Stat

WILEY

ORIGINAL ARTICLE OPEN ACCESS

Non-Elliptical Dimension Reduction in Survival Regression

Minjee Kim | Minjeong Kim | Jae Keun Yoo

Department of Statistics, Ewha Womans University, Seoul, Republic of Korea

Correspondence: Jae Keun Yoo (peter.yoo@ewha.ac.kr)

Received: 4 March 2026 | **Revised:** 29 May 2026 | **Accepted:** 3 June 2026

Keywords: censored data | informative predictor subspace | non-elliptical distribution | sufficient dimension reduction | survival regression

ABSTRACT

Sufficient dimension reduction (SDR) in survival regression aims to identify low-dimensional structures that preserve the relationship between survival time and predictors. Classical SDR methods, such as sliced inverse regression (SIR), rely on strong assumptions such as linearity, constant variance and coverage conditions, which are often violated in practical data. To address this issue, we propose an SDR framework for survival regression under non-elliptical predictor distributions. The proposed fused Cox proportional hazards with transformation-mean methods estimate the *survival informative predictor subspace* by integrating three components: the Cox proportional hazards direction, single response transformations and kernel fusion over multiple slicing schemes. Numerical studies and real data analysis show that our proposed methods outperform existing SIR-based approaches, particularly when the linearity condition fails. They also remain robust across various sample sizes, predictor dimensions and censoring levels. Overall, the proposed fused methods offer a reliable tool for dimension reduction in survival regression.

1 | Introduction

Sufficient dimension reduction (SDR) seeks a low-dimensional linear transformation of predictors that preserves the regression information. In survival regression $T|\mathbf{X}$ with high-dimensional $\mathbf{X} \in \mathbb{R}^p$, a central goal of SDR is to find $\mathbf{B}$ such that

$$F_{T|\mathbf{X}}(\cdot) = F_{T|\mathbf{B}^\top\mathbf{X}}(\cdot), \tag{1}$$

where $F_{T|\mathbf{X}}(\cdot)$ stands for a conditional distribution function of the survival time T given $\mathbf{X}$, and $\mathbf{B}$ is a $p \times q$ matrix with $q \leq p$. Statement (1) indicates that the lower-dimensional, linearly transformed predictor $\mathbf{B}^\top\mathbf{X}$ can replace the original p-dimensional predictor $\mathbf{X}$ without loss of information on the regression $T|\mathbf{X}$. The equivalence in (1) can be rephrased as follows:

$$T \perp\!\!\!\perp \mathbf{X}|\mathbf{B}^\top\mathbf{X}, \tag{2}$$

where $\perp\!\!\!\perp$ denotes statistical independence.

We denote $\mathcal{S}(\mathbf{B})$ as the column space of a matrix $\mathbf{B}$. Among all possible dimension reduction subspaces for the survival regression T on $\mathbf{X}$, the intersection is called the central subspace, denoted by $\mathcal{S}_{T|\mathbf{X}}$. The central subspace is minimal and unique, and its dimension d is referred to as the structural dimension. Hereafter, let $\boldsymbol{\eta} \in \mathbb{R}^{p\times d}$ be a true orthonormal basis of $\mathcal{S}_{T|\mathbf{X}}$, where $\boldsymbol{\eta}^\top\mathbf{X}$ represents a true sufficient predictor.

Practically, recovering a matrix $\mathbf{B}$ that satisfies (1) is not directly feasible from the observed data of T and $\mathbf{X}$, because the true survival time T is not fully observable due to censoring. To accommodate this aspect of the survival regression, a nonparametric dimension reduction approach was introduced by Cook (2003). This method extended the sliced inverse regression (SIR; Li 1991) to censored data through a bivariate slicing of (Y, δ), where Y is the observed survival time and δ is the event indicator. Under this approach, Yoo and Lee (2011) further developed nonparametric methods to test predictor effects in survival regression. The Cook approach will be discussed in detail in Section 2.1.

Minjee Kim and Minjeong Kim contributed equally to this paper.

The Cook approach requires the following crucial assumption: $E(\mathbf{X}|\mathbf{B}^\top\mathbf{X})$ is linear in $\mathbf{B}^\top\mathbf{X}$. This condition is satisfied if $\mathbf{X}$ is elliptically distributed, and a multivariate normal distribution is one of them. When $\mathbf{X}$ is non-elliptical, the linearity condition may fail, resulting in degraded or misleading estimates (Li and Dong 2009). Therefore, it is a fundamental issue to examine whether the linearity condition holds.

A popular way to check the condition is to investigate a scatterplot matrix of the predictors. If pairwise relationships show marginal linearity, the condition seems satisfied, but unobserved nonlinearity among predictors may still exist. So if the linearity condition fails, one may power-transform or reweight the $\mathbf{X}$ to approximate ellipticity according to Cook (2003) and Li and Dong (2009). However, if p is high, power-transformation may not be feasible, and reweighting is computationally intensive due to the deletion of observations.

To overcome these limitations, we extend an SDR method for non-elliptical predictors to survival regression. Specifically, we introduce a target subspace called *survival informative predictor subspace* (survival IPS). The survival IPS extends the informative predictor subspace of Yoo (2016) to censored data and contains the central subspace proposed by Cook (2003).

Methodologically, there are three key steps in the estimation. First, an initial direction is obtained from the Cox proportional hazards model to guide the slicing of $\mathbf{X}$. Second, (Y, δ) are transformed into one-dimensional response using either reweighting or mean imputation, following Datta et al. (2007). Finally, we fuse information across multiple slicing schemes by combining several kernel matrices into a single fused kernel matrix. This fusing procedure inherits the robustness of fused SIR (Cook and Zhang 2014) while adapting it to non-elliptical survival data.

The key point is that, under censoring and non-elliptical predictors, the main difficulty is not only to reduce dimension but also to target a subspace that is well-defined and yields reliable estimation. The survival IPS provides such a target by avoiding reliance on the linearity condition while still retaining the regression information relevant to $T|\mathbf{X}$. The Cox-guided slicing, one-dimensional transformation of (Y, δ) and the fusion across slicing schemes improve stability by reducing imbalanced slices due to censoring and sensitivity to tuning choices.

The organization of the paper is as follows. Section 2 reviews existing SDR methods under censoring and non-elliptical predictors, as well as fused SIR approach. In Section 3, the survival IPS is newly defined, and its estimation procedure is developed. Numerical studies and real data example are presented in Section 4. We summarize our work in Section 5.

2 | Literature Review

2.1 | SDR in Survival Regression

In survival regression, we typically deal with survival time T, censoring time C and a set of predictors $\mathbf{X}$. The observed response is represented as $Y = T\delta + C(1-\delta)$, where $\delta = I(T < C)$ indicates event occurrence (1 = event, 0 = censored). Again, the primary aim in SDR for survival regression is to estimate the central subspace $\mathcal{S}_{T|\mathbf{X}}$. However, due to censoring, bivariate SIR (Cook 2003) instead focused on estimating $\mathcal{S}_{(Y,\delta)|\mathbf{X}}$ from the observable pair (Y, δ) on $\mathbf{X}$.

Unlike the commonly used $C \perp\!\!\!\perp T|\mathbf{X}$, bivariate SIR only requires weaker conditions. The fundamental condition is $(T, C) \perp\!\!\!\perp \mathbf{X}|\mathbf{B}^\top\mathbf{X}$, where $\mathbf{B}^\top\mathbf{X}$ are sufficient predictors for the regression of T on $\mathbf{X}$. This can be decomposed into the following sub-conditions (eq. 12; Cook 2003):

$$\text{(A1) } T \perp\!\!\!\perp \mathbf{X}|\mathbf{B}^\top\mathbf{X} \quad \text{and} \quad \text{(A2) } C \perp\!\!\!\perp \mathbf{X}|\,(\mathbf{B}^\top\mathbf{X}, T). \tag{3}$$

Here, **(A2)** assumes that censoring may depend on $\mathbf{X}$ only through $\mathbf{B}^\top\mathbf{X}$, without requiring marginal independence between T and C. Under these conditions, $\mathcal{S}_{(Y,\delta)|\mathbf{X}} \subseteq \mathcal{S}_{(T,C)|\mathbf{X}} = \mathcal{S}_{T|\mathbf{X}}$, suggesting that the sufficient predictors for the observable regression $(Y, \delta)|\mathbf{X}$ are also sufficient for $T|\mathbf{X}$.

In bivariate SIR (Cook 2003), the sample is first separated according to δ, and then Y is sliced within each group. Let h_1 denote the number of slices among events ($\delta = 1$) and h_0 among censored observations ($\delta = 0$). The total number of slices is $H = h_0 + h_1$, where the choice of (h_0, h_1) balances the detail and robustness. After the slicing, the analysis proceeds in the spirit of SIR: $\mathbf{X}$ is standardized and the inverse conditional mean is computed within each slice. Leading eigenvectors of the SIR kernel matrix estimate $\mathcal{S}_{(Y,\delta)|\mathbf{X}}$, and under **(A1)** and **(A2)** (3), target $\mathcal{S}_{T|\mathbf{X}}$.

Although effective, bivariate SIR has two notable limitations. First, its estimation can be sensitive to the choice of (h_0, h_1). Yoo et al. (2016) suggest that this can be alleviated by transforming (Y, δ) into a single response, Y_{r} or Y_{m}. The details of this transformation are described in Section 3.2. Second, the method assumes that $\mathbf{X}$ satisfies the linearity condition, which may not hold for non-elliptical distribution. This motivates the use of informative predictor subspace (Yoo 2016) framework, which will be discussed in the following Section 2.2.

2.2 | Informative Predictor Subspace

An *informative predictor subspace* (IPS; Yoo 2016) was introduced to address the limitations of classical SDR methods that rely on linearity, constant variance and coverage conditions. The key idea is to define a subspace that contains the central subspace even when these conditions fail. The development in this subsection is stated in the general SDR setting and is not specific to survival regression. It is reviewed here because it provides the conceptual basis for the survival IPS.

Let $\boldsymbol{\Theta}$ be a $p \times q$ orthonormal basis matrix with $q \leq p$ for a dimension reduction subspace of a general response $Y|\mathbf{X}$. The IPS associated with $\boldsymbol{\Theta}$ is defined as

$$\mathcal{S}^{\mathrm{IPS}}_{\boldsymbol{\Theta}} = \boldsymbol{\Sigma}^{-1}\mathcal{S}\{E(\mathbf{X}|\boldsymbol{\Theta}^\top\mathbf{X}) - E(\mathbf{X})\} \tag{4}$$

where $\boldsymbol{\Sigma} = \mathrm{cov}(\mathbf{X})$. This subspace captures the variation of $\mathbf{X}$ through its conditional means given sufficient predictors $\boldsymbol{\Theta}^\top\mathbf{X}$, and it satisfies $\mathcal{S}_{Y|\mathbf{X}} \subseteq \mathcal{S}^{\mathrm{IPS}}_{\boldsymbol{\Theta}}$. Hence, $\mathcal{S}^{\mathrm{IPS}}_{\boldsymbol{\Theta}}$ serves as a broader target space that retains the information on $Y|\mathbf{X}$.

Key properties of $\mathcal{S}^{\text{IPS}}_{\boldsymbol{\Theta}}$ are notable.

- R.IPS1: For a $p\times p$ nonsingular matrix $\mathbf{A}$, $\mathbf{A}^{-1}\mathcal{S}^{\text{IPS}}_{\boldsymbol{\Theta}}$ is an IPS for $Y|\mathbf{A}^\top\mathbf{X}$.
- R.IPS2: Suppose both $\mathcal{S}(\boldsymbol{\Theta}_1)$ and $\mathcal{S}(\boldsymbol{\Theta}_2)$ are dimension reduction subspaces for $Y|\mathbf{X}$, and $\mathcal{S}^{\text{IPS}}_{\boldsymbol{\Theta}_1}$ and $\mathcal{S}^{\text{IPS}}_{\boldsymbol{\Theta}_2}$ are the corresponding IPSs. If $\mathcal{S}(\boldsymbol{\Theta}_1)\subseteq\mathcal{S}(\boldsymbol{\Theta}_2)$, then $\mathcal{S}^{\text{IPS}}_{\boldsymbol{\Theta}_1}\subseteq\mathcal{S}^{\text{IPS}}_{\boldsymbol{\Theta}_2}$ hold.

Let $\mathbf{Z}=\boldsymbol{\Sigma}^{-1/2}(\mathbf{X}-E(\mathbf{X}))$ and $\boldsymbol{\Theta}_{\mathbf{Z}}$ spans a dimension reduction subspace for $Y|\mathbf{Z}$. R.IPS1 implies that $\mathcal{S}^{\text{IPS}}_{\boldsymbol{\Theta}}=\boldsymbol{\Sigma}^{-1/2}\mathcal{S}^{\text{IPS}}_{\boldsymbol{\Theta}_{\mathbf{Z}}}$. So, $\mathbf{Z}$ can be used in its sample estimation, preserving all properties of $\mathcal{S}^{\text{IPS}}_{\boldsymbol{\Theta}}$. R.IPS2 indicates that the IPS for true basis $\boldsymbol{\eta}$, which spans $\mathcal{S}_{Y|\mathbf{X}}$, is minimal among all possible IPSs. It is called the *central IPS*, and denoted by $\mathcal{S}^{\text{IPS}}_{Y|\mathbf{X}}=\boldsymbol{\Sigma}^{-1}\mathcal{S}\{E(\mathbf{X}|\boldsymbol{\eta}^\top\mathbf{X})-E(\mathbf{X})\}$. When the linearity condition holds for $\boldsymbol{\eta}^\top\mathbf{X}$, then $\mathcal{S}^{\text{IPS}}_{Y|\mathbf{X}}=\mathcal{S}_{Y|\mathbf{X}}$. Thus, the primary objective of Yoo (2016) is to estimate $\mathcal{S}^{\text{IPS}}_{Y|\mathbf{X}}$, which contains $\mathcal{S}_{Y|\mathbf{X}}$ and can recover it when the linearity condition holds.

Having set the target to $\mathcal{S}^{\text{IPS}}_{Y|\mathbf{X}}$, estimating $E(\mathbf{X}|\boldsymbol{\eta}^\top\mathbf{X})=E[E(\mathbf{X}|Y,\ ,\boldsymbol{\eta}^\top\mathbf{X})|\boldsymbol{\eta}^\top\mathbf{X}]$ is crucial. However, because $\boldsymbol{\eta}$ is unknown, Yoo (2016) proposed two estimation methods to replace $\boldsymbol{\eta}^\top\mathbf{X}$ in a proper manner: clustering conditional mean (CCM) and *partially informative conditional mean* (PCM), both of which construct a suitable categorization of the unknown conditioning variable. Among them, PCM has shown empirical stability.

PCM constructs a partially informative basis $\widehat{\boldsymbol{\eta}}_{\mathbf{P}}=(\widehat{\boldsymbol{\eta}}_{\text{OLS}},\widehat{\boldsymbol{\eta}}_{\text{rpHd}})\in\mathbb{R}^{p\times 2}$ using OLS and residual-based principal Hessian directions (rpHd; Cook 1998). In practice, PCM categorizes $\widehat{\boldsymbol{\eta}}_{\mathbf{P}}^\top\mathbf{X}$ into h_c groups, and slices Y into h_y bins within each category. After that, calculate conditional means $\overline{\mathbf{Z}}_{c,h}$ within each slice, and construct the kernel matrix $\widehat{\mathbf{M}}=\sum_{c=1}^{h_c}\sum_{h=1}^{h_y}\frac{n_{c,h}}{n}\overline{\mathbf{Z}}_{c,h}\overline{\mathbf{Z}}_{c,h}^\top$ where $n_{c,h}$ is the sample size of each slice. The leading eigenvectors of $\widehat{\mathbf{M}}$ then provide the estimated IPS directions.

In short, the IPS framework is designed to contain the central subspace without requiring linearity or constant variance conditions. Building upon this, we extend the IPS concept to survival regression, where censoring further complicates the estimation. Details of survival IPS will be discussed in Section 3.

2.3 | Fused SIR in Survival Analysis

The idea of fusing kernel matrices across different slicing schemes was first proposed by Cook and Zhang (2014). Specifically, for a maximal slice number h, fused SIR defines

$$\mathbf{M}^{(h)}_{\text{FSIR}}=(\mathbf{M}^{(2)}_{\text{SIR}},\ldots,\mathbf{M}^{(h)}_{\text{SIR}})$$

where $\mathbf{M}^{(k)}_{\text{SIR}}$ denotes the kernel matrix of SIR with k slices. Because $\mathbf{M}^{(h)}_{\text{FSIR}}$ is a non-decreasing sequence of $\mathbf{M}^{(k)}_{\text{SIR}}$, it follows that $\mathcal{S}_{Y|\mathbf{Z}}=\mathcal{S}(\mathbf{M}^{(k)}_{\text{SIR}})\subseteq\mathcal{S}(\mathbf{M}^{(h)}_{\text{FSIR}})$ where $k=2,\ldots,h$. Under the usual identification assumption, $\boldsymbol{\Sigma}^{-1/2}\mathcal{S}(\mathbf{M}^{(h)}_{\text{FSIR}})=\mathcal{S}_{Y|\mathbf{X}}$. This fused SIR mitigates the sensitivity of SIR to the number of slices and achieves more stable estimates.

For survival regression, Yoo (2017) extended fused SIR by introducing two slicing orders for (Y,δ): **(i)** dividing δ into two groups and then slicing Y, **(ii)** slicing Y first and then dividing δ within each Y slice. Yoo (2017) also considered fusing across different slice numbers for (Y,δ), for example, when the maximal slice number is 10, the fused kernel incorporates results from slice numbers 4, 6, 8, and 10. Furthermore, Yoo (2019) applied fused SIR to the diffuse large–B-cell lymphoma (DLBCL) dataset (Rosenwald et al. 2002), demonstrating its empirical usefulness in survival prediction.

Finally, Um and Yoo (2020) proposed a *selective fusing* strategy for IPS estimation that combines homogeneous kernel families within the same estimation method, rather than mixing across different methods. In this framework, kernels derived from CCM are fused only with other CCM kernels, and likewise for PCM. Accordingly, it provides a theoretical justification for fusing PCM-type kernel matrices in our proposed approach.

3 | Non-Elliptical Dimension Reduction in Survival Regression

Our target is the *survival IPS*, defined through IPS with the true basis $\boldsymbol{\eta}$ of $\mathcal{S}_{T|\mathbf{X}}$. Under **(A1)** and **(A2)** (3), we have $\mathcal{S}_{(T,C)|\mathbf{X}}=\mathcal{S}_{T|\mathbf{X}}$. Following Cook (2003), we regard the observed data $(Y_i,\delta_i,\mathbf{X}_i),\ i=1,\ldots,n$, as being constructed from n i.i.d. realizations of $(T,C,\mathbf{X})$, with (Y,δ) defined as in Section 2.1. Because (Y,δ) is a measurable function of (T,C), it follows that $\mathcal{S}_{(Y,\delta)|\mathbf{X}}\subseteq\mathcal{S}_{(T,C)|\mathbf{X}}$. Thus, the observable regression $(Y,\delta)|\mathbf{X}$ cannot contain more dimension reduction information than the latent regression $(T,C)|\mathbf{X}$. Cook (2003) noted that equality is normally expected in practice, whereas proper containment would require carefully balanced conditions. A concrete example supporting this interpretation is given by Wen (2010): when the censoring indicator δ contributes no additional dimension reduction direction, $\mathcal{S}_{(Y,\delta)|\mathbf{X}}$ and $\mathcal{S}_{(T,C)|\mathbf{X}}$ coincide. Hereafter, we assume

$$\mathcal{S}_{(Y,\delta)|\mathbf{X}}=\mathcal{S}_{(T,C)|\mathbf{X}}=\mathcal{S}_{T|\mathbf{X}}. \tag{5}$$

With this assumption, the three formulations share a common target subspace, which motivates our use of the survival IPS.

Proposition 1. *Define the survival IPS by* $\mathcal{S}^{\text{IPS}}_{T|\mathbf{X}}=\boldsymbol{\Sigma}^{-1}\mathcal{S}\{E(\mathbf{X}|\boldsymbol{\eta}^\top\mathbf{X})-E(\mathbf{X})\}$, *i.e., the central IPS for* $T|\mathbf{X}$. *Then, at the population level,*

$$\mathcal{S}_{T|\mathbf{X}}\subseteq\mathcal{S}^{\text{IPS}}_{(Y,\delta)|\mathbf{X}}=\mathcal{S}^{\text{IPS}}_{(T,C)|\mathbf{X}}=\mathcal{S}^{\text{IPS}}_{T|\mathbf{X}}. \tag{6}$$

Proof. By the definition of IPS, $\mathcal{S}_{T|\mathbf{X}} \subseteq \mathcal{S}^{\text{IPS}}_{T|\mathbf{X}}$. Using the same $\boldsymbol{\eta}$ that spans $\mathcal{S}_{T|\mathbf{X}}$, the quantity $\boldsymbol{\Sigma}^{-1}\mathcal{S}\{E(\mathbf{X}|\boldsymbol{\eta}^{\mathrm{T}}\mathbf{X}) - E(\mathbf{X})\}$ does not depend on conditioning through T, (T, C) or (Y, δ). Indeed, by the law of iterated expectations, $E(\mathbf{X}|\boldsymbol{\eta}^{\mathrm{T}}\mathbf{X}) = E\left[E(\mathbf{X}|\boldsymbol{\eta}^{\mathrm{T}}\mathbf{X}, T, C)|\boldsymbol{\eta}^{\mathrm{T}}\mathbf{X}\right] = E\left[E(\mathbf{X}|\boldsymbol{\eta}^{\mathrm{T}}\mathbf{X}, Y, \delta)|\boldsymbol{\eta}^{\mathrm{T}}\mathbf{X}\right]$, so the resulting IPS is identical in all three cases. □

Therefore, under assumption (5), estimating $\mathcal{S}^{\text{IPS}}_{(Y,\delta)|\mathbf{X}}$ provides an identifiable route to covering the central subspace $\mathcal{S}_{T|\mathbf{X}} = \mathcal{S}(\boldsymbol{\eta})$. Moreover, we refer to this common IPS in Proposition 1 as the survival IPS, which motivates our *non-elliptical dimension reduction for survival regression*.

Although PCM for IPS estimation follows a two-step scheme—predictor categorization followed by response slicing—the second step must be modified in survival settings to the double slicing of (Y, δ). Under random censoring, however, some categories may contain few events or censored observations, resulting in small or even empty slices. To address these issues, the next subsections **(i)** induce more balanced predictor categories for $\mathbf{X}$ and **(ii)** use one-dimensional transformations of (Y, δ) instead of double-slicing.

3.1 | A Partially Informative Direction for Predictor Categorization

To estimate the survival IPS, a feasible and informative surrogate for the unknown conditioning variable $\boldsymbol{\eta}^{\mathrm{T}}\mathbf{X}$ is needed for predictor categorization. As in Yoo (2016), PCM constructs partially informative predictors $\widehat{\boldsymbol{\eta}}_{\mathbf{P}}^{\mathrm{T}}\mathbf{X}$ with $\widehat{\boldsymbol{\eta}}_{\mathbf{P}} = (\widehat{\boldsymbol{\eta}}_{\text{OLS}}, \widehat{\boldsymbol{\eta}}_{\text{rpHd}})$ (see Section 2.2). In survival regression, however, the direct use of $\widehat{\boldsymbol{\eta}}_{\mathbf{P}}^{\mathrm{T}}\mathbf{X}$ is problematic: OLS does not account for censoring, and rpHd is not directly applicable to the censored response (Y, δ).

To retain the spirit of partial informativeness in a survival setting, we use a Cox proportional hazards (CPH) direction as a feasible surrogate for predictor categorization under censoring. Let $\widehat{\boldsymbol{\beta}}_{\mathrm{c}} \in \mathbb{R}^{p\times 1}$ be the CPH coefficient estimated from $(Y, \delta)|\mathbf{X}$. We then categorize predictors by slicing the one-dimensional score $\widehat{\boldsymbol{\beta}}_{\mathrm{c}}^{\mathrm{T}}\mathbf{X}$ into h_X groups. The CPH direction is not intended to reproduce the specific roles of OLS and rpHd; rather, it is used to provide survival-relevant information for constructing predictor categories when the original partially informative basis is not directly available. This construction makes use of the semiparametric structure of the Cox model, which captures the relationship between T and $\mathbf{X}$ without assuming a baseline distribution.

Even if the proportional hazards assumption is misspecified, the CPH direction still serves as a partially informative surrogate for predictor categorization, leading to a loss of estimation accuracy rather than a failure of the procedure. In addition, slicing along $\widehat{\boldsymbol{\beta}}_{\mathrm{c}}^{\mathrm{T}}\mathbf{X}$ helps produce more balanced predictor groups under censoring, preventing the loss of entirely censored subsets and stabilizing the subsequent kernel estimation.

3.2 | Single Slicing of (Y, δ) Through Transformation

Even when predictor categories are formed by the CPH score, the double-slicing of (Y, δ) can leave some slices sparse or empty. To obtain more balanced slices, we transform the bivariate response into a single variable and then slice it into h_y bins. Datta et al. (2007) suggest two transformation methods for right-censored data: reweighting of uncensored observations and mean imputation for censored ones. Following Datta et al. (2007), these transformations are introduced under the log-time regression setting of an accelerated failure time (AFT) model, $\log T_i = \beta_0 + \mathbf{X}_i^{\mathrm{T}}\boldsymbol{\beta} + \sigma\varepsilon_i$, where β_0 is an intercept, $\boldsymbol{\beta}$ is a regression coefficient vector, and ε_i is an error term. The observed response is $Y_i = T_i\delta_i + C_i(1-\delta_i)$, so when $\delta_i = 0$, only C_i is observed and true T_i is unobserved.

Let $S(t) = \Pr(T > t)$ denote the survival function. By the Kaplan–Meier estimator,

$$\hat{S}(t) = \prod_{t_{(i)} \le t}\left\{1 - \frac{\Delta N(t_{(i)})}{R(t_{(i)})}\right\},$$

where $t_{(1)} < \cdots < t_{(m)}$ are ordered event times, $\Delta N(t_{(i)}) = \#\{j: Y_j = t_{(i)}, \delta_j = 1\}$ is event count at $t_{(i)}$, and $R(t_{(i)}) = \#\{j: Y_j \ge t_{(i)}\}$ is the risk set just before $t_{(i)}$.

3.2.1 | Reweighting

To correct the bias from censoring, rarer uncensored observations at larger times receive larger weights. To be specific, the uncensored observations are weighted by the inverse of censoring probability, whereas censored ones contribute 0 (Robins and Rotnitzky 1992; Robins and Finkelstein 2000; Satten and Datta 2001; Satten et al. 2001).

Let $S_c(t) = \Pr(C > t)$ denote the survival function of censoring time. By the Kaplan–Meier estimator,

$$\hat{S}_c(t) = \prod_{\tau_{(i)} \le t}\left\{1 - \frac{\Delta N_c(\tau_{(i)})}{R(\tau_{(i)})}\right\}, \tag{7}$$

where $\tau_{(1)} < \cdots < \tau_{(m_c)}$ are ordered censoring times and $\Delta N_c(\tau_{(i)}) = \#\{j: Y_j = \tau_{(i)}, \delta_j = 0\}$ is censoring count at $\tau_{(i)}$. By applying inverse weighting,

$$Y_{\mathrm{r},i} = \frac{\delta_i \log(Y_i)}{\hat{S}_c(Y_i -)},$$

where "–" denotes a left limit. Thus, uncensored observations ($\delta_i = 1$) are weighted by $1/\hat{S}_c(Y_i -)$, otherwise contribute 0. In the above log-time regression framework, under $C \perp\!\!\!\perp (T, \mathbf{X})$, $E(Y_{\mathrm{r},i}|\mathbf{X}) \approx E(\log T_i|\mathbf{X})$ (Koul et al. 1981; Datta et al. 2007). Thus, Y_{r} is used as a one-dimensional transformed response to facilitate stable slicing under censoring.

3.2.2 | Mean Imputation

Under right-censoring, the usual moment-based sample means are generally inconsistent. To address this, Datta (2005) propose imputing censored observations by their conditional expectation.

We use $\hat{S}(t)$ from the above notation. For censored observations ($\delta_i = 0$), compute

$$Y_i^* = \hat{S}(Y_i)^{-1} \sum_{t_{(j)} > Y_i} \log(t_{(j)}) \Delta \hat{S}(t_{(j)}), \tag{8}$$

where $\Delta \hat{S}(t_{(j)})$ is the jump size of $\hat{S}$ at time $t_{(j)}$. If the largest observed time is censored, we adopt the convention of treating it as an event to close the Kaplan–Meier tail, ensuring that the sum over $t_{(j)} > Y_i$ is well-defined. Finally, define the imputed response

$$Y_{\mathrm{m},i} = \delta_i \log Y_i + (1 - \delta_i) Y_i^*.$$

Under the same log-time regression framework, $E(Y_{\mathrm{m},i}|\mathbf{X}) \approx E(\log T_i|\mathbf{X})$. Thus, Y_{m} is also used as a one-dimensional transformed response for slicing under censoring.

We emphasize that the transformed responses Y_{r} and Y_{m} are introduced to obtain more balanced one-dimensional slicing variables under censoring. Because they are motivated through approximations to $E(\log T|\mathbf{X})$, their role is primarily mean-based on the log-time scale. Nevertheless, the corresponding mean structure is still contained in the central subspace of $T|\mathbf{X}$ (Cook and Li 2002). The transformed slicing step is used within the survival IPS framework, whose target is an upper bound of $\mathcal{S}_{T|\mathbf{X}}$. If a location regression holds for $\log T|\mathbf{X}$, then the corresponding central mean subspace coincides with the full central subspace; however, such a condition is not required for our development.

3.3 | Fused CPH With Transformation-Mean Method

Estimation based on a single choice of slice numbers can be unstable; heavy censoring often yields unbalanced slices even after transforming (Y, δ) to Y_{r} or Y_{m}. To reduce sensitivity to the number of slices, we fuse kernel information over multiple slicing schemes. We next describe the estimation target based on predictor and response slicing, the recommended slice-number choices and the fused kernel estimation used in the proposed *fused CPH with transformation-mean methods*.

Now, the proposed procedure is based on the following population-level quantity induced by a single response given the CPH direction:

$$\boldsymbol{\Sigma}^{-1} \mathcal{S}\{E(E(\mathbf{X}|Y_{\mathrm{r}},\ \boldsymbol{\beta}_{\mathrm{c}}^\top \mathbf{X})|\boldsymbol{\beta}_{\mathrm{c}}^\top \mathbf{X})\}$$
$$\text{or}\ \ \boldsymbol{\Sigma}^{-1} \mathcal{S}\{E(E(\mathbf{X}|Y_{\mathrm{m}},\ \boldsymbol{\beta}_{\mathrm{c}}^\top \mathbf{X})|\boldsymbol{\beta}_{\mathrm{c}}^\top \mathbf{X})\}.$$

Following the partial informativeness logic of Yoo (2016), this quantity is expected to serve as an upper bound of the survival IPS $\mathcal{S}_{T|\mathbf{X}}^{\mathrm{IPS}}$.

Unlike standard SDR, where slicing is mainly used to discretize the response, the slicing of $\hat{\boldsymbol{\beta}}_{\mathrm{c}}^\top \mathbf{X}$ here serves a different role. It provides the predictor categorization required in the spirit of PCM for IPS estimation, by using $\hat{\boldsymbol{\beta}}_{\mathrm{c}}^\top \mathbf{X}$ as a feasible and partially informative surrogate for the unknown conditioning variable $\boldsymbol{\eta}^\top \mathbf{X}$. Within each predictor group, slicing Y_{r} or Y_{m} then discretizes the transformed response. This is also useful under non-elliptical predictors, because predictor categorization can reveal conditional mean patterns that may be obscured if one relies on response slicing alone.

We recommend choosing h_X and h_y from the set $\{2, 3, 4\}$ rather than using arbitrary numbers of slices, where h_X denotes the number of slices for $\hat{\boldsymbol{\beta}}_{\mathrm{c}}^\top \mathbf{X}$ and h_y denotes the number of slices for Y_{r} or Y_{m}. Here, the total number of slices is $h_X \times h_y$. If arbitrary slice numbers are allowed, then unless the sample size is very large, over-categorization may produce sparse or empty slices and lead to severe imbalances. This in turn inflates the variability of the resulting kernel matrix estimation. For this reason, we exclude $(h_X, h_y) = (4, 4)$ and use $\mathcal{G} = \{(2,2), (2,3), (2,4), (3,2), (3,3), (3,4), (4,2), (4,3)\}$. This choice follows the practical recommendations for PCM and fused estimation in Yoo (2016) and Um and Yoo (2020).

Let $\tilde{X} \in \{1, \ldots, h_X\}$ denote the slice index from $\hat{\boldsymbol{\beta}}_{\mathrm{c}}^\top \mathbf{X}$, and within each predictor slice, let $\tilde{Y} \in \{1, \ldots, h_y\}$ denote the slice index from Y_{r} or Y_{m}. We use standardized predictors $\hat{\mathbf{Z}}_i = \hat{\boldsymbol{\Sigma}}^{-1/2}(\mathbf{X}_i - \overline{\mathbf{X}})$ where $\hat{\boldsymbol{\Sigma}}$ is the sample covariance matrix of $\mathbf{X}$ and $\overline{\mathbf{X}}$ is the sample mean vector. For a fixed $(h_X, h_y) \in \mathcal{G}$, we then form slice means $\overline{\mathbf{Z}}_{c,h} = \hat{E}(\hat{\mathbf{Z}}_i|\tilde{X} = c, \tilde{Y} = h)$ within each slice and $n_{c,h}$ denotes its sample size. The corresponding kernel matrix is the weighted sum of outer products of slice means over all $h_X \times h_y$ slices:

$$\widehat{\mathbf{M}}_{(h_X,h_y)} = \sum_{c=1}^{h_X} \sum_{h=1}^{h_y} \frac{n_{c,h}}{n} \overline{\mathbf{Z}}_{c,h} \overline{\mathbf{Z}}_{c,h}^\top.$$

To mitigate slice number sensitivity, we aggregate information across all $(h_X, h_y) \in \mathcal{G}$. The fused kernel matrix is

$$\widehat{\mathbf{M}}_{\mathrm{F}} = \sum_{(h_X,h_y) \in \mathcal{G}} \widehat{\mathbf{M}}_{(h_X,h_y)}.$$

Let $\hat{\boldsymbol{\gamma}}_1^d = \{\hat{\boldsymbol{\gamma}}_1, \ldots, \hat{\boldsymbol{\gamma}}_d\}$ be the eigenvectors corresponding to the first d largest eigenvalues of $\widehat{\mathbf{M}}_{\mathrm{F}}$. Finally, we back-transform the estimated directions to the original scale as $\hat{\boldsymbol{\eta}} = \hat{\boldsymbol{\Sigma}}^{-1/2} \hat{\boldsymbol{\gamma}}_1^d$, which spans the estimated survival IPS. The resulting $\hat{\boldsymbol{\eta}}^\top \mathbf{X}$ serve as sufficient predictors for subsequent analysis.

We consider three variants that differ only in the response transform: **(i)** fused CPH with reweighting transformation-mean method (fC-rtm) with Y_{r}; **(ii)** fused CPH with mean imputation transformation-mean method (fC-mtm) with Y_{m}; **(iii)** fused CPH with all transformation-mean method (fC-atm),

which is obtained by summing two fused kernel matrices from **(i)** and **(ii)**.

For clarity and interpretability, we focus on $d = 1$ throughout the paper, corresponding to a single-index reduction. The CPH coefficient $\widehat{\boldsymbol{\beta}}_c$ is used to guide slicing under censoring, and the proposed fused estimator can be extended to $d > 1$ by extracting the leading d eigenvectors of the fused kernel matrix. As a simple adequacy check for $d = 1$, let $\widehat{\lambda}_1 \geq \widehat{\lambda}_2 \geq \cdots$ denote the eigenvalues of the fused kernel matrix. A pronounced drop after the first eigenvalue, such as a large value of $\widehat{\lambda}_1/\widehat{\lambda}_2$, suggests that a one-dimensional structure is reasonable.

Algorithm 1. Fused Cox PH with transformation-mean method.

1. Fit a CPH model to obtain $\widehat{\boldsymbol{\beta}}_c$ and construct $\widehat{\boldsymbol{\beta}}_c^\top \mathbf{X}$.
2. Transform (Y, δ) to Y_r or Y_m.
3. For each choice of $h_X \in \{2, 3, 4\}$, slice $\widehat{\boldsymbol{\beta}}_c^\top \mathbf{X}$ into h_X slices to define $\tilde{X}$.
4. For each $\tilde{X}$ and corresponding choice of $h_y \in \{2, 3, 4\}$, slice Y_r or Y_m into h_y slices to define $\tilde{Y}$.
5. Obtain $\widehat{\mathbf{Z}}_i = \widehat{\boldsymbol{\Sigma}}^{-1/2}(\mathbf{X}_i - \overline{\mathbf{X}})$ for $i = 1, \ldots, n$.
6. Compute $\overline{\mathbf{Z}}_{c,h} = \widehat{E}\,(\widehat{\mathbf{Z}}_i | \tilde{X} = c, \tilde{Y} = h)$ within each slice.
7. Calculate the kernel matrix for each pair of (h_X, h_y):

$$\widehat{\mathbf{M}}_{(h_X,h_y)} = \sum_{c=1}^{h_X} \sum_{h=1}^{h_y} \frac{n_{c,h}}{n} \overline{\mathbf{Z}}_{c,h} \overline{\mathbf{Z}}_{c,h}^\top.$$

8. Form the fused kernel matrix

$$\widehat{\mathbf{M}}_F = \sum_{(h_X,h_y) \in \mathcal{G}} \widehat{\mathbf{M}}_{(h_X,h_y)}.$$

9. Let $\widehat{\boldsymbol{\gamma}}_1^d = \{\widehat{\boldsymbol{\gamma}}_1, \ldots, \widehat{\boldsymbol{\gamma}}_d\}$ be the first d eigenvectors of $\widehat{\mathbf{M}}_F$.
10. Set $\widehat{\boldsymbol{\eta}} = \widehat{\boldsymbol{\Sigma}}^{-1/2} \widehat{\boldsymbol{\gamma}}_1^d$, which spans the estimated survival IPS.
11. (fC-atm only) Run Steps 2–8 with Y_r and Y_m, respectively; sum the two resulting $\widehat{\mathbf{M}}_F$ and then apply Steps 9–10.

4 | Numerical Studies and Data Analysis

4.1 | Numerical Studies

We consider two versions of the predictor and three survival regressions based on the AFT model. Predictors are designed to compare whether the linearity or coverage conditions are satisfied or not. For the elliptical predictor, $\mathbf{X}_E = (X_1, \ldots, X_p)^\top$ with $(X_1, \ldots, X_p) \overset{iid}{\sim} N(0, 1)$. For the non-elliptical predictor, which was inspired by Yoo (2016), $\mathbf{X}_{Ne} = (X_1, \ldots, X_p)^\top$ is constructed by $U_1 \sim \text{Unif}(0, 1)$, $e \sim \text{Unif}(-0.5, 0.5)$, $U_2 = \log(U_1) + e$, $(W_1, W_2, W_3) \overset{iid}{\sim} N(0, 1)$, and $X_1 = U_1 + U_2$, $X_2 = U_2 + 0.7\ W_2 + 0.3\ W_3$, $X_3 = W_1$, $X_4 = U_1 + W_2$, $X_5 = W_3 - W_2$, $(X_6, \ldots, X_{10}) \overset{iid}{\sim} \exp(1)$, $(X_{11}, \ldots, X_p) \overset{iid}{\sim} N(0, 1)$, with the convention that empty index ranges are omitted. For each predictor design, we assess ellipticity via the Mahalanobis QQ plot (Mardia 1970): $D_i^2 = (\mathbf{X}_i - \overline{\mathbf{X}})^\top \widehat{\boldsymbol{\Sigma}}^{-1} (\mathbf{X}_i - \overline{\mathbf{X}})$ versus χ_p^2 quantiles. Figure 1 shows approximate linear alignment for $\mathbf{X}_E$, whereas $\mathbf{X}_{Ne}$ exhibits a clear lack of ellipticity.

With this predictor setting, we generate (Y, δ) from the following three models. Models I and II follow log-normal structures and Model III follows a Weibull structure. For Models II and III, we introduce nonlinearity between T and $\mathbf{X}$ through trigonometric products.

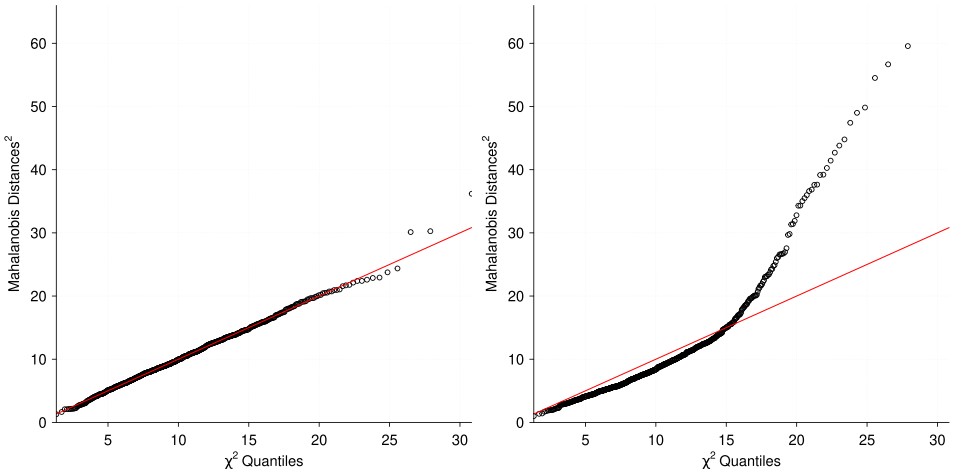
$$\textbf{Model I}: \ T|\mathbf{X} \overset{iid}{\sim} \exp\{2\boldsymbol{\eta}_1^\top \mathbf{X} + 0.1\varepsilon\},\ \varepsilon \sim N(0, 1),$$

$$\textbf{Model II}: \ T|\mathbf{X} \overset{iid}{\sim} \exp\{f(\boldsymbol{\eta}_1^\top \mathbf{X}) + 0.1\varepsilon\},\ \varepsilon \sim N(0, 1),$$

$$f(\boldsymbol{\eta}_1^\top \mathbf{X}) = \sin(\boldsymbol{\eta}_1^\top \mathbf{X}) + \cos(\boldsymbol{\eta}_1^\top \mathbf{X}),$$

$$\textbf{Model III}: \ T|\mathbf{X} \overset{iid}{\sim} \text{Weibull}(2, \exp\{f(\boldsymbol{\eta}_2^\top \mathbf{X})\}),$$

$$f(\boldsymbol{\eta}_2^\top \mathbf{X}) = \sin(\boldsymbol{\eta}_2^\top \mathbf{X}) + \cos(\boldsymbol{\eta}_2^\top \mathbf{X}),$$



(a) $\mathbf{X}_E$ (b) $\mathbf{X}_{Ne}$

FIGURE 1 | Mahalanobis QQ plots (Mardia 1970) comparing elliptical predictor $\mathbf{X}_E$ and non-elliptical predictor $\mathbf{X}_{Ne}$.

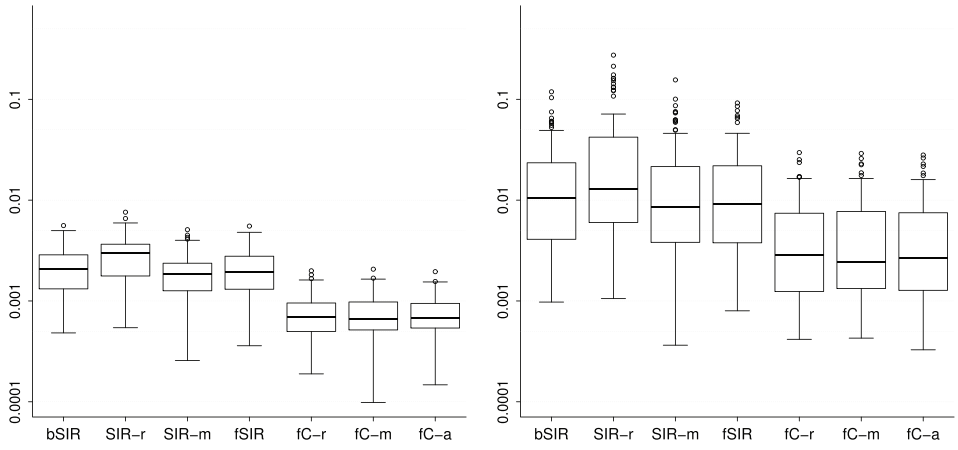
where a censoring variable C is also generated from log-normal distribution $\exp\{N(c_0, 2)\}$ so that $C \perp\!\!\!\perp (T, \mathbf{X})$, $Y = T\delta + C(1-\delta)$ and $\delta = I(T < C)$.

The true directions are $\boldsymbol{\eta}_1 = \left(1/\sqrt{10}\right)(2, -1, 0, -1, 2, 0, \ldots, 0)^\top$ and $\boldsymbol{\eta}_2 = \left(1/\sqrt{5}\right)(1, -1, -1, 1, 1, 0, \ldots, 0)^\top$. Constant c_0 from C is tuned to achieve target censoring rates. Unless otherwise noted, we set the sample size and predictor dimension as $(n, p) = (800, 10)$, the censoring rate at 50% and the structural dimension to $d = 1$.

The three proposed fused CPH methods—fC-rtm, fC-mtm and fC-atm—are applied for each model. Their performances are compared with bivariate SIR (Cook 2003), SIR with transformed response Y_r or Y_m (SIR-rtm and SIR-mtm; Yoo et al. 2016), and fused SIR for survival regression (Yoo 2017). In our method, we consider a maximum of 12 slices obtained from $\mathcal{G} = \{(2,2), (2,3), (2,4), (3,2), (3,3), (3,4), (4,2), (4,3)\}$ and aggregate the information. For fair comparison, we set the number of slices to 12 for bivariate SIR, SIR-rtm and SIR-mtm. For fused SIR, we fuse the two kernel matrices from Section 2.3, **(i)** dividing δ and then slicing Y; **(ii)** slicing Y and then dividing δ. We also set the maximal slice number for (Y, δ) to 12 so that the results from slice numbers 4, 6, 8, 10 and 12 are incorporated.

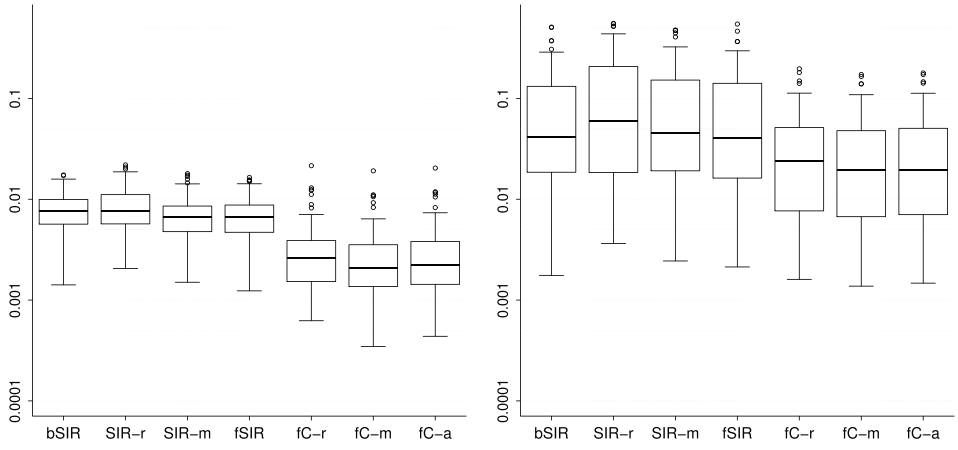
To measure how well the true $\boldsymbol{\eta}$ is estimated, we consider trace distance r_D from trace correlation (Hooper 1959): $r(\hat{\boldsymbol{\eta}}, \boldsymbol{\eta}) = \sqrt{\text{trace}\left(\hat{\boldsymbol{\eta}}(\hat{\boldsymbol{\eta}}^\top\hat{\boldsymbol{\eta}})^{-1}\hat{\boldsymbol{\eta}}^\top\boldsymbol{\eta}(\boldsymbol{\eta}^\top\boldsymbol{\eta})^{-1}\boldsymbol{\eta}^\top\right)}$, where $r_D(\hat{\boldsymbol{\eta}}, \boldsymbol{\eta}) = 1 - r(\hat{\boldsymbol{\eta}}, \boldsymbol{\eta})$. Smaller r_D implies better estimation and ranges from 0 to 1.

Figures 2–4 show boxplots of $r_D(\hat{\boldsymbol{\eta}}, \boldsymbol{\eta})$ for Models I–III over 100 replicates, and the corresponding median values are reported in Table 1. Across all scenarios, the fused CPH family attains



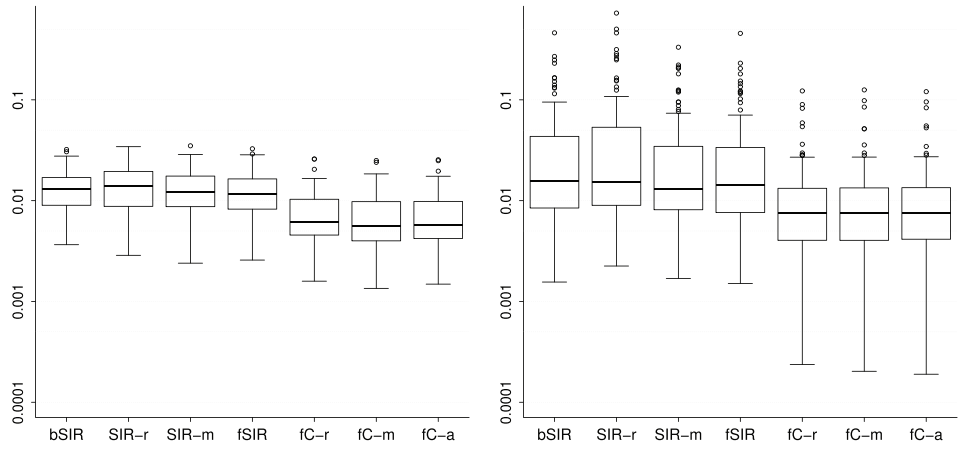
(a) Model I with $\mathbf{X}_E$ (b) Model I with $\mathbf{X}_{Ne}$

FIGURE 2 | Boxplots of trace distance $r_D(\hat{\boldsymbol{\eta}}, \boldsymbol{\eta})$ for Model I under $(n, p) = (800, 10)$ and 50% censoring. The y-axis is shown on a logarithmic scale. Method: bSIR = bivariate SIR (Cook 2003); SIR-r = SIR-rtm (Yoo et al. 2016); SIR-m = SIR-mtm (Yoo et al. 2016); fSIR = fused SIR (Yoo 2017); fC-r = fC-rtm; fC-m = fC-mtm; fC-a = fC-atm.



(a) Model II with $\mathbf{X}_E$ (b) Model II with $\mathbf{X}_{Ne}$

FIGURE 3 | Boxplots of trace distance $r_D(\hat{\boldsymbol{\eta}}, \boldsymbol{\eta})$ for Model II under $(n, p) = (800, 10)$ and 50% censoring. The y-axis is shown on a logarithmic scale. Method: bSIR = bivariate SIR (Cook 2003); SIR-r = SIR-rtm (Yoo et al. 2016); SIR-m = SIR-mtm (Yoo et al. 2016); fSIR = fused SIR (Yoo 2017); fC-r = fC-rtm; fC-m = fC-mtm; fC-a = fC-atm.



(a) Model III with $\mathbf{X}_{\mathrm{E}}$ (b) Model III with $\mathbf{X}_{\mathrm{Ne}}$

FIGURE 4 | Boxplots of trace distance $r_D(\widehat{\boldsymbol{\eta}}, \boldsymbol{\eta})$ for Model III under $(n, p) = (800, 10)$ and 50% censoring. The y-axis is shown on a logarithmic scale. Method: bSIR = bivariate SIR (Cook 2003); SIR-r = SIR-rtm (Yoo et al. 2016); SIR-m = SIR-mtm (Yoo et al. 2016); fSIR = fused SIR (Yoo 2017); fC-r = fC-rtm; fC-m = fC-mtm; fC-a = fC-atm.

TABLE 1 | Median trace distance $r_D(\widehat{\boldsymbol{\eta}}, \boldsymbol{\eta})$ for Models I–III under $(n, p) = (800, 10)$ and 50% censoring, corresponding to Figures 2–4. Method: bSIR = bivariate SIR (Cook 2003); SIR-r = SIR-rtm (Yoo et al. 2016); SIR-m = SIR-mtm (Yoo et al. 2016); fSIR = fused SIR (Yoo 2017); fC-r = fC-rtm; fC-m = fC-mtm; fC-a = fC-atm.

	$\mathbf{X}_{\mathrm{E}}$							$\mathbf{X}_{\mathrm{Ne}}$						
	bSIR	**SIR-r**	**SIR-m**	**fSIR**	**fC-r**	**fC-m**	**fC-a**	**bSIR**	**SIR-r**	**SIR-m**	**fSIR**	**fC-r**	**fC-m**	**fC-a**
Model I	0.0021	0.0030	0.0018	0.0019	0.0007	0.0007	0.0007	0.0106	0.0129	0.0085	0.0092	0.0028	0.0024	0.0027
Model II	0.0076	0.0077	0.0066	0.0067	0.0026	0.0021	0.0022	0.0413	0.0599	0.0451	0.0405	0.0239	0.0196	0.0197
Model III	0.0130	0.0141	0.0122	0.0117	0.0061	0.0056	0.0057	0.0156	0.0153	0.0131	0.0143	0.0076	0.0076	0.0076

the smallest medians and the tightest interquartile ranges of r_D. Under the elliptical designs (Figures 2a–4a), all methods estimate $\boldsymbol{\eta}$ similarly, yet the fused approaches are less dependent on the choice of slice numbers and provide slightly better estimates. In contrast, under the non-elliptical designs (Figures 2b–4b), SIR-based methods become highly variable as the linearity condition fails, leading to wide ranges of r_D, whereas the fused CPH variants remain stable. SIR-rtm shows the largest variability in Figure 3b; this is expected because the reweighting transformation assigns zero values to censored observations, leading to unbalanced slice sizes. In the same settings, fC-rtm remains stable by jointly slicing $\widehat{\boldsymbol{\beta}}_c^{\top}\mathbf{X}$ and Y_{r}, which helps balance slice sizes.

To assess sensitivity, we vary $(n, p) \in \{500, 800, 1100\} \times \{5, 10, 15, 20\}$ and tune c_0 to target censoring rates of 30%, 50% and 70%. To examine a small sample setting where n and p are relatively close, we additionally report $n = 100$ for 30% and 50% censoring. Tables 2–4 summarize the mean trace distance of fC-atm for Models I–III over 100 replications. With $\mathbf{X}_{\mathrm{Ne}}$, mean trace distance increases relative to the elliptical cases, but the degradation is modest; for $n \geq 500$, the corresponding trace correlation typically exceeds 0.95. Figure 5 further shows the trends of r_D for Model II over the full censoring grid (10%, 30%, 50%, 70%, 90%): larger n uniformly reduces r_D, heavier censoring inflates r_D although notable at 90%. Overall, fC-atm remains stable across (n, p) settings and censoring rates.

4.2 | Data Analysis: DLBCL Data

We analyse the DLBCL (DLBCL; Rosenwald et al. 2002) cDNA microarray dataset of `exprLym`, which contains 3581 genes measured from 190 patients, available in `Biobase` R package (Huber et al. 2015). For each patient, the survival time ranges from 0 to 21.8 years, with an overall censoring rate of 41.05%. Yoo (2019) demonstrated that fused SIR effectively captures gene expression structures on this data, supporting its validity for survival SDR. For preprocessing, we (i) discard genes with more than 10% missing values, (ii) impute remaining gene-wise missings by the gene mean and (iii) standardize each gene to zero mean and unit variance.

We adopt threefold cross-validation with training size $n_{\mathrm{tr}} = 127$ and test size $n_{\mathrm{te}} = 63$ per fold and implement a two-stage reduction. Because p is much larger than n, we first compute the leading k principal component (PC) scores from the training set. Let $\widehat{\boldsymbol{\Omega}}_k \in \mathbb{R}^{3581 \times k}$ denote the rotation matrix. We form $\widehat{\boldsymbol{\Omega}}_k^{\top}\mathbf{X}_{\mathrm{tr}}$ and project the test set as $\widehat{\boldsymbol{\Omega}}_k^{\top}\mathbf{X}_{\mathrm{te}}$. To examine how the initial PC re-

TABLE 2 | Mean trace distance over 100 replications (standard deviation in parentheses) of fC-atm for Model I across different n, p and censoring rates.

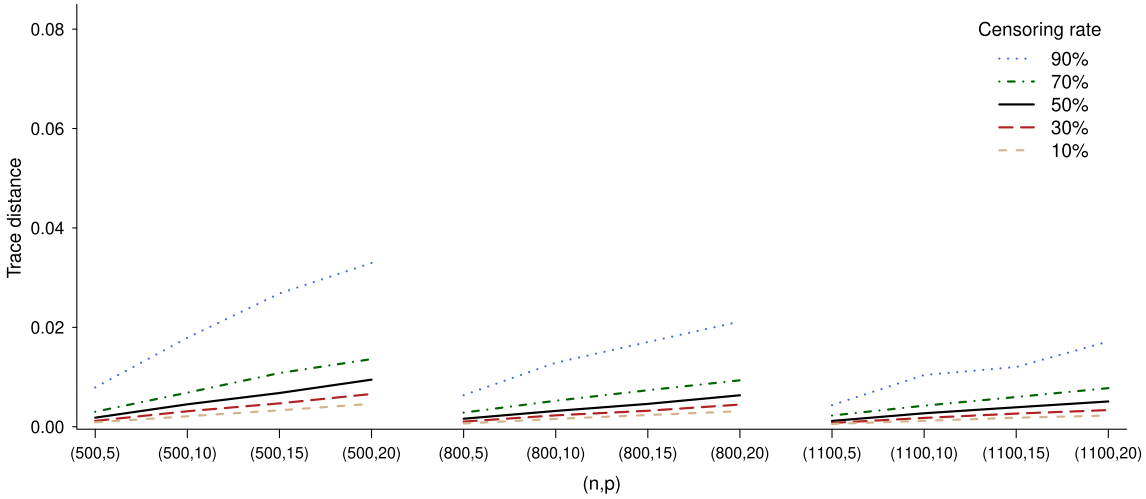
	Model I with X_E				Model I with X_{Ne}			
n	p=5	p=10	p=15	p=20	p=5	p=10	p=15	p=20
				30% censoring				
100	0.0025 (0.0018)	0.0059 (0.0033)	0.0097 (0.0048)	0.0143 (0.0064)	0.0318 (0.0475)	0.0403 (0.0508)	0.0425 (0.0485)	0.0485 (0.0585)
500	0.0004 (0.0004)	0.0010 (0.0005)	0.0016 (0.0007)	0.0021 (0.0007)	0.0057 (0.0060)	0.0063 (0.0066)	0.0074 (0.0087)	0.0073 (0.0072)
800	0.0003 (0.0002)	0.0006 (0.0003)	0.0009 (0.0003)	0.0013 (0.0005)	0.0041 (0.0049)	0.0044 (0.0045)	0.0047 (0.0044)	0.0054 (0.0053)
1100	0.0002 (0.0002)	0.0004 (0.0002)	0.0007 (0.0003)	0.0009 (0.0003)	0.0023 (0.0037)	0.0026 (0.0035)	0.0027 (0.0033)	0.0031 (0.0031)
				50% censoring				
100	0.0030 (0.0021)	0.0077 (0.0042)	0.0120 (0.0058)	0.0180 (0.0074)	0.0411 (0.0592)	0.0470 (0.0590)	0.0525 (0.0628)	0.0546 (0.0631)
500	0.0005 (0.0004)	0.0012 (0.0007)	0.0019 (0.0008)	0.0026 (0.0009)	0.0066 (0.0084)	0.0074 (0.0090)	0.0085 (0.0090)	0.0084 (0.0085)
800	0.0003 (0.0002)	0.0008 (0.0004)	0.0011 (0.0004)	0.0016 (0.0006)	0.0047 (0.0055)	0.0052 (0.0059)	0.0055 (0.0053)	0.0056 (0.0050)
1100	0.0002 (0.0002)	0.0006 (0.0003)	0.0008 (0.0003)	0.0011 (0.0004)	0.0030 (0.0042)	0.0033 (0.0037)	0.0031 (0.0035)	0.0037 (0.0036)
				70% censoring				
500	0.0006 (0.0004)	0.0014 (0.0007)	0.0022 (0.0008)	0.0031 (0.0010)	0.0076 (0.0092)	0.0073 (0.0091)	0.0079 (0.0079)	0.0089 (0.0080)
800	0.0004 (0.0003)	0.0009 (0.0003)	0.0013 (0.0005)	0.0018 (0.0007)	0.0056 (0.0060)	0.0063 (0.0066)	0.0065 (0.0062)	0.0066 (0.0064)
1100	0.0003 (0.0002)	0.0006 (0.0003)	0.0009 (0.0004)	0.0013 (0.0004)	0.0037 (0.0050)	0.0031 (0.0034)	0.0038 (0.0047)	0.0043 (0.0043)

TABLE 3 | Mean trace distance over 100 replications (standard deviation in parentheses) of fC-atm for Model II across different n, p and censoring rates.

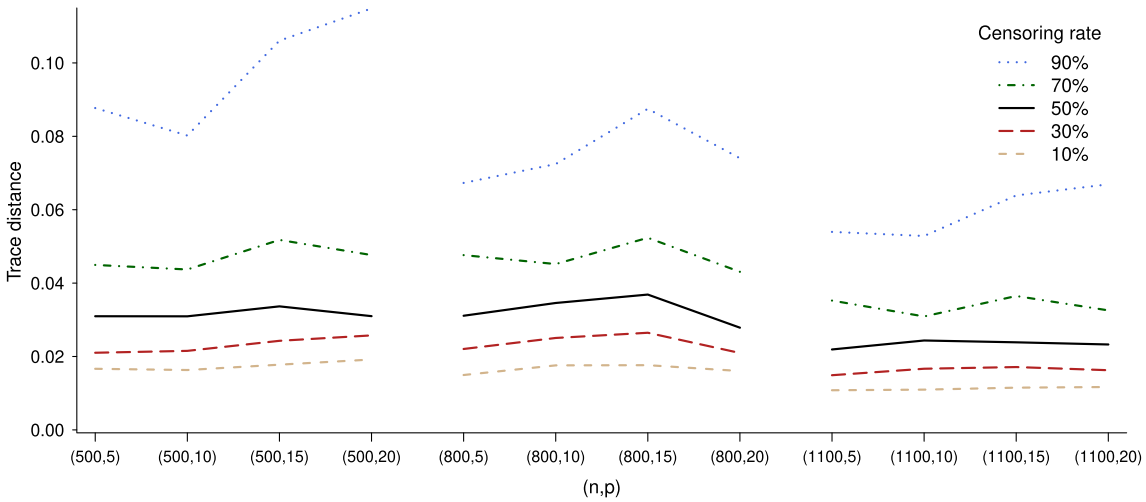
	Model II with X_E				Model II with X_{Ne}			
n	p=5	p=10	p=15	p=20	p=5	p=10	p=15	p=20
				30% censoring				
100	0.0060 (0.0059)	0.0151 (0.0112)	0.0260 (0.0149)	0.0372 (0.0217)	0.0885 (0.1238)	0.0941 (0.1349)	0.1143 (0.1238)	0.1153 (0.1253)
500	0.0012 (0.0013)	0.0031 (0.0026)	0.0047 (0.0036)	0.0066 (0.0049)	0.0210 (0.0281)	0.0215 (0.0228)	0.0243 (0.0273)	0.0258 (0.0275)
800	0.0010 (0.0013)	0.0023 (0.0020)	0.0032 (0.0023)	0.0045 (0.0028)	0.0221 (0.0275)	0.0251 (0.0314)	0.0265 (0.0318)	0.0210 (0.0209)
1100	0.0008 (0.0010)	0.0018 (0.0017)	0.0026 (0.0024)	0.0034 (0.0025)	0.0149 (0.0267)	0.0167 (0.0277)	0.0171 (0.0280)	0.0163 (0.0222)
				50% censoring				
100	0.0088 (0.0094)	0.0211 (0.0192)	0.0398 (0.0289)	0.0620 (0.0520)	0.1073 (0.1373)	0.1086 (0.1423)	0.1552 (0.1694)	0.1824 (0.1850)
500	0.0018 (0.0022)	0.0045 (0.0051)	0.0068 (0.0054)	0.0095 (0.0083)	0.0310 (0.0447)	0.0309 (0.0309)	0.0337 (0.0367)	0.0310 (0.0308)
800	0.0016 (0.0018)	0.0032 (0.0029)	0.0046 (0.0029)	0.0063 (0.0044)	0.0311 (0.0390)	0.0346 (0.0374)	0.0369 (0.0416)	0.0279 (0.0284)
1100	0.0012 (0.0014)	0.0027 (0.0029)	0.0039 (0.0037)	0.0051 (0.0040)	0.0219 (0.0399)	0.0244 (0.0355)	0.0239 (0.0334)	0.0233 (0.0280)
				70% censoring				
500	0.0030 (0.0037)	0.0069 (0.0069)	0.0108 (0.0100)	0.0136 (0.0108)	0.0450 (0.0653)	0.0437 (0.0486)	0.0518 (0.0537)	0.0476 (0.0438)
800	0.0028 (0.0039)	0.0052 (0.0050)	0.0073 (0.0051)	0.0093 (0.0063)	0.0476 (0.0587)	0.0452 (0.0428)	0.0524 (0.0539)	0.0431 (0.0421)
1100	0.0023 (0.0032)	0.0042 (0.0047)	0.0060 (0.0054)	0.0078 (0.0068)	0.0352 (0.0479)	0.0309 (0.0383)	0.0365 (0.0427)	0.0326 (0.0390)

TABLE 4 | Mean trace distance over 100 replications (standard deviation in parentheses) of fC-atm for Model III across different n, p and censoring rates.

	Model III with X_E				Model III with X_{Ne}			
n	***p*=5**	***p*=10**	***p*=15**	***p*=20**	***p*=5**	***p*=10**	***p*=15**	***p*=20**
				30% censoring				
100	0.0270 (0.0264)	0.0617 (0.0407)	0.1016 (0.0652)	0.1385 (0.0703)	0.1387 (0.2475)	0.1713 (0.2264)	0.2379 (0.2836)	0.1745 (0.2249)
500	0.0042 (0.0035)	0.0099 (0.0049)	0.0143 (0.0057)	0.0193 (0.0079)	0.0138 (0.0271)	0.0163 (0.0204)	0.0210 (0.0485)	0.0224 (0.0262)
800	0.0026 (0.0022)	0.0058 (0.0037)	0.0090 (0.0042)	0.0122 (0.0053)	0.0066 (0.0078)	0.0094 (0.0120)	0.0125 (0.0168)	0.0123 (0.0120)
1100	0.0019 (0.0014)	0.0043 (0.0030)	0.0066 (0.0034)	0.0090 (0.0034)	0.0049 (0.0068)	0.0084 (0.0172)	0.0083 (0.0083)	0.0102 (0.0112)
				50% censoring				
100	0.0356 (0.0359)	0.0830 (0.0600)	0.1310 (0.0821)	0.1770 (0.0929)	0.1709 (0.2735)	0.1688 (0.2166)	0.2790 (0.2897)	0.2451 (0.2676)
500	0.0056 (0.0050)	0.0124 (0.0065)	0.0186 (0.0075)	0.0253 (0.0110)	0.0172 (0.0318)	0.0225 (0.0293)	0.0259 (0.0474)	0.0291 (0.0344)
800	0.0034 (0.0028)	0.0076 (0.0046)	0.0122 (0.0054)	0.0161 (0.0072)	0.0088 (0.0150)	0.0129 (0.0185)	0.0188 (0.0570)	0.0146 (0.0119)
1100	0.0023 (0.0019)	0.0059 (0.0046)	0.0090 (0.0053)	0.0121 (0.0047)	0.0080 (0.0150)	0.0097 (0.0164)	0.0113 (0.0128)	0.0121 (0.0133)
				70% censoring				
500	0.0084 (0.0067)	0.0180 (0.0110)	0.0276 (0.0144)	0.0387 (0.0179)	0.0324 (0.0638)	0.0317 (0.0486)	0.0398 (0.0578)	0.0425 (0.0504)
800	0.0050 (0.0041)	0.0115 (0.0074)	0.0175 (0.0080)	0.0246 (0.0105)	0.0127 (0.0222)	0.0226 (0.0281)	0.0266 (0.0534)	0.0292 (0.0319)
1100	0.0037 (0.0029)	0.0087 (0.0060)	0.0135 (0.0073)	0.0183 (0.0088)	0.0128 (0.0259)	0.0147 (0.0230)	0.0175 (0.0211)	0.0184 (0.0247)



(a) Model II with $\mathbf{X}_{\mathrm{E}}$



(b) Model II with $\mathbf{X}_{\mathrm{Ne}}$

FIGURE 5 | Mean trace distance of fC-atm for Model II across different (n, p) settings, at censoring rates 10%, 30%, 50%, 70% and 90%.

duction affects predictive performance, we report results for $k \in \{20, 40, 100\}$. The choice $k = 40$ is particularly motivated by Li and Li (2004), who analysed the same DLBCL data and reported no strong violation of the linearity condition in the reduced predictors with 40 PCs.

For each k, we compare four one-dimensional approaches: full CPH, bivariate SIR, fused SIR and fC-atm, where full CPH denotes the Cox model fitted on all k PC scores $\widehat{\mathbf{\Omega}}_k^{\top}\mathbf{X}_{\mathrm{tr}}$. Performance is assessed by time-dependent areas under the ROC curves, evaluated over follow-up times from 1 to 10 years on both training and test folds.

Figures 6–8 report the time-dependent areas under the ROC curves for $k = 20$, 40 and 100, respectively. On the training folds, full CPH generally performs well across the three choices of k, which is expected because it is fitted directly on all k PC scores. On the test folds, fC-atm shows the best out-of-sample discrimination at $k = 20$ in Figure 6b, whereas both full CPH and fused SIR are lower. This suggests that, for smaller k, non-elliptical features may remain more pronounced in the reduced predictors, so the practical advantage of our proposed method becomes clearer. At $k = 40$ in Figure 7b, the predictive performances of full CPH, fused SIR and fC-atm become closer, with fC-atm remaining slightly better. This is consistent with the analysis of Li and Li (2004) based on 40 PCs, a setting more favourable to fused SIR. At $k = 100$ in Figure 8b, fused SIR gives the best out-of-sample performance, whereas fC-atm and full CPH remain broadly comparable. Across all three choices of k, bivariate SIR is generally less competitive than the other methods on both training and test folds. Thus, the data example illustrates that the proposed method offers a practical advantage when the linearity condition is less plausible in the reduced predictors while still showing robust performance as k increases.

5 | Discussion

In this paper, we addressed SDR for survival regression under non-elliptical predictors. Our target was the survival IPS, which contains the central subspace without requiring the linearity, constant variance or coverage conditions.

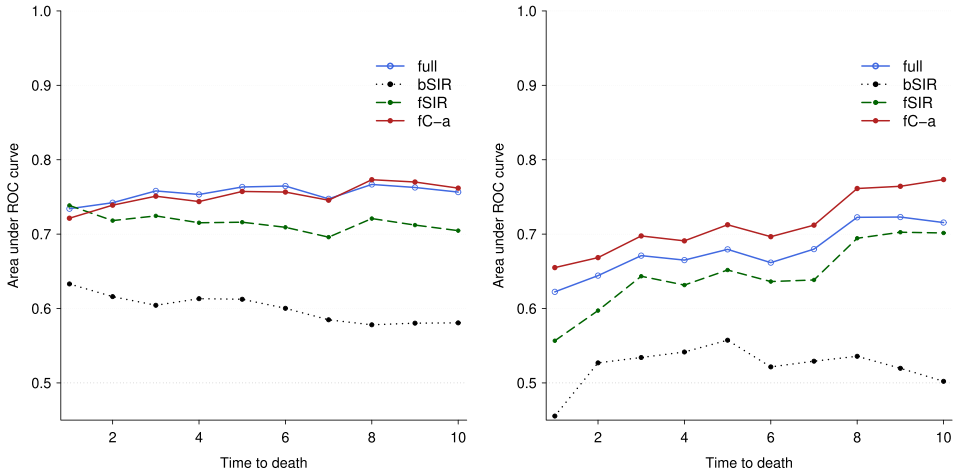


FIGURE 6 | For the DLBCL dataset, time-dependent areas under the ROC curves for $k = 20$ in the training and test folds. Method: full = full CPH; bSIR = bivariate SIR (Cook 2003); fSIR = fused SIR (Yoo 2017); fC-a = fC-atm.

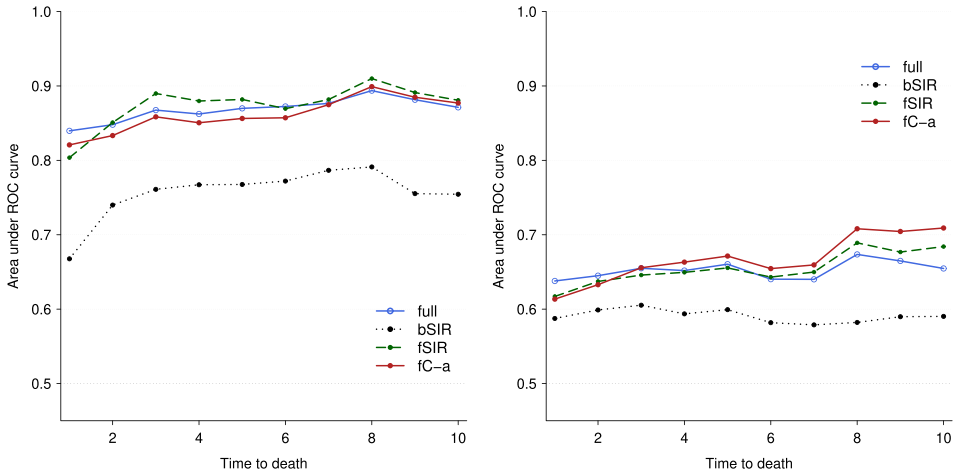


FIGURE 7 | For the DLBCL dataset, time-dependent areas under the ROC curves for $k = 40$ in the training and test folds. Method: full = full CPH; bSIR = bivariate SIR (Cook 2003); fSIR = fused SIR (Yoo 2017); fC-a = fC-atm.

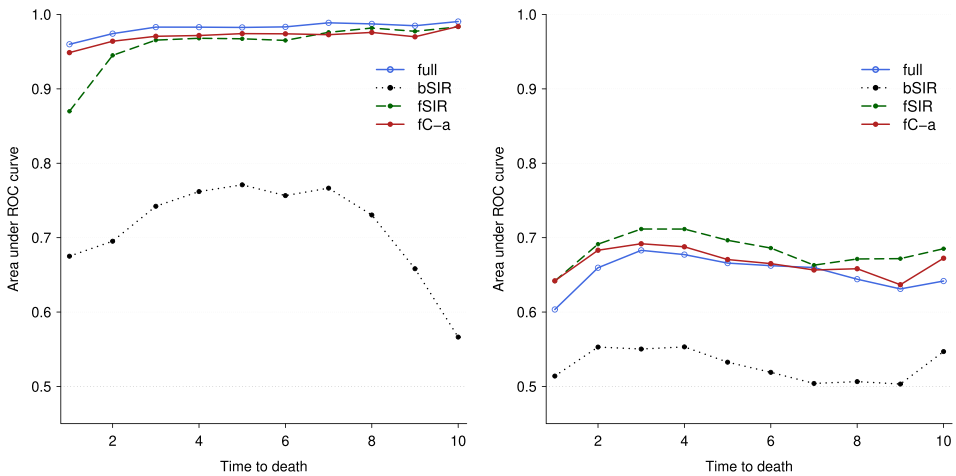


FIGURE 8 | For the DLBCL dataset, time-dependent areas under the ROC curves for $k = 100$ in the training and test folds. Method: full = full CPH; bSIR = bivariate SIR (Cook 2003); fSIR = fused SIR (Yoo 2017); fC-a = fC-atm.

We proposed three estimators—fC-rtm, fC-mtm and fC-atm—that **(i)** form predictor categories along a Cox direction; **(ii)** one-dimensional transformations of bivariate response (Y, δ) (reweighting or mean imputation) to obtain more balanced slices; and **(iii)** fuse kernel matrices across multiple numbers of slices to reduce sensitivity.

Numerical studies demonstrated that the fused CPH methods consistently outperform existing SIR-based competitors, particularly under non-elliptical predictors. Sensitivity analysis further confirmed that performance is robust across (n, p) and censoring rates, with only mild degradation under heavy censoring. In the DLBCL study, our proposed methods also showed a practical advantage when the reduced predictors were less favourable to the linearity condition while remaining robust across different PC dimensions.

Therefore, the fused CPH family provides a practical approach to SDR in survival regression. One limitation is that the current paper focuses on a single-index structure with $d = 1$ and does not provide a formal dimension selection theory for $d > 1$. Extending the methodology to higher structural dimensions and developing selection procedures are important directions for future research. Another promising extension is to adapt the survival IPS framework to more complex survival settings, such as competing risks.

Author Contributions

Minjee Kim contributed to manuscript writing, derived theoretical results and conducted simulation studies. Minjeong Kim contributed to methodology development and edited the paper. Jae Keun Yoo designed the study, derived theoretical results and wrote the original draft.

Acknowledgements

For Jae Keun Yoo, this work was supported by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Korean Ministry of Education (RS-2026-25472445). For Minjee Kim, this work was supported by Institute of Information & communications Technology Planning & Evaluation (IITP) grant funded by the Korea government (MSIT) (RS-2022-00143911, AI Excellence Global Innovative Leader Education Program).

Conflicts of Interest

The authors declare no conflicts of interest.

Data Availability Statement

The data that support the findings of this study are openly available in bioconductor at https://bioconductor.org/packages/release/bioc/html/Biobase.html, reference number https://doi.org/10.1038/nmeth.3252.

References

Cook, R. D. 1998. "Principal Hessian Directions Revisited." *Journal of the American Statistical Association* 93, no. 441: 84–94.

Cook, R. D. 2003. "Dimension Reduction and Graphical Exploration in Regression Including Survival Analysis." *Statistics in Medicine* 22, no. 9: 1399–1413.

Cook, R. D., and B. Li. 2002. "Dimension Reduction for Conditional Mean in Regression." *Annals of Statistics* 30, no. 2: 455–474.

Cook, R. D., and X. Zhang. 2014. "Fused Estimators of the Central Subspace in Sufficient Dimension Reduction." *Journal of the American Statistical Association* 109, no. 506: 815–827.

Datta, S. 2005. "Estimating the Mean Life Time Using Right Censored Data." *Statistical Methodology* 2, no. 1: 65–69.

Datta, S., J. Le-Rademacher, and S. Datta. 2007. "Predicting Patient Survival From Microarray Data by Accelerated Failure Time Modeling Using Partial Least Squares and Lasso." *Biometrics* 63, no. 1: 259–271.

Hooper, J. W. 1959. "Simultaneous Equations and Canonical Correlation Theory." *Econometrica: Journal of the Econometric Society* 27, no. 2: 245–256.

Huber, W., V. J. Carey, R. Gentleman, et al. 2015. "Orchestrating High-Throughput Genomic Analysis With Bioconductor." *Nature Methods* 12, no. 2: 115–121.

Koul, H., V. Susarla, and J. Van Ryzin. 1981. "Regression Analysis With Randomly Right-Censored Data." *Annals of Statistics* 9, no. 6: 1276–1288.

Li, B., and Y. Dong. 2009. "Dimension Reduction for Nonelliptically Distributed Predictors." *Annals of Statistics* 37, no. 3: 1272–1298.

Li, K. C. 1991. "Sliced Inverse Regression for Dimension Reduction." *Journal of the American Statistical Association* 86, no. 414: 316–327.

Li, L., and H. Li. 2004. "Dimension Reduction Methods for Microarrays With Application to Censored Survival Data." *Bioinformatics* 20, no. 18: 3406–3412.

Mardia, K. V. 1970. "Measures of Multivariate Skewness and Kurtosis With Applications." *Biometrika* 57, no. 3: 519–530.

Robins, J. M., and D. M. Finkelstein. 2000. "Correcting for Noncompliance and Dependent Censoring in an Aids Clinical Trial With Inverse Probability of Censoring Weighted (IPCW) Log-Rank Tests." *Biometrics* 56, no. 3: 779–788.

Robins, J. M., and A. Rotnitzky. 1992. "Recovery of Information and Adjustment for Dependent Censoring Using Surrogate Markers." In *AIDS Epidemiology: Methodological Issues*, edited by N. P. Jewell, K. Dietz, and V. T. Farewell, 297–331. Birkhäuser.

Rosenwald, A., G. Wright, W. C. Chan, et al. 2002. "The Use of Molecular Profiling to Predict Survival After Chemotherapy for Diffuse Large-B-Cell Lymphoma." *New England Journal of Medicine* 346, no. 25: 1937–1947.

Satten, G. A., and S. Datta. 2001. "The Kaplan-Meier Estimator as an Inverse-Probability-of-Censoring Weighted Average." *American Statistician* 55, no. 3: 207–210.

Satten, G. A., S. Datta, and J. Robins. 2001. "Estimating the Marginal Survival Function in the Presence of Time Dependent Covariates." *Statistics & Probability Letters* 54, no. 4: 397–403.

Um, H. Y., and J. K. Yoo. 2020. "Fused Clustering Mean Estimation of Central Subspace." *Journal of the Korean Statistical Society* 49, no. 2: 350–363.

Wen, X. M. 2010. "On Sufficient Dimension Reduction for Proportional Censorship Model With Covariates." *Computational Statistics & Data Analysis* 54, no. 8: 1975–1982.

Yoo, J. K. 2016. "Sufficient Dimension Reduction Through Informative Predictor Subspace." *Statistics* 50, no. 5: 1086–1099.

Yoo, J. K. 2017. "Fused Sliced Inverse Regression in Survival Analysis." *Communications for Statistical Applications and Methods* 24, no. 5: 533–541.

Yoo, J. K. 2019. "Analysis of Microarray Right-Censored Data Through Fused Sliced Inverse Regression." *Scientific Reports* 9, no. 1: 15094.

Yoo, J. K., S. J. Kim, B. S. Seo, H. Shin, and S. A. Sim. 2016. "Dimension Reduction for Right-Censored Survival Regression: Transformation

Approach." *Communications for Statistical Applications and Methods* 23, no. 3: 259–268.

Yoo, J. K., and K. Lee. 2011. "Model-Free Predictor Tests in Survival Regression Through Sufficient Dimension Reduction." *Lifetime Data Analysis* 17, no. 3: 433–444.