Statistica Sinica Preprint No: SS-2021-0250	
Title	Sufficient Variable Screening for Ultrahigh-Dimensional Right Censored Data via Independence Measures
Manuscript ID	SS-2021-0250
URL	http://www.stat.sinica.edu.tw/statistica/
DOI	10.5705/ss.202021.0250
Complete List of Authors	Baoying Yang, Qingcong Yuan and Xiangrong Yin
Corresponding Authors	Baoying Yang
E-mails	yangbaoying@swjtu.edu.cn
Notice: Accepted version subject to English editing.	

Sufficient variable screening for ultrahigh-dimensional right censored data via independence measures

Baoying Yang[1], Qingcong Yuan[2,4], and Xiangrong Yin[3]

[1]*Southwest Jiaotong University,* [2]*Miami University,*

[3]*University of Kentucky and* [4]*Biostatistics and Programming*

Abstract: **We develop two sufficient variable screening procedures utilizing the newly proposed censored distance correlation measures for ultrahigh-dimensional right censored data. Compared to many existing methods, our procedures more effectively detect active predictors that are marginally independent of the response. They are also model-free and robust against model misspecification. Through simulations and real data analysis, we demonstrate the distinct advantages of our proposed procedures over existing variable screening methods.**

Key words and phrases: Distance correlation; Feature screening; Independence

Corresponding author: Baoying Yang, Department of Statistics, College of Mathematics, Southwest Jiaotong University, Chengdu, Sichuan 611756, China. E-mail:yangbaoying@swjtu.edu.cn

measure.

1. Introduction

Variable screening has become increasingly important in various research fields. The renaissance of variable screening began with the sure independent screening (SIS) method of Fan and Lv (2008), which is based on the marginal Pearson correlation and is specifically tailored for linear regressions with Gaussian predictors and responses. Following the pioneering work of SIS, many methods have been proposed over the last two decades, using either model-specific or model-free approaches for ultrahigh-dimensional data. See Fan and Song (2010), Liu, Li and Wu (2014), Chang, Tang and Wu (2013), Zhu et al. (2011), Li, Zhong and Zhu (2012), Mai and Zou (2015), Shao and Zhang (2014), among many others.

In the context of ultrahigh-dimensional survival data analysis, the response is the time to an event that is often subject to censoring. Censoring brings more difficulties and challenges for the feature screening of ultrahigh-dimensional data. To address these complexities, a variety of ultrahigh-dimensional screening techniques designed to survival outcomes have flourished. For example, Fan, Feng and Wu (2010) investigated the SIS method for the Cox proportional hazards model by ranking variables according to

their respective univariate partial log-likelihoods (ISIS$_C$). He, Wang and Hong (2013) developed a quantile-adaptive model-free screening procedure based on the disparity between unconditional and conditional quantiles for each covariate in the presence of right censored data with heterogeneity (QaSIS). Song et al. (2014) devised a marginal screening procedure that utilizes an inverse probability-of-censoring weighted version of Kendall's τ (CRIS). Zhou and Li (2017) proposed a censored version of the SIRS method (Zhu et al., 2011) using the inverse probability-of-censoring weighting (CSIR) method. Zhang, Liu and Wu (2017) introduced a screening method based on the correlation between the cumulative distribution function $F(y)$ and each covariate (CRSIS). Yan, Tang and Zhao (2017) considered Spearman rank correlation screening which assesses the correlation between the distributions $F(y)$ and $F(X_\alpha)$ of each covariate (SVSIR). Liu, Zhang and Zhao (2018) extended the fused Kolmogorov filter proposed by Mai and Zou (2015) to handle right censored survival data (KM), while Chen, Chen and Wang (2018) proposed a robust feature screening procedure based on a distance correlation measurement (RCDCS). Additionally, Hong, et al. (2018a) developed the integrated powered density screening method (IPOD), and Hong, Kang and Li (2018b) introduced L_q-norm learning to handle ultrahigh-dimensional censored data (LQ).

The aforementioned techniques utilized for ultrahigh-dimensional censored data, with the exception of Fan, Feng and Wu (ISIS$_C$; 2010), are all marginal screening methods. Despite their popularity, these marginal screening techniques, which assess the dependence between the response and individual covariates, have inherent limitations. As mentioned in Fan and Lv (2008), some crucial variables may be marginally uncorrelated but jointly correlated with the response. Marginal screening procedures may fail to detect such active predictors. This issue is also relevant to censored data. To address this issue, based on previous investigations and experiences, Hong, Kang and Li (2018b) and Chen (2018) proposed harnessing the conditional correlation between individual covariates and survival time when some variables are known to be active. However, practical scenarios often lack prior information or may contain erroneous information. Fan, Feng and Wu (2010) introduced the censored iterative sure independence screening (ISIS$_C$) approach as an alternative when prior information is absent. Nevertheless, the theoretical foundations for the ISIS$_C$ method remain elusive. These facts motivate us to explore new procedures, following Yang, Yin and Zhang (2019), to overcome the challenges of existing marginal screening approaches, relax the need for prior information, and seek theoretical justifications.

In this paper, we focus on right-censored survival data and introduce

two innovative sufficient variable screening procedures, inspired by insights from the literature on sufficient dimension reduction. Our procedures translate the conditional independence measures into alternative, manageable independence measures. Importantly, our approach is model-free and robust against model misspecification. Significantly, the proposed procedures are advantages when some active predictors are marginally independent of the response, while many existing methods based on marginal relationships fail.

The structure of this paper is as follows. Section 2 presents the preliminaries of the sufficient variable screening concept and revisits the distance correlation measure. In Section 3, we propose a censored distance correlation measure for censored data using the inverse probability-of-censoring weighting method. In Section 4, we develop one-stage and two-stage sufficient variable screening procedures. The theoretical properties of our procedures are explored in Section 5, and different methods are compared through simulations in Section 6. A real data application is presented in Section 7. A short discussion is included in Section 8. All proofs can be found in the supplementary materials.

Throughout this paper, we assume that Y represents the univariate survival time, and $\mathbf{X} = (X_1, \cdots, X_p)^\top$ is a $p \times 1$ vector. The notation

$\mathbf{U} \perp\!\!\!\perp \mathbf{V}|\mathbf{W}$ indicates that given $\mathbf{W}$, $\mathbf{U}$ and $\mathbf{V}$ are independent.

2. Preliminaries and Distance Correlation Review

2.1 Some preliminaries

The following presents the definition of sufficient variable selection from Yin and Hilafu (2015).

Definition 1. *If there is a $p \times q$ matrix $\mathbf{A}$ with $q \leq p$, wherein the columns of $\mathbf{A}$ are composed of unit vectors, denoted as $\mathbf{e}_\alpha$, with the α^{th} element equal to 1, such that $Y \perp\!\!\!\perp \mathbf{X}|\mathbf{A}^\top\mathbf{X}$, then the column space of $\mathbf{A}$ is referred to as a variable selection space. The intersection of all such spaces, if it adheres to the aforementioned conditional independence condition, is called the central variable selection space, denoted by $\mathcal{S}^V_{Y|\mathbf{X}}$.*

Let $\mathbf{X}_\mathcal{D}$ represent the set of elements within $\mathbf{X}$ that are involved in $\mathcal{S}^V_{Y|\mathbf{X}}$, and let $\mathbf{X}_{\bar{\mathcal{D}}}$ denote its complement, where $\mathcal{D}$ and $\bar{\mathcal{D}}$ are the respective index sets. Yin and Hilafu (2015) discussed the existence and uniqueness of $\mathcal{S}^V_{Y|\mathbf{X}}$. In this paper, we assume that $\mathcal{S}^V_{Y|\mathbf{X}}$ always exists. Consequently, Definition 1 is equivalent to asserting $Y \perp\!\!\!\perp \mathbf{X}_{\bar{\mathcal{D}}}|\mathbf{X}_\mathcal{D}$, which, in turn, implies that $\Pr(Y|\mathbf{X}_\mathcal{D}, \mathbf{X}_{\bar{\mathcal{D}}}) = \Pr(Y|\mathbf{X}_\mathcal{D})$. Therefore, we can eliminate $\mathbf{X}_{\bar{\mathcal{D}}}$ without sacrificing any relevant regression information. However, identi-

fying $\mathbf{X}_{\mathcal{D}}$ directly through the utilization of the conditional independence $Y \perp\!\!\!\perp \mathbf{X}_{\bar{\mathcal{D}}} | \mathbf{X}_{\mathcal{D}}$ seems to be an infeasible task because it is difficult to decide which variables should be included in the set $\mathbf{X}_{\mathcal{D}}$ and how many of them. To address this problem effectively, inspired by the subsequent proposition, Yang, Yin and Zhang (2019) explored a sufficient variable screening approach aiming to identify a larger subset of predictors that includes the active predictors set. We retain the notation $\mathbf{X}_{\mathcal{D}}$ to represent this target active set for variable screening.

Proposition 1. *Let* $\mathbf{X}$, $\mathbf{X}_1$ *and* $\mathbf{X}_2$ *be random vectors, and* $\mathbf{X}^\top = (\mathbf{X}_1^\top, \mathbf{X}_2^\top)$. *Then, either statement (i) or statement (ii) implies statement (iii):*

(i) $(Y, \mathbf{X}_2) \perp\!\!\!\perp \mathbf{X}_1$;

(ii) $\mathbf{X}_1 \perp\!\!\!\perp \mathbf{X}_2 | Y$ *and* $Y \perp\!\!\!\perp \mathbf{X}_1$;

(iii) $Y \perp\!\!\!\perp \mathbf{X}_1 | \mathbf{X}_2$.

It is important to note that statement (iii), representing the ultimate objective of sufficient variable screening, is necessarily true if either statement (i) or statement (ii) holds. Consequently, it becomes essential to develop methodologies for testing both statements (i) and (ii). To achieve this, Yang, Yin and Zhang (2019) devised two sufficient variable screening approaches based on statements (i) and (ii), referred to as one-stage sufficient variable screening and two-stage sufficient variable screening, respec-

tively. Note that marginal screening methods, such as SIS (Fan and Lv, 2008) and its variations, rely on the second part of statement (ii), and therefore, are insufficient in their ability to imply (iii).

2.2 Review of distance correlation

Let $(\mathbf{U}, \mathbf{V})$ represent random vectors and $(\mathbf{U}', \mathbf{V}')$ and $(\mathbf{U}'', \mathbf{V}'')$ denote independent copies of $(\mathbf{U}, \mathbf{V})$. Assume that $(\mathbf{U}_1, \mathbf{V}_1), \ldots, (\mathbf{U}_n, \mathbf{V}_n)$ is a random sample of $(\mathbf{U}, \mathbf{V})$. Székely, Rizzo and Bakirov (2007) articulated the distance covariance $\mathrm{dcov}^2(\mathbf{U}, \mathbf{V})$ as follows:

$$\mathrm{dcov}^2(\mathbf{U}, \mathbf{V}) = S_1 + S_2 - 2S_3,$$

where $S_1 = \mathrm{E}\left[\|\mathbf{U} - \mathbf{U}'\|_{d_u}\|\mathbf{V} - \mathbf{V}'\|_{d_v}\right]$, $S_2 = \mathrm{E}\|\mathbf{U} - \mathbf{U}'\|_{d_u}\mathrm{E}\|\mathbf{V} - \mathbf{V}'\|_{d_v}$, and $S_3 = \mathrm{E}\left[\|\mathbf{U} - \mathbf{U}'\|_{d_u}\|\mathbf{V} - \mathbf{V}''\|_{d_v}\right]$. Here, $\|\cdot\|_{d_u}$ and $\|\cdot\|_{d_v}$ represent the Euclidean norms in dimensions d_u and d_v respectively. Similar definitions can be encountered throughout the paper. The distance correlation (DC) between $\mathbf{U}$ and $\mathbf{V}$ is defined as follows:

$$\mathrm{dcorr}^2(\mathbf{U}, \mathbf{V}) = \mathrm{dcov}^2(\mathbf{U}, \mathbf{V})/\sqrt{\mathrm{dcov}^2(\mathbf{U}, \mathbf{U}')\mathrm{dcov}^2(\mathbf{V}, \mathbf{V}')}.$$

When $\mathbf{U}$ and $\mathbf{V}$ take values in a separable Hilbert space, Lyons (2013) pointed out that the DC measure defined for Euclidean spaces in Székely, Rizzo and Bakirov (2007) is still applicable to separable Hilbert

spaces. A noteworthy property of the DC measure is its ability to assess all types of dependence between $\mathbf{U}$ and $\mathbf{V}$ in arbitrary dimensions. The DC value equals to zero if and only if $\mathbf{U}$ and $\mathbf{V}$ are independent. This property makes DC particularly suitable for variable screening in ultrahigh-dimensional data.

The corresponding sample versions of S_1, S_2, and S_3 are given by

$$\hat{S}_1 = n^{-2} \sum_{i,j=1}^{n} \|\mathbf{U}_i - \mathbf{U}'_j\|_{d_u} \|\mathbf{V}_i - \mathbf{V}'_j\|_{d_v}, \quad \hat{S}_2 = n^{-4} \sum_{i,j=1}^{n} \|\mathbf{U}_i - \mathbf{U}'_j\|_{d_u} \sum_{i,j=1}^{n} \|\mathbf{V}_i - \mathbf{V}'_j\|_{d_v},$$

$$\hat{S}_3 = n^{-3} \sum_{i,j,\ell=1}^{n} \|\mathbf{U}_i - \mathbf{U}'_j\|_{d_u} \|\mathbf{V}_i - \mathbf{V}''_\ell\|_{d_v}.$$

By substituting these expressions into the formula of $\mathrm{dcorr}^2(\mathbf{U}, \mathbf{V})$, we obtain an estimator of DC.

3. New Dependency Measures for Right Censored Data

In the realm of ultrahigh-dimensional right-censored survival data, let Y represent the survival time and C represent the censoring time. We denote the observed time as T, where $T = \min(Y, C)$, and let $\delta = \mathbf{I}(Y < C)$ be the censoring indicator. To ensure identifiability under the random censoring scheme (Lu and Li, 2011), we assume that $Y \perp\!\!\!\perp C|\mathbf{X}$, where $\mathbf{X}$ represents the covariates. Our primary objective is to accurately identify the active predictor indices, denoted as $\mathcal{D}$, based on the aforementioned DC measure. How-

ever, it is important to note that we observe T instead of Y. Ignoring the censored observations can lead to significant inaccuracies, as highlighted in (Tsiatis, 2007). One solution is to use the inverse of the censoring probability to reweight the uncensored observations (Robins, Rotnitzky and Zhao , 1994). In this section, we introduce three inverse probability-weighted measures to assess the dependence between the predictors and right censored response.

3.1 An inverse probability weighted marginal utility

Following Székely, Rizzo and Bakirov (2007), we define the population distance covariance between X_α and T as $\text{dcov}^2_\alpha(X_\alpha, T) = S_1 + S_2 - 2S_3$, with

$$S_1 = E\left[\frac{\delta\delta'}{G(T|\mathbf{X})G(T'|\mathbf{X}')}\|X_\alpha - X'_\alpha\|_{d_{X_\alpha}}\|T - T'\|_{d_T}\right],$$

$$S_2 = E\|X_\alpha - X'_\alpha\|_{d_{X_\alpha}} E\left[\frac{\delta\delta'}{G(T|\mathbf{X})G(T'|\mathbf{X}')}\|T - T'\|_{d_T}\right],$$

$$S_3 = E\left[\frac{\delta\delta'\delta''}{G(T|\mathbf{X})G(T'|\mathbf{X}')G(T''|\mathbf{X}'')}\|X_\alpha - X'_\alpha\|_{d_{X_\alpha}}\|T - T''\|_{d_T}\right].$$

Here, $G(t|\mathbf{X}) = Pr(C > t|\mathbf{X})$ represents the conditional survival function of censoring time C given $\mathbf{X}$, and (X'_α, T') and (X''_α, T'') denote independent and identical copies of (X_α, T). Throughout the paper, we define $0/0 = 0$ to ensure that S_1, S_2, and S_3 well-defined. Following the meticulous calculations in Appendix A.1, we demonstrate that the censored version

of the distance covariance is equivalent to that determined with a known response,

$$\mathrm{dcov}_\alpha^2(X_\alpha, T) = \mathrm{dcov}_\alpha^2(X_\alpha, Y). \tag{3.1}$$

The DC measure between X_α and T is defined as

$$\mathrm{dcorr}^2(X_\alpha, T) = \ \mathrm{dcov}^2(X_\alpha, T)/\sqrt{\mathrm{dcov}^2(X_\alpha, X'_\alpha)\mathrm{dcov}^2(T, T')}.$$

At the sample level, our observations consist of $\{(\mathbf{X}_i, T_i, \delta_i), i = 1, \cdots, n\}$ instead of $\{(\mathbf{X}_i, Y_i), i = 1, \cdots, n\}$. The respective sample estimators for S_1, S_2 and S_3 are given as follows:

$$\widehat{S}_1 = \frac{1}{n^2}\sum_{i,j=1}^{n} \frac{\delta_i\delta'_j}{\widehat{G}_n(T_i|\mathbf{X}_i)\widehat{G}_n(T'_j|\mathbf{X}_j)}\|X_{\alpha,i} - X'_{\alpha,j}\|_{d_{X_\alpha}}\|T_i - T'_j\|_{d_T},$$

$$\widehat{S}_2 = \frac{1}{n^2}\sum_{i,j=1}^{n} \|X_{\alpha,i} - X'_{\alpha,j}\|_{d_{X_\alpha}} \frac{1}{n^2}\sum_{i,j=1}^{n} \frac{\delta_i\delta'_j}{\widehat{G}_n(T_i|\mathbf{X}_i)\widehat{G}_n(T'_j|\mathbf{X}_j)}\|T_i - T'_j\|_{d_T},$$

$$\widehat{S}_3 = \frac{1}{n^3}\sum_{i,j,l=1}^{n} \frac{\delta_i\delta'_j\delta''_l}{\widehat{G}_n(T_i|\mathbf{X}_i)\widehat{G}_n(T'_j|\mathbf{X}'_j)\widehat{G}_n(T''_l|\mathbf{X}''_l)}\|X_{\alpha,i} - X'_{\alpha,j}\|_{d_{X_\alpha}}\|T_i - T''_l\|_{d_T},$$

where $\widehat{G}_n(t|\mathbf{X})$ is the sample estimator of $G(t|\mathbf{X})$.

To estimate $G(t|\mathbf{X})$, the simplest approach assumes that $G(t|\mathbf{X}) = G(t)$. A similar assumption was employed in He, Wang and Hong (2013), Song et al. (2014), Zhou and Li (2017), and Zhang et al. (2018). The classical Kaplan-Meier method can be applied to estimate $G(t)$. Additional methods to estimate $G(t|\mathbf{X})$ along with their respective conditions and results are detailed in Appendix A.2.

3.2 An inverse probability weighted one-stage utility

For one-stage sufficient variable screening, Proposition 1 (i) holds that $(Y, \mathbf{X}_{-\alpha}) \perp\!\!\!\perp X_\alpha$, where $\mathbf{X}_{-\alpha} = (X_1, \ldots, X_{\alpha-1}, X_{\alpha+1}, \ldots, X_p)$. Denote $\mathbf{U} = X_\alpha$, $\mathbf{V} = (Y, \mathbf{X}_{-\alpha})$, and $\mathbf{W} = (T, X_{-\alpha})$. The distance covariance can be defined as follows: $\mathrm{dcov}_\alpha^{*\,2}(\mathbf{U}, \mathbf{W}) = S_1 + S_2 - 2S_3$, where S_j, $j = 1$, 2 and 3, are as follows:

$$S_1 = E\left[\frac{\delta\delta'}{G(T|\mathbf{X})G(T'|\mathbf{X}')}\|\mathbf{U} - \mathbf{U}'\|_{d_\mathbf{U}}\|\mathbf{W} - \mathbf{W}'\|_{d_\mathbf{W}}\right],$$

$$S_2 = E\|\mathbf{U} - \mathbf{U}'\|_{d_\mathbf{U}} E\left[\frac{\delta\delta'}{G(T|\mathbf{X})G(T'|\mathbf{X}')}\|\mathbf{W} - \mathbf{W}'\|_{d_\mathbf{W}}\right],$$

$$S_3 = E\left[\frac{\delta\delta'\delta''}{G(T|\mathbf{X})G(T'|\mathbf{X}')G(T''|\mathbf{X}'')}\|\mathbf{U} - \mathbf{U}'\|_{d_\mathbf{U}}\|\mathbf{W} - \mathbf{W}''\|_{d_\mathbf{W}}\right],$$

where $(\mathbf{U}', \mathbf{W}')$, $(\mathbf{U}'', \mathbf{W}'')$ are independent and identical copies of $(\mathbf{U}, \mathbf{W})$. According to the detailed calculations in Appendix A.1, we can prove that

$$\mathrm{dcov}_\alpha^{*\,2}(\mathbf{U}, \mathbf{W}) = \mathrm{dcov}_\alpha{}^2(\mathbf{U}, \mathbf{V}). \tag{3.2}$$

The sample counterparts $\widehat{S}_1$, $\widehat{S}_2$, and $\widehat{S}_3$ are given by:

$$\widehat{S}_1 = \frac{1}{n^2}\sum_{i,j=1}^{n}\frac{\delta_i\delta_j'}{\widehat{G}_n(T_i|\mathbf{X}_i)\widehat{G}_n(T_j'|\mathbf{X}_j')}\|\mathbf{U}_i - \mathbf{U}_j'\|_{d_\mathbf{U}}\|\mathbf{W}_i - \mathbf{W}_j'\|_{d_\mathbf{W}},$$

$$\widehat{S}_2 = \frac{1}{n^2}\sum_{i,j=1}^{n}\|\mathbf{U}_i - \mathbf{U}_j'\|_{d_\mathbf{U}}\frac{1}{n^2}\sum_{i,j=1}^{n}\frac{\delta_i\delta_j'}{\widehat{G}_n(T_i|\mathbf{X}_i)\widehat{G}_n(T_j'|\mathbf{X}_j')}\|\mathbf{W}_i - \mathbf{W}_j'\|_{d_\mathbf{W}},$$

$$\widehat{S}_3 = \frac{1}{n^3}\sum_{i,j,l=1}^{n}\frac{\delta_i\delta_j'\delta_l''}{\widehat{G}_n(T_i|\mathbf{X}_i)\widehat{G}_n(T_j'|\mathbf{X}_j')\widehat{G}_n(T_l''|\mathbf{X}_l'')}\|\mathbf{U}_i - \mathbf{U}_j'\|_{d_\mathbf{U}}\|\mathbf{W}_i - \mathbf{W}_l''\|_{d_\mathbf{W}}.$$

3.3 An inverse probability weighted two-stage utility

For two-stage sufficient variable screening, Proposition 1 (ii) applies that $X_\alpha \perp\!\!\!\perp \mathbf{X}_{-\alpha}|Y$ and $Y \perp\!\!\!\perp X_\alpha$. The assessment of the conditional independence $X_\alpha \perp\!\!\!\perp \mathbf{X}_{-\alpha}|Y$ is not trivial; hence, we adopt a slicing approach. If Y is uncensored, we can slice it directly. A general partition of the real number line is defined as $\mathcal{J} = \{[l_{s-1}, l_s) : l_{s-1} < l_s, s = \{1, \cdots, S\}, \bigcup_{s=1}^{S}[l_{s-1}, l_s) \setminus l_0 = R\}$, where l_s represents the $\frac{s}{S}^{\text{th}}$ sample quantiles of Y, $l_0 = -\infty$ and $l_S = \infty$. Each interval $[l_{s-1}, l_s)$ is referred to as an s-slice. Note that the first slice, (l_0, l_1), is open, but we slightly abuse the notation for consistency across all slices. However, if Y is censored and (T, δ) is observed, we can only slice the observations based on T and refer to the partition as the aforementioned $\mathcal{J}$.

The population version of the censored two-stage distance covariance is denoted as $\text{dcov}_\alpha^{**2}\{(X_\alpha, \mathbf{X}_{-\alpha})|T\}$. For ease of computation, we approximate it based on the fact that $E(\text{dcov}_\alpha^{**2}\{(X_\alpha, \mathbf{X}_{-\alpha})|T\}) \approx S^{-1}\sum_{s=1}^{S}\text{dcov}_\alpha^{**2}\{(X_\alpha, \mathbf{X}_{-\alpha}) \mid T \in [l_{s-1}, l_s)\}$, where $\text{dcov}_\alpha^{**2}\{(X_\alpha, \mathbf{X}_{-\alpha})|T \in [l_{s-1}, l_s)\} = S_{1,s} + S_{2,s} - 2S_{3,s}$, with $S_{j,s}$, for $j = 1, 2$ and 3 defined below:

$$S_{1,s} = E\Big[\frac{\delta\delta'}{G(T|\mathbf{X})G(T'|\mathbf{X}')}\|X_\alpha - X'_\alpha\|_{d_{X_\alpha}}\|\mathbf{X}_{-\alpha} - \mathbf{X}'_{-\alpha}\|_{d_{\mathbf{X}_{-\alpha}}}\times \mathbf{1}\{T \in [l_{s-1}, l_s)\}\mathbf{1}\{T' \in [l_{s-1}, l_s)\}\Big]/ E\Big[\frac{\delta\delta'}{G(T|\mathbf{X})G(T'|\mathbf{X}')}\mathbf{1}\{T \in [l_{s-1}, l_s)\}\mathbf{1}\{T' \in [l_{s-1}, l_s)\}\Big],$$

$$S_{2,s} = E\Big[\frac{\delta\delta'}{G(T|\mathbf{X})G(T'|\mathbf{X}')}\|X_\alpha - X'_\alpha\|_{d_{X_\alpha}}\mathbf{1}\{T \in [l_{s-1}, l_s)\}\mathbf{1}\{T' \in [l_{s-1}, l_s)\}\Big]\times E\Big[\frac{\delta\delta'}{G(T|\mathbf{X})G(T'|\mathbf{X}')}\|\mathbf{X}_{-\alpha} - \mathbf{X}'_{-\alpha}\|_{d_{\mathbf{X}_{-\alpha}}}\mathbf{1}\{T \in [l_{s-1}, l_s)\}\mathbf{1}\{T' \in [l_{s-1}, l_s)\}\Big]/ E\Big[\frac{\delta\delta'}{G(T|\mathbf{X})G(T'|\mathbf{X}')}\mathbf{1}\{T \in [l_{s-1}, l_s)\}\mathbf{1}\{T' \in [l_{s-1}, l_s)\}\Big],$$

$$S_{3,s} = E\Big[\frac{\delta\delta'\delta''}{G(T|\mathbf{X})G(T'|\mathbf{X}')G(T''|\mathbf{X}'')}\|X_\alpha - X'_\alpha\|_{d_{X_\alpha}}\|\mathbf{X}_{-\alpha} - \mathbf{X}''_{-\alpha}\|_{d_{\mathbf{X}_{-\alpha}}}\times \mathbf{1}\{T \in [l_{s-1}, l_s)\}\mathbf{1}\{T' \in [l_{s-1}, l_s)\}\mathbf{1}\{T'' \in [l_{s-1}, l_s)\}\Big]/ E\Big[\frac{\delta\delta'\delta''}{G(T|\mathbf{X})G(T'|\mathbf{X}')G(T''|\mathbf{X}'')}\times \mathbf{1}\{T \in [l_{s-1}, l_s)\}\mathbf{1}\{T' \in [l_{s-1}, l_s)\}\mathbf{1}\{T'' \in [l_{s-1}, l_s)\}\Big],$$

where $(\mathbf{X}', T', \delta')$ and $(\mathbf{X}'', T'', \delta'')$ are independent and identical copies of $(\mathbf{X}, T, \delta)$. After some calculation (Appendix A.1), we obtain that

$$dcov_\alpha^{**2}\{(X_\alpha, \mathbf{X}_{-\alpha})|T \in [l_{s-1}, l_s)\} = dcov_\alpha{}^2\{(X_\alpha, \mathbf{X}_{-\alpha})|Y \in [l_{s-1}, l_s)\}. \tag{3.3}$$

The indices of the observations belonging to slice s are denoted as $\mathcal{V}_s = \{k, T_k \in [l_{s-1}, l_s), k = 1, 2, \cdots, n\}$, and the empirical estimator $\hat{S}_{1,s}$ of $S_{1,s}$

is

$$\widehat{S}_{1,s} = \frac{1}{n^2} \sum_{i,j \in \mathcal{V}_s} \frac{\delta_i \delta_j'}{\widehat{G}_n(T_i|\mathbf{X}_i)\widehat{G}_n(T_j'|\mathbf{X}_j')} \|X_{(\alpha,i)} - X'_{(\alpha,j)}\| \|\mathbf{X}_{-(\alpha,i)} - \mathbf{X}'_{-(\alpha,j)}\| / \left(\frac{1}{n^2} \sum_{i,j \in \mathcal{V}_s} \frac{\delta_i \delta_j'}{\widehat{G}_n(T_i|\mathbf{X}_i)\widehat{G}_n(T_j'|\mathbf{X}_j')} \right),$$

where the subscripts (α, i) and (α, j) represent the i-th and j-th elements of a specific slice, respectively. The estimations for $S_{2,s}$, $S_{3,s}$ can be obtained in a similar manner as $S_{1,s}$.

4. Sufficient Variable Screening Procedures for Right Censored Data

We now present three algorithms based on Proposition 1 as follows. Let $\mathcal{I}\{A, B\}$ represent a population-level independence measure between two random vectors A and B.

(i) *Censored Marginal Feature Screening Procedure (CFS_M)*: This procedure calculates the marginal relationship as follows: $u_\alpha = \mathcal{I}(T, X_\alpha) = dcorr_\alpha{}^2(T, X_\alpha)$.

(ii) *Censored One-Stage Sufficient Variable Screening Procedure* ($CSVS_1$): This procedure builds upon statement (i), and computes $u_\alpha^* = \mathcal{I}\{(T, \mathbf{X}_{-\alpha}), X_\alpha\} = dcorr_\alpha^{*2}\{(T, \mathbf{X}_{-\alpha}), X_\alpha\}$.

(iii) *Censored Two-Stage Sufficient Variable Screening Procedure* ($CSVS_2$): This method extends from statement (ii), and computes $u_\alpha^{**} = \mathcal{I}\{(X_\alpha, \mathbf{X}_{-\alpha})|T\} = dcorr_\alpha^{**2}\{(X_\alpha, \mathbf{X}_{-\alpha})|T\}$. Additionally, it approximates u_α^{**} as $u_\alpha^{\mathcal{J}} = \sum_{s=1}^{S} u_{\alpha,s}^{\mathcal{J}}/S$, where $u_{\alpha,s}^{\mathcal{J}} = \mathcal{I}\{(X_\alpha, \mathbf{X}_{-\alpha}) \mid T \in [l_{s-1}, l_s)\} = dcorr_\alpha^{**2}\{(X_\alpha, \mathbf{X}_{-\alpha})|T \in [l_{s-1}, l_s)\}$.

Sample estimates of the pivotal quantities are denoted as $\widehat{u}_\alpha$, $\widehat{u}_\alpha^*$, and $\widehat{u}_\alpha^{**}$, respectively.

The sufficient variable screening algorithm is as follows:

Censored Sufficient Variable Screening (CSVS):

1. Calculate $\widehat{u}_\alpha = \widehat{\mathcal{I}}\{T, X_\alpha\}$ for $\alpha \in \{1, \ldots, p\}$. The estimate $\mathbf{X}_{\hat{\mathcal{D}}_1}$ is the set of X_α's with the largest d_1 values of $\widehat{u}_\alpha$, where d_1 is the preselected value.

2(a). Calculate $\widehat{u}_\alpha^* = \widehat{\mathcal{I}}\{(T, \mathbf{X}_{-\alpha}), X_\alpha\}$ for $\alpha \in \{1, \ldots, p\}$. The estimate $\mathbf{X}_{\hat{\mathcal{D}}_2}$ is the set of X_α's with the largest d_2 values of $\widehat{u}_\alpha^*$ but not in $\mathbf{X}_{\hat{\mathcal{D}}_1}$ from Step 1.

2(b). Alternatively, slice T into S nonoverlapping slices. Then calculate $\widehat{u}_\alpha^{**} = \sum_{s=1}^{S} \widehat{u}_{\alpha,s}^{\mathcal{J}}/S$ with $\widehat{u}_{\alpha,s}^{\mathcal{J}} = \widehat{\mathcal{I}}\{(X_\alpha, \mathbf{X}_{-\alpha})|T \in [l_{s-1}, l_s)\}$. The estimate $\mathbf{X}_{\hat{\mathcal{D}}_2}$ is the set of X_α's with the largest d_2 values of $\widehat{u}_\alpha^{**}$ but not in $\mathbf{X}_{\hat{\mathcal{D}}_1}$ from Step 1.

3. The final estimate $\mathbf{X}_{\hat{\mathcal{D}}}$ is the union of these two sets, i.e., $\mathbf{X}_{\hat{\mathcal{D}}_1} \cup \mathbf{X}_{\hat{\mathcal{D}}_2}$.

In the aforementioned algorithm, Step 2(a) and Step 2(b) follow separate paths guided by statement (i) and statement (ii) to achieve sufficient variable screening. In contrast, the traditional marginal screening method (CFS_M) focuses only on Step 1. The CSVS_1 algorithm encompasses both Step 1 and Step 2(a), while the CSVS_2 algorithm incorporates Step 1 along with Step 2(b). The additional step (2(a) or 2(b)) ensures the sufficiency of the selected features. It is noteworthy that in the CSVS_1 algorithm, we integrate CFS_M due to its practical significance, as the marginal relation typically plays an important role. Note that in the CSVS_2 algorithm, the number of slices S can be subjective. A general recommendation is to avoid an excessive number of slices, as it may render a small number of observations unstable within each slice. For further insights on selecting the appropriate value of S, additional information can be found in Yang, Yin and Zhang (2019) and Li (1991). From a practical standpoint, the choice of suitable values for d_1 and d_2 is also essential. As suggested by Yang, Yin and Zhang (2019), we opted for $S = 2$ and set $d = \lfloor n/\ln(n) \rfloor$, $d_1 = \lfloor 0.95d \rfloor$ and $d_2 = d - d_1$ in our implementation, which seems effective.

The performance of the censored marginal screening procedure CFS_M can be assessed using other existing measurements, such as correlation rank

screening (CRSIS; Zhang, Liu and Wu, 2017), Spearman rank correlation screening (SVSIR; Yan, Tang and Zhao, 2017), and censored rank independence screening (CRIS; Song et al., 2014). In this paper, we illustrate the proposed censored one-stage (CSVS_1) and two-stage (CSVS_2) algorithms using the newly proposed censored distance correlation (CDC) in Section 2. The advantage of CDC is that it can measure the dependence between two random vectors with arbitrary dimensions, while the abovementioned screening measures can only assess the dependency between univariate random variables.

5. Theoretical Properties

We now study the theoretical properties of the measurements $\widehat{u}_\alpha$, $\widehat{u}^*_\alpha$, and $\widehat{u}^{**}_\alpha$ in the screening procedures.

Note that $u_\alpha = \mathcal{I}(T, X_\alpha) = 0$ if and only if T is independent of X_α. It guarantees that u_α effectively separates and ranks the active predictors over the inactive ones, i.e., $\max_{\alpha\in\bar{\mathcal{D}}} u_\alpha < \min_{\alpha\in\mathcal{D}} u_\alpha$. Consequently, the quantity u_α serves as a valuable tool for feature screening. However, it is also important to consider the separating capability of the empirical version $\widehat{u}_\alpha$ and its sure screening property. The theoretical properties of $\widehat{u}_\alpha$ are presented in Theorem 1, and the following conditions are needed:

($C1$): There exists a positive constant s_0 such that for all $0 < s \leq 2s_0$,

$$E\{exp(s||\mathbf{X}||^2_{d_\mathbf{X}})\} < \infty \quad E\{exp(s||T||^2_{d_T})\} < \infty,$$

where $d_\mathbf{X}$ (and d_T) are the dimensions of $\mathbf{X}$ (and T). In our case, similar to Li, Zhong and Zhu (2012), when both $\mathbf{X}$ and T are uniformly bounded or conform to a multivariate normal, condition ($C1$) holds true.

($C2$): $\min_{\alpha \in \mathcal{D}} u_\alpha \geq 2cn^{-\nu}$, for $0 \leq \nu < 1/2$, and $c > 0$.

Condition ($C2$) closely resembles Condition (C2) in Li, Zhong and Zhu (2012). Essentially, this condition implies that u_α cannot be too small in the active set $\mathcal{D}$. Condition ($C2$) reflects the signal strength exhibited by individual active predictors, thereby controlling the probability error rate in the selection of active predictors (Zhu et al., 2011).

To ensure that the Kaplan-Meier estimator and its inverse function are well-behaved, Condition (C3) is necessary. This requirement is common in the survival analysis literature, under the assumption that $G(T|\mathbf{X}) = G(t)$. Similar conditions for alternative scenarios to estimate $G(T|\mathbf{X})$ are presented in Appendix A.2.

($C3$): If $G(t \mid \mathbf{X}) = G(t)$, then $Pr(t \leq Y \leq C) \geq \tau_0 > 0$ for $t \in [0, T_{\max}]$, where $T_{\max}$ is the maximum follow-up time. In addition, $sup\{t : Pr(Y > t) > 0\} \geq sup\{t : Pr(C > t) > 0\}$ and $G'_\alpha(t)$, the first derivative of $G_\alpha(t)$, is uniformly bounded.

Theorem 1. *Let* $0 < (\nu + \gamma) < 1/2$. *Under conditions (C1) and (C3), there exist positive constants* $c, c_1, c_2 > 0$ *such that*

$$\begin{aligned} & Pr(\max_{1\le\alpha\le p} |\widehat{u}_\alpha - u_\alpha| \ge cn^{-\nu}) \\ & \le O\left\{p\left[\exp\left(-c_1 n^{1-2(\nu+\gamma)}\right) + n\exp\left(-c_2 n^{\gamma}\tau_0^2\right)\right]\right\}. \end{aligned} \tag{5.4}$$

Furthermore, under condition (C2), let $\zeta = \min_{\alpha\in\mathcal{D}_1} u_\alpha - \max_{\alpha\in\bar{\mathcal{D}}_1} u_\alpha$, *i.e.* $\zeta \ge 2cn^{-\nu} \ge 0$. *Then we have*

$$\begin{aligned} & Pr(\max_{\alpha\in\bar{\mathcal{D}}_1} \widehat{u}_\alpha < \min_{\alpha\in\mathcal{D}_1} \widehat{u}_\alpha) \\ & \ge 1 - O\left\{2p\left[\exp\left(-c_1 n^{1-2(\nu+\gamma)}\right) + n\exp\left(-c_2 n^{\gamma}\tau_0^2\right)\right]\right\}. \end{aligned} \tag{5.5}$$

Under conditions (C1), (C2), and (C3), we have that

$$\begin{aligned} & Pr(\mathcal{D} \subseteq \widehat{\mathcal{D}}) \\ & \ge 1 - O(s_n\left[\exp\left(-c_1 n^{1-2(\nu+\gamma)}\right) + n\exp\left(-c_2 n^{\gamma}\tau_0^2\right)\right]), \end{aligned} \tag{5.6}$$

where s_n *is the cardinality of* $\mathcal{D}$.

The result (5.4) presented in Theorem 1 indicates that $\widehat{u}_\alpha$ is rank consistent. This indicates that the maximum disparities between $\widehat{u}_\alpha$ and u_α for all $\alpha = 1, \cdots, p$ can be controlled within a certain level. Its overwhelming probability is allowed to converge to zero at an exponential rate in terms of n and p. Equations (5.5) and (5.6) jointly guarantee that $\widehat{u}_\alpha$ consistently ranks active predictors above inactive ones in terms of probability. Furthermore, it confirms that true active predictors survive with a probability

approaching one at an exponential rate in terms of p and n. Theorem 1 sheds light on the relationship between p and n in the context of rank consistency and sure screening properties. If we balance the two terms on the right-hand side of (5.4) and choose the optimal order $\gamma = (1-2\nu)/3$, the relationship between p and n can be expressed as $p = o(\exp(cn^{(1-2\nu)/3}))$. Consequently, (5.4) becomes

$$Pr(\max_{1\le\alpha\le p} |\widehat{u}_\alpha - u_\alpha| \ge cn^{-\nu}) \le O\left\{p\left[\exp\left(-c_1 n^{1-2(\nu+\gamma)}\right)\right]\right\}.$$

If we make an additional assumption that $\mathbf{X}$ is uniformly bounded in terms of p, then (5.4) further simplifies to

$$Pr(\max_{1\le\alpha\le p} |\widehat{u}_\alpha - u_\alpha| \ge cn^{-\nu}) \le O\left\{p\left[\exp\left(-c_1 n^{1-2\nu}\right)\right]\right\}.$$

For a similar conclusion, refer to Li, Zhong and Zhu (2012).

To investigate the separating capacity of the empirical version $\widehat{u}^*_\alpha$ and its sure screening property, including conditions (C1) and (C3), we also need condition $(C2^*)$.

$(C2^*)$: $\min_{\alpha\in\mathcal{D}^*} u^*_\alpha \ge 2cn^{-\nu}$, for $0 \le \nu < 1/2$, and $c > 0$.

Theorem 2. *Let $0 < (\nu+\gamma) < 1/2$. Under conditions $(C1)$ and $(C3)$, there exist positive constants $c, c_1, c_2 > 0$ such that*

$$\begin{aligned} &Pr(\max_{1\le\alpha\le p} |\widehat{u}^*_\alpha - u^*_\alpha| \ge cn^{-\nu}) \\ &\le O\left\{p\left[\exp\left(-c_1 n^{1-2(\nu+\gamma)}\right) + n\exp\left(-c_2 n^\gamma \tau_0^2\right)\right]\right\}. \end{aligned} \tag{5.7}$$

Furthermore, under condition $(C2^*)$*, where* $\zeta^* = \min_{\alpha\in\mathcal{D}^*} u^*_\alpha - \max_{\alpha\in\bar{\mathcal{D}}^*} u^*_\alpha$*, i.e.* $\zeta^* \geq 2cn^{-\nu} \geq 0$*. Then we have*

$$\begin{aligned} &Pr(\max_{\alpha\in\bar{\mathcal{D}}^*} \widehat{u}^*_\alpha < \min_{\alpha\in\mathcal{D}^*} \widehat{u}^*_\alpha) \\ &\geq 1 - O\left\{2p\left[\exp\left(-c_1 n^{1-2(\nu+\gamma)}\right) + n\exp\left(-c_2 n^\gamma \tau_0^2\right)\right]\right\}. \end{aligned} \tag{5.8}$$

Under conditions $(C1)$*,* $(C2^*)$*, and* $(C3)$*, we have that*

$$\begin{aligned} &Pr(\mathcal{D}^* \subseteq \widehat{\mathcal{D}}^*) \\ &\geq 1 - O(s_n\left[\exp\left(-c_1 n^{1-2(\nu+\gamma)}\right) + n\exp\left(-c_2 n^\gamma \tau_0^2\right)\right]), \end{aligned} \tag{5.9}$$

where s_n *is the cardinality of* $\mathcal{D}^*$*.*

The following Theorem 3 presents the theoretical properties of $\widehat{u}^{**}_\alpha$. To prove Theorem 3, we require the following additional conditions and Lemma 5 in Appendix A.3.

$(C2^{**})$: $\min_{\alpha\in\mathcal{D}^{**}} u^{**}_\alpha \geq 2cn^{-\nu}$, $0 \leq \nu < 1/2$, and $c > 0$.

$(C4^{**})$: After slicing T, the sample $\mathbf{X}_1, \cdots, \mathbf{X}_n$ remains an independent observation within each slice.

$(C5^{**})$: For any interval $[b_1, b_2)$, we have

$$\begin{aligned} \inf_{y\in[b_1,b_2)} \mathcal{I}\{(X_\alpha, \mathbf{X}_{-\alpha})|T = y\} &\leq \mathcal{I}\{(X_\alpha, \mathbf{X}_{-\alpha})|T \in [b_1, b_2)\} \\ &\leq \sup_{y\in[b_1,b_2)} \mathcal{I}\{(X_\alpha, \mathbf{X}_{-\alpha})|T = y\}. \end{aligned}$$

Furthermore, for any $\epsilon > 0$, if $1/S - \epsilon \leq Pr(T \in [b_1, b_2)) \leq 1/S + \epsilon$, then

for any $y_1, y_2 \in [b_1, b_2)$,

$$|\mathcal{I}\{(X_\alpha, \mathbf{X}_{-\alpha})|T = y_1\} - \mathcal{I}\{(X_\alpha, \mathbf{X}_{-\alpha})|T = y_2\}| \leq \epsilon/2.$$

Note that the assumption $(C4^{**})$ appears strong; nevertheless, it is a relatively common assumption. For instance, in sufficient dimension reduction, Li (1991), Gannoun and Saracco (2003), and Cook and Ni (2005) make this assumption; and in screening methods such as Mai and Zou (2015), the assumption is made as well. With this condition, we can obtain the rank consistency of $\widehat{u}^{\mathcal{J}}_{\alpha,s}$ in each slice. Condition $(C5^{**})$ is equivalent to the condition (C2) in Mai and Zou (2015), which will be used to provide a necessary exponential inequality. Condition $(C5^{**})$ ensures $\mathcal{I}\{(\mathbf{X}_{-\alpha}, X_\alpha)|T \in [b_1, b_2)\}$ with slices to accurately approximate the goal $\mathcal{I}\{(\mathbf{X}_{-\alpha}, X_\alpha)|T\}$.

Theorem 3. *Let* $0 < (\nu + \gamma) < 1/2$. *Under conditions* $(C1)$, $(C3)$, $(C4^{**})$, *and* $(C5^{**})$, *there exist positive constants* $c, c_1, c_2, c_3 > 0$ *such that*

$$\begin{aligned} &Pr(\max_{1\leq\alpha\leq p} |\widehat{u}^{**}_\alpha - u^{**}_\alpha| \geq cn^{-\nu}) \\ &\leq O(p[\exp\left(-c_1\tau_0^4 n^{1-2(\gamma+\nu)}\right) + n\exp\left(-c_2 n^\gamma \tau_0^2\right) + \exp\left(-c_3 n^{1-2\nu}\right)]). \end{aligned} \tag{5.10}$$

Furthermore, under condition $(C2^{**})$, *let* $\zeta^{**} = \min_{\alpha\in\mathcal{D}^{**}} u^{**}_\alpha - \max_{\alpha\in\bar{\mathcal{D}}^{**}} u^{**}_\alpha$,

i.e. $\zeta^{**} \geq 2cn^{-\nu} \geq 0$. *Then we have*

$$Pr(\max_{\alpha \in \widehat{\mathcal{D}}^{**}} \widehat{u}_{\alpha}^{**} < \min_{\alpha \in \mathcal{D}^{**}} \widehat{u}_{\alpha}^{**})$$
$$\geq 1 - O(2p[\exp\left(-c_1 \tau_0^4 n^{1-2(\gamma+\nu)}\right) + n \exp\left(-c_2 n^{\gamma} \tau_0^2\right) + \exp\left(-c_3 n^{1-2\nu}\right)]). \tag{5.11}$$

Under conditions $(C1)$, $(C2^{**})$, $(C3)$, $(C4^{**})$, *and* $(C5^{**})$, *we have that*

$$Pr(\mathcal{D}^{**} \subseteq \widehat{\mathcal{D}}^{**})$$
$$\geq 1 - O(s_n[\exp\left(-c_1 \tau_0^4 n^{1-2(\gamma+\nu)}\right) + n \exp\left(-c_2 n^{\gamma} \tau_0^2\right) + \exp\left(-c_3 n^{1-2\nu}\right)]), \tag{5.12}$$

where s_n *is the cardinality of* $\mathcal{D}^{**}$.

Proofs of Theorems 1, 2, and 3 are in Appendix A.3.

6. Numerical Studies

In this section, we evaluate the performance of our proposed approaches through simulations. We utilize the CDC, CDC_1, and CDC_2 approaches to represent the marginal procedure (CFS_M), one-stage procedure (CSVS_1) and two-stage procedure (CSVS_2), respectively, by employing the newly introduced censored dependency measurements. In our simulation, we estimate $G(T|\mathbf{X})$ using the unconditional Kaplan-Meier estimator $G(t)$ for the sake of simplicity. We compare our methods with the following ap-

proaches from the literature: QaSIS (He, Wang and Hong, 2013), a quantile adaptive model-free variable screening, where we set the quantile to 0.5; CSIR (Zhou and Li, 2017); RCDCS (Chen, Chen and Wang, 2018); CRSIS (Zhang, Liu and Wu, 2017); SVSIR (Yan, Tang and Zhao, 2017); CRIS (Song et al., 2014); KM (Liu, Zhang and Zhao, 2018; Mai and Zou, 2015); IPOD (Hong, et al., 2018a); LQ (Hong, Kang and Li, 2018b); and ISIS_C (Fan, Feng and Wu , 2010).

For each simulated model, we set $n = \{100, 200\}$ and $p = \{1000, 2000\}$. The presented outcomes are based on 500 replicates and rely on $\mathcal{P}_s$ and $\mathcal{P}_a$, which represent the proportions of individually active predictors being selected and the proportion of all active predictors being selected, respectively. Note that the results tend to be more favorable when $\mathcal{P}_s$ and $\mathcal{P}_a$ approach 1.

Example 1. *Suppose that the survival time T follows a Cox proportional hazards model with a conditional hazard function similar to that of Zhang, Liu and Wu (2017), Chen, Chen and Wang (2018), and Liu, Zhang and Zhao (2018):*

$$h(t|\mathbf{X}) = h_0(t) \exp\left(\mathbf{X}^\top \beta\right),$$

where $h_0(t) = (t - 0.5)^2$, $\mathbf{X} = (X_1, \ldots, X_p)^\top \sim N(0, \Sigma)$, and $\Sigma = (\sigma_{ij})_{p\times p}$ with $\sigma_{ij} = 0.8^{|i-j|}$. We set $\beta^\top = (0.35, 0.35, 0.35, 0.35, 0.35, 0, \ldots, 0)$, i.e.,

Table 1: Simulation results of $\mathcal{P}_s$ and $\mathcal{P}_a$ for Example 1.

	$\mathcal{P}_s$					$\mathcal{P}_a$	$\mathcal{P}_s$					$\mathcal{P}_a$
	X_1	X_2	X_3	X_4	X_5	ALL	X_1	X_2	X_3	X_4	X_5	ALL
	$n=100, p=1000, CR\approx 0.019$						$n=200, p=1000, CR\approx 0.019$					
CDC	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
CDC_1	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
CDC_2	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
QaSIS	0.976	0.998	1.000	1.000	0.986	0.960	1.000	1.000	1.000	1.000	1.000	1.000
CSIR	0.722	0.858	0.876	0.852	0.698	0.516	0.992	0.998	1.000	1.000	0.996	0.988
RCDCS	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
CRSIS	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
SVSIR	0.982	1.000	1.000	1.000	0.986	0.970	1.000	1.000	1.000	1.000	1.000	1.000
CRIS	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
KM	0.998	1.000	0.998	1.000	0.994	0.994	0.996	1.000	1.000	1.000	0.998	0.996
IPOD	1.000	1.000	1.000	1.000	0.994	0.994	1.000	1.000	1.000	1.000	1.000	1.000
LQ	0.994	0.990	0.986	0.986	0.996	0.952	1.000	1.000	1.000	1.000	1.000	1.000
$ISIS_C$	0.868	0.884	0.922	0.892	0.840	0.508	0.976	0.982	0.976	0.966	0.962	0.862

only the first five predictors are active. The censoring time follows a uniform distribution $C \sim U(0, 100)$.

We also consider the survival time T from the transformation model in Example 1 in Appendix A.4 of the supplementary file. The outcomes presented in Table 1 (and Tables 1-2 in Appendix A.4) reveal that the CDC, CDC_1, and CDC_2 methods exhibit similar performance compared to other existing screening methods. One may question the necessity of using the proposed procedures. To explore this further, we consider the following example.

Example 2. *For each of the four models in this example, we assume* $\mathbf{X} \sim N(0, \Sigma_p)$, *where* $\Sigma_p = (\sigma_{ij})$. *Specifically,* $\sigma_{ii} = 1$ *for* $i = 1, \cdots, p$, $\sigma_{i4} = \sigma_{4j} = \rho^{\frac{7}{10}}$ *for* $i \neq 4$, *and* $\sigma_{ij} = \rho$ *for* $i \neq j, i \neq 4, j \neq 4$. *Additionally, we*

assume that $\epsilon \sim N(0,1)$ *and is independent of* $\mathbf{X}$. *Under this setup,* X_4 *is marginally independent of* Y, *while it becomes an active predictor when* $\rho \neq 0$. *For each model, censoring time* C *is independently generated from two different normal distributions, resulting in censoring rates of approximately* $0.01 \sim 0.03$ *and* $0.1 \sim 0.3$, *respectively. The models and the distributions of* C *are outlined as follows:*

(a). $Y = \max\{X_1+X_2+X_3-3\rho^{\frac{7}{10}}e^{(X_4-\frac{1}{2})}+10+\epsilon, 0\}$, $C \sim N(13,1), N(10,1)$

(b). $Y = \max\{X_1+X_2+X_3-\frac{3}{5}\rho^{\frac{7}{10}}X_4^3+10+\epsilon, 0\}$, $C \sim N(15,1), N(12,1)$

(c). $Y = \max\{X_1+X_2+X_3-3\rho^{\frac{7}{10}}X_4+10+\epsilon, 0\}$, $C \sim N(14,1), N(11,1)$

(d). $\log Y = X_1+X_2+X_3-3\rho^{\frac{7}{10}}X_4+\epsilon$, $C \sim N(100,1), N(7,1)$

The predictors X_4 in models (a), (b), and (c) correspond to exponential, cubic, and linear terms, respectively. Model (d) represents a log-linear regression model. Due to space constraints, only a portion of the results for model (a) are presented in Table 2, while the results for the other simulations and models can be found in Tables 3-9 within Appendix A.4. Across all methods, performance is generally improved under moderate correlation ($\rho = 0.5$) compared to scenarios with high correlation $\rho = 0.8$. As expected, the simulation results clearly indicate that methods exclusively focusing on the marginal relationship between $\mathbf{X}$ and Y are highly unlikely to detect X_4. Both ISIS_C and the proposed CDC_1 and CDC_2 procedures demonstrate a

high probability of identifying X_4. However, ISIS_C performs suboptimally when the correlation is strong ($\rho = 0.8$).

When the censoring rate is low, the CDC_1 and CDC_2 methods exhibit superior performance compared to the other methods. They not only select the active variable X_4 but also enhance their ability to screen the marginal variables (selecting X_1, X_2, and X_3). In high censoring rate situations, the CDC_1 and CDC_2 approaches are more likely to identify X_4 with higher probability than ISIS_C, particularly when the correlation is strong ($\rho = 0.8$), although their ability to screen marginal variables is poor. This behavior is expected because our proposed methods employ the inverse probability-censoring weighted technique to address censoring. Essentially, the proposed methods leverage noncensored observations in their calculation, which may result in some efficiency loss when the censoring proportion is high. A similar conclusion can be found in Song et al. (2014).

The preceding observation inspired us to combine the existing marginal screening techniques with our proposed one-stage (CSVS_1) and two-stage (CSVS_2) methodologies when the censoring rate is higher. Note that we still choose the CDC measure as the dependence measure in both the CSVS_1 and CSVS_2 approaches. This choice is based on its capacity to assess the dependence relationship between two random vectors of arbitrary dimen-

Table 2: Simulation results of $\mathcal{P}_s$ and $\mathcal{P}_a$ for Example 2 (a)

	$r=0.5$					$r=0.8$				
	$\mathcal{P}_s$				$\mathcal{P}_a$	$\mathcal{P}_s$				$\mathcal{P}_a$
	X_1	X_2	X_3	X_4	ALL	X_1	X_2	X_3	X_4	ALL
$n=200, p=1000, d=[n/log(n)], CR \approx 0.01 \sim 0.03$										
CDC	0.998	1.000	1.000	0.034	0.034	0.972	0.980	0.988	0.048	0.042
CDC_1	0.998	1.000	1.000	1.000	0.998	0.972	0.978	0.988	1.000	0.938
CDC_2	0.998	1.000	1.000	1.000	0.998	0.972	0.978	0.988	1.000	0.938
QaSIS	0.946	0.950	0.974	0.698	0.600	0.314	0.358	0.334	0.762	0.040
CSIR	0.688	0.734	0.722	0.010	0.004	0.400	0.440	0.440	0.002	0.000
RCDCS	1.000	1.000	1.000	0.004	0.004	0.964	0.960	0.962	0.012	0.008
CRSIS	1.000	1.000	1.000	0.002	0.002	0.964	0.932	0.922	0.000	0.000
SVSIR	1.000	1.000	1.000	0.000	0.000	0.982	0.976	0.980	0.000	0.000
CRIS	1.000	1.000	1.000	0.000	0.000	0.980	0.982	0.984	0.000	0.000
KM	0.978	0.984	0.974	0.098	0.092	0.556	0.574	0.578	0.172	0.024
IPOD	0.990	0.990	0.986	0.088	0.084	0.614	0.624	0.634	0.178	0.036
LQ	0.980	0.976	0.988	0.160	0.140	0.302	0.356	0.308	0.252	0.006
$ISIS_C$	0.990	1.000	1.000	0.993	0.986	0.716	0.723	0.720	0.870	0.440
$n=200, p=1000, d=[n/log(n)], CR \approx 0.1 \sim 0.3$										
CDC	0.796	0.802	0.802	0.026	0.010	0.842	0.834	0.818	0.050	0.032
CDC_1	0.794	0.800	0.800	0.942	0.460	0.832	0.828	0.814	0.990	0.546
CDC_2	0.794	0.800	0.800	1.000	0.478	0.832	0.828	0.814	1.000	0.552
QaSIS	0.634	0.618	0.620	0.460	0.132	0.234	0.308	0.282	0.696	0.028
CSIR	0.714	0.774	0.760	0.010	0.004	0.396	0.448	0.438	0.002	0.000
RCDCS	1.000	0.998	1.000	0.008	0.008	0.950	0.948	0.944	0.010	0.006
CRSIS	0.998	1.000	1.000	0.002	0.002	0.930	0.912	0.894	0.000	0.000
SVSIR	1.000	1.000	1.000	0.000	0.000	0.974	0.972	0.978	0.000	0.000
CRIS	0.998	1.000	1.000	0.002	0.002	0.978	0.978	0.984	0.000	0.000
KM	0.988	0.994	0.978	0.088	0.088	0.562	0.576	0.580	0.162	0.018
IPOD	0.988	0.992	0.986	0.086	0.084	0.624	0.622	0.632	0.180	0.032
LQ	0.966	0.976	0.970	0.148	0.126	0.312	0.340	0.314	0.254	0.006
$ISIS_C$	0.990	0.983	0.983	0.986	0.956	0.656	0.670	0.620	0.793	0.300

sions, whereas other existing screening methods are limited to univariate random variables.

In essence, within the algorithm, we combine the marginal screening methods with the CDC_1 and CDC_2 procedures. For example, RCDCS_1^M represents the combination of RCDCS with the CDC_1 procedure, and RCDCS_2^M represents the amalgamation of RCDCS with the CDC_2 procedure. The results presented in Table 3 (and Tables 10-13 in Appendix A.4) demonstrate that the combined procedures exhibit enhanced performance, particularly when the censoring rate is higher.

Drawing upon the simulation results, it is evident that the proposed CDC_1 and CDC_2 methodologies generally perform well when the censoring rate is low or the sample size n is relatively large. Consequently, we make the following recommendations: When the censoring rate is low (below 5%), we advocate using the CDC_1 or CDC_2 procedure. Conversely, when the censoring rate is high (exceeding 5%), we recommend employing a procedure that combines the CDC_1 or CDC_2 approach with the RCDCS, CRSIS, SVSIR, or CRIS method. Beyond the aforementioned considerations, it is essential to acknowledge that achieving a more accurate screening result through sufficient variable screening procedures requires a more extensive computational duration.

Table 3: Combination of existing methods and one-stage, two-stage algorithms:

Example 2 (a)

	$r=0.5$					$r=0.8$				
	$\mathcal{P}_s$				$\mathcal{P}_a$	$\mathcal{P}_s$				$\mathcal{P}_a$
	X_1	X_2	X_3	X_4	ALL	X_1	X_2	X_3	X_4	ALL
	$n=200, p=1000, d=[n/log(n)], CR \approx 0.1 \sim 0.3$									
QaSIS_1^M	0.622	0.610	0.610	0.952	0.270	0.226	0.284	0.276	0.998	0.032
QaSIS_2^M	0.622	0.610	0.610	1.000	0.274	0.226	0.284	0.276	1.000	0.032
CSIR_1^M	0.706	0.764	0.750	0.942	0.380	0.386	0.438	0.432	0.988	0.062
CSIR_2^M	0.706	0.764	0.748	1.000	0.394	0.386	0.438	0.432	1.000	0.062
RCDCS_1^M	1.000	0.996	1.000	0.930	0.926	0.946	0.946	0.938	0.990	0.826
RCDCS_2^M	1.000	0.996	1.000	1.000	0.996	0.946	0.946	0.938	1.000	0.836
CRSIS_1^M	0.998	1.000	1.000	0.940	0.938	0.924	0.904	0.892	0.988	0.732
CRSIS_2^M	0.998	1.000	1.000	1.000	0.998	0.924	0.904	0.892	1.000	0.742
SVSIR_1^M	1.000	1.000	1.000	0.94	0.94	0.974	0.970	0.978	0.988	0.910
SVSIR_2^M	1.000	1.000	1.000	1.000	1.000	0.974	0.970	0.978	1.000	0.922
CRIS_1^M	0.998	1.000	1.000	0.94	0.938	0.976	0.978	0.984	0.988	0.926
CRIS_2^M	0.998	1.000	1.000	1.000	0.998	0.976	0.978	0.984	1.000	0.938
KM_1^M	0.988	0.994	0.978	0.946	0.912	0.55	0.568	0.566	0.994	0.178
KM_2^M	0.988	0.994	0.978	1.000	0.962	0.55	0.568	0.566	1.000	0.182
IPOD_1^M	0.988	0.99	0.986	0.946	0.912	0.608	0.62	0.620	0.994	0.246
IPOD_2^M	0.988	0.99	0.986	1.000	0.964	0.608	0.62	0.620	1.000	0.252
LQ_1^M	0.964	0.976	0.970	0.948	0.868	0.300	0.336	0.300	0.992	0.040
LQ_2^M	0.964	0.976	0.970	1.000	0.914	0.300	0.336	0.300	1.000	0.042

7. Real Data Analysis

In this paper, we analyze two real-world datasets. One is a neuroblastoma dataset, while the other one is a diffuse large-B-cell lymphoma (DLBCL) dataset. The first dataset analysis is included in the main text, while the other one is in the appendix due to limited space.

Neuroblastoma is a malignant pediatric tumor originated from neural crest elements of the sympathetic nervous system. It predominantly affects the pediatric population and can even manifest in infancy. Our aim is to identify the genes that influence patient survival, as studied by Fan, Feng and Wu (2010). They examined neuroblastoma data using an initial variable screening stage followed by the application of the smoothly clipped absolute deviation (SCAD) penalty stage, which involves tuning parameter selection through cross-validation. Thus, the selection of different tuning parameters may result in significantly different sets of selected genes. To ensure a fair comparison, we focus on the screening stage. All gene expression levels are standardized to have a mean of zero and a standard deviation of one during the exploratory data analysis.

The normalized neuroblastoma data is obtained from the MicroArray Quality Control phase-II (MAQC-II) project and can be accessed through the ArrayExpress database (`https://www.ebi.ac.uk/arrayexpress/`; Ac-

cession: E-TABM-38). The dataset consists of 255 patients aged from 0 to 296 months at the time of diagnosis, with a median age of 15 months. Among these patients, 160 patients exhibit gene expression at 10163 probe sites. The censoring rate is approximately 36%. Our focus is on the association between gene expression and overall survival time.

Following Zhou and Li (2017), we randomly split the dataset into a training set with $n_1 = 106$ patients and a testing set with the remaining $n_2 = 54$ patients. Various screening procedures are employed for the training data, and the top $d = [n_1/log(n_1)] = 22$ genes are ultimately retained. Detailed information about the gene IDs selected using various variable screening methods on the training set can be found in Appendix A.4. Table 4 summarizes the gene IDs that are commonly selected across multiple screening methods. Notably, genes 5926 and 447 are commonly selected by the marginal method CDC, as well as by RCDCS$_1^M$, CRSIS$_1^M$, SVSIR$_1^M$, and CRIS$_1^M$. Similarly, genes 2292 and 6278 are commonly selected by the variants RCDCS$_2^M$, CRSIS$_2^M$, SVSIR$_2^M$,CRIS$_2^M$, and CDC$_2$. This suggests that the aforementioned genes have a close relationship with overall survival and warrant further exploration. However, these genes are not selected by the marginal screening methods.

Using the testing data, we evaluate the prediction performance of these

screening approaches based on the $d = 22$ selected genes. Since the true model is unknown, we adopt the Cox proportional-hazards model as a working model. Following Bair and Tibshirani (2004), patients can be classified into low-risk and high-risk subgroups. The log-rank test is employed to test the null hypothesis of no disparity in survival between these subgroups, and the p-values are presented in the final column of Table 5. Based on the p-values, all screening methods successfully identify subgroups. The C-statistic (Uno et al., 2011) is employed to evaluate the predictive power of the resultant models and is defined as $C_n = Pr(\beta^t \mathbf{X}_i > \beta^t \mathbf{X}_j | T_i < T_j)$, where $\beta^t \mathbf{X}_i$ is the risk score of the i-th patient. A higher C-statistic indicates better predictive power. Heuristically, $C_n = 0.5$ implies no predictive power. Table 5 presents the C-statistic values, along with their 95% lower and upper bounds. The standard deviation (SD) of the C-statistic is obtained through a perturbation resampling method with 200 replicates.

As shown in Table 5, the marginal screening methods, namely, CSIR, CRSIS, CRIS and LQ, yield overall higher C-statistic values compared to the other methods. However, due to the high censoring rate in real data, the performance of CDC is not as excellent as the aforementioned methods, although it remains competitive. Conversely, the performance of ISIS_C lags behind, partly due to its design for parametric models and its vulnerabil-

Table 4: Gene IDs that are commonly selected by screening methods.

Gene IDs	Methods
{5926,447}	{CDC,RCDCS$_1^M$, CRSIS$_1^M$, SVSIR$_1^M$,CRIS$_1^M$}
{2292,6278}	{RCDCS$_2^M$, CRSIS$_2^M$, SVSIR$_2^M$,CRIS$_2^M$, CDC$_2^M$}
{3633}	{RCDCS, CRSIS, CRIS ,IPOD, KM, and LQ}
{870}	{RCDCS, SVSIR, CSIR,LQ}
{1977}	{RCDCS, CRSIS, ISIS}
{8312}	{RCDCS, CRSIS, SVSIR}
{5746}	{RCDCS, CRSIS, CSIR}
{8727}	{RCDCS, KM, IPOD}
{2397, 5960}	{RCDCS, CRSIS}
{3846}	{RCDCS, SVSIR}
{10375}	{ SVSIR, CDC}
{7207}	{RCDCS,CRIS}
{659 }	{ RCDCS, CDC}
{2353}	{ CRSIS, CDC}
{1000}	{ SVSIR, CRSIS}
{8814 }	{ KM, LQ}

ity to model misspecification. In addition, most combination procedures, where a marginal procedure is coupled with a one-stage or two-stage procedure, outperform the marginal methods. For example, measures combined with the one-stage procedure, such as RCDCS, SVSIR, and IPOD. exhibit slightly superior performance in terms of a higher C-statistic value compared to the marginal methods. Similarly, the QaSIS, RCDCS, KM, and IPOD procedures, which are coupled with the two-stage procedure, demonstrate slightly enhanced performance compared to the marginal methods. These results confirm that the proposed one-stage and two-stage procedures enhance not only variable screening performance but also the overall performance when fused with other marginal screening methods.

Table 5: The C-statistic for different screening methods.

	C-statistic	standard error	lower bound of C	upper bound of C	p value
QaSIS	0.895	0.099	0.702	1.089	4.013e-05
QaSIS_1^M	0.888	0.182	0.532	1.245	1.427e-03
QaSIS_2^M	0.913	0.085	0.747	1.079	1.204e-05
CSIR	0.992	0.153	0.693	1.291	2.016e-04
CSIR_1^M	0.952	0.188	0.583	1.320	1.041e-04
CSIR_2^M	0.961	0.191	0.587	1.335	8.443e-05
RCDCS	0.865	0.168	0.536	1.194	2.431e-07
RCDCS_1^M	0.891	0.147	0.602	1.179	2.487e-05
RCDCS_2^M	0.891	0.104	0.686	1.095	3.472e-07
CRSIS	0.971	0.138	0.700	1.241	2.731e-03
CRSIS_1^M	0.966	0.152	0.669	1.264	4.981e-05
CRSIS_2^M	0.965	0.135	0.700	1.230	2.635e-05
SVSIR	0.908	0.146	0.621	1.195	1.425e-03
SVSIR_1^M	0.928	0.161	0.612	1.244	2.297e-03
SVSIR_2^M	0.899	0.217	0.474	1.324	6.419e-04
CRIS	1.000	0.004	0.992	1.008	6.513e-06
CRIS_1^M	1.000	0.003	0.992	1.007	4.221e-06
CRIS_2^M	1.000	0.002	0.997	1.003	1.739e-05
KM	0.933	0.154	0.631	1.234	3.191e-04
KM_1^M	0.934	0.153	0.633	1.234	3.192e-04
KM_2^M	0.938	0.168	0.610	1.267	8.983e-06
IPOD	0.900	0.143	0.620	1.179	1.824e-05
IPOD_1^M	0.907	0.106	0.699	1.115	1.956e-07
IPOD_2^M	0.904	0.113	0.684	1.125	5.547e-06
LQ	1.000	0.003	0.995	1.005	1.128e-03
LQ_1^M	0.972	0.099	0.778	1.166	1.302e-04
LQ_2^M	0.973	0.154	0.670	1.275	2.968e-05
CDC	0.913	0.124	0.669	1.157	2.042e-06
CDC_1^M	0.913	0.102	0.712	1.114	2.042e-06
CDC_2^M	0.901	0.084	0.738	1.065	8.794e-07
ISIS_c	0.892	0.097	0.702	1.081	6.131e-06

8. Discussion

In this paper, we propose two censored sufficient variable screening algorithms (CSVS_1 and CSVS_2), specifically designed for ultrahigh-dimensional right censored data. These innovative methods are particularly useful for detecting active predictors that exhibit marginal independence from the response variable, a task where conventional marginal methodologies struggle. Although the iterative ISIS_C algorithm introduced by Fan, Feng and Wu (2010) may yield similar results to those of CSVS_1 and CSVS_2 in certain scenarios, it typically lacks theoretical underpinnings. In contrast, our proposed procedures draw upon the extensive literature on sufficient dimension reduction (Yin and Hilafu, 2015), providing them with a robust theoretical foundation for the selection of all active predictors. Illustratively, we employ the censored version of the distance covariance measure (CDC) to demonstrate the sure screening properties of our proposed methods.

Based on simulation studies, we demonstrate that when the censoring rate is low, our CDC_1 and CDC_2 methods outperform existing methods. Conversely, when the censoring rate is high, the CDC_1 and CDC_2 procedures outshine in identifying informative variables that maintain marginal independence from the response variable, especially when the correlation is high $\rho = 0.8$. However, it is worth noting that their effectiveness in marginal

screening is hampered by the use of the inverse probability-censoring weights. In light of these findings, we are motivated to propose a hybrid approach that combines the strengths of marginal screening methods with CDC_1 and CDC_2, yielding superior performance. Developing other approaches to handling a high censoring rate for sufficient variable screening is a potential research direction.

Beyond the aforementioned advantages, it is imperative to highlight that CSVS_1 and CSVS_2 are illustrated through the DC measure, as it can handle dependence relationships between two random vectors of arbitrary dimensions. However, it is essential to recognize that DC represents a specialized instance of the Hilbert-Schmidt independence criterion (HSIC), a framework based on the embedding of probability distributions into reproducing kernel Hilbert spaces (RKHS, Gretton et al. (2005)). While it is conceivable to extend the purview of CSVS_1 and CSVS_2 to encompass HSIC, such an extension is by no means a trivial task. Empirical evidence from Yang et al. (2019)'s simulations suggests that the proposed method based on HSIC may not be sufficiently robust in ultrahigh-dimensional settings. Inspired by the work of He et al. (2021), we are encouraged to explore feature screening procedures rooted in HSIC, incorporating diverse kernel functions and considering the interplay between p and n. Future research

regarding this topic could be conducted.

Supplementary Material

Supplementary material is available that includes proofs of theorems, additional simulation instances, real data analysis, and corresponding results.

Acknowledgments

The authors would like to thank the Editor, an Associate Editor and two referees for their valuable comments and helpful suggestions, which greatly improved the paper, and Byran Smucker (Department of Statistics, Miami University) for reading the final version.

References

Bair, E. and Tibshirani, R. (2004). Semi-supervised methods to predict patient survival from gene expression data. *PLOS Biology* **2**, E108.

Chang, J., Tang, C. Y. and Wu, Y.(2013). Marginal empirical likelihood and sure independence feature screening. *Annals of statistics* **41**, 2123–2148.

Chen, X. (2018). Model-free conditional feature screening for ultra-high dimensional right censored data. *Journal of Statistical Computation and Simulation* **88**, 2425–2446.

Chen, X., Chen, X. and Wang, H.(2018). Robust feature screening for ultra-high dimensional right censored data via distance correlation. *Computational Statistics and Data Analysis* **119**, 118–138.

Cook, R. D. and Ni, L. (2005). Sufficient dimension reduction via inverse regression: A minimum discrepancy approach. *Journal of the American Statistical Association* **100**, 410–428.

Fan, J., Feng, Y. and Wu, Y. (2010). High-dimensional variable selection for Cox proportional hazards model. In *Borrowing strength: Theory powering applications–a Festschrift for Lawrence D. Brown*(Edited by Berger, J. O., Cai, T. T. and Johnstone, I. M.), 70–86. Institute of Mathematical Statistics Collections.

Fan, J. and Lv. J. (2008). Sure independence screening for ultrahigh dimensional feature space. *Journal of the Royal Statistical Society: Series B* **70**, 849–911.

Fan, J. and Song, R. (2010). Sure independence screening in generalized linear models with np-dimensionality. *The Annals of Statistics* **38**, 3567–3604.

Gannoun, A. and Saracco, J. (2003). An asymptotic theory for SIR α method. *Statistica Sinica* **13**, 297–310.

Gretton, A., Bousquet, O., Smola, A. and Schölkopf, B. (2005) Measuring statistical dependence with Hilbert-Schmidt norms. In*International Conference on Algorithmic Learning Theory*(Edited by Jain, S., Simon, H.U., Tomita, E.), 63-77. Lecture Notes in Computer Science (LNAI,volume 3734), Springer, Berlin, Heidelberg.

He, X., Wang, L. and Hong, H. G. (2013). Quantile-adaptive model-free variable screening for high-dimensional heterogeneous data. *Annals of Statistics* **41**, 342–369.

He, T., Zhong, P. L., Cui, Y. and Mandrekarong, V. (2021). Unified Tests for Nonparametric Functions in RKHS With Kernel Selection and Regularization. *Statistica Sinica* **33**, 919–944.

Hong, H. G., Chen, X. R., Christiani, D. C. and Li, Y. (2018). Integrated powered density: Screening ultrahigh dimensional covariates with survival outcomes. *Biometrics* **74**, 421–429.

Hong, H. G., Kang, J. and Li, Y. (2018). Conditional screening for ultra-high dimensional covariates with survival outcomes. *Lifetime data analysis* **24**, 45–71.

Li, K. C. (1991). Sliced inverse regression for dimension reduction. *Journal of the American Statistical Association* **86**, 316–327.

Li, R., Zhong, W. and Zhu, L. (2012). Feature screening via distance correlation learning. *Journal of the American Statistical Association* **107**, 1129–1139.

Liu, J., Li, R. and Wu, R. (2014). Feature selection for varying coefficient models with ultrahigh-dimensional covariates. *Journal of the American Statistical Association* **109**, 266–274.

Liu, J., Zhang, J. and Zhao, X. (2018). A new nonparametric screening method for ultrahigh-dimensional survival data. *Computational Statistics and Data Analysis* **119**, 74–85.

Lyons, R. (2013). Distance covariance in metric spaces. *Annals of Probability* **41**, 3284–3305.

Lu, S. H. and Li. K. (2011). Sufficient dimension reduction for censored regressions. *Biometrics* **67**, 513–523.

Mai, Q. and Zou. H. (2015). The fused kolmogorov filter: A nonparametric model-free screening method. *The Annals of Statistics* **43**, 1471–1497.

Robins, J. M., Rotnitzky, A. and Zhao, L. P. (1994). Estimation of regression coefficients when some regressors are not always observed. *Journal of the American statistical Association* **89**, 846–866.

Shao, X. and Zhang. J. (2014). Martingale difference correlation and its use in high-dimensional variable screening. *Journal of the American Statistical Association* **109**, 1302–1318.

Song, R., Lu, W., Ma. S. and Jeng, X. J. (2014). Censored rank independence screening for high-dimensional survival data. *Biometrika* **101**, 799–814.

Székely, G., Rizzo, M. and Bakirov, N. K. (2007). Measuring and testing dependence by correlation of distances. *The Annals of Statistics* **35**, 2769–2794.

Tsiatis, A. (2007). *Semiparametric Theory and Missing Data*. 1st Edition. Springer Series in Statistics.

Uno, H., Cai, T., Pencina, M. J., D'Agostino, R. B. and Wei, L. J. (2011). On the c-statistics for evaluating overall adequacy of risk prediction procedures with censored survival data. *Statistics in Medicine* **30**, 1105–1117.

Yan, X., Tang, N. and Zhao, X. (2017). The spearman rank correlation screening for ultrahigh dimensional censored data. *arXiv*:1702.02708.

Yang, B., Yin, X. and Zhang, N. (2019). Sufficient variable selection using independence measures for continuous response. *Journal of Multivariate Analysis* **173**, 480–493.

Yin, X. and Hilafu. H. (2015). Sequential sufficient dimension reduction for large p, small n problems. *Journal of the Royal Statistical Society: Series B* **77**, 879–892.

Zhang, J., Liu, Y. and Wu, Y. (2017). Correlation rank screening for ultrahigh-dimensional survival data. *Computational Statistics and Data Analysis* **108**, 121–132.

Zhang, J., Yin, G., Liu, Y. and Wu, Y. (2018). Censored cumulative residual independent screening for

ultrahigh-dimensional survival data. *Lifetime data analysis* **24**, 273–292.

Zhou, T. and Zhu, L. (2017). Model-free feature screening for ultrahigh dimensional censored regression. *Statistics and Computing* **27**, 947–961.

Zhu, L. P., Li, L., Li, R. and Zhu, L. X. (2011). Model-free feature screening for ultrahigh-dimensional data. *Journal of the American Statistical Association* **106**, 1464–1475.

Baoying Yang, Department of Statistics, College of Mathematics, Southwest Jiaotong University, Chengdu, Sichuan 611756, China.

E-mail: (yangbaoying@swjtu.edu.cn)

QingCong Yuan, Department of Statistics, Miami University, Oxford, OH 45056.

AND Biostatistics and Programming, Sanofi, USA.

E-mail: (qingcong.yuan@gmail.com)

Xiangrong Yin, Department of Statistics, University of Kentucky, Lexington, KY 40536.

Professor Xiangrong Yin passed away in August of 2020. He had made significant contributions to the development of this paper.