The Annals of Statistics
2023, Vol. 51, No. 5, 1965–1988
https://doi.org/10.1214/23-AOS2313

EFFICIENT ESTIMATION OF THE MAXIMAL ASSOCIATION BETWEEN MULTIPLE PREDICTORS AND A SURVIVAL OUTCOME

BY TZU-JUNG HUANG[1,a], ALEX LUEDTKE[2,b] AND IAN W. MCKEAGUE[3,c]

[1]*Vaccine and Infectious Disease Division, Fred Hutchinson Cancer Research Center,* [a]*thuang2@fredhutch.org*
[2]*Department of Statistics, University of Washington,* [b]*aluedtke@uw.edu*
[3]*Department of Biostatistics, Columbia University,* [c]*im2131@columbia.edu*

This paper develops a new approach to post-selection inference for screening high-dimensional predictors of survival outcomes. Post-selection inference for right-censored outcome data has been investigated in the literature, but much remains to be done to make the methods both reliable and computationally-scalable in high dimensions. Machine learning tools are commonly used to provide *predictions* of survival outcomes, but the estimated effect of a selected predictor suffers from confirmation bias unless the selection is taken into account. The new approach involves the construction of semiparametrically efficient estimators of the linear association between the predictors and the survival outcome, which are used to build a test statistic for detecting the presence of an association between any of the predictors and the outcome. Further, a stabilization technique reminiscent of bagging allows a normal calibration for the resulting test statistic, which enables the construction of confidence intervals for the maximal association between predictors and the outcome and also greatly reduces computational cost. Theoretical results show that this testing procedure is valid even when the number of predictors grows superpolynomially with sample size, and our simulations support this asymptotic guarantee at moderate sample sizes. The new approach is applied to the problem of identifying patterns in viral gene expression associated with the potency of an antiviral drug.

1. Introduction. The problem of identifying associations between high-dimensional predictors and a survival outcome is of great interest in the biomedical sciences. In virology, for example, the potency of an antiviral drug (in controlling viral replication) is typically assessed in terms of a type of survival time outcome, and it is important to identify associations between patterns of viral gene expression and the drug's potency (Gilbert et al. (2017)). In cancer genomics, patterns of patients' gene expression can also influence survival time outcomes. Diffuse large B-cell lymphoma, for instance, has been studied with the aim of identifying such patterns from massive collections of gene-expression data (Rosenwald et al. (2002); Bøvelstad, Nygård and Borgan (2009)). In earlier work (Huang, McKeague and Qian (2019)), we introduced an approach to this general problem based on marginal accelerated failure time modeling. In the present paper, we expand this approach to provide a semiparametrically efficient and more computationally tractable method that can handle the screening of extremely large numbers of predictors (as is typical with gene expression data).

Our approach is based on marginal screening of the predictors, and specifying the link between the survival outcome and the predictors by a general semiparametric accelerated failure time (AFT) model that does not make any distributional assumption on the error term. The error term is merely taken to be uncorrelated with the predictors (i.e., the so-called *assumption lean* linear model setting). Let T be the (log-transformed) time-to-event outcome, and $\boldsymbol{U} = (U_1, \ldots, U_p)^T$ denote a p-dimensional vector of predictors. Note that $p = p_n$ can grow

Received January 2023.
MSC2020 subject classifications. Primary 62N03, 62G10; secondary 62G20.
Key words and phrases. Marginal screening, post-selection inference, semiparametric efficiency.

with n, but we omit the subscript n throughout for notational simplicity unless otherwise stated. The AFT model takes the form

$$T = \alpha_0 + \boldsymbol{U}^T \boldsymbol{\beta}_0 + \varepsilon, \tag{1}$$

where $\alpha_0 \in \mathbb{R}$ is an intercept, $\boldsymbol{\beta}_0 \in \mathbb{R}^p$ is a vector of slope parameters and ε is the zero-mean error term that is uncorrelated with $\boldsymbol{U}$. The transformed survival outcome T is possibly right-censored by C, and we only observe $X = \min\{T, C\}$ and $\delta = 1(T \leq C)$. The problem is to test the global null $\boldsymbol{\beta}_0 = 0$. We emphasize that model (1) is locally nonparametric (van der Laan and Robins (2003)) and holds without distributional assumptions (such as independent errors) apart from mild moment conditions, by defining the second term as the L_2-projection of T onto the linear span of $U_1, \ldots, U_p$.

An especially attractive feature of the AFT model is that the marginal association between T and each predictor can be represented directly in terms of a correlation, and does not require any structural assumptions. This allows us to reduce the high-dimensional screening problem (involving all p components of $\boldsymbol{\beta}_0$) to a single test of whether the most correlated predictor with T is significant. A popular approach to the screening of predictors in survival analysis is to use relative or excess conditional hazard function representations of associations. However, the AFT approach has the advantage that a lack of any marginal correlation implies the absence of all correlation between T and $\boldsymbol{U}$ (under the mild assumption that the covariance matrix of $\boldsymbol{U}$ is invertible); in the hazard-rate setting, there is no such connection and the semiparametric model needs to hold for testing methods to be useful.

Koul, Susarla and Van Ryzin (1981) (henceforth, KSV) introduced the technique of inversely weighting the observed outcomes by the Kaplan–Meier estimator of the censoring distribution, enabling the use of standard least squares estimators from the uncensored linear model. Subsequently, two additional sophisticated methods were proposed to fit the semiparametric AFT model. The Buckley–James estimator (Buckley and James (1979); Ritov (1990)) replaces the censored survival outcome by the conditional expectation of T given the data. The rank-based method is an estimating equation approach formulated in terms of the partial likelihood score function (Tsiatis (1990); Lai and Ying (1991a,b); Ying (1993); Jin et al. (2003)). A difficulty with the Buckley–James and rank-based methods is that they fail to preserve a direct link with the AFT, which is essential for marginal screening based on correlation. Our new marginal screening test will rely on finding an asymptotically efficient estimator of each marginal slope parameter; this will have a considerable advantage in terms of efficiency over the marginal screening method based on the KSV estimator (Huang, McKeague and Qian (2019)).

The marginal KSV estimators stem from regressing the estimated synthetic response $Y = \delta X / \hat{G}_n(X)$ on successive components of $\boldsymbol{U}$, where Y is regarded as an inverse probability weighted estimate, and $\hat{G}_n$ is the standard Kaplan–Meier estimator of the survival function of C. Under independent censoring, the use of least squares estimators, treating Y as a response variable, is justified in view of the uniform consistency of $\hat{G}_n$ under mild conditions (e.g., when the distribution functions of T and C have no common jumps; see Stute and Wang (1993)). Independent censoring is a common assumption in the high-dimensional screening of predictors for survival outcomes (He, Wang and Hong (2013); Song et al. (2014); Li et al. (2016)).

1.1. *Key contributions*. We now outline the various novel steps involved in developing our proposed test. In Huang, McKeague and Qian (2019), we showed that $|\operatorname{Corr}(U_k, T)|$ and $|\operatorname{Corr}(U_k, \tilde{Y})|$, for $k = 1, \ldots, p$, are maximized at a common index k. This was used to justify replacing T by $\tilde{Y} = \delta X / G(X)$ (and in turn its estimate Y) in the empirical version of the U_k-specific slope parameter $\Psi_k(P) = \operatorname{Cov}_P(U_k, T) / \operatorname{Var}_P(U_k)$, for use as a test statistic for the global null hypothesis $\boldsymbol{\beta}_0 = 0$ in the sense that it suffices to test $\Psi_k(P) = 0$ for all k.

Our aim now is to replace this test statistic by one that is more efficient (when k is treated as fixed), and also that is easier to calibrate taking the selection of k into account. Writing $\tilde{\mathbb{P}}_n$ as the empirical distribution of $(U_k, \tilde{Y})$, when G is known, the influence function of $\Psi_k(\tilde{\mathbb{P}}_n)$ will be derived from that of the sample correlation coefficient in the uncensored case (Devlin, Gnanadesikan and Kettenring (1975)). This will lead to the influence function $\mathrm{IF}_k(O|P)$ of the (inefficient) KSV estimator $\hat{\Psi}_k(\tilde{\mathbb{P}}_n)$ that replaces the unknown G in $\Psi_k(\tilde{\mathbb{P}}_n)$ by $\hat{G}_n$. This derivation will be based on the influence function of $\hat{G}_n$ and some empirical process and Slutsky-type arguments. The next step is to project $\mathrm{IF}_k(\cdot|P)$ onto the tangent space of the observation model to obtain an efficient influence function IF_k^*, which in turn will lead to an asymptotically efficient one-step estimator of $\Psi_k(P)$. This will be accomplished in part using results of van der Laan and Robins (2003) and van der Laan, Gill and Robins (2000). The one-step estimator takes the form $S_k(\hat{P}_n, \mathbb{P}_n) = \Psi_k(\hat{P}_n) + \mathbb{P}_n \mathrm{IF}_k^*(\cdot|\hat{P}_n)$, where $\hat{P}_n$ is a plug-in estimator of the various features of P that appear in IF_k^*, $\Psi_k(\hat{P}_n) = \hat{\Psi}_k(\tilde{\mathbb{P}}_n)$, and $\mathbb{P}_n$ is the empirical distribution of the data (acting as an expectation operator). Estimation of those features will involve estimation of the function $E[\tilde{Y}|U_k = u, X \geq s]$ as in van der Laan and Hubbard (1998), the empirical distribution of the predictor U_k and a local Kaplan–Meier estimator of the conditional censoring distribution given U_k.

Our main contribution is a method to calibrate the test. This will be done by introducing a stabilized version of $S_k(\hat{P}_n, \mathbb{P}_n)$ that "smooths out" the implicit selection of k, along the lines of Luedtke and van der Laan (2018) in the uncensored case. This stabilized version is constructed by taking a weighted average over subsamples, and is asymptotically equivalent to a martingale sum (provided $\hat{P}_n$ has no effect asymptotically), which leads to a standard normal limit even under growing dimensions, when $p = p_n \to \infty$ and $\log(p_n)/n^{1/4} \to 0$. Although the stabilized one-step estimator can have slightly diminished power compared with its unstabilized counterpart when k is given, at least in the uncensored case (Luedtke and van der Laan (2018)), it vastly reduces computational cost by avoiding the need for a double bootstrap (Huang, McKeague and Qian (2019)).

It would be interesting to extend the proposed methodology to handle measures of association other than a slope parameter, providing theoretical guarantees for our method in such settings would require substantial additional arguments. However, even for the maximal slope parameter we consider, our theoretical results build on dozens of lemmas whose proofs rely on carefully studying the form of remainder terms that arise in a first-order expansion of the estimated parameter. While similar results could presumably be established for other association parameters, the technical details of the needed arguments may change considerably and studying them would be greatly beyond the scope of this paper.

The most challenging step in validating our method (i.e., showing asymptotic normality of the test statistic) involves finding an exponential tail bound for a "remainder term" involving a collection of martingale integrals with integrands falling in a general class $\mathcal{H}_n$ of càglàd functions of bounded variation that depends on sample size. Although the task of bounding families of martingale integrals with *predictable* integrands commonly arises in standard survival analysis settings, the difficulty here is that we have *unpredictable* integrands (a case in which standard martingale inequalities fail). The unpredicability of such integrands is a novel aspect of the problem that, to our knowledge, has not arisen elsewhere in the survival analysis literature, but that arises here because $\hat{P}_n$ involves unpredictable elements relative to the subsamples used in the stabilization.

More specifically, let $N_n(s)$ be the aggregated counting process given by the number of censored observations occurring by time s, as s varies over the follow-up period $\mathcal{T}$, and denote its compensator by A_n and corresponding martingale $\bar{M} = N_n - A_n$. Then we show (under suitable conditions) that the upper bound on the class of martingale integrals given by

$$\sup_{h \in \mathcal{H}_n} \left| \int_{\mathcal{T}} h(s)\, d\bar{M}(s) \right|$$

is asymptotically negligible as $n \to \infty$. This result is general enough to give control over unpredictable integrands with respect to the natural filtration of the martingale, provided they can be viewed as belonging to the class of deterministic functions $\mathcal{H}_n$. In this way, bracketing entropy techniques are shown to be applicable by adapting a probability inequality bound for a family of counting process integrals due to van de Geer (1995).

In practice, the implementation of our stabilized one-step estimator to screen predictors of dimension $p = 10^6$ based on data of $n = 500$ on a single-core laptop only takes one minute. Hence, our proposed test enjoys both statistical and computational efficiency. Further, it provides an asymptotically valid confidence interval for the slope parameter of the selected predictor. As far as we know, no other competing method provides all of these features in the setting of high-dimensional marginal screening for survival outcomes.

1.2. *Prior literature.* Variable selection methods for right-censored survival data are widely available, although formal testing procedures are far less prevalent. For example, variants of regularized Cox regression have been studied by Tibshirani (1997); Fan and Li (2002); Bunea and McKeague (2005); Zhang and Lu (2007); Bøvelstad, Nygård and Borgan (2009); Engler and Li (2009); Antoniadis, Fryzlewicz and Letué (2010); Binder, Porzelius and Schumacher (2011); Wu (2012); Sinnott and Cai (2016). Penalized AFT models have been considered by Huang, Ma and Xie (2006); Datta, Le-Rademacher and Datta (2007); Johnson (2008); Johnson, Lin and Zeng (2008); Cai, Huang and Tian (2009); Huang and Ma (2010); Bradic, Fan and Jiang (2011); Ma and Du (2012); Li, Dicker and Zhao (2014). These methods ensure the consistency of variable selection only (i.e., the oracle property), and do not address the issue of post-selection inference. Fang, Ning and Liu (2017) have established asymptotically valid confidence intervals for a preconceived regression parameter in a high-dimensional Cox model after variable selection on the remaining predictors, but this does not apply to marginal screening (where no regression parameter is singled out, a priori). Yu, Bradic and Samworth (2021) recently constructed valid confidence intervals for the regression parameters in high-dimensional Cox models, but their approach also does not apply to marginal screening because it is predicated on the presence of active predictors (and also preselection of parameters of interest). Zhong, Hu and Li (2015) considered the problem for preconceived regression parameters within a high-dimensional additive risk model. Chai et al. (2019) considered the same problem in a high-dimensional AFT model. Taylor and Tibshirani (2018) proposed a method of finding post-selection corrected p-values and confidence intervals for the Cox model based on conditional testing. However, to the best of our knowledge, their method has not been explored theoretically (except in the uncensored linear regression setting with fixed design and normal errors; see Lockhart et al. (2014)). Statistical methods for variable selection based on marginal screening for survival data have been studied by Fan, Feng and Wu (2010), who extended sure independence screening to survival outcomes based on the Cox model. Their method applies to the selection of components of ultrahigh dimensional predictors, but no formal testing is available. Other relevant references include Zhao and Li (2012), Gorst-Rasmussen and Scheike (2013), He, Wang and Hong (2013), Song et al. (2014), Zhao and Li (2014), Hong, Kang and Li (2018), Li et al. (2016), Hong et al. (2018), Pan et al. (2019), Xia, Li and Fu (2019), Hong et al. (2020) and Liu, Chen and Li (2020).

1.3. *Overview.* The article is organized as follows. In Section 2, we formulate the estimation problem and introduce background material on semiparametric efficiency. The one-step efficient estimator of the target parameter is developed in Section 3 in the case of a single predictor. In Section 4, we develop an asymptotic normality result for calibrating the proposed test statistic that takes selection of the predictor into account. Various competing methods

are discussed in Section 5. Numerical results reported in Section 6 show that the proposed approach has favorable performance compared with these competing methods. In Section 7, we present an application using data on viral gene expression as related to the potency of an anti-retroviral drug for the treatment of HIV-1. Concluding remarks are given in Section 8. Proofs are placed in the Appendix and in the Supplementary Material (Huang, Luedtke and McKeague (2023)).

2. Preliminaries. First we recall the standard survival analysis model with independent right censorship. Let T and C denote a (log-transformed) survival time and censoring time, respectively, and suppose we observe n i.i.d. copies of $O = (X, \delta, \boldsymbol{U}) \sim P$, where $X = \min\{T, C\}$, $\delta = 1(T \le C)$ and $\boldsymbol{U} = (U_k, k = 1, \ldots, p)$ is a p-vector of predictors. We denote the joint distribution of $(T, \boldsymbol{U})$ by Q and the censoring distribution by G, and we also assume throughout that the censoring time C is independent of $(T, \boldsymbol{U})$. Though this joint independence assumption will be stronger than needed, it will greatly simplify the developments when $\boldsymbol{U}$ is of large dimension relative to sample size. The distribution P belongs to the statistical model $\mathcal{M}$, which is the collection of distributions P_1 parameterized by (Q_1, G_1) such that P_1 has density with respect to an appropriate dominating measure ν given by

$$\frac{dP_1}{d\nu}(x, \delta, \boldsymbol{u}) = \big[q_1(x|\boldsymbol{u})G_1(C \ge x)\big]^{\delta}\big[Q_1(T \ge x|\boldsymbol{U} = \boldsymbol{u})g_1(x)\big]^{1-\delta}q_1(\boldsymbol{u}),$$

where q_1 and g_1 are the densities of Q_1 and G_1 with respect to ν. Let the follow-up period be $\mathcal{T} = (-\infty, \tau]$. The sample space is denoted by $\mathcal{X} = \mathcal{T} \times \{0, 1\} \times \mathbb{R}^p$ and the empirical distribution on this space is denoted $\mathbb{P}_n$. Moreover, for a distribution P_1 on the support of O and a function f mapping from a realization of O to $\mathbb{R}^d$, we let $P_1 f \equiv P_1 f(O) \equiv \int f(o)\, dP_1(o)$.

Our approach to marginal screening is based on an estimator of the maximal (absolute) slope parameter from fitting a marginal linear regression of the survival outcome T against each predictor U_k. That is, we target the parameter

$$\Psi(P) \equiv \max_{k=1,\ldots,p} \big|\Psi_k(P)\big| \tag{2}$$

and the indices that attain this maximum, where $\Psi_k : \mathcal{M} \to \mathbb{R}$ is given by

$$\Psi_k(P) = \frac{\mathrm{Cov}_P(U_k, T)}{\mathrm{Var}_P(U_k)}. \tag{3}$$

In general, $\Psi_k(P)$ is not equal to the marginal regression coefficient of U_k in (1). Throughout, we assume that U_k and T have nondegenerate finite second moments. Further, in order for the target parameter to be proportional to the maximal absolute (Pearson) correlation, we implicitly assume that all the U_k are prestandardized to have unit variance—this assumption only plays an interpretive role in the sequel. The parameter $\Psi_k(P)$ can be identified in terms of the conditional mean lifetime $E[T|U_k]$ and the marginal distribution of U_k. Indeed,

$$\mathrm{Cov}_P(U_k, T) = \mathrm{Cov}_P\big(U_k, E[T|U_k]\big). \tag{4}$$

The proposed one-step estimator of $\Psi_k(P)$ that we will develop also involves estimation of G.

We will need some general concepts from semiparametric efficiency theory (e.g., Pfanzagl (1990)). Suppose we observe a general random vector $O \sim P$. Let $L_0^2(P)$ denote the Hilbert space of P-square integrable functions with mean zero. Consider a smooth one-dimensional family of probability measures $\{P_\epsilon\}$ passing through P and having score function $k \in L_0^2(P)$ at $\epsilon = 0$. The tangent space $\mathbf{T}^{\mathcal{M}}(P)$ is the $L_0^2(P)$-closure of the linear span of all such score functions k. For example, if nothing is known about P, then $P_\epsilon(do) = (1 + \epsilon k(o))P(do)$

is such a submodel for any bounded function k with mean zero (provided ϵ is sufficiently small), so $\mathbf{T}^{\mathcal{M}}(P)$ is seen to be the whole of $L_0^2(P)$ in this case.

Let $\psi : \mathcal{M} \to \mathbb{R}$ be a parameter that is pathwise differentiable at P: there exists $g \in L_0^2(P)$ such that $\lim_{\epsilon \to 0}(\psi(P_\epsilon) - \psi(P))/\epsilon = \langle g, k \rangle$, for any smooth submodel $\{P_\epsilon\}$ with score function k, as above, where $\langle \cdot, \cdot \rangle$ is the inner product in $L_0^2(P)$. The function g is called a gradient (or influence function) for ψ; the projection IF_ψ of any gradient on the tangent space $\mathbf{T}^{\mathcal{M}}(P)$ is unique and is known as the canonical gradient (or efficient influence function). The supremum of the Cramér–Rao bounds for all submodels (the information bound) is given by the second moment of $\mathrm{IF}_\psi(O)$. Furthermore, the influence function of any regular and asymptotically linear estimator must be a gradient (Proposition 2.3 in Pfanzagl (1990)).

A one-step estimator is an empirical bias correction of a naïve plug-in estimator in the direction of a gradient of the parameter of interest (Pfanzagl (1982)); when this gradient is the canonical gradient, then this results in an efficient estimator under some regularity conditions. A one-step estimator for $\psi(P)$ is constructed as follows. First, one obtains an initial estimate $\hat{P}$ of P. For any gradient $D(\hat{P})$ of the parameter ψ evaluated at $\hat{P}$, by the definition of the gradient this initial estimate satisfies

$$\psi(\hat{P}) - \psi(P) = -PD(\hat{P}) + \mathrm{Rem}_\psi(\hat{P}, P),$$

where $\mathrm{Rem}_\psi(\hat{P}, P)$ is negligible if $\hat{P}$ is close to P in an appropriate sense. As $D(P)$ has mean zero under P, we expect that $PD(\hat{P})$ is close to zero if D is continuous in its argument and $\hat{P}$ is close to P. However, the rate of convergence of $PD(\hat{P})$ to zero as sample size grows may be slower than $n^{-1/2}$. The one-step estimator aims to improve $\psi(\hat{P})$ and achieve $n^{1/2}$-consistency and asymptotically normality by adding an empirical estimate $\mathbb{P}_n D(\hat{P})$ of its deviation from $\psi(P)$. By the above, the one-step estimator $\hat{\psi} \equiv \psi(\hat{P}) + \mathbb{P}_n D(\hat{P})$ satisfies the expansion

$$\hat{\psi} - \psi(P) = (\mathbb{P}_n - P)D(\hat{P}) + \mathrm{Rem}_\psi(\hat{P}, P).$$

Under an empirical process and $L^2(P)$ consistency condition on $D(\hat{P})$, the leading term on the right-hand side is asymptotically equivalent to $(\mathbb{P}_n - P)D(P)$, which converges in distribution to a mean-zero Gaussian limit with consistently estimable covariance. The construction of this one-step estimator is generally nonunique because there is generally more than one gradient for ψ; this is true in our setting when $\psi = \Psi_k$ and we assume that C is independent of $(T, \boldsymbol{U})$. To minimize the variance of the Gaussian limit, then $D(\hat{P})$ can generally be chosen to be equal to the canonical gradient of ψ at $\hat{P}$, since under conditions the mean-square limit of the efficient influence function at $\hat{P}$ will be equal to the efficient influence function at P.

3. One-step estimator with a single predictor. Restricting attention to the case of a single predictor U_k for any given k, in this section we develop an asymptotically efficient one-step estimator of $\Psi_k(P)$. To this end, we need to estimate the involving features of P in the development of the one-step estimator, and introduce an estimator $\hat{P}_n$ that consists of the following three items:

(i) $\mathbb{Q}_n$: the empirical distribution used to estimate the marginal distribution of the given predictor U_k, denoted by Q_u.

(ii) $\hat{G}_n(\cdot|u)$: the Kaplan–Meier estimator of the censoring distribution conditional on $c(U_k) = c(u)$, with c a fixed, user-defined coarsening so that $c(U_k)$, $U_k \sim Q_u$, is a finitely supported discrete random variable. Here, $\hat{G}_n$ is defined as a maximum likelihood estimator in a model whose tangent space for G is denoted by $\mathbf{T}^*(G)$. The corresponding estimator of

the conditional cumulative hazard function is

$$\hat{\Lambda}_n(\cdot|u) = \int_0^{\cdot} \frac{1(Y_n(u,s) > 0)}{Y_n(u,s)} N_n(u, ds),$$

where

$$N_n(u,s) = \sum_{i=1}^{n} 1\big(X_i \le s, \delta_i = 0, c(U_{k,i}) = c(u)\big), Y_n(u,s) = \sum_{i=1}^{n} 1\big(X_i \ge s, c(U_{k,i}) = c(u)\big)$$

denote the stratified basic counting process and the size of the risk set at time s, respectively.

(iii) $\hat{E}_n(u,s,k)$: an estimator of $E[\tilde{Y}|U_k = u, X \ge s]$ under $\tilde{P}$ (the joint distribution of $(U_k, \tilde{Y})$), which is restricted to take values in some P-Donsker family of uniformly bounded functions of $(u,s) \in \mathbb{R} \times \mathcal{T}$ with given k. Note that along with (ii), this is equivalent to estimating $E[T|U_k = u, X \ge s]$, according to the equality $E[T|U_k = u, X \ge s] = G(s)E[\tilde{Y}|U_k = u, X \ge s]$. We suppress the argument s if $s = -\infty$, namely using $\hat{E}_n(u,k)$ to estimate $E[\tilde{Y}|U_k = u]$ that is equal to $E[T|U_k = u]$. Also, we require that, for the given k and each u, the process $s \mapsto \hat{E}_n(u,s,k)$ is predictable with respect to the filtration

$$\sigma\big\{N_n(\cdot, s'), Y_n(\cdot, s'), U_{k,i}, i = 1, \ldots, n, s' \le s \in \mathcal{T}\big\}. \tag{5}$$

In view of (4), which can be expressed in terms of $E[\tilde{Y}|U_k]$, Q_u and G, we see that $\Psi_k(P)$ can be estimated by plugging-in $\hat{P}_n$:

$$\Psi_k(\hat{P}_n) = \frac{\mathrm{Cov}_{\mathbb{Q}_n}(U_k, \hat{E}_n(U_k,k))}{\mathrm{Var}_{\mathbb{Q}_n}(U_k)}.$$

From (13) in Appendix A, with the dependence on k made explicit, we find that the influence function of $\Psi_k(P)$ is $\mathrm{IF}_k^*(\cdot|P) = \mathrm{IF}_k^{ipw}(\cdot|P) - \mathrm{IF}_k^{car}(\cdot|P)$, where $\mathrm{IF}_k^{ipw}(\cdot|P)$ and $\mathrm{IF}_k^{car}(\cdot|P)$ can be expressed in terms of the features of P by

$$\begin{aligned}
&\mathrm{IF}_k^{ipw}(\cdot|P):\\
&\quad o \mapsto \frac{(u - Q_u[U_k])(\tilde{y} - Q_u[E[\tilde{Y}|U_k]])}{\mathrm{Var}_{Q_u}(U_k)} - \frac{\mathrm{Cov}_{Q_u}(U_k, E[\tilde{Y}|U_k])}{\mathrm{Var}^2_{Q_u}(U_k)}\big(u - Q_u[U_k]\big)^2;\\
&\mathrm{IF}_k^{car}(\cdot|P):\\
&\quad o \mapsto \frac{(u - Q_u[U_k])}{\mathrm{Var}_{Q_u}(U_k)} \int_{\mathcal{T}} E[\tilde{Y}|U_k = u, X \ge s]\big\{1(x \in ds, \delta = 0) - 1(x \ge s)\, d\Lambda(s)\big\}.
\end{aligned} \tag{6}$$

Here, the form of $\mathrm{IF}_k^{car}(\cdot|P)$ is derived by inserting

$$E[T|U_k = u, X \ge s] = G(s)E[\tilde{Y}|U_k = u, X \ge s]$$

into (12) of Appendix A. This implies that $\mathrm{IF}_k^*(\cdot|P) = \mathrm{IF}_k^{ipw}(\cdot|P) - \mathrm{IF}_k^{car}(\cdot|P)$ is represented by the introduced features of P, and enables the estimation of $\mathrm{IF}_k^*(\cdot|P)$ in terms of $\hat{P}_n$.

The one-step estimator is then given by

$$\begin{aligned}
S_k(\mathbb{P}_n, \hat{P}_n) &= \Psi_k(\hat{P}_n) + \mathbb{P}_n\, \mathrm{IF}_k^*(\cdot|\hat{P}_n)\\
&= \Psi_k(\hat{P}_n) + \frac{\mathbb{P}_n(U_k - \mathbb{Q}_n[U_k])(\delta X/\hat{G}_n(X|U_k) - \mathbb{Q}_n[\hat{E}_n(U_k,k)])}{\mathrm{Var}_{\mathbb{Q}_n}(U_k)}\\
&\quad - \Psi_k(\hat{P}_n) - \mathbb{P}_n\, \mathrm{IF}_k^{car}(\cdot|\hat{P}_n)\\
&= \frac{\mathbb{P}_n(U_k - \mathbb{Q}_n[U_k])Y}{\mathrm{Var}_{\mathbb{Q}_n}(U_k)}\\
&\quad - \frac{1}{\mathrm{Var}_{\mathbb{Q}_n}(U_k)} \mathbb{P}_n\bigg[\big(U_k - \mathbb{Q}_n[U_k]\big) \int_{\mathcal{T}} \hat{E}_n(U_k, s, k) \hat{M}(ds|U_k)\bigg],
\end{aligned} \tag{7}$$

where $\hat{M}(ds|u) = 1(X \in ds, \delta = 0, c(U_k) = c(u)) - 1(X \geq s, c(U_k) = c(u))d\hat{\Lambda}_n(s|u)$. For the second equality, note that the second term in $\mathbb{P}_n \mathrm{IF}_k^{ipw}(\cdot|\hat{P}_n)$ is precisely $\Psi_k(\hat{P}_n)$. The third equality holds by $\mathbb{P}_n(U_k - \mathbb{Q}_n[U_k])\mathbb{Q}_n[\hat{E}_n(U_k, k)] = \mathbb{Q}_n(U_k - \mathbb{Q}_n[U_k])\mathbb{Q}_n[\hat{E}_n(U_k, k)] = 0$ and $Y = \delta X/\hat{G}_n(X|U_k)$.

Under the following conditions, we next establish the asymptotic linearity of this one-step estimator. The proof of this result and relevant remarks are given in Appendix B and Section S5 of the Supplementary Material (Huang, Luedtke and McKeague (2023)).

(A.1) Each predictor U_k has bounded support and is nondegenerate.

(A.2) The survival function of the censoring, G, is continuous and $G(\tau) > 0$.

(A.3) There is a positive probability of a subject still being at risk at the end of follow-up for each subgroup defined by the coarsening of U_k: $\mathrm{P}(X \geq \tau, c(U_k) = c(u)) > 0$.

(A.4) There exists a uniformly-bounded, nonrandom function $(u, s) \mapsto \bar{E}(u, s, k)$ that is indexed by k and left-continuous in s, such that given k, $\mathbb{E}\{|\hat{E}_n(u, s, k) - \bar{E}(u, s, k)|\} = o(n^{-1/4})$ for each (u, s) and $\sup_{(u,s)} |\hat{E}_n(u, s, k) - \bar{E}(u, s, k)|$ is bounded in probability, where $\mathbb{E}$ denotes the expectation over $O_1, \ldots, O_n$.

Condition (A.1) is a mild requirement on the marginal distributions of the predictors; they are not required to be continuous and they can have different supports. Condition (A.4) imposes a mild stability control on the behavior of the estimator $\hat{E}_n$. When $\bar{E}(U_k, k) \neq E[\tilde{Y}|U_k]$, where $\bar{E}(u, k) = \bar{E}(u, -\infty, k)$ and $\bar{E}$ is from (A.4), we need a correction to IF_k^* given by

$$\mathrm{IF}_k^\dagger(\cdot|\bar{E}, P) : o \mapsto \frac{\mathrm{Cov}_{Q_u}(U_k, \bar{E}(U_k, k) - E[\tilde{Y}|U_k])}{\mathrm{Var}_{Q_u}^2(U_k)}\big[\big(u - Q_u[U_k]\big)^2 - \mathrm{Var}_{Q_u}(U_k)\big].$$

To introduce the following theorem, we need additional notation—let $\Pi(\cdot|S)$ denote the projection operator onto a closed linear subspace $S \subseteq L_0^2(P)$ and $\perp$ denote the orthogonal complement in $L_0^2(P)$.

THEOREM 3.1. *Under conditions* (A.1), (A.2), (A.3) *and* (A.4),

$$S_k(\mathbb{P}_n, \hat{P}_n) - \Psi_k(P) = [\mathbb{P}_n - P]\Pi\big(\mathrm{IF}_k^*(\cdot|\bar{E}, Q_u, G) + \mathrm{IF}_k^\dagger(\cdot|\bar{E}, P)|\mathbf{T}^*(G)^\perp\big) + o_p\big(n^{-1/2}\big),$$

where $\mathbf{T}^*(G)$ *is the tangent space defined in* (*ii*) *above.*

Suppose that $\hat{E}_n(u, s, k)$ consistently estimates the true conditional residual life function, that is, $\bar{E}(u, s, k) = E[\tilde{Y}|U_k = u, X \geq s]$ for all (u, s) with given k, in which case $\mathrm{IF}_k^\dagger(\cdot|\bar{E}, P) = 0$. Recall from Appendix A that $\mathrm{IF}_k^*(\cdot|\bar{E}, Q_u, G) = \Pi(\mathrm{IF}_k^{ipw}(\cdot|P)|\mathbf{T}^{car}(G)^\perp)$, with $\mathbf{T}^{car}(G)$ as defined in (11) thereof and, therefore, $\mathrm{IF}_k^*(\cdot|\bar{E}, Q_u, G) \in \mathbf{T}^{car}(G)^\perp$. Furthermore, $\mathbf{T}^*(G) \subseteq \mathbf{T}^{car}(G)$, and so $\mathbf{T}^*(G)^\perp \supseteq \mathbf{T}^{car}(G)^\perp$. Consequently, we know that $\mathrm{IF}_k^*(\cdot|\bar{E}, Q_u, G) \in \mathbf{T}^*(G)^\perp$, which yields that

$$\Pi\big(\mathrm{IF}_k^*(\cdot|\bar{E}, Q_u, G)|\mathbf{T}^*(G)^\perp\big) = \mathrm{IF}_k^*(\cdot|\bar{E}, Q_u, G) \equiv \mathrm{IF}_k^*(\cdot|P).$$

Hence, $S_k(\mathbb{P}_n, \hat{P}_n)$ is asymptotically linear with the influence function equal to the efficient influence function, that is, $S_k(\mathbb{P}_n, \hat{P}_n)$ is asymptotically efficient. Furthermore, under regularity conditions, the empirical variance of $\mathrm{IF}_k^*(O|\hat{P}_n)$ converges to the variance of $\mathrm{IF}_k^*(O|P)$ under $O \sim P$, that is, to the variance of the one-step estimator.

4. Stabilized one-step estimator. For the inference of the target parameter $\Psi(P)$ defined in (2), we need to incorporate the selection of the most informative predictor into the one-step estimator developed in (7). We adapt the stabilization approach of Luedtke and van der Laan (2018) to leverage data for the predictor selection and to introduce the resulting

variation into the inference procedure. The idea is first to randomly order the data, and consider subsamples consisting of the first j observations for $j = q_n, \ldots, n-1$, where $\{q_n\}$ is some positive integer sequence such that both q_n and $n - q_n$ tend to infinity. Based on the subsample of size j, an estimator of the label of the most informative predictor is

$$k_j = \arg\max_{k=1,\ldots,p} |\Psi_{\hat{G}_j,k}(\mathbb{P}_j)| \equiv \arg\max_{k=1,\ldots,p} \left| \frac{\mathrm{Cov}_{\mathbb{P}_j}(U_k, \delta X/\hat{G}_j(X))}{\mathrm{Var}_{\mathbb{P}_j}(U_k)} \right|, \tag{8}$$

where $\hat{G}_j$ is the usual Kaplan–Meier estimator of G based on the subsample of size j, and $\mathbb{P}_j$ is the empirical distribution of this subsample.

A reduced version of the earlier condition (A.3) is now understood without any predictors as

(A.5) There is a positive probability of a subject still being at risk at the end of follow-up: $\mathrm{P}(X \geq \tau) > 0$.

The stabilized one-step estimator of $\Psi(P)$ is then given by

$$S_n^* = \frac{1}{n - q_n} \sum_{j=q_n}^{n-1} w_{nj} m_j S_{k_j}(\delta_{O_{j+1}}, \hat{P}_{nj}), \tag{9}$$

where $m_j \in \{-1, 1\}$ is the sign of $\Psi_{\hat{G}_j,k_j}(\mathbb{P}_j)$, S_{k_j} refers to (7) with the predictor U_k now being U_{k_j} and $\hat{P}_{nj} \equiv (\hat{E}_j, \mathbb{Q}_j, \hat{G}_n)$ that refers to $\hat{P}_n$ based on only the first j observations to estimate part of the parameters of P. Here, $\delta_{O_{j+1}}$ is the Dirac measure putting unit mass at O_{j+1}, $w_{nj} \equiv \bar{\sigma}_n/\hat{\sigma}_{nj}$ with $\bar{\sigma}_n = \{(n - q_n)^{-1} \sum_{j=q_n+1}^{n} (1/\hat{\sigma}_{nj})\}^{-1}$,

$$\hat{\sigma}_{nj}^2 = \frac{1}{j} \sum_{i=1}^{j} \left\{ m_j \, \mathrm{IF}_{k_j}^*(O_i | \hat{P}_{nj}) - \frac{1}{j} \sum_{i=1}^{j} m_j \, \mathrm{IF}_{k_j}^*(O_i | \hat{P}_{nj}) \right\}^2,$$

and $\mathrm{IF}_{k_j}^*$ is IF_k^* with the predictor taken as U_{k_j}.

Note that $\hat{P}_{nj}$ in the stabilized one-step estimator S_n^* involves subsamples. As we will see from simulation studies, however, using the full-sample estimator $\hat{P}_n$ instead, considerably improves the performance of S_n^* in small samples.

The following 95% confidence interval for $\Psi(P)$ is justified by the asymptotic normality of S_n^* given in Theorem 4.1 below:

$$[\mathrm{LB}_n, \mathrm{UB}_n] = \left[S_n^* \pm 1.96 \frac{\bar{\sigma}_n}{\sqrt{n - q_n}} \right],$$

and the two-sided p-value is

$$2(1 - \Phi(|\sqrt{n - q_n} S_n^*/\bar{\sigma}_n|)),$$

where Φ is the cumulative distribution function of $\mathcal{N}(0, 1)$.

THEOREM 4.1. *Suppose the number of predictors $p = p_n$ satisfies $\log(p_n)/n^{1/4} \to 0$, and the smallest subsample size q_n used for stabilization satisfies $n - q_n \to \infty$, $n/q_n = O(1)$ and $q_n^{1/4}/\log(n \vee p_n) \to \infty$. Assume (A.1), (A.2), (A.5), the asymptotic stability conditions (A.7)–(A.8) that are stated just before the proof in Appendix C, and the nondegeneracy condition*

(A.6) $\mathrm{Var}_{Q_u}(U_k)$, $\mathrm{Var}(U_k 1(X \geq s))$ *and* $\mathrm{Var}(\mathrm{IF}_k^*(O|P))$ *are bounded away from zero and infinity, as functions of $k \in \{1, \ldots, p_n\}$ and $s \in \mathcal{T}$.*

Then S_n^* is an asymptotically normal estimator of $\Psi(P)$:

$$\sqrt{n-q_n}\bar{\sigma}_n^{-1}\big[S_n^* - \Psi(P)\big] \xrightarrow{d} \mathcal{N}(0,1).$$

The proof is postponed to Appendix C. Note that condition (A.7) in Appendix C removes the need to include $\mathrm{IF}^{\dagger}$ in IF^* when constructing S_n^*. In practice, it is advisable to pre-standardize each predictor (as is commonly recommended in the variable selection literature) to provide scale invariance; the above result is given in terms of the unstandardized predictors for simplicity of presentation.

The stabilized one-step estimator is reminiscent of bagging, the aggregation of multiple weak learners constructed from subsets of the data (in this case, S_{k_j} for $j \geq q_n$). The value of q_n determines how many weak learners are collected ($n - q_n$ of them) and plays the role of a tuning parameter. Taking a smaller q_n is expected to reduce variability in the performance of S_n^*, but taking too small value of q_n leads to overfitting. In practice, we recommend setting $q_n = n/2$ (which satisfies the conditions in Theorem 4.1) as a reasonable trade-off, although in practice it is advisable to run the analysis for a few values of q_n and compare the results.

REMARK 4.2 (Estimation of the conditional residual life function). Along the lines of van der Laan and Hubbard (1998), such an estimator $\hat{E}_j(u,s,k)$ can be constructed by regressing Y on U_k using only a subsample $\{O_i = (X_i, \delta_i, \boldsymbol{U}_i), i : X_i \geq s, i \leq j\}$ in the fashion of Koul, Susarla and Van Ryzin (1981), leading to

$$\begin{aligned}\hat{E}_j(u,s,k) &= \mathbb{P}_j\big[Y1(X \geq s)\big] \\ &\quad + \frac{\mathrm{Cov}_{\mathbb{P}_j}(U_k 1(X \geq s), Y1(X \geq s))}{\mathrm{Var}_{\mathbb{P}_j}(U_k 1(X \geq s))}\big(u - \mathbb{P}_j\big[U_k 1(X \geq s)\big]\big).\end{aligned} \tag{10}$$

Note that for the given k and each u, the process $s \mapsto \hat{E}_j(u,s,k)$ is unpredictable with respect to the filtration defined in (5). This is the unpredictability issue referred to in the Introduction.

REMARK 4.3 (Additional stabilization). A practical issue in implementing a test based on the stabilized one-step estimator is variation due to the ordering of the data. Ordering the data in a different way can change the value of S_n^* and the resulting p-value, making the result difficult to reproduce. One way to address this issue is to use a Bonferroni correction of the minimal p-value resulting from $R \geq 1$ random orderings of the data, taking into account the trade-off in terms of computational cost (which grows proportionally to R). Then the null is rejected if the minimum of the p-values obtained from the R random orderings is less than 5% (after Bonferroni correction for R-fold multiple testing). We refer to the resulting method as the *R-fold stabilized* one-step test. In practice, we recommend setting $R = 10$.

REMARK 4.4 (Computational cost). The stabilized one-step estimator does not involve subsampling in the usual sense, for example, in the way it is used in bootstrapping, forming say 1000 independent subsamples, each without replacement, although the idea is similar in some respects. We only make use of *certain* subsamples, namely those formed from the first j observations in a given ordering of the data (for a small number if random orderings of the data), as j runs from q_n to $n-1$. More specifically, our test statistic S_n^* is a weighted average of statistics combining each one of these subsamples with its successive observation, for a total of $n - q_n$ terms. This averaging stabilizes the asymptotic behavior of S_n^*, and consequently simplifies the calibration of our test and CIs (via an asymptotically normal limit that is proved using martingale methods). Standard subsampling is computationally much more intensive, requiring say 1000 subsamples, and in our high-dimensional setting becomes computationally prohibitive.

5. Competing methods.

Marginal Cox models with Bonferroni correction (*Bonferroni Cox*). This procedure uses marginal Cox models for linking the survival outcome T to each predictor U_k, $k = 1, \ldots, p$. Provided the asymptotic normality of the maximum partial likelihood estimator (Andersen and Gill (1982)), we conduct a Z-test with Bonferroni correction to investigate whether each marginal regression coefficient is zero or not.

Marginal one-step estimators with Bonferroni correction (*Bonferroni one-step*). For each predictor $U_k, k = 1, \ldots, p$, the marginal test statistic is $B_k \equiv \sqrt{n} S_k(\mathbb{P}_n, \hat{P}_n)/\hat{\sigma}_k$, where $\hat{\sigma}_k^2$ is the sample second moment of $\mathrm{IF}_k^*(O|\hat{P}_n)$. Marginal testing over all k with Bonferroni correction controls the familywise error rate of the global null hypothesis $\boldsymbol{\beta}_0 = \mathbf{0}$. This method is theoretically supported by Theorem 3.1.

One-step estimator with the selected predictor (*naive one-step*). With the label k of the most correlated predictor U_k estimated by $\hat{k}_n \equiv \arg\max_{k=1,\ldots,p} |\Psi_{\hat{G}_n,k}(\mathbb{P}_n)|$, where $\hat{G}_n$ and $\mathbb{P}_n$ are as defined below (8) but based on the full sample instead, the test statistic turns to be $B_{\hat{k}_n}$. When simply taking $\hat{k}_n$ as given, it implies that $B_{\hat{k}_n}$ follows an asymptotically standard normal null distribution. We include this estimator to showcase the consequence of ignoring the selection bias that results from such naive use of $\hat{k}_n$.

Oracle one-step estimator (*oracle one-step*). In this case, the label k of the most correlated predictor U_k is given, and the test statistic is simply B_k, which has an asymptotically standard normal null distribution. Assuming knowledge of k is of course unrealistic, but this estimator serves as a benchmark against which the other methods can be compared.

6. Simulation results. In this section, we report the results of simulation studies evaluating the performance of the stabilized one-step estimator (with $q_n = n/2$) in comparison with the competing methods in Section 5. The log-transformed survival times are generated under one of the following AFT scenarios:

Model N: $T = \varepsilon$;
Model A1: $T = U_1/4 + \varepsilon$;
Model A2: $T = \sum_{k=1}^{p} \beta_k U_k + \varepsilon$ with $\beta_1 = \cdots = \beta_5 = 0.15$, $\beta_6 = \cdots = \beta_{10} = -0.1$, $\beta_k = 0$ for $k \geq 11$.

The noise ε is distributed either as $\mathcal{N}(0, 1)$ independently of $\boldsymbol{U}$, or as $\mathcal{N}(0, 0.7(|U_1| + 0.7))$ (conditionally on $\boldsymbol{U}$). The predictors $\boldsymbol{U}$ have a p-dimensional multivariate normal distribution with unit variances and an exchangeable correlation structure such that $\mathrm{Corr}(U_k, U_j) = 0.75$, $k \neq j$. In Model N, there is no active predictor, while there is only a single active predictor in Model A1. In Model A2, there are ten active predictors, each having weaker influence than the single predictor in Model A1; the most correlated predictor is not unique in this model. The censoring time C is taken to be the logarithm of an exponential random variable with rate parameters that give either light censoring (10%) or heavy censoring (30%). Here, we just consider light censoring; results for the heavy censoring case are given in the Supplementary Material (Huang, Luedtke and McKeague (2023)). For each data generating scenario, we fix the sample size at $n = 500$, and consider 5 values of p of the form 10^a (for $a = 2, 3, \ldots, 6$). A nominal significance level of 5% is used throughout. The Kaplan–Meier estimator $\hat{G}_n$ is used in S_n^*, as justified by the independent censoring assumption; although a more sophisticated conditional Kaplan–Meier estimator could be used instead, doing so would involve an additional computational cost.

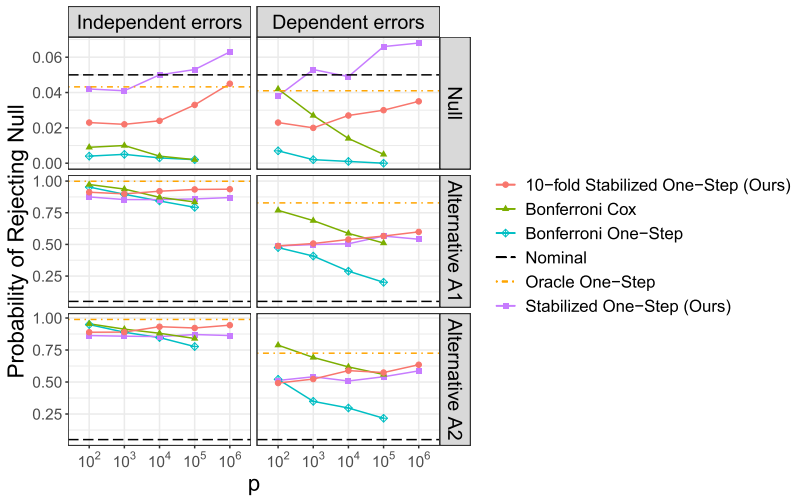

FIG. 1. *Empirical rejection rates based on* 1000 *samples* ($n = 500$) *generated from models with independent and dependent errors under light censoring* (10%), *for* p *in the range* 10^2–10^6, *using the full sample to obtain* $\hat{P}_n$. *The panels tagged by Null give the results under the null model, while those tagged by Alternative* A1 *and Alternative* A2 *display the results under two different alternative models.*

Empirical rejection rates based on 1000 Monte Carlo replications under the various scenarios are displayed in Figure 1, using the full sample-based $\hat{P}_n$ to estimate the features of P. The panels for Null show type I error rates under Model N for independent and dependent errors, respectively, with the nominal level of 5% shown by the horizontal black dashed line and the type I error of the oracle one-step estimator shown by the horizontal orange dot–dashed line. The results of the naive one-step estimator is as highly anticonservative as expected (results not shown). Similarly, the panels for (Alternative A1, Alternative A2) show the power under Model A1 and Model A2, with independent and dependent errors, respectively.

The left panels of Figure 1 show the results for independent errors. The stabilized one-step estimator provides the closest-to-nominal type-I error (apart from the oracle one-step estimator). The right panels of Figure 1 give the results in the case of dependent errors, and show that the variants of stabilized one-step estimators outperform other methods in power—the Bonferroni one-step method when p is larger than 1000 and the Bonferroni Cox method when p is larger than 10^4. These two competing methods are favored in terms of power over the stabilized one-step estimator in the case of independent errors, but not in the case of dependent errors. The computational cost of the Bonferroni one-step and the Bonferroni Cox methods is prohibitive for $p = 10^6$ and they are not included.

In the same figure, we examine the effect of multiple random orderings on the performance of the stabilized one-step estimator (taking $R = 10$ in the R-fold stabilized one-step test). The type-I error is now always below 5%, and the power remains the same or even improved.

It is of interest to understand how power changes with signal strength. To this end, we consider the following two alternative models in which the signal strength is represented by η (we restrict attention to $p = 10^4$ and $p = 10^5$, $n = 500$ under light censoring).

Model A3: $T = \eta U_1 + \varepsilon$ with the value of η from 0.1 to 0.2 in increments of 0.025.

Model A4: $T = \sum_{k=1}^{p} \beta_k U_k + \varepsilon$ with $\beta_1 = \cdots = \beta_{10} = \eta$, $\beta_k = 0$ for $k \geq 11$, and the value of η from 0.01 to 0.02 in increments of 0.0025.

Note that Model A3 reduces the signal strength from 0.25 as in Model A1 to 0.2 or below, while Model A4 attenuates the signal in Model A2 by a factor of ten. Figures 2 and 3 show that Bonferroni Cox and the stabilized one-step estimator share similar power behavior for

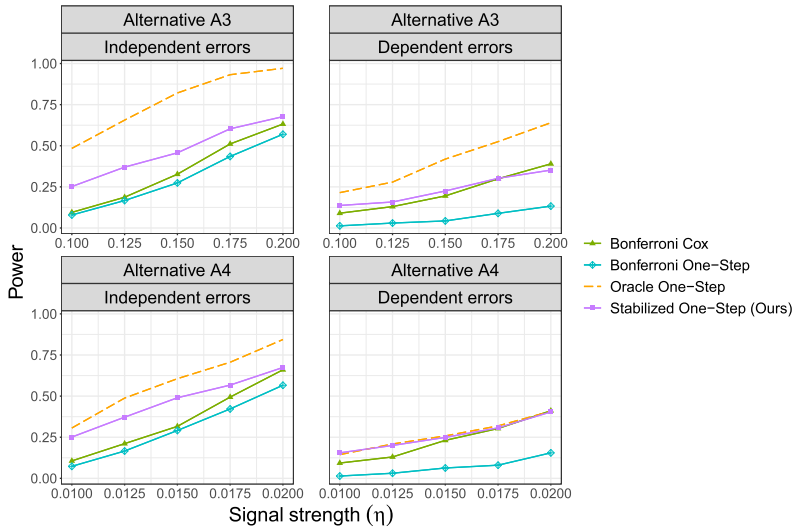

FIG. 2. *Empirical rejection rates based on* 1000 *samples of* $(n, p) = (500, 10^4)$ *generated from alternative models A*3 *and A*4 *with varying values of signal strength* η *along with independent and dependent errors under light censoring* (10%).

$p = 10^4$ (except in the case of relatively low signal strength), but the stabilized one-step estimator clearly outperforms Bonferroni Cox when $p = 10^5$.

REMARK 6.1 (Interpretation of $\Psi_k(P)$). The values of the marginal regression coefficients (3) are of course determined by the corresponding regression coefficients in the AFT model (1) that generates the data. Specifically, in our simulation models A1 and A3 (that have independent errors), the regression coefficient of the (only) active predictor in the AFT model agrees with (3). However, in Model A2 (with independent errors) it can be shown that the marginal regression coefficients are

$$\Psi_k(P) = \frac{\mathrm{Cov}_P(U_k, T)}{\mathrm{Var}_P(U_k)} = \begin{cases} 0.225 & \text{if } k = 1, \ldots, 5, \\ 0.1625 & \text{if } k = 6, \ldots, 10, \end{cases}$$

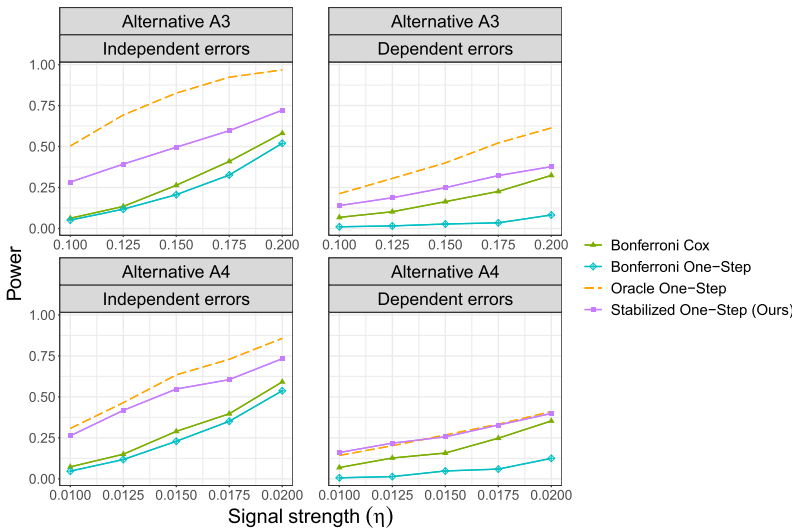

FIG. 3. *As in Figure* 2, *except for* $p = 10^5$.

in contrast with the AFT model regression coefficients 0.15 and -0.1 for $k = 1, \ldots, 5$ and $k = 6, \ldots, 10$, respectively. Note that the marginal regression coefficients of U_k, $k = 6, \ldots, 10$, are positive, whereas the corresponding AFT model coefficients are negative; this is essentially due to high correlation between predictors. Further, in Model A4, in which $T = \eta \sum_{k=1}^{10} U_k + \varepsilon$ (with independent errors and regression coefficients η), the corresponding marginal regression coefficients are 7.75η.

7. Application to viral replication data. A widely used measure of the potency of an antiviral drug is the concentration needed to achieve a 50% reduction of the (in vitro) rate of viral replication (IC_{50}, in units of μg/mL). In this application, we treat IC_{50} as a survival time outcome of interest. If the virus is highly resistant, then a 50% reduction in viral replication rate may not be observed, resulting in a right-censored outcome. We consider the antiviral VRC01, an antibody for HIV-1 that is currently being evaluated in a Phase 2b trial for the prevention of HIV-1 infection (Gilbert et al. (2017); Magaret et al. (2019)). In this case, a reduction in viral replication is thought to be caused by a VRC01-mediated neutralization, and the lower IC_{50}, the more sensitive the virus is to VRC01-mediated neutralization.

Data on a total of 624 pseudoviruses were retrieved from the CATNAP database (Yoon et al. (2015)). We restrict attention to a subgroup of size ($n = 611$) after removing 13 pseudoviruses with unreliable IC_{50} measurements. The censoring rate of IC_{50} is 16% in the analyzed data set. The 611 pseudoviruses are of 24 subtypes, where Subtypes B and C are predominant over others: 293 of them (48.0%) belong to Subtype C and 81 (13.3%) are of Subtype B. In terms of geographic regions where the viruses originate, 126 of 611 pseudoviruses are from Asia (20.6%), 96 from Europe or the Americas (15.7%), 170 from Northern Africa (27.8%) and 219 from Southern Africa (35.9%). Pseudoviruses of different subtypes may present varied gene expression, so we analyze the data for Subtypes B and C separately. We set the end of follow-up τ to the 90th percentile of IC_{50} in each data set with the corresponding sample size as indicated in Table 1. We aim to investigate whether the potency of VRC01 depends on HIV-1 proteomic characteristics that are presented in lieu of the envelope (Env) amino acid (AA) sequence features. Data on 817 features are available:

1. Binary features: indicating whether a particular AA sequence appears at a particular position, or whether a position is the starting site of some given enzymatic process. There are 799 features of this type.
2. Count features: representing total numbers of enzyme-directed chemical reactions observed to take place within a given region, or the total length of the aligned sequences over a region. There are 18 count features.

To simplify interpretation of the effects of binary or count features, we carry out separate analyses for each type. Binary features and binary interactions are included when their incidence rates fall in the range 5–95%. All count features are first standardized, then all pairwise

TABLE 1

Numbers of binary and count features for Subtypes B and C; proportion included among all possible are given in parentheses

	Binary effects		Count effects	
	Main	Interaction	Main	Interaction
Subtype B ($n = 81$)	220 (28%)	11,642 (48%)	17 (94%)	136 (100%)
Subtype C ($n = 293$)	252 (32%)	12,698 (40%)	17 (94%)	136 (100%)

TABLE 2

Results of applying the Bonferroni Cox and the Bonferroni one-step methods as well as the stabilized one-step estimator to data on Subtypes B and C; the numbers of binary and count features are denoted p_{bin} and p_{count}, respectively

		Binary effects		Count effects	
$(n, p_{\text{bin}}, p_{\text{count}})$	Method	95% CI	p-value	95% CI	p-value
Subtype B	Bonferroni Cox	NA	0.08	NA	0.04
(81, 11,862, 153)	Bonferroni One-Step	NA	0.01	NA	0.04
	Stabilized One-Step	(7.8, 9.5)	< 0.001	(12.7, 13.5)	< 0.001
Subtype C	Bonferroni Cox	NA	< 0.001	NA	0.34
(293, 12,950, 153)	Bonferroni One-Step	NA	< 0.001	NA	0.91
	Stabilized One-Step	(10.4, 23.1)	< 0.001	(3.3, 5.0)	< 0.001

interactions of these features are also standardized. Although throughout we are using correlation as the measure of association, for binary and count features other association measures may be reasonable (but are beyond the scope of the present analysis). The total number of features included in the analysis varies by feature type as well as viral subtype, as given in Table 1.

As discussed at the end of Section 6, in implementing the stabilized one-step estimator we recommend using the full sample to estimate the parameters of P, along with R random orderings of the data. Histograms of p-values based on 1000 random orderings of the data with respect to different virus subtypes are given in Figures S.1–S.2 in the Supplementary Material (Huang, Luedtke and McKeague (2023)). These histograms show the strong dependence of the p-value on the random ordering. In Table 2, we report the p-values of the stabilized one-step estimator and 95% confidence intervals using $q_n = n/2$ and $R = 200$ random orderings of the data (separated by virus subtype). The reported confidence interval corresponds to the minimal p-value. This table also presents the results of applying the Bonferroni Cox and the Bonferroni one-step methods to the original data. Though both of the Bonferroni Cox and the Bonferroni one-step methods also return significant results, except in the case of binary features and Subtype B or the case of count features and Subtype C, these two methods generally yield more conservative conclusions than the stabilized one-step estimator, as expected.

The confidence intervals based on the stabilized one-step estimator in Table 2 represent changes in IC_{50} (in units of μg/mL) due to the presence of the identified binary feature, or due to a unit increase in the identified count feature. Genetic descriptions of these identified features are provided in Table S.1 of Section S7 in the Supplementary Material (Huang, Luedtke and McKeague (2023)).

8. Discussion. Though we have focused on using the correlation as the marginal association measure, our results can be extended to a wide range of other measures. For example, more robust association measures that have been used in marginal screening include quantile correlation (Ma, Li and Tsai (2017)) and copula-based correlation (Xia and Li (2021)). The key requirements on such a measure are that (a) it be pathwise differentiable in the full data model where censoring is not present, and (b) the efficient influence function be nondegenerate in this full data setting. Requirement (a) is sufficient for the association measure to be pathwise differentiable even in the presence of right censoring, thereby allowing the construction of a one-step estimator for each marginal association. Requirement (b) enables the use of first-order asymptotics to study the behavior of these one-step estimators; without this condition, the centered estimator, scaled by the square root of sample size, would converge weakly to zero rather than to a nondegenerate mean-zero normal distributed random variable.

Requirement (b) also ensures that the variance estimates used for standardization in the stabilized one-step estimator do not converge to zero, a condition that is required by existing theory for this estimator. Examples of parameters satisfying (a) and (b) include the Spearman correlation, odds ratio and model-agnostic hazard ratio (Whitney, Shojaie and Carone (2019)). Examples of parameters satisfying (a), but not (b), include the nonparametric R^2, distance correlation (Székely, Rizzo and Bakirov (2007)) and maximum mean discrepancy (Smola, Gretton and Borgwardt (2006)).

APPENDIX A: DERIVATION OF THE ONE-STEP ESTIMATOR

First, consider $\Psi_k(P)$ for a given k, which can be reexpressed via (3) as

$$\Psi_{k,G}(P) \equiv \frac{\mathrm{Cov}_P(U_k, \delta X/G(X))}{\mathrm{Var}_P(U_k)},$$

explicitly in terms of its dependence on the survival function of the censoring (G). For notational simplicity, we suppress the dependence on k, and write $\Psi_G(P)$ and U, both here and in the corresponding proofs in the sequel. Following the proof in Section S3.1 concerning inverse probability weighting when G is known, the influence function of $\Psi_G(\mathbb{P}_n)$ is

$$\mathrm{IF}^{ipw}(\cdot|P)\colon o \mapsto \frac{(u - P[U])(\tilde{y} - P[T])}{\mathrm{Var}_P(U)} - \frac{\mathrm{Cov}_P(U,T)}{\mathrm{Var}_P^2(U)}\big(u - P[U]\big)^2,$$

where $\tilde{y} = \delta x/G(x)$. In Section S3.2, it can be shown that plugging-in the Kaplan–Meier estimator of G into $\Psi_G(\mathbb{P}_n)$ leads to improved efficiency, even when G is known; this estimator is regular and asymptotically linear with influence function $\mathrm{IF}(\cdot|P)$ as given in Theorem S3.1. However, as will become apparent, $\mathrm{IF}(\cdot|P) \in \mathbf{T}^{nu}(G)^{\perp}$ does not fall in the tangent space $\mathbf{T}^{\mathcal{M}}(P)$ at P in the model $\mathcal{M}$, where $\mathbf{T}^{nu}(G)$ is as defined in (S3.6) and $\perp$ denotes the orthogonal complement in $L_0^2(P)$. Therefore, we need to project $\mathrm{IF}(\cdot|P)$ onto $\mathbf{T}^{\mathcal{M}}(P)$ to obtain an efficient influence function IF*. Once we have access to IF*, it will be then be feasible to construct an asymptotically efficient one-step estimator of $\Psi(P)$.

To compute this projection, despite our assumption of independent censoring, it is convenient to consider the broader coarsening-at-random (CAR) model $\mathcal{M}^{car} \supseteq \mathcal{M}$. Under $\mathcal{M}^{car}$, G is viewed as a survival function for C conditionally on U, and this survival function may depend nontrivially on U. Since we have assumed that C is independent of (T, U) for the particular distribution that generated our data, this conditional survival function is equal to the marginal survival function $G(\cdot)$ for that distribution. This observation slightly simplifies the expression for the tangent space for G in $\mathcal{M}^{car}$, which is given by

$$\mathbf{T}^{car}(G) = \left\{ \int_{\mathcal{T}} H(U,s)\, dM(s) \big| H\colon \mathbb{R} \times \mathcal{T} \to \mathbb{R} \right\}, \tag{11}$$

where H is any measurable function for which the integral has finite variance, and $dM(s) = 1(X \in ds, \delta = 0) - 1(X \geq s)d\Lambda(s)$ with $\Lambda(\cdot)$ as the cumulative hazard function corresponding to $G(\cdot)$ with respect to the filtration $\mathcal{F}_s = \sigma\{1(X \leq s', \delta = 0), 1(X \geq s'), U, s' \leq s \in \mathcal{T}\}$. See Example 1.12 in van der Laan and Robins (2003) for further details. Moreover, since $\mathcal{M} \subseteq \mathcal{M}^{car}$, $\mathbf{T}^{nu}(G) \subseteq \mathbf{T}^{car}(G)$.

To obtain the efficient influence function IF*, we could project $\mathrm{IF}(\cdot|P)$ onto $\mathbf{T}^{\mathcal{M}}(P)$. To compute this projection, it will be useful to first show that $\mathbf{T}^{car}(G)^{\perp} \subseteq \mathbf{T}^{\mathcal{M}}(P)$. This can be shown as follows. Let $\mathbf{T}^{car}(Q)$ and $\mathbf{T}^{\mathcal{M}}(Q)$, respectively, denote the tangent space generated by local fluctuations of P in the CAR and $\mathcal{M}$ models that modify Q but leave G unchanged. Because both CAR and $\mathcal{M}$ could induce a (locally) nonparametric model for Q, $\mathbf{T}^{car}(Q) = \mathbf{T}^{\mathcal{M}}(Q)$. Furthermore, because P factorizes as a product of mappings of the variation-independent components Q and G, we can write (i) $\mathbf{T}^{car}(P) = \mathbf{T}^{car}(Q) \oplus \mathbf{T}^{car}(G)$

and (ii) $\mathbf{T}^{\mathcal{M}}(P) = \mathbf{T}^{\mathcal{M}}(Q) \oplus \mathbf{T}^{\mathcal{M}}(G)$, as orthogonal sums. By (i) and the fact that $\mathbf{T}^{car}(P) = L_0^2(P)$, we have $\mathbf{T}^{car}(Q) = \mathbf{T}^{car}(G)^{\perp}$. Hence, by (ii) and the fact that $\mathbf{T}^{car}(Q) = \mathbf{T}^{\mathcal{M}}(Q)$, we find that $\mathbf{T}^{car}(G)^{\perp} \subseteq \mathbf{T}^{\mathcal{M}}(P)$. Using $\Pi(\cdot|S)$ to denote the projection operator onto a closed linear subspace $S \subseteq L_0^2(P)$, we have

$$\begin{aligned}
\mathrm{IF}^*(\cdot|P) &= \Pi\big(\mathrm{IF}(\cdot|P)|\mathbf{T}^{\mathcal{M}}(P)\big) \\
&= \Pi\big(\Pi\big(\mathrm{IF}(\cdot|P)|\mathbf{T}^{car}(G)\big) + \Pi\big(\mathrm{IF}(\cdot|P)|\mathbf{T}^{car}(G)^{\perp}\big)|\mathbf{T}^{\mathcal{M}}(P)\big) \\
&= \Pi\big(\Pi\big(\mathrm{IF}(\cdot|P)|\mathbf{T}^{car}(G)\big)|\mathbf{T}^{\mathcal{M}}(P)\big) + \Pi\big(\Pi\big(\mathrm{IF}(\cdot|P)|\mathbf{T}^{car}(G)^{\perp}\big)|\mathbf{T}^{\mathcal{M}}(P)\big) \\
&= \Pi\big(\mathrm{IF}(\cdot|P)|\mathbf{T}^{nu}(G)\big) + \Pi\big(\mathrm{IF}(\cdot|P)|\mathbf{T}^{car}(G)^{\perp}\big),
\end{aligned}$$

where the final equality uses that (1) $\mathbf{T}^{\mathcal{M}}(P) \cap \mathbf{T}^{car}(G) = \mathbf{T}^{nu}(G)$; (2) $\mathbf{T}^{car}(G)^{\perp} \subseteq \mathbf{T}^{\mathcal{M}}(P)$. By Lemma S4.4, we know that $\mathrm{IF}(\cdot|P) \in \mathbf{T}^{nu}(G)^{\perp}$, implying that the first term on the right-hand side is zero. The same lemma tells us that $\mathrm{IF}(\cdot|P) = \Pi(\mathrm{IF}^{ipw}(\cdot|P)|\mathbf{T}^{nu}(G)^{\perp})$, so the above display, along with the fact that $\mathbf{T}^{nu}(G) \subseteq \mathbf{T}^{car}(G)$, implies that

$$\mathrm{IF}^*(\cdot|P) = \Pi\big(\mathrm{IF}^{ipw}(\cdot|P)|\mathbf{T}^{car}(G)^{\perp}\big).$$

It remains to project $\mathrm{IF}^{ipw}(\cdot|P)$ onto $\mathbf{T}^{car}(G)^{\perp}$. The projection of $\mathrm{IF}^{ipw}(\cdot|P)$ onto $\mathbf{T}^{car}(G)$ is given by

$$\begin{aligned}
&\mathrm{IF}^{car}(\cdot|P): \\
&\quad o \mapsto \frac{(u - P[U])}{\mathrm{Var}(U)} \int_{\mathcal{T}} E[T|U = u, X \geq s] \frac{1(x \in ds, \delta = 0) - 1(x \geq s)\, d\Lambda(s)}{G(s)}
\end{aligned} \tag{12}$$

by Proposition 5.4 of van der Laan, Gill and Robins (2000). Its projection onto $\mathbf{T}^{car}(G)^{\perp}$ is then given by

$$\mathrm{IF}^*(\cdot|P) = \mathrm{IF}^{ipw}(\cdot|P) - \mathrm{IF}^{car}(\cdot|P). \tag{13}$$

In general, $\mathrm{IF}(\cdot|P)$ is not equivalent to $\mathrm{IF}^*(\cdot|P)$, so $\mathrm{IF}(\cdot|P)$ does not fall in the tangent space $\mathbf{T}^{\mathcal{M}}(P)$ at P for the model $\mathcal{M}$ and, therefore, plugging-in the Kaplan–Meier estimator of G into $\Psi_G(\mathbb{P}_n)$ is inefficient; see Section S3.2 for further discussion.

APPENDIX B: PROOF OF THEOREM 3.1

PROOF. In this theorem, our objective is to show that, under regularity conditions, $S(\mathbb{P}_n, \hat{P}_n)$ is asymptotically linear with influence function

$$\Pi\big\{\mathrm{IF}^*(\cdot|\bar{E}, Q_u, G)|\mathbf{T}^*(G)^{\perp}\big\}.$$

Let $\hat{P}_n' = (\hat{E}_n, \mathbb{Q}_n, G)$ denote the estimate of P but with $\hat{G}_n$ replaced by the true censoring distribution G. Note also that $S(\mathbb{P}_n, \hat{P}_n') = \Psi(\hat{P}_n) + \mathbb{P}_n\,\mathrm{IF}^*(\cdot|\hat{E}_n, \mathbb{Q}_n, G)$, where equations (2)–(4) imply that Ψ does not depend on the censoring distribution so that $\Psi(\hat{P}_n) = \Psi(\hat{P}_n')$. We have that

$$\begin{aligned}
S(\mathbb{P}_n, \hat{P}_n) - \Psi(P) &= S\big(\mathbb{P}_n, \hat{P}_n'\big) - \Psi(P) + S(\mathbb{P}_n, \hat{P}_n) - S\big(\mathbb{P}_n, \hat{P}_n'\big) \\
&= S\big(\mathbb{P}_n, \hat{P}_n'\big) - \Psi(P) + \mathbb{P}_n\big[\mathrm{IF}^*(\cdot|\hat{E}_n, \mathbb{Q}_n, \hat{G}_n) - \mathrm{IF}^*(\cdot|\hat{E}_n, \mathbb{Q}_n, G)\big] \\
&= S\big(\mathbb{P}_n, \hat{P}_n'\big) - \Psi(P) + P\big[\mathrm{IF}^*(\cdot|\hat{E}_n, \mathbb{Q}_n, \hat{G}_n) - \mathrm{IF}^*(\cdot|\hat{E}_n, \mathbb{Q}_n, G)\big] \\
&\quad + [\mathbb{P}_n - P]\big[\mathrm{IF}^*(\cdot|\hat{E}_n, \mathbb{Q}_n, \hat{G}_n) - \mathrm{IF}^*(\cdot|\hat{E}_n, \mathbb{Q}_n, G)\big].
\end{aligned} \tag{14}$$

The last term on the right-hand side of (14) is $o_p(n^{-1/2})$ by (S5.4.2) of Lemma S5.4, with the required conditions verified in Lemmas S5.1–S5.3. Moreover, Lemma S5.5 shows that

$$P\big[\mathrm{IF}^*(\cdot|\hat{E}_n, \mathbb{Q}_n, \hat{G}_n) - \mathrm{IF}^*(\cdot|\bar{E}, Q_u, \hat{G}_n) + \mathrm{IF}^*(\cdot|\bar{E}, Q_u, G) - \mathrm{IF}^*(\cdot|\hat{E}_n, \mathbb{Q}_n, G)\big] = o_p\big(n^{-1/2}\big),$$

which would simplify the middle term in (14) as follows. At last, Lemma S5.6 gives the asymptotic linearity of the term $S(\mathbb{P}_n, \hat{P}_n') - \Psi(P)$ in (14).

To simplify the middle term on the right-hand side in (14), we further decompose it as

$$\begin{aligned} &P\big[\mathrm{IF}^*(\cdot|\hat{E}_n, \mathbb{Q}_n, \hat{G}_n) - \mathrm{IF}^*(\cdot|\hat{E}_n, \mathbb{Q}_n, G)\big] \\ &\quad = P\big[\mathrm{IF}^*(\cdot|\bar{E}, Q_u, \hat{G}_n) - \mathrm{IF}^*(\cdot|\bar{E}, Q_u, G)\big] \\ &\qquad + P\big[\mathrm{IF}^*(\cdot|\hat{E}_n, \mathbb{Q}_n, \hat{G}_n) - \mathrm{IF}^*(\cdot|\bar{E}, Q_u, \hat{G}_n) + \mathrm{IF}^*(\cdot|\bar{E}, Q_u, G) - \mathrm{IF}^*(\cdot|\hat{E}_n, \mathbb{Q}_n, G)\big], \end{aligned}$$

with the last line of the above display shown as $o_p(n^{-1/2})$ by Lemma S5.5 as mentioned earlier. Expressing the first term of the above display using the notation $\Phi(\tilde{G}) = P[\mathrm{IF}^*(\cdot|\bar{E}, Q_u, \tilde{G})]$, for $\tilde{G}$ equal to $\hat{G}_n$ or G, upon inserting them back into (14) we have that

$$S(\mathbb{P}_n, \hat{P}_n) - \Psi(P) = S\big(\mathbb{P}_n, \hat{P}_n'\big) - \Psi(P) + \Phi(\hat{G}_n) - \Phi(G) + o_p\big(n^{-1/2}\big),$$

together with the previously developed results.

In Lemma S5.6, $S(\mathbb{P}_n, \hat{P}_n')$ is shown to be a regular asymptotically linear estimator of $\Psi(P)$ with influence function $\mathrm{IF}^* + \mathrm{IF}^\dagger$ in the model $\mathcal{M}(G)$ with G known. Further, because we specified that $\hat{G}_n$ is estimated via maximum likelihood, the delta method can be used to show that $\Phi(\hat{G}_n)$ is an asymptotically efficient estimator of $\Phi(G)$ in the model used for G (with tangent space given by $\mathbf{T}^*(G)$). Combining the above results, we have verified all the required conditions of Theorem 2.3 in van der Laan and Robins (2003), from which we conclude that

$$S(\mathbb{P}_n, \hat{P}_n) - \Psi(P) = [\mathbb{P}_n - P]\Pi\big\{\mathrm{IF}^*(\cdot|\bar{E}, Q_u, G) + \mathrm{IF}^\dagger(\cdot|\bar{E}, P)|\mathbf{T}^*(G)^\perp\big\} + o_p\big(n^{-1/2}\big),$$

and the proof is complete. □

REMARK B.1. If $\bar{E}(u) \neq E[\tilde{Y}|U = u]$ for some u, it is still possible that $\mathrm{IF}^\dagger(\cdot|\bar{E}, P) = 0$. For instance, if we assume that $\Psi(P)$ does not depend on the limit of $\hat{E}_n(u)$ to ensure identifiability, then recalling the definition of $\Psi(P)$ in (2), we have $\mathrm{Cov}_{Q_u}(U, \bar{E}(U)) = \mathrm{Cov}_{Q_u}(U, E[\tilde{Y}|U])$, and thus $\mathrm{IF}^\dagger(\cdot|\bar{E}, P) = 0$.

REMARK B.2. Regardless of whether $\mathrm{IF}^\dagger(\cdot|\bar{E}, P) = 0$ or not, the variance of the influence function of $S(\mathbb{P}_n, \hat{P}_n)$ is no larger than that of $\mathrm{IF}^*(\cdot|\bar{E}, Q_u, G) + \mathrm{IF}^\dagger(\cdot|\bar{E}, P)$ because projections only decrease the variance. For any real number m, define

$$\mathrm{IF}_m^\dagger(\cdot|\bar{E}, P) : o \mapsto \frac{\mathrm{Cov}_{Q_u}(U, \bar{E}(U)) - m}{\mathrm{Var}_{Q_u}^2(U)}\big[(u - Q_u[U])^2 - \mathrm{Var}_{Q_u}(U)\big].$$

When $m = m^* \equiv \mathrm{Cov}_{Q_u}(U, E[\tilde{Y}|U])$, we have $\mathrm{IF}_{m^*}^\dagger(\cdot|\bar{E}, P) = \mathrm{IF}^\dagger(\cdot|\bar{E}, P)$. Note also that $|m^*| \leq c\,\mathrm{Var}_{Q_u}^{1/2}(E[\tilde{Y}|U]) \leq M$ since $\mathrm{Var}_{Q_u}^{1/2}(U) \leq c$, where the upper bounds c and M exist according to (A.1) and (A.2). Therefore, we see the variance of $\mathrm{IF}^*(\cdot|\bar{E}, Q_u, G) + \mathrm{IF}^\dagger(\cdot|\bar{E}, P)$ is upper bounded by

$$\sup_{-M \leq m \leq M} P\big(\big[\mathrm{IF}^*(\cdot|\bar{E}, Q_u, G) + \mathrm{IF}_m^\dagger(\cdot|\bar{E}, P)\big]^2\big). \tag{15}$$

For each given m, the variance of $\mathrm{IF}^*(\cdot|\bar{E}, Q_u, G) + \mathrm{IF}^{\dagger}_m(\cdot|\bar{E}, P)$ can be estimated via the sample variance of the same quantity but with the unknown parameters replaced by the corresponding estimates. Therefore, the quantity in (15) can be estimated by taking the supremum of these estimates over $m \in [-M, M]$, and we denote this resulting estimate by $\sigma_n^{\dagger 2}$. Then $\sigma_n^{\dagger 2}$ is also a valid upper bound of the sample variance of the influence function of $S(\mathbb{P}_n, \hat{P}_n)$, so the Wald-type confidence intervals constructed using $\sigma_n^{\dagger 2}$ will have conservative coverage asymptotically. An alternative is to construct a (conservative) confidence interval using the bootstrap percentile t-method.

APPENDIX C: PROOF OF THEOREM 4.1

Before proceeding to the proof, we make the following asymptotic stability assumptions:

(A.7) $\hat{E}_n$ defined in (10) for a given predictor U_k consistently estimates (pointwise in its arguments) the true conditional mean residual life function $E_0(u, s, k) \equiv E[\tilde{Y}|U_k = u, X \geq s]$, which is assumed to be uniformly bounded and left continuous in s, and with k_j defined in (8),

$$E\Big[\sup_{(j,s)\in\{q_n,\ldots,n\}\times\mathcal{T}} \big|E_0(U_{k_j}, s, k_j) - E_0(U_{k_{j-1}}, s, k_{j-1})\big|\Big] = o\big(n^{-1/2}\big).$$

(A.8) If the slope parameter $\Psi_k(P) \neq 0$ for some k, there exists a sufficiently large $c > 0$ and a sequence of nonempty subsets $\mathcal{K}_n^* \subseteq \mathcal{K}_n = \{1, \ldots, p_n\}$ such that

$$\inf_{k\in\mathcal{K}_n^*} \left|\frac{\mathrm{Cov}(U_k, T)}{\mathrm{Var}(U_k)}\right| - \sup_{l\in\mathcal{K}_n\setminus\mathcal{K}_n^*} \left|\frac{\mathrm{Cov}(U_l, T)}{\mathrm{Var}(U_l)}\right| \geq c\sqrt{\frac{\log(n \vee p_n)}{q_n}},$$

where the supremum over $l \in \mathcal{K}_n \setminus \mathcal{K}_n^*$ is defined to be 0 if $\mathcal{K}_n^* = \mathcal{K}_n$, and

$$\mathrm{Diam}\big(\mathcal{K}_n^*\big) \equiv \sup_{k,l\in\mathcal{K}_n^*} \left|\left|\frac{\mathrm{Cov}(U_k, T)}{\mathrm{Var}(U_k)}\right| - \left|\frac{\mathrm{Cov}(U_l, T)}{\mathrm{Var}(U_l)}\right|\right| = o\big(n^{-1/2}\big).$$

Condition (A.7) guarantees that differences in selected predictors when the used subsample size changes from $j-1$ to j have asymptotically negligible effect on the corresponding mean residual life functions. Condition (A.8) ensures a separation of the sequence of alternative hypotheses from the global null in an asymptotic sense. The first part of this condition extends the usual root-n rate decay restriction made on local alternatives studied for hypothesis tests (Davidson and MacKinnon (1987)) to our setting, where the number of predictors grows with sample size. The second part of this condition can be viewed as a type of identifiability condition for the set of predictors $\mathcal{K}_n^*$ that are most associated with T.

The proof below is developed in a special case of taking a user-defined coarsening c so that $c(U)$ is a degenerate random variable, which reasonably reduces $\hat{G}_n(\cdot|u)$ to $\hat{G}_n(\cdot)$, and this is supported by the independent censoring assumption.

PROOF. We will show the asymptotic normality of $\sqrt{n - q_n}\bar{\sigma}_n^{-1}[S_n^* - \Psi(P)]$, where $\Psi(P)$ is defined in (2). From the expression S_n^* in (9), the desired result will follow from the limiting distribution of

$$\frac{1}{\sqrt{n - q_n}} \sum_{j=q_n}^{n-1} \hat{\sigma}_{nj}^{-1} m_j\big[S_{k_j}(\delta_{O_{j+1}}, P) - \Psi_{k_j}(P)\big],$$

with certain remainder terms shown to be asymptotically negligible in Lemmas S6.19–S6.22, based on concentration results and supportive preliminaries developed in Lemmas S6.1–S6.18.

We start by introducing a decomposition of the stabilized one-step estimator. The distribution of P is identified by (E_0, Q_u, G), where $E_0(u,s,k) \equiv E[\tilde{Y}|U_k = u, X \geq s]$. Replacing in various ways each feature of P by its estimator introduced in Section 3 gives $\hat{P}_{nj} = (\hat{E}_j, \mathbb{Q}_j, \hat{G}_n)$; $\hat{P}''_{nj} = (E_0, \mathbb{Q}_j, \hat{G}_n)$ and $\hat{P}'''_{nj} = (E_0, \mathbb{Q}_j, G)$. Therefore, we are able to decompose the statistic of interest as

$$
\begin{aligned}
&\sqrt{n-q_n}\bar{\sigma}_n^{-1}\big[S_n^* - \Psi(P)\big] \\
&\quad= \frac{1}{\sqrt{n-q_n}} \sum_{j=q_n}^{n-1} \hat{\sigma}_{nj}^{-1} m_j \big[S_{k_j}(\delta_{O_{j+1}}, \hat{P}_{nj}) - S_{k_j}\big(\delta_{O_{j+1}}, \hat{P}''_{nj}\big)\big] \\
&\qquad+ \frac{1}{\sqrt{n-q_n}} \sum_{j=q_n}^{n-1} \hat{\sigma}_{nj}^{-1} m_j \big[S_{k_j}\big(\delta_{O_{j+1}}, \hat{P}''_{nj}\big) - S_{k_j}\big(\delta_{O_{j+1}}, \hat{P}'''_{nj}\big)\big] \\
&\qquad+ \frac{1}{\sqrt{n-q_n}} \sum_{j=q_n}^{n-1} \hat{\sigma}_{nj}^{-1} m_j \big[S_{k_j}\big(\delta_{O_{j+1}}, \hat{P}'''_{nj}\big) - S_{k_j}(\delta_{O_{j+1}}, P)\big] \\
&\qquad+ \frac{1}{\sqrt{n-q_n}} \sum_{j=q_n}^{n-1} \hat{\sigma}_{nj}^{-1} m_j \big[S_{k_j}(\delta_{O_{j+1}}, P) - \Psi_{k_j}(P)\big] \\
&\qquad+ \frac{1}{\sqrt{n-q_n}} \sum_{j=q_n}^{n-1} \hat{\sigma}_{nj}^{-1} m_j \big[\Psi_{k_j}(P) - \Psi(P)\big] \equiv \mathrm{(I)} + \mathrm{(II)} + \mathrm{(III)} + \mathrm{(IV)} + \mathrm{(V)}.
\end{aligned} \tag{16}
$$

Using Lemmas S6.19–S6.22 mentioned above, in conjunction with (A.7)–(A.8), gives the asymptotic negligibility of (I), (II), (III) and (V). Therefore, the remaining task is to show the asymptotic normality of (IV). Modifying the expression in (7), with $(\mathbb{P}_n, \hat{P}_n)$ replaced by $(\delta_{O_{j+1}}, P)$ and with the predictor taken as U_{k_j}, gives that $S_{k_j}(\delta_{O_{j+1}}, P) = \Psi_{k_j}(P) + \mathrm{IF}^*_{k_j}(O_{j+1}|P)$. This further implies that

$$
\mathrm{(IV)} = \frac{1}{\sqrt{n-q_n}} \sum_{j=q_n}^{n-1} \hat{\sigma}_{nj}^{-1} m_j \,\mathrm{IF}^*_{k_j}(O_{j+1}|P),
$$

where $\mathrm{IF}^*_{k_j}(\cdot|P)$ is $\mathrm{IF}^*_k(\cdot|P)$ with the predictor U_k taken as U_{k_j}. Decompose (IV) into two terms:

$$
\frac{1}{\sqrt{n-q_n}} \sum_{j=q_n}^{n-1} \left[\frac{\sigma_{nj}}{\hat{\sigma}_{nj}} - 1\right] \frac{m_j}{\sigma_{nj}} \mathrm{IF}^*_{k_j}(O_{j+1}|P) + \frac{1}{\sqrt{n-q_n}} \sum_{j=q_n}^{n-1} \frac{m_j}{\sigma_{nj}} \mathrm{IF}^*_{k_j}(O_{j+1}|P). \tag{17}
$$

Note that $\hat{\sigma}_{nj}$ in the first term above involves the partial-sample estimator $\hat{P}_{nj}$.

Let $\lesssim$ denote "bounded above up to a universal multiplicative constant that does not depend on (j,n)," and $\sigma^2_{nj} \equiv \int \mathrm{IF}^*_{k_j}(o|P)^2 \, dP(o) = \mathrm{Var}(\mathrm{IF}^*_{k_j}(O|P))$. Note that σ^2_{nj} is bounded away from zero by (A.6), which implies that $\min_{k\in\mathbb{N}} \mathrm{Var}(\mathrm{IF}^*_k(O|P))$ is bounded away from zero; namely, there exists some constant $\epsilon > 0$ so that $\sigma^2_{nj} \geq \min_{k\in\mathbb{N}} \mathrm{Var}(\mathrm{IF}^*_k(O|P)) \geq \epsilon$. The first term in the decomposition of (IV) in (17) is seen to be of order $o_p(1)$ as follows. Let $\mathcal{O}_{nj} \equiv \sigma(\{O_1, \ldots, O_j\}, \hat{G}_n)$, and

$$
H_{nj} \equiv \frac{1}{\sqrt{n-q_n}} \left[\frac{\sigma_{n,j-1}}{\hat{\sigma}_{n,j-1}} - 1\right] \frac{m_{j-1}}{\sigma_{n,j-1}} \mathrm{IF}^*_{k_{j-1}}(O_j|P);
$$

the first term in the decomposition of (IV) in (17) is equal to $\sum_{j=q_n}^{n-1} H_{n,j+1}$. Note that $E|H_{nj}| < \infty$, using the fact that σ^2_{nj} is bounded away from zero by (A.6) and so is $\hat{\sigma}^2_{nj}$ by

(S6.11.1) of Lemma S6.11, and also that H_{nj} is $\mathcal{O}_{nj}$-measurable. Along with

$$E[H_{n,j+1}|\mathcal{O}_{nj}] = \frac{1}{\sqrt{n-q_n}} E\left[\left(\frac{\sigma_{nj}}{\hat{\sigma}_{nj}} - 1\right)\frac{m_j}{\sigma_{nj}} E\big[\mathrm{IF}^*_{k_j}(O_{j+1}|P)|\mathcal{O}_{nj}\big]\right] = 0,$$

$\{(H_{nj}, \mathcal{O}_{nj}), j = q_n + 1, \ldots, n\}$ is a martingale difference sequence. Moreover, we have that

$$|H_{n,j+1}| \lesssim \sqrt{\frac{\log(n \vee p_n)}{q_n(n-q_n)}} \equiv B_n. \tag{18}$$

Then Chebyshev's inequality implies that for $\varepsilon > 0$,

$$\begin{aligned}
\mathrm{P}\left(\left|\sum_{j=q_n}^{n-1} H_{n,j+1}\right| \geq \varepsilon\right) &\leq \varepsilon^{-2} E\left[\left(\sum_{j=q_n}^{n-1} H_{n,j+1}\right)^2\right] \\
&= \varepsilon^{-2}\left(\sum_{j=q_n}^{n-1} E\big[H^2_{n,j+1}\big] + 2 \sum_{q_n \leq i < j \leq n-1} E\big[H_{n,i+1} E[H_{n,j+1}|\mathcal{O}_{nj}]\big]\right) \\
&= \varepsilon^{-2} \sum_{j=q_n}^{n-1} E\big[H^2_{n,j+1}\big] \leq \varepsilon^{-2}(n-q_n) B_n^2 \to 0;
\end{aligned}$$

this result disposes of the first term in the decomposition of (IV) in (17).

Observe that the second term in (17) is a sum of martingale differences because $E[m_j \mathrm{IF}^*_{k_j}(O_{j+1}|P)|O_1, \ldots, O_j] = 0$. Therefore, it converges in distribution to standard normal by the martingale central limit theorem for triangular arrays (e.g., Theorem 2 in Gaenssler, Strobel and Stute (1978)) under the following conditions:

$$\frac{1}{n-q_n} \sum_{j=q_n}^{n-1} E\left[\frac{[\mathrm{IF}^*_{k_j}(O_{j+1}|P)]^2}{\sigma^2_{nj}} \Big| O_1, \ldots, O_j\right] \xrightarrow{p} 1;$$

$$\frac{1}{n-q_n} \sum_{j=q_n}^{n-1} E\left[\frac{[\mathrm{IF}^*_{k_j}(O_{j+1}|P)]^2}{\sigma^2_{nj}} 1\left(\left|\frac{\mathrm{IF}^*_{k_j}(O_{j+1}|P)}{\sigma_{nj}}\right| > \epsilon_0\sqrt{n-q_n}\right) \Big| O_1, \ldots, O_j\right] \xrightarrow{p} 0$$

for every $\epsilon_0 > 0$. The first condition follows from the definition of σ^2_{nj}, which implies that each term in the summation is identically equal to 1. The second condition holds because $\mathrm{IF}^*_k(\cdot|P)$ is uniformly bounded over k in view of (13) and (6) (giving the expression for IF^*_k), (A.1)–(A.5), (A.6)–(A.7) and further, σ^2_{nj} is assumed to be uniformly bounded away from zero by (A.6). □

Acknowledgments. We thank Peter Gilbert for suggesting the application in Section 7.

Funding. AL was supported by the National Institutes of Health (NIH) under award number DP2-LM013340.

IWM was supported by NIH under award 1R01 AG062401 and by the National Science Foundation (NSF) under award DMS-2112938. The content is solely the responsibility of the authors and does not necessarily represent the official views of NIH or NSF.

SUPPLEMENTARY MATERIAL

Supplement to "Efficient estimation of the maximal association between multiple predictors and a survival outcome" (DOI: 10.1214/23-AOS2313SUPP; .zip). Proofs, technical details, R code and additional results involving the simulation studies and the real data application are placed in the supplement (Huang, Luedtke and McKeague (2023)).

REFERENCES

ANDERSEN, P. K. and GILL, R. D. (1982). Cox's regression model for counting processes: A large sample study. *Ann. Statist.* **10** 1100–1120. MR0673646

ANTONIADIS, A., FRYZLEWICZ, P. and LETUÉ, F. (2010). The Dantzig selector in Cox's proportional hazards model. *Scand. J. Stat.* **37** 531–552. MR2779635 https://doi.org/10.1111/j.1467-9469.2009.00685.x

BINDER, H., PORZELIUS, C. and SCHUMACHER, M. (2011). An overview of techniques for linking high-dimensional molecular data to time-to-event endpoints by risk prediction models. *Biom. J.* **53** 170–189. MR2897395 https://doi.org/10.1002/bimj.201000152

BØVELSTAD, H. M., NYGÅRD, S. and BORGAN, Ø. (2009). Survival prediction from clinico-genomic models—a comparative study. *BMC Bioinform.* **10** Article 413.

BRADIC, J., FAN, J. and JIANG, J. (2011). Regularization for Cox's proportional hazards model with NP-dimensionality. *Ann. Statist.* **39** 3092–3120. MR3012402 https://doi.org/10.1214/11-AOS911

BUCKLEY, J. and JAMES, I. (1979). Linear regression with censored data. *Biometrika* **66** 429–436.

BUNEA, F. and MCKEAGUE, I. W. (2005). Covariate selection for semiparametric hazard function regression models. *J. Multivariate Anal.* **92** 186–204. MR2102251 https://doi.org/10.1016/j.jmva.2003.09.006

CAI, T., HUANG, J. and TIAN, L. (2009). Regularized estimation for the accelerated failure time model. *Biometrics* **65** 394–404. MR2751463 https://doi.org/10.1111/j.1541-0420.2008.01074.x

CHAI, H., ZHANG, Q., HUANG, J. and MA, S. (2019). Inference for low-dimensional covariates in a high-dimensional accelerated failure time model. *Statist. Sinica* **29** 877–894. MR3931392

DATTA, S., LE-RADEMACHER, J. and DATTA, S. (2007). Predicting patient survival from microarray data by accelerated failure time modeling using partial least squares and LASSO. *Biometrics* **63** 259–271. MR2345596 https://doi.org/10.1111/j.1541-0420.2006.00660.x

DAVIDSON, R. and MACKINNON, J. G. (1987). Implicit alternatives and the local power of test statistics. *Econometrica* **55** 1305–1329. MR0923463 https://doi.org/10.2307/1913558

DEVLIN, S. J., GNANADESIKAN, R. and KETTENRING, J. R. (1975). Robust estimation and outlier detection with correlation coefficients. *Biometrika* **62** 531–545.

ENGLER, D. and LI, Y. (2009). Survival analysis with high-dimensional covariates: An application in microarray studies. *Stat. Appl. Genet. Mol. Biol.* **8** Art. 14. MR2476392 https://doi.org/10.2202/1544-6115.1423

FAN, J., FENG, Y. and WU, Y. (2010). High-dimensional variable selection for Cox's proportional hazards model. In *Borrowing Strength: Theory Powering Applications—a Festschrift for Lawrence D. Brown*. *Inst. Math. Stat. (IMS) Collect.* **6** 70–86. IMS, Beachwood, OH. MR2798512

FAN, J. and LI, R. (2002). Variable selection for Cox's proportional hazards model and frailty model. *Ann. Statist.* **30** 74–99. MR1892656 https://doi.org/10.1214/aos/1015362185

FANG, E. X., NING, Y. and LIU, H. (2017). Testing and confidence intervals for high dimensional proportional hazards models. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **79** 1415–1437. MR3731669 https://doi.org/10.1111/rssb.12224

GAENSSLER, P., STROBEL, J. and STUTE, W. (1978). On central limit theorems for martingale triangular arrays. *Acta Math. Acad. Sci. Hung.* **31** 205–216. MR0471030 https://doi.org/10.1007/BF01901971

GILBERT, P. B., JURASKA, M., DECAMP, A. C., KARUNA, S., EDUPUGANTI, S., MGODI, N. et al. (2017). Basis and statistical design of the passive HIV-1 antibody mediated prevention (AMP) test-of-concept efficacy trials. *Stat. Commun. Infec. Dis.* **9** 20160001. MR3743441 https://doi.org/10.1515/scid-2016-0001

GORST-RASMUSSEN, A. and SCHEIKE, T. (2013). Independent screening for single-index hazard rate models with ultrahigh dimensional features. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **75** 217–245. MR3021386 https://doi.org/10.1111/j.1467-9868.2012.01039.x

HE, X., WANG, L. and HONG, H. G. (2013). Quantile-adaptive model-free variable screening for high-dimensional heterogeneous data. *Ann. Statist.* **41** 342–369. MR3059421 https://doi.org/10.1214/13-AOS1087

HONG, H. G., CHEN, X., CHRISTIANI, D. C. and LI, Y. (2018). Integrated powered density: Screening ultrahigh dimensional covariates with survival outcomes. *Biometrics* **74** 421–429. MR3825328 https://doi.org/10.1111/biom.12820

HONG, H. G., CHEN, X., KANG, J. and LI, Y. (2020). The L_q-norm learning for ultrahigh-dimensional survival data: An integrative framework. *Statist. Sinica* **30** 1213–1233. MR4257530 https://doi.org/10.5705/ss.202017.0537

HONG, H. G., KANG, J. and LI, Y. (2018). Conditional screening for ultra-high dimensional covariates with survival outcomes. *Lifetime Data Anal.* **24** 45–71. MR3742906 https://doi.org/10.1007/s10985-016-9387-7

HUANG, J. and MA, S. (2010). Variable selection in the accelerated failure time model via the bridge method. *Lifetime Data Anal.* **16** 176–195. MR2608284 https://doi.org/10.1007/s10985-009-9144-2

HUANG, J., MA, S. and XIE, H. (2006). Regularized estimation in the accelerated failure time model with high-dimensional covariates. *Biometrics* **62** 813–820. MR2247210 https://doi.org/10.1111/j.1541-0420.2006.00562.x

HUANG, T.-J, LUEDTKE, A. and MCKEAGUE, I. W (2023). Supplement to "Efficient estimation of the maximal association between multiple predictors and a survival outcome." https://doi.org/10.1214/23-AOS2313SUPP

HUANG, T.-J., MCKEAGUE, I. W. and QIAN, M. (2019). Marginal screening for high-dimensional predictors of survival outcomes. *Statist. Sinica* **29** 2105–2139. MR3970349

JIN, Z., LIN, D. Y., WEI, L. J. and YING, Z. (2003). Rank-based inference for the accelerated failure time model. *Biometrika* **90** 341–353. MR1986651 https://doi.org/10.1093/biomet/90.2.341

JOHNSON, B. A. (2008). Variable selection in semiparametric linear regression with censored data. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **70** 351–370. MR2424757 https://doi.org/10.1111/j.1467-9868.2008.00639.x

JOHNSON, B. A., LIN, D. Y. and ZENG, D. (2008). Penalized estimating functions and variable selection in semiparametric regression models. *J. Amer. Statist. Assoc.* **103** 672–680. MR2435469 https://doi.org/10.1198/016214508000000184

KOUL, H., SUSARLA, V. and VAN RYZIN, J. (1981). Regression analysis with randomly right-censored data. *Ann. Statist.* **9** 1276–1288. MR0630110

LAI, T. L. and YING, Z. (1991a). Large sample theory of a modified Buckley–James estimator for regression analysis with censored data. *Ann. Statist.* **19** 1370–1402. MR1126329 https://doi.org/10.1214/aos/1176348253

LAI, T. L. and YING, Z. (1991b). Rank regression methods for left-truncated and right-censored data. *Ann. Statist.* **19** 531–556. MR1105835 https://doi.org/10.1214/aos/1176348110

LI, J., ZHENG, Q., PENG, L. and HUANG, Z. (2016). Survival impact index and ultrahigh-dimensional model-free screening with survival outcomes. *Biometrics* **72** 1145–1154. MR3591599 https://doi.org/10.1111/biom.12499

LI, Y., DICKER, L. and ZHAO, S. D. (2014). The Dantzig selector for censored linear regression models. *Statist. Sinica* **24** 251–268. MR3183683

LIU, Y., CHEN, X. and LI, G. (2020). A new joint screening method for right-censored time-to-event data with ultra-high dimensional covariates. *Stat. Methods Med. Res.* **29** 1499–1513. MR4106953 https://doi.org/10.1177/0962280219864710

LOCKHART, R., TAYLOR, J., TIBSHIRANI, R. J. and TIBSHIRANI, R. (2014). A significance test for the lasso. *Ann. Statist.* **42** 413–468. MR3210970 https://doi.org/10.1214/13-AOS1175

LUEDTKE, A. R. and VAN DER LAAN, M. J. (2018). Parametric-rate inference for one-sided differentiable parameters. *J. Amer. Statist. Assoc.* **113** 780–788. MR3832226 https://doi.org/10.1080/01621459.2017.1285777

MA, S. and DU, P. (2012). Variable selection in partly linear regression model with diverging dimensions for right censored data. *Statist. Sinica* **22** 1003–1020. MR2987481 https://doi.org/10.5705/ss.2010.267

MA, S., LI, R. and TSAI, C.-L. (2017). Variable screening via quantile partial correlation. *J. Amer. Statist. Assoc.* **112** 650–663. MR3671759 https://doi.org/10.1080/01621459.2016.1156545

MAGARET, C. A., BENKESER, D. C., WILLIAMSON, B. D., BORATE, B. R., CARPP, L. N., GEORGIEV, I. S., SETLIFF, I., DINGENS, A. S., SIMON, N. et al. (2019). Prediction of VRC01 neutralization sensitivity by HIV-1 gp160 sequence features. *PLoS Comput. Biol.* **15** e1006952. https://doi.org/10.1371/journal.pcbi.1006952

PAN, W., WANG, X., XIAO, W. and ZHU, H. (2019). A generic sure independence screening procedure. *J. Amer. Statist. Assoc.* **114** 928–937. MR3963192 https://doi.org/10.1080/01621459.2018.1462709

PFANZAGL, J. (1982). *Contributions to a General Asymptotic Statistical Theory. Lecture Notes in Statistics* **13**. Springer, New York. With the assistance of W. Wefelmeyer. MR0675954

PFANZAGL, J. (1990). *Estimation in Semiparametric Models: Some Recent Developments. Lecture Notes in Statistics* **63**. Springer, New York. MR1048589 https://doi.org/10.1007/978-1-4612-3396-1

RITOV, Y. (1990). Estimation in a linear regression model with censored data. *Ann. Statist.* **18** 303–328. MR1041395 https://doi.org/10.1214/aos/1176347502

ROSENWALD, A., WRIGHT, G., CHAN, W. C., CONNORS, J. M., CAMPO, E., FISHER, R. I., GASCOYNE, R. D., MULLER-HERMELINK, H. K., SMELAND, E. B. et al. (2002). The use of molecular profiling to predict survival after chemotherapy for diffuse large-B-cell lymphoma. *N. Engl. J. Med.* **346** 1937–1947. https://doi.org/10.1056/NEJMoa012914

SINNOTT, J. A. and CAI, T. (2016). Inference for survival prediction under the regularized Cox model. *Biostatistics* **17** 692–707. MR3604274 https://doi.org/10.1093/biostatistics/kxw016

SMOLA, A. J., GRETTON, A. and BORGWARDT, K. (2006). Maximum mean discrepancy. In 13*th International Conference*, *ICONIP* 2006, *Hong Kong*, *China*, *October* 3–6, 2006: *Proceedings*.

SONG, R., LU, W., MA, S. and JENG, X. J. (2014). Censored rank independence screening for high-dimensional survival data. *Biometrika* **101** 799–814. MR3286918 https://doi.org/10.1093/biomet/asu047

STUTE, W. and WANG, J.-L. (1993). The strong law under random censorship. *Ann. Statist.* **21** 1591–1607. MR1241280 https://doi.org/10.1214/aos/1176349273

SZÉKELY, G. J., RIZZO, M. L. and BAKIROV, N. K. (2007). Measuring and testing dependence by correlation of distances. *Ann. Statist.* **35** 2769–2794. MR2382665 https://doi.org/10.1214/009053607000000505

TAYLOR, J. and TIBSHIRANI, R. (2018). Post-selection inference for ℓ_1-penalized likelihood models. *Canad. J. Statist.* **46** 41–61. MR3767165 https://doi.org/10.1002/cjs.11313

TIBSHIRANI, R. (1997). The lasso method for variable selection in the Cox model. *Stat. Med.* **16** 385–395. https://doi.org/10.1002/(sici)1097-0258(19970228)16:4<385::aid-sim380>3.0.co;2-3

TSIATIS, A. A. (1990). Estimating regression parameters using linear rank tests for censored data. *Ann. Statist.* **18** 354–372. MR1041397 https://doi.org/10.1214/aos/1176347504

VAN DE GEER, S. (1995). Exponential inequalities for martingales, with application to maximum likelihood estimation for counting processes. *Ann. Statist.* **23** 1779–1801. MR1370307 https://doi.org/10.1214/aos/1176324323

VAN DER LAAN, M. J., GILL, R. D. and ROBINS, J. M. (2000). Locally efficient estimation in censored data models: Theory and examples Technical Report, Division of Biostatistics, Univ. California, Berkeley, CA.

VAN DER LAAN, M. J. and HUBBARD, A. E. (1998). Locally efficient estimation of the survival distribution with right-censored data and covariates when collection of data is delayed. *Biometrika* **85** 771–783. MR1666754 https://doi.org/10.1093/biomet/85.4.771

VAN DER LAAN, M. J. and ROBINS, J. M. (2003). *Unified Methods for Censored Longitudinal Data and Causality*. *Springer Series in Statistics*. Springer, New York. MR1958123 https://doi.org/10.1007/978-0-387-21700-0

WHITNEY, D., SHOJAIE, A. and CARONE, M. (2019). Comment: Models as (deliberate) approximations [MR4048582; MR4048583]. *Statist. Sci.* **34** 591–598. MR4048590 https://doi.org/10.1214/19-STS747

WU, Y. (2012). Elastic net for Cox's proportional hazards model with a solution path algorithm. *Statist. Sinica* **22** 271–294. MR2933176 https://doi.org/10.5705/ss.2010.107

XIA, X. and LI, J. (2021). Copula-based partial correlation screening: A joint and robust approach. *Statist. Sinica* **31** 421–447. MR4270391 https://doi.org/10.5705/ss.20

XIA, X., LI, J. and FU, B. (2019). Conditional quantile correlation learning for ultrahigh dimensional varying coefficient models and its application in survival analysis. *Statist. Sinica* **29** 645–669. MR3931382

YING, Z. (1993). A large sample study of rank estimation for censored regression data. *Ann. Statist.* **21** 76–99. MR1212167 https://doi.org/10.1214/aos/1176349016

YOON, H., MACKE, J., WEST, A. P. JR, FOLEY, B., BJORKMAN, P. J., KORBER, B. et al. (2015). CATNAP: A tool to compile, analyze and tally neutralizing antibody panels. *Nucleic Acids Res.* **43**.

YU, Y., BRADIC, J. and SAMWORTH, R. J. (2021). Confidence intervals for high-dimensional Cox models. *Statist. Sinica* **31** 243–267.

ZHANG, H. H. and LU, W. (2007). Adaptive Lasso for Cox's proportional hazards model. *Biometrika* **94** 691–703. MR2410017 https://doi.org/10.1093/biomet/asm037

ZHAO, S. D. and LI, Y. (2012). Principled sure independence screening for Cox models with ultra-high-dimensional covariates. *J. Multivariate Anal.* **105** 397–411. MR2877525 https://doi.org/10.1016/j.jmva.2011.08.002

ZHAO, S. D. and LI, Y. (2014). Score test variable screening. *Biometrics* **70** 862–871. MR3295747 https://doi.org/10.1111/biom.12209

ZHONG, P.-S., HU, T. and LI, J. (2015). Tests for coefficients in high-dimensional additive hazard models. *Scand. J. Stat.* **42** 649–664. MR3391684 https://doi.org/10.1111/sjos.12127