<td>3.78</td>
</tr>
<tr>
<td>PC 2</td>
<td>0.83 (0.38)</td>
<td>0.40 (0.28)</td>
<td>1.85</td>
</tr>
<tr>
<td>PC 3</td>
<td>1.09 (0.31)</td>
<td>0.69 (0.35)</td>
<td>1.34</td>
</tr>
<tr>
<td>PC 4</td>
<td>1.16 (0.26)</td>
<td>0.77 (0.31)</td>
<td>1.08</td>
</tr>
<tr>
<td>32 Dim</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PC 1</td>
<td>0.35 (0.22)</td>
<td>0.13 (0.03)</td>
<td>6.99</td>
</tr>
<tr>
<td>PC 2</td>
<td>0.74 (0.36)</td>
<td>0.31 (0.19)</td>
<td>3.14</td>
</tr>
<tr>
<td>PC 3</td>
<td>0.98 (0.30)</td>
<td>0.54 (0.27)</td>
<td>1.98</td>
</tr>
<tr>
<td>PC 4</td>
<td>1.02 (0.24)</td>
<td>0.64 (0.21)</td>
<td>1.36</td>
</tr>
</tbody>
</table>
uncorrelated $P(T^{(j)} \leq \bar{T} | T^{(j)} \leq \bar{T})$ will be higher than $P(T^{(j')} \leq \bar{T})$. This in turn means that when $T^{(j)}$ and $T^{(j')}$ are independent $P(\tilde{N}_{1}^{(j)}(t) = 1 | P(\tilde{N}_{1}^{(j)}(t) = 1)$ will be larger than $P(\tilde{N}_{1}^{(j')}(t) = 1)$, leading to a positive correlation between $\tilde{N}_{1}^{(j)}(t)$ and $\tilde{N}_{1}^{(j')}(t)$. The effect on $\tilde{M}_{1}^{(j)}(t)$ and $\tilde{M}_{1}^{(j')}(t)$ is not as large because it will not introduce correlation between $\tilde{\Lambda}_{1}^{(j)}(t \land \bar{T}^{(j)})$ and $\tilde{\Lambda}_{1}^{(j')}(t \land \bar{T}^{(j')})$ to the same degree. Because of this in our simulation set up the separation between the first and second eigenvalues for $\tilde{R}N(t)$ is larger than for $\tilde{R}M(t)$, which leads to more precise estimates of the first principal direction. More information on this issue, and specifically the correlation between $\tilde{N}^{(j)}(t)$ and $\tilde{N}^{(j')}(t)$, even when $T^{(j)}$ and $T^{(j')}$ are uncorrelated can be found in the supplementary materials.

5. MPACT trial

The MPACT trial was a clinical trial that ran from 2009 to 2013 in which 861 patients with metastatic pancreatic cancer were randomized to be treated with either the standard of care, gemcitabine, or a novel medication, paclitaxel [Von Hoff et al. 2013]. The data for the 430 patients randomized to standard of care is available through Project Data Sphere®.

During this trial, there were nine adverse events that occurred in at least 50 of the patients: abdominal pain, anemia, constipation, decreased appetite, fatigue, nausea, neutropenia, thrombocytopenia, and vomiting. For each adverse event the failure time is the time from randomization to the first occurrence of that event. Patients who left the study due to death or disease progression before having a given adverse event were considered to have a competing event. Patients who left the study before having a given adverse event for any other reason were considered to be censored. Table 2 gives an overview of the nine different events. The event rate is the proportion subjects who had the given adverse event,
while the censoring rate is the proportion of subjects who were censored. The median and mean event times are calculated just among those who had the given event of interest. We can see that the event rate ranges from 0.13 to 0.27, while the censoring rate ranges from 0.28 to 0.35. In addition, the distribution of all nine event types is positively skewed with the mean larger than the median, indicating most subjects who have a given event have it early in the study with a few having the event later in the study.

<table>
<thead>
<tr>
<th>Event Type</th>
<th>Event Rate</th>
<th>Censoring Rate</th>
<th>Median Time (Days)</th>
<th>Mean Time (Days)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Abdominal Pain</td>
<td>0.14</td>
<td>0.34</td>
<td>36</td>
<td>57.31</td>
</tr>
<tr>
<td>Anemia</td>
<td>0.26</td>
<td>0.29</td>
<td>30</td>
<td>59.10</td>
</tr>
<tr>
<td>Constipation</td>
<td>0.13</td>
<td>0.35</td>
<td>18.5</td>
<td>43.09</td>
</tr>
<tr>
<td>Decreased Appetite</td>
<td>0.14</td>
<td>0.33</td>
<td>22.5</td>
<td>46.50</td>
</tr>
<tr>
<td>Fatigue</td>
<td>0.27</td>
<td>0.29</td>
<td>19</td>
<td>39.96</td>
</tr>
<tr>
<td>Nausea</td>
<td>0.20</td>
<td>0.33</td>
<td>15</td>
<td>35.86</td>
</tr>
<tr>
<td>Neutropenia</td>
<td>0.26</td>
<td>0.28</td>
<td>23</td>
<td>53.43</td>
</tr>
<tr>
<td>Thrombocytopenia</td>
<td>0.18</td>
<td>0.31</td>
<td>19</td>
<td>69.68</td>
</tr>
<tr>
<td>Vomiting</td>
<td>0.14</td>
<td>0.35</td>
<td>24</td>
<td>41.95</td>
</tr>
</tbody>
</table>

In order to better understand the relationship between these nine adverse events we conduct survival PCA for the nine events. We used the martingale correlation matrix because the martingales contain information on when an event occurred in addition to whether it occurred by a given time. Because patients are subject to both censoring due to leaving the study as well as the competing risk of death or progression of disease we use the competing risk martingale correlation matrix. $\hat{RM}(t)$ was estimated between $t = [30, 30, \ldots]^T$ and $t = [360, 360, \ldots]^T$ in the increment of one day. Day 360 was chosen as a final time point because by that time 420 of the 430 patients had left the study, while day 30 was chosen as the starting point to ensure a sufficient number of events had occurred to get reasonable precision for the estimates. The principal component loadings based on $\hat{RM}(t)$ at $t =$
are presented in Table 3. Figure 1 shows the directions for the first two principal components plotted over time between day 30 and 360. The line type is based on the following clinically-defined grouping. Constitutional (C): Fatigue; Gastrointestinal (G): Abdominal pain, constipation, decreased appetite, nausea, vomiting; Hematologic (H): Anemia, neutropenia, thrombocytopenia.

Table 3: Principal component directions at day 360 and proportion of variance explained for each principal component using estimates based on martingale correlation matrix

<table>
<thead>
<tr>
<th></th>
<th>PC1</th>
<th>PC2</th>
<th>PC3</th>
<th>PC4</th>
<th>PC5</th>
<th>PC6</th>
<th>PC7</th>
<th>PC8</th>
<th>PC9</th>
</tr>
</thead>
<tbody>
<tr>
<td>Abdominal pain</td>
<td>0.38</td>
<td>-0.21</td>
<td>-0.18</td>
<td>0.03</td>
<td>0.05</td>
<td>0.73</td>
<td>-0.41</td>
<td>-0.26</td>
<td>-0.04</td>
</tr>
<tr>
<td>Anemia</td>
<td>0.09</td>
<td>0.42</td>
<td>0.56</td>
<td>0.51</td>
<td>-0.39</td>
<td>0.26</td>
<td>-0.02</td>
<td>0.15</td>
<td>-0.05</td>
</tr>
<tr>
<td>Constipation</td>
<td>0.25</td>
<td>0.26</td>
<td>0.36</td>
<td>-0.16</td>
<td>0.80</td>
<td>0.10</td>
<td>0.09</td>
<td>0.21</td>
<td>-0.11</td>
</tr>
<tr>
<td>Decreased appetite</td>
<td>0.42</td>
<td>-0.16</td>
<td>-0.42</td>
<td>0.33</td>
<td>-0.03</td>
<td>-0.09</td>
<td>0.11</td>
<td>0.68</td>
<td>-0.16</td>
</tr>
<tr>
<td>Fatigue</td>
<td>0.43</td>
<td>0.00</td>
<td>-0.06</td>
<td>0.47</td>
<td>0.17</td>
<td>-0.28</td>
<td>0.28</td>
<td>-0.53</td>
<td>0.34</td>
</tr>
<tr>
<td>Nausea</td>
<td>0.46</td>
<td>0.14</td>
<td>0.06</td>
<td>-0.35</td>
<td>-0.29</td>
<td>-0.21</td>
<td>0.12</td>
<td>-0.27</td>
<td>-0.65</td>
</tr>
<tr>
<td>Neutropenia</td>
<td>-0.10</td>
<td>0.55</td>
<td>-0.45</td>
<td>-0.14</td>
<td>-0.06</td>
<td>0.40</td>
<td>0.54</td>
<td>-0.04</td>
<td>0.09</td>
</tr>
<tr>
<td>Thrombocytopenia</td>
<td>-0.01</td>
<td>0.60</td>
<td>-0.34</td>
<td>0.09</td>
<td>0.10</td>
<td>-0.30</td>
<td>-0.65</td>
<td>-0.04</td>
<td>-0.01</td>
</tr>
<tr>
<td>Vomiting</td>
<td>0.44</td>
<td>0.10</td>
<td>0.14</td>
<td>-0.48</td>
<td>-0.27</td>
<td>-0.07</td>
<td>-0.09</td>
<td>0.22</td>
<td>0.64</td>
</tr>
<tr>
<td>Proportion Variance</td>
<td>0.26</td>
<td>0.15</td>
<td>0.11</td>
<td>0.10</td>
<td>0.10</td>
<td>0.08</td>
<td>0.08</td>
<td>0.06</td>
<td>0.05</td>
</tr>
</tbody>
</table>

Based on Figure 1, it is apparent that the largest loadings in the first principal component are gastrointestinal and constitutional events, and the loadings for these events all go in the same direction. The largest loadings for the second principal component are the hematologic events which also all go in the same direction. Therefore the first principal component is driven by the occurrence of gastrointestinal events and fatigue, and the second principal component is driven by the occurrence of hematologic events. This shows that martingales for gastrointestinal and constitutional events tend to be correlated with martingales for other gastrointestinal and constitutional events, while martingales for hematologic events tend to be correlated with martingales for other hematologic events. Together the first two principal components explain close to 40% of the total variance, which is consistent over time.
Figure 1: Principal component direction loadings from day 30 to 360 for first two principal components using martingale correlation matrix estimates. Line types indicate constitutional, gastrointestinal and hematologic event types.

Unlike the first two principal components, the third through ninth principal components do not have a straightforward interpretation. In addition, as can be seen in the proportion of variance explained in Table 3, the proportion of variance explained for the third through seventh principal components are similar. This is true for all the time points estimated, and caused issues with potentially crossing eigenvalues. In order to make them comparable over time, instead of ordering the principal directions based on proportion of variance explained the order was chosen to minimize the sum of the angles between principal directions across time. First day 360 was chosen as a reference date, and all the principal directions were ordered in descending order of the associated eigenvalues. We can define $\hat{v}_{jt}$ to be the $jth$ principal direction for $t = [t, t, \ldots]^T$. For all days other than 360 the ordering of the principal directions was chosen in order to minimize $\sum_{j=1}^{9} \text{Angle}(\hat{v}_{jt}, \hat{v}_{j360})$. This meant for some days
the ordering of the principal directions would be changed. A figure of the principal directions using this method for ordering can be found in the supplementary materials.

In addition to simple analysis of the principal component loadings we used the first two principal component scores in a Cox PH model with death or progression of disease as the outcome. We follow the methods described in Section 3 to estimate the principal component scores. Because the loadings and interpretation of the first two principal components are consistent over time we include the principal component scores as time varying covariates. The first principal component can be thought of as a measure of how often and how early a subject had gastrointestinal and constitutional events, with those subjects who have had more and earlier gastrointestinal and constitutional events tending to have a higher principal component 1 score. Similarly the second principal component can be thought of as a measure of how often and how early a subject had hematologic events. The principal component scores for the first two principal components were then used as time varying covariates in a Cox PH model with time until death or progression of disease as the outcome of interest. In addition to the principal component scores, the age in years, sex, and Karnofsky performance status were included as covariates. The Karnofsky performance status is a numeric measure of the general well being of cancer patients, and ranged from 70 to 100 within the sample (Karnofsky et al., 1948). One subject was dropped from the model due to missing Karnofsky performance status.

Table 4 gives the estimated hazard ratios for the Cox PH model. The hazard ratio for the first principal component is estimated to be 1.62 and has a p-value of $4.28 \times 10^{-7}$. This indicates that for those subjects who have had more and earlier gastrointestinal events by a given timepoint have a higher estimated hazard of death holding other covariates constant.
This result is consistent with the notion that as-yet-undocumented disease progression may manifest as fatigue or gastrointestinal adverse events. This would mean that subjects with a higher PC 1 score have a higher hazard for death or progression of disease, as we see in the results. The hazard ratio for the second principal component is less than one, but the p-value is equal to 0.46. The estimated coefficient for PC 2 indicates that those subjects who have had more and earlier hematologic events by a given time point have a lower hazard of death holding other covariates constant. However given the p-value we cannot rule out a null or opposite result. PC 2 having a hazard ratio below 1 is consistent with hematologic adverse events indicating greater exposure to the drug (due to increased absorption or decreased degradation) or a greater sensitivity to its effects. In this case we might expect subjects with more and earlier hematologic events to have a lower hazard of death or progression of disease. This is consistent with the estimated hazard ratio for PC 2 being below 1, however the p-value indicates that this may just be due to random variation in the MPACT trial sample. As a sensitivity analysis we estimated the principal components using the counting process correlation matrix instead of the martingale correlation matrix, and estimated the same Cox PH model with the counting process correlation principal component scores. These results led to similar conclusions as the results using the martingale correlation matrix and can be found in the Supplementary materials.

6. Discussion

In this paper we show how PCA can be defined for multivariate survival data in the presence of censoring by using either the counting processes or the corresponding martingales defined by each event type. We build on previous results for bivariate survival data to show how
Table 4: Estimated hazard ratios for Cox PH models including first two martingale correlation PC scores

<table>
<thead>
<tr>
<th>PC</th>
<th>HR</th>
<th>P-Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>PC1</td>
<td>1.62</td>
<td>4.28E-07</td>
</tr>
<tr>
<td>PC2</td>
<td>0.92</td>
<td>0.46</td>
</tr>
<tr>
<td>Age</td>
<td>0.99</td>
<td>0.04</td>
</tr>
<tr>
<td>Sex:Male</td>
<td>1.01</td>
<td>0.94</td>
</tr>
<tr>
<td>Karnofsky Performance Status</td>
<td>0.98</td>
<td>0.02</td>
</tr>
</tbody>
</table>

N 429
Events 264

to estimate the full covariance and correlation matrices for either the counting processes or martingales at a given time point. In addition we are able to extend this to the semi-competing risk setting where each of the different event types are subject to a competing risk as well as an independent censoring time. For both the standard censoring only setting and the semi-competing risk setting we are able to show that the estimators for the martingale and counting process covariance or correlation matrices converge to a mean zero Gaussian process when properly normalized.

We also show that the loadings for the principal components based on the martingales and counting processes can be consistently estimated through eigendecomposition of the corresponding covariance or correlation matrix. The corresponding estimates are shown to be consistent and asymptotically normal. A subject’s principal component score can only be estimated up to the time they are censored, however this still allows for principal component scores to be used as covariates in Cox PH regression.

The results for this paper are all for a fixed number of non-competing failure times, \( p \). Given the increasing importance of 'big data' it is also of interest to consider the high-dimensional setting where \( p \to \infty \). It would be important to define a set of conditions under which the estimates of the martingale and counting process covariance and correlation
matrices will converge in probability to the corresponding true matrices. One way to do this would be by using maximal inequalities as defined in Kosorok (2008). This would allow us to show convergence even as $p \to \infty$, so long as $n$ grows at a fast enough rate relative to $p$. However, in order to use maximal inequalities, it is necessary to sufficiently bound the rate at which the martingale or counting process covariance and correlation estimates converge to the true value for each pair of failure times. This is a stronger condition than the consistency and asymptotic normality that we have shown in this paper. An example of the type of bound needed can be found in Theorem 1 of Bitouzé et al. (1999), which defines a Dvoretzky–Kiefer–Wolfowitz type inequality for the Kaplan-Meier estimator. To our knowledge, no similar result has been shown for any known bivariate survival function estimator. If such a result can be shown to hold then consistency of the martingale and counting process correlation and covariance estimates in a high-dimensional setting would follow.

The usefulness of this method is shown using data from the comparator arm of the MPACT trial for patients with pancreatic cancer. We estimate the principal components based on the martingale correlation matrix for nine different adverse events experienced by patients in the trial, and are able to define medically relevant groupings of these events. An R package which implements these methods is available at github.com/blangworthy/survPCA. One area for future research is to further consider ordering and potential changing of principal components over time. In the case where two or more principal components have similar eigenvalues the ordering may change over time due to random noise. It may be of interest to investigate how to identify when changes in principal component loadings and eigenvalues over time are due to random noise or an actual change in true underlying covariance or
correlation matrix. Additionally the described methods do not allow for controlling for covariates. For standard PCA it is possible to estimate conditional PCA by using model residuals after regressing out a set of shared covariates. A similar idea may be possible using a Cox PH model for each different event type using a shared set of covariates and then estimating PCA using martingale residuals. One difference between this method and the survival PCA we present in this paper is the martingale residuals are typically calculated at the event or censoring time for each subject, which can be different across subjects and event types in the multivariate setting. However martingales are always mean zero so covariance measures between them may give a meaningful measure of the association between different event types.

Supplementary Materials

The online supplementary materials include additional simulation results as well as additional analysis of the data from the MPACT trial.

Acknowledgements

The authors would like to thank Dr. Yu Cheng from the University of Pittsburgh for supplying code to compute the Dabrowska bivariate survival function estimator.

Appendix

Proof of Theorem 1 Define \( \theta_C = [S^{(1)}, \ldots, S^{(p)}, \Lambda^{(1)}, \ldots, \Lambda^{(p)}, S^{(1,2)}, \ldots, S^{(p-1,p)}]^T \) and \( \hat{\theta}_C = [\hat{S}^{(1)}, \ldots, \hat{S}^{(p)}, \hat{\Lambda}^{(1)}, \ldots, \hat{\Lambda}^{(p)}, \hat{S}^{(1,2)}, \ldots, \hat{S}^{(p-1,p)}]^T \) where \( \hat{S}^{(j)} \) is the Kaplan-Meier estimator, \( \hat{\Lambda}^{(j)} \) is the Nelson-Aalen estimator. We will assume \( \hat{S}^{(j,j')} \) is the Dabrowska estimator,
but other bivariate survival estimators are possible. Based on previously shown results for the Kaplan-Meier, Nelson-Aalen, and Dabrowska estimators, (see Kosorok (2008) for further details on the Kaplan-Meier and Nelson-Aalen estimators and Cheng et al. (2007) for further details on the Dabrowska estimator)

\[ \sqrt{n}[\hat{\theta}_C - \theta_C] \to Z_{\theta_C}, \]

where \( Z_{\theta_C} = [Z_{S_1}, \ldots, Z_{S_p}, Z_{A_1}, \ldots, Z_{A_p}, Z_{S_{1,2}}, \ldots, Z_{S_{p-1,p}}]^T \) is a \( 2p + (p^2 - p)/2 \) dimensional mean 0 Gaussian process. Using this set up we can show the results for \( \hat{CN} \) and \( \hat{CM} \)

1. \( \hat{CN} \): Consider the mapping

\[ \phi_{CN_{jj'}}(\theta_C) = S^{(j)}_j - S^{(j)}_j S^{(j')}_. \]

Using this mapping

\[ \sqrt{n}[\text{Vec}(\hat{CN}) - \text{Vec}(CN)] = \sqrt{n} \begin{bmatrix} \phi_{CN_{11}}(\hat{\theta}_C) - \phi_{CN_{11}}(\theta_C) \\ \phi_{CN_{12}}(\hat{\theta}_C) - \phi_{CN_{12}}(\theta_C) \\ \vdots \\ \phi_{CN_{p-1p}}(\hat{\theta}_C) - \phi_{CN_{p-1p}}(\theta_C) \\ \phi_{CN_{pp}}(\hat{\theta}_C) - \phi_{CN_{pp}}(\theta_C) \end{bmatrix}. \]

The Hadamard derivative of \( \phi_{CN_{jj'}}(\theta_C) \) in the direction of \( Z_{\theta_C} \) is

\[ \phi'_{\theta_CCN_{jj'}}(Z) = S^{(j)}_j Z_{S_j} + S^{(j')}_j Z_{S_{j,j'}} - Z_{S_{j,j'}}. \]
Therefore by the functional delta method, Theorem 2.8 in Kosorok (2008),
\[ \sqrt{n}[\text{Vec}(\hat{C}N) - \text{Vec}(CN)] \rightsquigarrow Z_{CN} \] where \( Z_{CN} \) is a \( p^2 \) dimensional mean 0 Gaussian process.

2. \( \hat{CM} \): Consider the mapping

\[ \phi_{CMjj'}(\theta_C)(t_j, t_{j'}) = S^{(j,j')}(t_j, t_{j'}) - 1 + \int_0^{t_j} S^{(j,j')}(s_j^-, t_{j'}) \lambda^{(j)}(s_j) ds_j + \int_0^{t_{j'}} S^{(j,j')}(t_j, s_{j}^-) \lambda^{(j')}(s_{j'}) ds_{j'} + \int_0^{t_j} \int_0^{t_{j'}} S^{(j,j')}(s_j^-, s_{j'}^-) \lambda^{(j)}(s_j) \lambda^{(j')}(s_{j'}) ds_j ds_{j'}. \] (6.18)

Using this mapping

\[
\sqrt{n}[\text{Vec}(\hat{CM}) - \text{Vec}(CM)] = \sqrt{n} \begin{bmatrix}
\phi_{CM11}(\hat{\theta}_C) - \phi_{CM11}(\theta_C) \\
\phi_{CM12}(\hat{\theta}_C) - \phi_{CM12}(\theta_C) \\
\vdots \\
\phi_{CMp-1p}(\hat{\theta}_C) - \phi_{CMp-1p}(\theta_C) \\
\phi_{CMpp}(\hat{\theta}_C) - \phi_{CMpp}(\theta_C)
\end{bmatrix}.
\]

As with the results for \( \hat{CN} \), the desired result will follow if it can be shown that \( \phi_{CMjj'}(\theta_C)(t_j, t_{j'}) \) is Hadamard differentiable in the direction of \( Z_{\theta_C} \). In order to do this we can consider each of the five parts on the right hand side of Equation (6.18) separately. It is straightforward to show that \( S^{(j,j')}(t_j, t_{j'}) \) and \(-1\) are hadamard differentiable. \( \int_0^{t_j} S^{(j,j')}(s_j^-, t_{j'}) \lambda^{(j)}(s_j) ds_j \) and \( \int_0^{t_{j'}} S^{(j,j')}(t_j, s_{j}^-) \lambda^{(j')}(s_{j'}) ds_{j'} \) can both be shown to be Hadamard differentiable through Lemma 12.3 from Kosorok (2008), and \( \int_0^{t_j} \int_0^{t_{j'}} S^{(j,j')}(s_j^-, s_{j'}^-) \lambda^{(j)}(s_j) \lambda^{(j')}(s_{j'}) ds_j ds_{j'} \) can be shown to be Hadamard differ-
entiable through Lemma A5 from Cheng et al. (2007). Therefore by the functional
delta method \( \sqrt{n}[\text{Vec}(\hat{CM}) - \text{Vec}(CM)] \Rightarrow Z_{CM} \) where \( Z_{CM} \) is a \( p^2 \) dimensional mean
0 Gaussian process.

In order to show the results for \( \hat{RN} \) and \( \hat{RM} \) we define \( \theta_R = [S^{(1)}, \ldots, S^{(p)}, CN^{(1,1)}, CN^{(1,2)}, \ldots, CN^{(p,p)}, CM^{(1,1)}, CM^{(1,2)}, \ldots, CM^{(p,p)}]^T \) and
\( \hat{\theta}_R = [\hat{S}^{(1)}, \ldots, \hat{S}^{(p)}, \hat{CN}^{(1,1)}, \hat{CN}^{(1,2)}, \ldots, \hat{CN}^{(p,p)}, \hat{CM}^{(1,1)}, \hat{CM}^{(1,2)}, \ldots, \hat{CM}^{(p,p)}]^T \). Then
using results from above
\[ \sqrt{n}[\hat{\theta}_R - \theta_R] \Rightarrow Z_{\theta_R}, \]
where \( Z_{\theta_R} = [Z_{S_1}, \ldots, Z_{S_p}, Z_{CN_{1,1}}, Z_{CN_{1,2}}, \ldots, Z_{CN_{p,p}}, Z_{CM_{1,1}}, Z_{CM_{1,2}}, \ldots, Z_{CM_{p,p}}]^T \) is a \( p+2p^2 \)
dimensional mean 0 Gaussian process. From here we can show the results for \( \hat{RN} \) and \( \hat{RM} \)
using the functional delta method as above.

1. \( \hat{RN} \) Define
\[ \phi_{RNjj'}(\theta_R) = \frac{CN^{(jj')}}{\sqrt{S^{(j)}(1 - S^{(j)})}\sqrt{S^{(j')}(1 - S^{(j')})}}. \]

Using this mapping
\[ \sqrt{n}[\text{Vec}(\hat{RN}) - \text{Vec}(RN)] = \sqrt{n} \begin{bmatrix} \phi_{RN11}(\hat{\theta}_R) - \phi_{RN11}(\theta_R) \\ \phi_{RN12}(\hat{\theta}_R) - \phi_{RN12}(\theta_R) \\ \vdots \\ \phi_{RNp-1p}(\hat{\theta}_R) - \phi_{RNp-1p}(\theta_R) \\ \phi_{RNpp}(\hat{\theta}_R) - \phi_{RNpp}(\theta_R) \end{bmatrix}. \]

The Hadamard differentiability of \( \phi_{RNjj'}(\theta_R) \) in the direction of \( Z_{\theta_R} \) can be shown
through repeated application of the chain rule and the following results,
• If \( \phi_1(A) = A^2 \) then \( \phi'_{A1}(\alpha) = 2A\alpha \).

• If \( \phi_2(A) = \sqrt{A} \) then \( \phi'_{A2}(\alpha) = \frac{\alpha}{2\sqrt{A}} \).

• If \( \phi_3(A, B) = \frac{A}{B} \) then \( \phi'_{AB3}(\alpha, \beta) = \frac{Ba - A\beta}{B^2} \).

Therefore by the functional delta method \( \sqrt{n}[\text{Vec}(\hat{R}N) - \text{Vec}(RN)] \Rightarrow Z_{RN} \) where \( Z_{RN} \) is a \( p^2 \) dimensional mean 0 Gaussian process.

2. \( \hat{RM} \) Results are similar to those for \( \hat{RN} \) if we consider the mapping

\[
\phi_{RMjj'}(\theta_R) = \frac{CM^{jj'}}{\sqrt{1 - S^{jj'}} \sqrt{1 - S^{jj'}_p}}
\]

**Proof of Theorem 2** Similar to the proof of Theorem 1 we define

\[
\hat{\theta}_C = [F^{(1)}_1, ..., F^{(p)}_1, F^{(1,1)}_{11}, F^{(1,2)}_{11}, ..., F^{(p,p)}_{11}, F^{(1,1)}_{12}, F^{(1,2)}_{12}, ..., F^{(p,p)}_{12}, F^{(1,1)}_{21}, F^{(1,2)}_{21}, ..., F^{(p,p)}_{21}, F^{(1,1)}_{22}, F^{(1,2)}_{22}, ..., F^{(p,p)}_{22}, \hat{\Lambda}^{(1)}_1, ..., \hat{\Lambda}^{(j)}_1]_T \quad \text{and} \quad \hat{\theta}_C = [\hat{F}^{(1)}_1, ..., \hat{F}^{(p)}_1, \hat{F}^{(1,1)}_{11}, \hat{F}^{(1,2)}_{11}, ..., \hat{F}^{(p,p)}_{11}, \hat{F}^{(1,1)}_{12}, \hat{F}^{(1,2)}_{12}, ..., \hat{F}^{(p,p)}_{12}, \hat{F}^{(1,1)}_{21}, \hat{F}^{(1,2)}_{21}, ..., \hat{F}^{(p,p)}_{21}, \hat{F}^{(1,1)}_{22}, \hat{F}^{(1,2)}_{22}, ..., \hat{F}^{(p,p)}_{22}, \hat{\Lambda}^{(1)}_1, ..., \hat{\Lambda}^{(j)}_1]_T \]

Using results from Cheng et al. (2007) it can be shown that

\[
\sqrt{n}[\hat{\theta}_C - \hat{\theta}_C] \Rightarrow Z_{\hat{\theta}_C},
\]

where \( Z_{\hat{\theta}_C} = [Z_{F11}, ..., Z_{F1p}, Z_{F111}, ..., Z_{F1111}, Z_{F1112}, ..., Z_{F11121}, Z_{F1211}, Z_{F1212}, ..., Z_{F12pp}, Z_{F2111}, ..., Z_{F21111}, Z_{F21112}, ..., Z_{F21121}, Z_{F2211}, ..., Z_{F222pp}, Z_{\Lambda_1}, ..., Z_{\Lambda_p}]_T \) is a \( 2p + 3p^2 \) dimensional mean 0 Gaussian process. The proofs for \( \hat{C}N \) and \( \hat{C}M \) follow in a similar manner to \( \hat{C}N \) and \( \hat{C}M \).
1. \( \tilde{CN} \): The relevant mapping is for \( \tilde{CN} \)

\[
\phi_{\tilde{CN}jj'}(\tilde{\theta}_C) = F_{11}^{(j,j')} - F_1^{(j)}F_1^{(j')}
\]

Using this mapping the desired results can be shown using similar methods to \( \tilde{CN} \) in the proof of Theorem 1 above.

2. \( \tilde{CM} \): The relevant mapping for \( \tilde{CM} \) is

\[
\phi_{\tilde{CM}jj'}(\tilde{\theta}_C)(t) = \tilde{\Lambda}_1^{(j)}(t)\tilde{\Lambda}_1^{(j')}(t)\tilde{\gamma}^{(j,j')}(t, t) + \int_{0}^{t} \tilde{\Lambda}_1^{(j)}(s)\tilde{\Lambda}_1^{(j')}(s)F_{22}^{(j,j')}(ds)
+ \int_{0}^{t} \int_{0}^{s} \{1 - \tilde{\Lambda}_1^{(j)}(s_1)\}\{1 - \tilde{\Lambda}_1^{(j')}(s_2)\}F_{11}^{(j,j')}(ds_1, ds_2)
+ \int_{0}^{t} \{1 - \tilde{\Lambda}_1^{(j)}(s)\}\{\tilde{\Lambda}_1^{(j')}(t)\}F_1^{(j')}(ds) - F_{11}^{(j,j')}(ds, t) - F_{12}^{(j,j')}(ds, t)
+ \int_{0}^{t} \{1 - \tilde{\Lambda}_1^{(j')}(s)\}\{\tilde{\Lambda}_1^{(j)}(t)\}F_1^{(j)}(ds) - F_{11}^{(j,j')}(t, ds) - F_{21}^{(j,j')}(t, ds)
+ \int_{0}^{t} \int_{s_1}^{t} \{1 - \tilde{\Lambda}_1^{(j)}(s_1)\}\{\tilde{\Lambda}_1^{(j')}(s_2)\}F_{12}^{(j,j')}(ds_1, ds_2)
+ \int_{0}^{t} \int_{s_1}^{t} \{1 - \tilde{\Lambda}_1^{(j')}(s_1)\}\{\tilde{\Lambda}_1^{(j)}(s_2)\}F_{12}^{(j',j)}(ds_1, ds_2),
\]

This mapping can be shown to be Hadamard differentiable in the direction of \( Z_{\tilde{\theta}_C} \) through repeated application of Lemma 12.3 from [Kosorok (2008)] and Lemma A5 from [Cheng et al. (2007)]. The results then follow using similar methods to the proof for \( \tilde{CM} \) in Theorem 1 above.

The results for \( \tilde{RN} \) and \( \tilde{RM} \) can be obtained by defining \( \tilde{\theta}_R = [F_1^{(1)}, \ldots, F_1^{(p)}, \tilde{CN}^{(1,1)}, \tilde{CN}^{(1,2)}, \ldots, \tilde{CN}^{(p,p)}, \tilde{CM}^{(1,1)}, \tilde{CM}^{(1,2)}, \ldots, \tilde{CM}^{(p,p)}]^T \) and

\[
\hat{\theta}_R = [\hat{F}_1^{(1)}, \ldots, \hat{F}_1^{(p)}, \hat{CN}^{(1,1)}, \hat{CN}^{(1,2)}, \ldots, \hat{CN}^{(p,p)}, \hat{CM}^{(1,1)}, \hat{CM}^{(1,2)}, \ldots, \hat{CM}^{(p,p)}]^T.
\]
ing the results from above

\[ \sqrt{n}[\hat{\theta}_R - \bar{\theta}_R] \rightsquigarrow Z_{\hat{\theta}_R}, \]

where \( Z_{\hat{\theta}_R} = [Z_{F11}, \ldots, Z_{F1p}, Z_{C\bar{N}11}, Z_{C\bar{N}12}, \ldots, Z_{C\bar{N}pp}, Z_{C\bar{M}11}, Z_{C\bar{M}12}, \ldots, Z_{C\bar{M}pp}]^T \) is a \( p + 2p^2 \) dimensional mean 0 Gaussian process. The results for \( \hat{R}_N \) and and \( \hat{R}_M \) can be shown using similar methods to \( \hat{R}_N \) and \( \hat{R}_M \) above using the two mappings below.

1. \( \hat{R}_N \): The relevant mapping is

\[ \phi_{\hat{R}Njj'}(\hat{\theta}_R) = \frac{C_{\bar{N}}^{(j,j')}}{\sqrt{F_1^{(j)}(1 - F_1^{(j)})} \sqrt{F_1^{(j')}(1 - F_1^{(j')})}}. \]

2. \( \hat{R}_M \): The relevant mapping is

\[ \phi_{\hat{R}Mjj'}(\hat{\theta}_R) = \frac{C_{\bar{M}}^{(j,j')}}{\sqrt{F_1^{(j)} F_1^{(j')}}}. \]

**Proof of Theorem 3** The consistency of \( \hat{V} \) can be established through a straightforward application of the Davis-Kahan Theorem, and the consistency of \( \hat{\Xi} \) can be established through Weyl's inequality. From here asymptotic normality can be established using steps similar to Theorem 13.5.1 in Anderson (2003), which derives the asymptotic variances of the eigenvectors and eigenvalues for the sample covariance matrix when data have a multivariate normal distribution and the sample covariance matrix has wishart distribution. However the steps can be straightforwardly extended to any asymptotically normal positive definite estimate of the covariance or correlation matrix as shown below.

Consider the transformation \( Q = V^T \hat{\Sigma} V \). Then by the Delta method \( \sqrt{n}(Q - \Xi) \rightarrow_d \)
\[ N(0, J_V \Psi \Sigma J_V^T) \] where \( J_V = V^T \otimes V^T \). It can be shown that \( \hat{\Sigma} \) and \( Q \) have the same singular values so \( Q \) can be represented as

\[ Q = G\hat{\Xi}G^T, \quad (6.19) \]

for orthogonal \( G \). \( G \) can be uniquely defined with the constraint \( g_{ii} \geq 0 \). Note that \( \hat{\Xi} = \hat{V}^T\hat{\Sigma}\hat{V} \), which together with the fact that \( V^T\hat{V} \) is orthogonal implies that every column of \( G \) is equal to \( \pm \) the corresponding column of \( V^T\hat{V} \). Because \( \hat{V} \to V \) the constraint that \( g_{ii} \geq 0 \) implies that \( G \to I \) where \( I \) the identity matrix. Define \( U = \sqrt{n}(Q - \Xi) \), \( D = \sqrt{n}(\hat{\Xi} - \Xi) \) and \( W = \sqrt{n}(Y - I) \). When we combine Equation \( (6.19) \) with \( GG^T = I \), and the conditions \( g_{ii} > 0 \) and \( \hat{\xi}_1, > \ldots, > \hat{\xi}_p \), we get a set of one to one functions from \( Q \) to \( G \) and \( \hat{\Xi} \) except on a set of measure zero which are continuously differentiable and have well defined inverses in the neighborhood of \( \hat{\Xi} = \Xi \) and \( G = I \). Therefore by the fact that \( U \) is asymptotically normal with mean zero, from the delta method \( W \) and \( D \) are also asymptotically normal with mean zero. Further because the column of \( \hat{V} \) is equal to \( \pm \) the corresponding column of \( VG \) and \( G \to I \), \( \sqrt{n}(\hat{V} - V) \) has the same limiting distribution as \( \sqrt{n}(VG - V) \) which is asymptotically normal with mean zero by the delta method.

In order to show that the limiting variances are functions of \( \Psi, \Sigma, V, \) and \( \Lambda \), using similar algebra from Theorem 13.5.1 in (Anderson, 2003) we get the following equalities

\[ U = W\Lambda + D + \Lambda W^T + o_p(1) \quad (6.20) \]

\[ 0 = W + W^T + o_p(1) \quad (6.21) \]

By combining results from Equations \( (6.20) \) and \( (6.21) \) and ignoring the \( o_p(1) \) terms we
obtain

\[ w_{ii} = 0 \]  \hspace{1cm} (6.22)

\[ d_{ii} = u_{ii} \]  \hspace{1cm} (6.23)

\[ w_{ij} = \frac{u_{ij}}{\lambda_j - \lambda_i} \text{ for } i \neq j \]  \hspace{1cm} (6.24)

The limiting distribution for \( W \) and \( D \) can be solved for using the limiting distribution of \( U \).

The limiting distribution of \( W \) can be used to find the limiting distribution of \( \sqrt{n}(\hat{V} - V) \).
References


First Author Affiliation: Department of Biostatistics, University of North Carolina at Chapel Hill
E-mail: langworthy.ben@gmail.com

Second Author Affiliation: Department of Biostatistics, University of North Carolina at Chapel Hill
E-mail: cai@bios.unc.edu

Third Author Affiliation: School of Medicine, University of North Carolina at Chapel Hill
E-mail: rcorty@gmail.com

Fourth Author Affiliation: Department of Biostatistics, University of North Carolina at Chapel Hill
E-mail: kosorok@bios.unc.edu

Fifth Author Affiliation: Department of Biostatistics, University of North Carolina at Chapel Hill
E-mail: jpfine1606@yahoo.com