



Published in final edited form as:

*Biometrics*. 2014 December ; 70(4): 862–871. doi:10.1111/biom.12209.

## Score test variable screening

**Sihai Dave Zhao** and

Department of Biostatistics and Epidemiology, University of Pennsylvania Perelman School of Medicine, Philadelphia, Pennsylvania, U.S.A

**Yi Li**

Department of Biostatistics, University of Michigan, Ann Arbor, Michigan, U.S.A

### Abstract

Variable screening has emerged as a crucial first step in the analysis of high-throughput data, but existing procedures can be computationally cumbersome, difficult to justify theoretically, or inapplicable to certain types of analyses. Motivated by a high-dimensional censored quantile regression problem in multiple myeloma genomics, this paper makes three contributions. First, we establish a score test-based screening framework, which is widely applicable, extremely computationally efficient, and relatively simple to justify. Secondly, we propose a resampling-based procedure for selecting the number of variables to retain after screening according to the principle of reproducibility. Finally, we propose a new iterative score test screening method which is closely related to sparse regression. In simulations we apply our methods to four different regression models and show that they can outperform existing procedures. We also apply score test screening to an analysis of gene expression data from multiple myeloma patients using a censored quantile regression model to identify high-risk genes.

### Keywords

High-dimensional data; Feature selection; Projected subgradient method; Score test; Variable screening

## 1 Introduction

High-dimensional datasets are now common in clinical genomics research. Though regularized estimation can consistently estimate sparse regression parameters even when  $p > n$  (Bühlmann et al., 2011), in practice these methods still perform poorly if  $p \gg n$  (Fan and Lv, 2008). Variable screening is crucial for quickly reducing tens of thousands of covariates to a more manageable size. Our interest in screening is motivated by our work with censored quantile regression in the study of the genomics of multiple myeloma, a blood cancer characterized by the hyperproliferation of plasma cells in the bone marrow. We are interested in identifying genes highly associated with the 10% quantile of the conditional

---

8 Supplementary Materials

Web Appendix A, which contains the theoretical justification of score screening and is referenced in Section 2, is available with this paper at the *Biometrics* website on Wiley Online Library. We also provide a zip file including an R implementation of our screening methods, a simulation example, and instructions.

survival distribution in order to better understand the biological basis of high-risk myeloma, in view of personalized medicine.

Perhaps the most popular screening framework is marginal screening, where each covariate is individually evaluated for association with the outcome and those with associations above some threshold are retained. Currently three major classes of marginal screening methods have been proposed. Wald screening retains covariates with the most significant marginal parameter estimates, and has been theoretically justified for generalized linear models and the Cox model (Fan and Lv, 2008; Fan and Song, 2010; Zhao and Li, 2012). Semiparametric screening assumes a functional form for the regression model but not for the probability distribution, and uses model-free statistics to quantify the associations between covariates and the outcome. Such methods have been proposed for single-index hazard models, linear transformation models, and general single-index models (Fan and Song, 2010; Zhu et al., 2011; Li et al., 2012a). Finally, nonparametric screening does not assume a functional form for the regression model and instead approximates it, using for example a B-spline basis. It retains covariates whose estimated functional relationships have the largest  $L_2$ -norms. Such methods have been studied for linear additive models and censored quantile regression (Fan et al., 2011; He et al., 2013). The distance correlation-based screening method of Li et al. (2012b) requires very few assumptions about both the regression model and the probability model. It is well-known that marginal screening can miss covariates that are only associated with the outcome conditional on other covariates. To address this difficulty, iterative versions of several of these procedures have been proposed, though without theoretical support.

However, there are several issues that make existing screening methods unsuitable for application to our multiple myeloma analysis. Wald screening using censored quantile regression estimators, such as those of Honore et al. (2002), Portnoy (2003), Peng and Huang (2008), or Wang and Wang (2009), has not been theoretically justified. Semiparametric screening is not appropriate because the probability model is actually critical in our case: we are interested only in genes that affect the 0.1 quantile, whereas semiparametric screening would identify genes that affect any quantile of the survival distribution. There were no nonparametric screening methods for censored quantile regression until very recently, with the work of He et al. (2013), but in practice their approach is still computationally cumbersome, especially for resampling or cross-validation procedures where screening must be repeated multiple times. There is also no efficient iterative screening procedure for this model.

To address these issues, we propose in this paper a marginal *score testing* framework, where we use score tests rather than Wald tests to effect variable screening. This has several advantages. First, score screening is a general approach which can be applied to any model that can be fit using an estimating equation, including censored quantile regression, as well as to semi- and nonparametric regression models. Second, theoretical justification for score test screening is much simpler than for other screening methods and generally requires only concentration inequalities. Third, because they only require fitting the null model, score tests are exceedingly computationally efficient. Finally, the score test perspective suggests a new

method for iterative screening that is easy to implement and is closely related to sparse regression, suggesting a possible approach to a theoretical justification.

In this paper we make three contributions. First, in Section 2 we propose marginal score test screening procedure and illustrate its application to several popular models. We give theoretical justifications for these procedures in Web Appendix A. Second, in Section 3 we propose a resampling-based method for choosing the number of covariates to retain after screening, based on the principle of reproducible screening. This procedure is only practical because score screening can be quickly computed. Third, in Section 4 we propose an iterative score test screening procedure based on projected subgradient methods from the numerical optimization literature. We illustrate our procedures on simulated data in Section 5, use them in our MM analysis in Section 6, conclude with a discussion in Section 7.

## 2 Score test screening

### 2.1 Method

Let  $\mathbf{X}_{ik} = (X_{ik1}, \dots, X_{ikp_n})^T$  be the vector of covariates measured at the  $k^{\text{th}}$  observation on the  $i^{\text{th}}$  subject, where  $k = 1, \dots, K_i$  and  $i = 1, \dots, n$ , and let  $\beta_0$  be a set of possibly infinite-dimensional parameters quantifying the association of the  $\mathbf{X}_{ik}$  with the outcome. For example, in linear models the outcome is a function of  $\mathbf{X}_{ik}^T \beta_0$  and  $\beta_0$  is a vector of scalar coefficients, and in additive models the outcome is a function of  $\sum_{j=1}^{p_n} f_j(X_{ikj})$  and  $\beta_0$  is the set of functions  $f_j$ . We will say that  $\beta_{0j} = 0$  implies that the  $j^{\text{th}}$  covariate is not functionally associated with the outcome and is thus unimportant, though this is a slight abuse of notation, as  $\beta_{0j}$  for irrelevant covariates would equal the scalar zero in linear models but the zero function in additive models. Finally, let  $\mathbf{U}(\beta)$  be an estimating equation for  $\beta_0$ , such that  $\mathbf{U}(\beta_0) \rightarrow \mathbf{0}$  in probability as  $n \rightarrow \infty$ .

Denote the set of important covariates by  $\mathcal{M} = \{j: \beta_{0j} \neq 0\}$ . We assume that its size  $|\mathcal{M}| = s_n$  is small and fixed or growing slowly. Our proposed marginal score test screening proceeds as follows:

1. Center and standardize each covariate to have mean 0 and variance 1.
2. For each covariate  $j$ , construct an estimating equation for  $\beta_{0j}$  assuming the marginal model that all other covariates are unrelated to the outcome. Denote this marginal estimating equation by  $U_j^M(\beta_j)$
3. Retain the parameters  $\hat{\mathcal{M}} = \{j: |U_j^M(0)| \geq \gamma_n\}$  for some threshold  $\gamma_n$ .

Each  $|U_j^M(0)|$  is the numerator of the score test statistic for  $H_0: \beta_{0j} = 0$  under the  $j^{\text{th}}$  marginal model and thus is a sensible screening statistic. We could also screen after dividing each  $U_j^M(0)$  by an estimate of its standard deviation. However, this would add computational complexity to our procedure, and even without doing so we will be able to achieve good results and give theoretical performance guarantees. In the presence of nuisance parameters, such as intercept terms, we propose using profiled score tests, where

we first estimate the nuisance parameters under the null model and then evaluate the  $U_j^M(0)$  fixing the value of nuisance terms at these estimates. To avoid theoretical difficulties we will assume that nuisance parameters are either known, or can be well-estimated in independent datasets, so that in the screening step they can be treated as constants.

In order for score screening to have desirable theoretical properties, we need the sample  $U_j^M(0)$  to quickly approach its population limit. Let  $u_j^M(\beta_j)$  be the limiting marginal estimating equation, such that  $U_j^M(\beta_j) \rightarrow u_j^M(\beta_j)$ .

**Condition 1**—For  $\kappa \in (0, 1/2)$  and  $c_2 > 0$ ,  $p_n P(|U_j^M(0) - u_j^M(0)| \geq c_2 n^{-\kappa}) \rightarrow 0$ .

In Web Appendix A we discuss the verification of Condition 1, which is often a simple consequence of a concentration inequality, and explicitly verify it for censored quantile regression. We also show that under this condition and a few other mild assumptions:

**Theorem 1**—If  $\gamma_n = c_1 n^{-\kappa/2}$ , then  $P(\mathcal{M} \subseteq \hat{\mathcal{M}}) \rightarrow 1$ .

**Theorem 2**—If  $\gamma_n = c_1 n^{-\kappa/2}$ , then  $P\{|\hat{\mathcal{M}}| \leq O(\sigma_{\max}^* n^{2\kappa})\} \rightarrow 1$ , where  $\sigma_{\max}^*$  is related to the largest singular value of the negative Jacobian of the limiting estimating equation.

Theorem 1 shows that marginal score testing can capture all of the important covariates with high probability. This holds even if  $p_n$  grows exponentially in  $n$ . Theorem 2 shows that the number of selected covariates is not too large, with high probability. For example, if  $\sigma_{\max}^*$  increased only polynomially in  $n$ ,  $|\hat{\mathcal{M}}|$  would increase polynomially, and the false positive rate would decrease quickly to zero.

## 2.2 Examples

When applied to the models studied thus far in the screening literature, score test screening gives procedures that are very similar to previously proposed procedures. Throughout this section we let  $K_i = 1$ , with each covariate vector  $\mathbf{X}_i = (X_{i1}, \dots, X_{ip_n})^T$ . We also assume that the  $X_{ij}$  have mean 0 and variance 1.

First consider the usual ordinary least squares model studied by Fan and Lv (2008), where  $Y_i$  is a continuous outcome. The full model is  $E(Y_i | \mathbf{X}_i) = \mathbf{X}_i^T \beta_0$ , so the  $j^{\text{th}}$  marginal score equation is  $U_j^M(\beta_j) = n^{-1} \sum_i X_{ij}(Y_i - X_{ij}\beta_j)$ . Score test screening then retains  $\hat{\mathcal{M}} = \{j: n^{-1} |\sum_i X_{ij} Y_i| \geq \gamma_n\}$ , which is exactly the correlation screening procedure originally proposed by Fan and Lv (2008).

Next consider the Cox model. Let  $T_i$  be the survival time,  $C_i$  the censoring time,  $Y_i = \min(T_i, C_i)$ ,  $\delta_i = I(T_i < C_i)$ ,  $\tilde{N}_i(s) = I(T_i > s, \delta_i = 1)$ , and  $\hat{Y}_i(s) = I(Y_i > s)$ . The marginal score equations are  $U_j^M(\beta_j) =$

$$\frac{1}{n} \sum_{i=1}^n \int \left\{ X_{ij} - \frac{\sum_{i=1}^n X_{ij} \exp(X_{ij} \beta_j) \tilde{Y}_{i(s)}}{\sum_{i=1}^n \exp(X_{ij} \beta_j) \tilde{Y}_{i(s)}} \right\} d\tilde{N}_i(s),$$

and  $\hat{\mathcal{M}} = \{j: |U_j^M(0)| \geq \gamma_n\}$ . This is exactly the screening procedure used by Gorst-Rasmussen and Scheike (2013).

Finally consider a nonparametric model, where we assume only that  $P(Y_i < y | \mathbf{X}_i)$  has a continuous distribution function  $F_0(y; \mathbf{X}_i, \boldsymbol{\beta}_0)$  whose dependence on  $\mathbf{X}_i$  is parametrized by  $\boldsymbol{\beta}_0$ . Conditional on  $\mathbf{X}_l$  and  $\mathbf{X}_m$ ,  $F_0(Y_l; \mathbf{X}_l, \boldsymbol{\beta}_0)$  and  $F_0(Y_m; \mathbf{X}_m, \boldsymbol{\beta}_0)$  are independent and identically distributed uniform random variables. This motivates defining  $\mathbf{U}(\boldsymbol{\beta}) =$

$$\frac{1}{n^2} \sum_{m=1}^n \sum_{l=1}^n \mathbf{X}_l \left[ I\{F_0(Y_l; \mathbf{X}_l, \boldsymbol{\beta}) < F_0(Y_m; \mathbf{X}_m, \boldsymbol{\beta})\} - \frac{1}{2} \right].$$

Since  $E\{\mathbf{U}(\boldsymbol{\beta}_0)\} = \mathbf{0}$ , this is an unbiased estimating equation for  $\boldsymbol{\beta}_0$ . Though it cannot be used to estimate  $\boldsymbol{\beta}_0$  because the functional form of  $F_0$  is unknown, it is still useful for constructing a screening procedure. The marginal score equations are  $U_j^M(\beta_j) =$

$$\frac{1}{n^2} \sum_{m=1}^n \sum_{l=1}^n X_{lj} \left[ I\{F_0(Y_l; X_{lj}, \beta_j) < F_0(Y_m; X_{mj}, \beta_j)\} - \frac{1}{2} \right].$$

When  $\beta_j = 0$ ,  $F_0(y; X_{lj}, 0)$  is a monotone function that does not depend on  $X_{lj}$ , which implies that  $U_j^M(0) = n^{-2} \sum_{lm} X_{lj} \{I(Y_l < Y_m) - 1/2\}$  and therefore

$\hat{\mathcal{M}} = \{j: |n^{-2} \sum_{lm} X_{lj} I(Y_l < Y_m)| \geq \gamma_n\}$ . This is very similar to proposal of Zhu et al.

(2011), who suggested  $\hat{\mathcal{M}} = \{j: n^{-1} \sum_m \{n^{-1} \sum_l X_{lj} I(Y_l < Y_m)\}^2 \geq \gamma_n\}$ .

Each of these screening procedures can be implemented as or more quickly than the corresponding Wald screening. In addition, the nonparametric screening procedure is impossible in the Wald framework. Each of these screening procedures can be theoretically justified by verifying Condition 1 and applying Theorems 1 and 2.

### 3 Reproducible screening threshold

In practice, it is unclear how best to choose the screening threshold  $\gamma_n$ . Fan and Lv (2008) suggested retaining the top  $n/\log n$  covariates. Zhao and Li (2012) proposed a method to choose  $\gamma_n$  based on the desired false positive rate of the set of retained covariates. Similarly, Zhu et al. (2011) suggested simulating auxiliary variables and setting the threshold so that no auxiliary variables are retained, and proved that this procedure controls the false positive rate of screening. Finally, He and Lin (2011) used the stability selection approach of Meinshausen and Bühlmann (2010) to retain covariates that are frequently selected when screening is performed on multiple subsamples of the data.

Though controlling the false positive rate is important, we believe that in practice the more relevant issue is the reproducibility of the screening procedure. Let  $\hat{\mathcal{M}}^{(j)}$  be the top  $j$  variables retained after screening our observed data. Suppose we had  $B$  other independent samples of the same size, from the same generating distribution, and let  $\mathcal{M}_b^{(j)}$  be the top  $j$  variables we retain after screening the  $b^{\text{th}}$  sample. Finally, let  $O_b^{(j)} = |\mathcal{M}_b^{(j)} \cap \hat{\mathcal{M}}^{(j)}|$  be the size of the overlap between the two sets. We would like to choose  $j$  such that the  $O_b^{(j)}$  are large on average, so that our screening results are reproducible across different samples. On the other hand, when  $j$  is large, the  $O_b^{(j)}$  will be large even if no variables were truly associated with the outcome, so reproducibility would be uninformative.

We propose comparing the size of the overlap to the number we would expect by chance under the null hypothesis that none of the  $p_n$  variables are associated with the outcome. The variables in  $\mathcal{M}_b^{(j)}$  can then be thought of as having been chosen at random. Conditional on the observed dataset, the  $O_b^{(j)}$  would therefore follow a hypergeometric distribution, with

$$E_{H_0}(O_b^{(j)}) = \frac{j^2}{p}, \quad \text{var}_{H_0}(O_b^{(j)}) = \frac{j^2(p-j)^2}{p^2(p-1)},$$

where the subscripts indicate that the expectation and variance are calculated under  $H_0$ . We propose to retain the top  $j$  variables such that the average of the  $O_b^{(j)}$  shows the greatest deviation from  $E_{H_0}(O_b^{(j)})$ , standardized by  $\text{var}_{H_0}(O_b^{(j)})$ .

Because we do not have  $B$  independent datasets, we approximate the  $\mathcal{M}_b^{(j)}$  using bootstrap samples of our observed data. Specifically, our threshold for reproducible screening is calculated as follows:

1. Choose a step size  $s$  and let  $\mathcal{J} = \{is : i \in \mathbb{N}, 1 \leq is \leq p_n\}$ .
2. For each  $j \in \mathcal{J}$ , screen the observed data to obtain  $\hat{\mathcal{M}}^{(j)}$ .
3. For each  $j \in \mathcal{J}$ , generate  $B$  bootstrap samples and screen the  $b^{\text{th}}$  sample to get  $\mathcal{M}_b^{(j)}$ .
4. Let  $O_b^{(j)} = |\mathcal{M}_b^{(j)} \cap \hat{\mathcal{M}}^{(j)}|$ .
5. Retain  $\hat{\mathcal{M}}^{(j^*)}$ , where

$$j^* = \arg \max_{j \in \mathcal{J}} \frac{B^{-1} \sum_b O_b^{(j)} - E_{H_0}(O_b^{(j)})}{\{B^{-1} \text{var}_{H_0}(O_b^{(j)})\}^{-1/2}}.$$

When the step size  $s = 1$ , we search for the optimal  $j$  across all  $j = \{1, \dots, p_n\}$ . In practice, to reduce computation time we can search over a smaller subset by taking a larger step size.

Our method is closely related to higher criticism thresholding (Donoho and Jin, 2008), but evaluates the reproducibility of each potential set of retained covariates, whereas higher criticism does not.

#### 4 Iterative score test screening

When the covariates are highly correlated, marginal screening may incur a large number of false positives, and may miss covariates that are only important conditional on other covariates. Fan and Lv (2008) and Fan et al. (2009) therefore proposed iterative screening: an initial set of covariates is first identified using marginal screening. Next a multivariate regularized selection procedure is used to further select a subset of these covariates. Finally the remaining covariates are again screened individually, but this time controlling for the covariates in the subset. All selected covariates are subjected to multivariate selection again, and the procedure iterates until some stopping rule is achieved.

However, this algorithm requires fitting regularized regression estimates at each step, which for complicated models can be difficult to implement and computationally intensive. Furthermore, its theoretical properties are very difficult to analyze. Zhu et al. (2011) proposed an alternative method which at each step performs marginal screening on the projections of each remaining covariate onto the orthogonal complement of the columns space of the already selected covariates. This method is akin to forward selection, so a covariate cannot be dropped from the selected set once it has been added.

Our score-test screening perspective suggests a new approach to iterative screening:

1. Set  $\boldsymbol{\beta}^{(0)} = \mathbf{0}$ .
2. For  $k = 1, \dots, K$ :
  - a. Let  $\mathbf{b}^{(k)} = \boldsymbol{\beta}^{(k-1)} - \alpha_k \mathbf{U}(\boldsymbol{\beta}^{(k-1)})$  for some step size  $\alpha_k$ .
  - b. Let  $\boldsymbol{\beta}^{(k)} = \Pi_R(\mathbf{b}^{(k)})$ , where  $\Pi_R: \mathbb{R}^{p_n} \rightarrow \mathbb{R}^{p_n}$  is the Euclidean projection onto the  $\ell_1$ -ball of radius  $R$ .
3. Retain covariates  $\hat{\mathcal{M}} = \{j; \beta_j^{(K)} \neq 0\}$ , where  $\beta_j^{(k)}$  is the  $j^{\text{th}}$  component of  $\boldsymbol{\beta}^{(k)}$ .

The intuition is that when  $k = 1$ , step 2(a) is equivalent to calculating the marginal score statistics  $U_j^M(0)$  and step 2(b) sets all but the largest of them to zero. Thus after a single iteration, this procedure is identical to score test screening. When  $k > 1$ , step 2(a) controls for the covariates selected in  $\boldsymbol{\beta}^{(k-1)}$  by using  $-\alpha_k \mathbf{U}(\boldsymbol{\beta}^{(k-1)})$  to update the importance of the covariate. Step 2(b) then again selects only the top covariates. In the ideal case where the sample size is infinite and  $\boldsymbol{\beta}^{(k-1)} = \boldsymbol{\beta}_0$ , step 2(a) gives  $\mathbf{b}^{(k)} = \boldsymbol{\beta}_0$  and step 2(b) selects the largest components of  $\boldsymbol{\beta}_0$ .

Our algorithm has several advantages. First, it does not require fitting any regularized regression estimates and is relatively computationally convenient. The evaluations of the  $\mathbf{U}(\boldsymbol{\beta}^{(k-1)})$  are quick to compute, and a simple algorithm for implementing the projection  $\Pi_R$  can be found in Daubechies et al. (2008), with a more efficient procedure proposed by Duchi

et al. (2008). Second, covariates can be dropped from the retained set as the iteration progresses, which is an improvement over forward selection. Third, our algorithm exactly corresponds to projected subgradient methods for minimizing nonsmooth functions. In fact, if  $\mathbf{U}(\boldsymbol{\beta})$  is the subdifferential of some loss function  $f(\boldsymbol{\beta})$ , it has been shown that

$$\lim_{k \rightarrow \infty} f(\boldsymbol{\beta}^{(k)}) = \inf_{\|\boldsymbol{\beta}\|_1 \leq R} f(\boldsymbol{\beta})$$

for certain choices of  $\alpha_k$  (Shor et al., 1985). The minimization problem on the right-hand side is exactly equivalent to the lasso (Tibshirani, 1996) with loss function  $f$ , and this links our iterative screening algorithm to sparse regression methods. Finally, when  $f$  is smooth, Agarwal et al. (2012) proved that  $\boldsymbol{\beta}^{(k)}$  converges to  $\boldsymbol{\beta}_0$  under certain conditions, and if a similar result holds for nonsmooth  $f$ , this connection may allow for a theoretical analysis of iterative score test screening.

There are three tuning parameters we must set when implementing iterative screening: the radius  $R$ , the step sizes  $\alpha_k$ , and the maximum number of iterations. We can choose  $R$  by either guessing the  $\ell_1$ -norm of the true  $\boldsymbol{\beta}_0$ . Since our algorithm can be viewed as a regression estimator, we can also minimize information criteria or cross-validated prediction errors. Since iterative screening tends to be time-consuming in high-dimensions, it is easiest to avoid cross-validation and to use a liberal guess for  $\|\boldsymbol{\beta}_0\|_1$ . To set the step sizes, one popular rule is to let the  $\alpha_k$  be square summable but not summable, with  $\alpha_k = \gamma/k$ . To choose  $\gamma$ , we first note that it can be shown that

$$\min_{k=1, \dots, K} f(\boldsymbol{\beta}^{(k)}) - \inf_{\|\boldsymbol{\beta}\|_1 \leq R} f(\boldsymbol{\beta}) \leq \frac{D^2 + G^2 \sum_{k=1}^K \alpha_k^2}{2 \sum_{k=1}^K \alpha_k}$$

where  $D$  is the Euclidean distance from  $\boldsymbol{\beta}^{(0)}$  to a point that minimizes  $f$  and  $G$  is an upper bound on  $\mathbf{U}(\boldsymbol{\beta}^{(k)})$  for all  $k$  (Shor et al., 1985). When  $\alpha_k = \gamma/k$ , this converges to zero as  $K \rightarrow \infty$ , but fixing  $K$  we can derive that the right-hand side is minimized at

$\gamma^2 = D^2 (G^2 \sum_{k=1}^K \alpha_k^2)^{-1} \rightarrow D^2 (G^2 \pi^2 / 6)^{-1}$ . We propose approximating  $D$  by  $R$  and  $G$  by  $\|\mathbf{U}(\boldsymbol{\beta}^{(0)})\|_2$  to get step sizes  $\alpha_k = R \sqrt{6} / \{k \pi \|\mathbf{U}(0)\|_2\}$ . Finally, the maximum number of iterations should ideally be as large as possible, with the speed of convergence depending on the restricted convexity and smoothness of  $f$  (Agarwal et al., 2012). In practice we stop after either  $\mathbf{U}(\boldsymbol{\beta}^{(k)}) \approx \mathbf{0}$ ,  $\boldsymbol{\beta}^{(k-1)} \approx \boldsymbol{\beta}^{(k)}$ , or  $K = 250$  iterations. Early stopping can be viewed as another way of regularizing the regression estimate  $\boldsymbol{\beta}$ .

## 5 Simulations

### 5.1 Settings

We illustrate our marginal and iterative score test screening on data simulated from four models, described below along with their corresponding estimating equations. We ran 100 simulations, each with  $n = 400$  observations and  $p_n = 10,000$  covariates. We compared our



methods to the semiparametric screening of Zhu et al. (2011), and when possible we also compared to Wald and nonparametric screening.

**Example 1 (accelerated failure time model)**—The accelerated failure time model is a useful alternative to the Cox model for survival outcomes (Wei, 1992) and posits that  $\log(T_i) = \mathbf{X}_i^T \boldsymbol{\beta}_0 + \epsilon_i$  where  $T_i$  are the survival times,  $\mathbf{X}_i$  are  $p_n \times 1$  covariate vectors, and  $\epsilon_i$  are independent of  $\mathbf{X}_i$ . We only observe follow-up times  $Y_i = \min(T_i, C_i)$  and censoring indicators  $\delta_i = I(T_i < C_i)$ , but the  $\boldsymbol{\beta}_0$  can be estimated using the estimating equation  $\mathbf{U}(\boldsymbol{\beta}) =$

$$n^{-1} \sum_{l=1}^n \sum_{m=1}^n (\mathbf{X}_m - \mathbf{X}_l) I\{e_l(\boldsymbol{\beta}) \leq e_m(\boldsymbol{\beta})\} \delta_i,$$

where  $e_i(\boldsymbol{\beta}) = \log(Y_i) - \mathbf{X}_i^T \boldsymbol{\beta}$  (Tsiatis, 1996; Jin et al., 2003; Cai et al., 2009).

Score test screening retains

$$\hat{\mathcal{M}} = \{j: |\sum_{lm} (X_{mj} - X_{lj}) I(Y_l \leq Y_m) \delta_l| \geq \gamma_n\},$$

and it is simple to verify Condition 1 for this procedure using Bernstein’s inequality for U-statistics (Hoeffding, 1963). We implemented Wald test screening using the estimator of Jin et al. (2003), available in the R package lss. Nonparametric screening has not been developed for this model.

We generated the covariates from a  $p$ -dimensional multivariate normal with a covariance matrix whose  $jk^{th}$  entry equaled  $0.8^{|j-k|}$ . We then let  $\beta_{0j} = 1.5$  for  $j = 5, 10, 15, 20, 25$  and  $j = 35, 40, 45, 50$ ,  $\beta_{0j} = -1.5$  for  $j = 30$ , and  $\beta_{0j} = 0$  for all other  $j$ . Under this construction the 30<sup>th</sup> covariate is marginally unimportant. We separated the nonzero entries of  $\boldsymbol{\beta}_0$  so that important covariates would be fairly correlated with a few unimportant covariates. Finally, we generated  $\epsilon_i$  from a standard normal distribution,  $T_i$  according to the model, and  $C_i$  from an exponential distribution with rate parameter 0.3 to give 30% censoring.

**Example 2 (linear censored quantile regression)**—For a quantile  $\tau \in (0, 1)$ , censored quantile regression models posit  $h(T_i) = \beta_{int}(\tau) + \mathbf{X}_i^T \boldsymbol{\beta}_0(\tau) + e_i(\tau)$ , where the intercept  $\beta_{int}(\tau)$  and the coefficients  $\boldsymbol{\beta}_0(\tau)$  depend on  $\tau$  and  $e_i(\tau)$  has  $\tau^{th}$  quantile equal to 0 conditional on  $\mathbf{X}_i$ . The  $h$  function is a known monotone transformation, and here we let it be the log function. In contrast to global models such as the Cox or accelerated failure time model, this censored quantile regression directly models the  $\tau$  conditional quantile and makes no assumptions about the other quantiles. Honore et al. (2002) proposed the estimating equation  $\mathbf{U}(\boldsymbol{\beta}) =$

$$\frac{1}{n} \sum_i \mathbf{X}_i \tau I\{h(Y_i) > \beta_{int} + \mathbf{X}_i^T \boldsymbol{\beta}\} - \frac{1}{n} \sum_i \mathbf{X}_i (1 - \tau) \hat{S}_{h(C)}\{h(Y_i)\}^{-1} I\{h(Y_i) \leq \beta_{int} + \mathbf{X}_i^T \boldsymbol{\beta}\} \delta_i \hat{S}_{h(C)}(\beta_{int} + \mathbf{X}_i^T \boldsymbol{\beta}),$$

where  $\hat{S}_{h(C)}$  is an estimate of  $S_{h(C)}(t) = P\{h(C_i) > t | \mathbf{X}_i\}$ . This estimate could be obtained by positing a regression model for  $h(C_i)$  conditional on the  $\mathbf{X}_i$ , but for theoretical and practical simplicity we will make the common assumption that  $C_i$  is completely independent of  $T_i$  and  $\mathbf{X}_i$  and use the Kaplan-Meier estimator (see for example Cheng et al. (1995), Uno et al. (2011), and He et al. (2013)).

Score test screening retains the parameters  $\{j: |U_j^M(0)| \geq \gamma_n\}$  where  $U_j^M(0) =$

$$\frac{1}{n} \sum_i X_{ij} \tau I\{h(Y_i) > \beta_{int}\} - \frac{1}{n} \sum_i X_{ij} (1 - \tau) \hat{S}_{h(C)}\{h(Y_i)\}^{-1} I\{h(Y_i) \leq \beta_{int}\} \delta_i \hat{S}_{h(C)}(\beta_{int}).$$

In Web Appendix A we verify Condition 1 for this screening procedure. To use score test screening, we first estimated the nuisance parameter  $\beta_{int}$  under the null model in an independently simulated dataset. We implemented Wald screening using the estimator of Peng and Huang (2008), available in the package `quantreg`. He et al. (2013) developed a nonparametric screening method for quantile regression, which we also applied.

We used the  $\beta_0$  and covariate structure as in Example 1, except that we thresholded each  $X_{ij}$  to have a magnitude of at most 2. We then

$\log(T_i) = \mathbf{X}_i^T \beta_0 + \epsilon_i \left\{ 9 + 1.5(X_{i,55} - X_{i,5}) / \Phi^{-1}(0.25) \right\}$ , where  $\epsilon_i$  followed a standard normal distribution. Under this construction covariates  $j \in \{5, \dots, 50\}$  are associated with the  $\tau = 0.5$  conditional quantile, and  $j \in \{10, \dots, 55\}$  are relevant to the  $\tau = 0.25$  conditional quantile. Here the 30<sup>th</sup> covariate is again marginally unimportant. Finally, we simulated  $C_i$  from an exponential distribution with rate 0.15 to give 30% censoring.

**Example 3 (nonlinear censored quantile regression)**—We generated survival times from a nonlinear censored quantile regression model adapted from Example 4 of He et al. (2013). If  $g_1(x) = x$ ,  $g_2(x) = (2x - 1)^2$ ,  $g_3(x) = \sin(2\pi x) / \{2 - \sin(2\pi x)\}$ ,  $g_4(x) = 0.1 \sin(2\pi x) + 0.2 \cos(2\pi x) + 0.3 \sin(2\pi x)^2 + 0.4 \cos(2\pi x)^3 + 0.5 \sin(2\pi x)^3$ , we simulated

$$\log(T_i) = 5g_1(X_{i,1}) + 5g_1(X_{i,2}) + 3g_2(X_{i,3}) + 3g_2(X_{i,4}) + 4g_3(X_{i,5}) + 4g_3(X_{i,6}) + 6g_4(X_{i,7}) + 6g_4(X_{i,8}) + \epsilon_i,$$

where  $\epsilon_i$  followed a standard normal distribution. We generated the  $\mathbf{X}_i$  as in Example 1 and  $\log(C_i)$  from an exponential distribution with rate 0.15 to give 30% censoring.

Under the null hypothesis the functions  $g_j = 0$  for all  $j$ , so the marginal estimating equations  $U_j^M(0)$  evaluated at zero are identical to those from Example 2. The theoretical justifications thus also follow from Example 2. We applied the nonparametric screening of He et al. (2013) as well, which was designed for this nonlinear setting.

**Example 4 (Cox model with measurement error)**—The Cox model is the most popular method for modeling the effect of covariates on survival, but in many cases the covariates may be measured with errors, where instead of observing  $\mathbf{X}_i$  we observe only  $\mathbf{W}_i$

$= \mathbf{X}_i + \epsilon_i$ . Not accounting for measurement error can result in bias, and to address this issue Song and Huang (2005) proposed the corrected score equation  $\mathbf{U}(\boldsymbol{\beta}) =$

$$\frac{1}{n} \sum_{i=1}^n \int \left[ \mathbf{W}_i + D(\boldsymbol{\beta}) - \frac{\sum_{i=1}^n \tilde{\mathbf{W}}_i(\boldsymbol{\beta}, s) \exp\{\tilde{\mathbf{W}}_i(\boldsymbol{\beta}, s)^T \boldsymbol{\beta}\} \tilde{Y}_i(s)}{\sum_{i=1}^n \exp\{\tilde{\mathbf{W}}_i(\boldsymbol{\beta}, s)^T \boldsymbol{\beta}\} \tilde{Y}_i(s)} \right] d\tilde{N}_i(s),$$

where  $\tilde{\mathbf{W}}_i(\boldsymbol{\beta}, s) = \mathbf{W}_i + D(\boldsymbol{\beta}) d\tilde{N}_i(s)$ ,  $D(\boldsymbol{\beta}) = E\{\epsilon_i \exp(\epsilon_i^T \boldsymbol{\beta})\} / E\{\exp(\epsilon_i^T \boldsymbol{\beta})\} - E(\epsilon_i)$ ,  $\tilde{N}_i(s) = I(T_i \leq s, \delta_i = 1)$  is the observed failure process, and  $\tilde{Y}_i(s) = I(Y_i \geq s)$  is the at-risk process. The  $D(\boldsymbol{\beta})$  term is unknown in general unless the distribution of  $\epsilon_i$  is known.

Under the null hypothesis of  $\boldsymbol{\beta}_0 = \mathbf{0}$ ,  $D(\mathbf{0}) = \mathbf{0}$ , so score test screening retains

$$\hat{\mathcal{M}} = \left[ j : \left| \frac{1}{n} \sum_{i=1}^n \int \left\{ W_{ij} - \frac{\sum_{i=1}^n W_{ij} \tilde{Y}_i(s)}{\sum_{i=1}^n \tilde{Y}_i(s)} \right\} d\tilde{N}_i(s) \right| \geq \gamma_n \right]$$

regardless of the distribution of  $\epsilon_i$ . Condition 1 can be verified using Lemmas 2 and 3 of Gorst-Rasmussen and Scheike (2013). Wald screening is not possible without knowing the distribution of  $\epsilon_i$ , and nonparametric screening has not been developed for this model.

We generated the covariates and set  $\boldsymbol{\beta}_0$  as in Example 1. We then generated the  $T_i$  from the usual Cox model with baseline hazard function equal to 1. Next we let  $\mathbf{W}_i = \mathbf{X}_i + \epsilon_i$ , where the  $\epsilon_i$  were independent of the  $\mathbf{X}_i$  and normally distributed with a compound symmetry covariance matrix with correlation parameter 0.5. We generated  $\log(C_i)$  from an exponential distribution with rate parameter 0.3 to give 30% censoring.

## 5.2 Results

These simulations were run on machines with 2 GHz Intel Xeon cores with 4GB of memory per core. Table 1 reports the average runtimes of these various screening methods and shows that our marginal score test procedure is by far the most computationally efficient. In Example 1 it is many orders of magnitude faster than Wald screening, and in Examples 2 and 3 it is 60 times faster than the nonparametric method of He et al. (2013). In each example it is also at least twice as fast as the semiparametric estimator of Zhu et al. (2011).

Table 2 compares score test screening to existing methods in terms of the minimum number of variables that need to be retained in order to capture all of the important covariates. All methods were comparable in Example 1. In Example 2, score screening was comparable to Wald screening and outperformed the nonparametric screening of He et al. (2013). Semiparametric screening performed the best but was unable to identify the fact that  $\beta_{0,5}$  was important only to the 0.5 quantile and  $\beta_{0,55}$  was important only to the 0.25 quantile. Screening was difficult for all methods in Example 3. In Example 4 the only two screening methods that could accommodate the unknown measurement error distribution were score and semiparametric screening, which performed similarly.

Table 3 compares the performance of our threshold for reproducible screening to the  $n/\log n$  rule of Fan and Lv (2008) and the auxiliary variables method of Zhu et al. (2011). To calculate our reproducible screening threshold we generated 100 bootstrap samples and searched for the optimal threshold  $j$  across  $j = \{10, 20, \dots, p_n\}$ . All methods performed well in Example 1, giving high true positive rates along with substantial dimension reduction. In Example 2 at  $\tau = 0.5$ , Wald and score screening gave the best true positive rates, but score screening had a higher true discovery rate and frequently retained fewer covariates. On the other hand, at  $\tau = 0.25$  the final model sizes after score screening were close to 2000. However, even retaining 2000 covariates still represents an 80% reduction in dimension. Screening procedures are designed to be followed by a second sparse regression step like lasso, and 2000 covariates is very manageable by these follow-up procedures. In Examples 3 and 4, score screening was able to retain very few covariates while still giving very high true positive rates.

Table 4 reports the performance of our iterative screening procedure from Section 4, which we applied to the parametric models in Examples 1 and 2 with  $R = 20$ . In those models the 30<sup>th</sup> covariate had a nonzero coefficient in the true model but was marginally unassociated with the outcome. In Example 1 iterative screening was able to capture that covariate in nearly all of the simulations. In Example 2, iterative screening was still to capture the variable after retaining only around 200–300 variables, as opposed to marginal score screening, which had to retain thousands of variables. However, the hidden covariate was only captured in very few simulations, indicating that variable screening for Example 2 is a difficult problem.

## 6 Data analysis

### 6.1 Analysis methods

We were interested in identifying genes highly associated with the 10% conditional quantile of the survival distribution of MM patients, because these genes are likely to be important in high-risk MM. Previous studies have searched for genes associated with patient survival (Shaughnessy et al., 2007; Decaux et al., 2008), but their analyses did not recognize that some genes may only affect certain quantiles of the conditional survival distribution.

We used gene expression and survival outcome data from newly diagnosed multiple myeloma patients who were recruited into clinical trials UARK 98-026 and UARK 2003-33, which studied the total therapy II (TT2) and total therapy III (TT3) treatment regimens, respectively. These data are described in Zhan et al. (2006) and Shaughnessy et al. (2007), and can be obtained through the MicroArray Quality Control Consortium II study (Shi et al., 2010), available on GEO (GSE24080). Gene expression profiling was performed using Affymetrix U133Plus2.0 microarrays, and we averaged the expression levels of probesets corresponding to the same gene, resulting in 33,326 covariates. We used the TT2 arm as a training set, giving us 340 subjects and 126 observed deaths, we validated the results on the TT3 arm.

To identify these high-risk genes we used the censored quantile regression of Honore et al. (2002), described earlier in Example 2 in Section 5.1, with the transformation function  $h =$

log. First, in the screening step we compared Wald screening with the estimator of Peng and Huang (2008), marginal score screening, the semiparametric method of Zhu et al. (2011), the nonparametric method of He et al. (2013), and iterative score screening. In the score screening procedures we estimated the nuisance intercept parameter from another MM dataset collected by Avet-Loiseau et al. (2009). For iterative score screening we set  $R = 20$ .

Second, to set a screening threshold we retained the top  $n/\log n$  covariates from Wald and nonparametric screening, used our reproducible screening threshold for score screening, and used the auxiliary variables procedure of Zhu et al. (2011) for semiparametric screening. For reproducible screening we generated 100 bootstrap samples and searched for the optimal threshold  $j$  across  $j = \{10, 20, \dots, p_n\}$ , as in the simulations.

Finally, we used the screened covariates to estimate regression models. To our knowledge there do not exist any computationally convenient procedures for censored quantile regression for arbitrary quantiles that can be computed in high-dimensions, so we used our projected subgradient method from Section 4 to serve as a regression estimator. We tuned the procedure by selecting the value of  $R$  that minimized an approximate Bayesian

Information Criterion, which we calculated as  $\|n\mathbf{U}(\hat{\beta}_R)\|_2^2 + \|\hat{\beta}_R\|_0 \log n$  with  $\mathbf{U}$  the estimating equation of Honore et al. (2002) and  $\hat{\beta}_R$  the regression estimate for a given value of  $R$ .

## 6.2 Results

Wald screening required 930 seconds, the nonparametric screening of He et al. (2013) required 240 seconds, iterative score screening required 84 seconds, the semiparametric screening of Zhu et al. (2011) required 44 seconds, and marginal score screening took only 5 seconds. Because of the computational efficiency of score screening, calculating the reproducible screening threshold required only 934 seconds, which was still just as fast as Wald screening.

Table 5 reports the genes selected in the final censored quantile regression models. Iterative and reproducible score screening behaved very similarly, giving nearly identical final models. However, they shared no genes in common with the results of the other screening methods. One possible reason is that the correlations between the selected genes were not low. For example, among the top 100 genes selected by Wald screening, 20% of the pairwise correlations were above 0.25 and the largest reached 0.73, and for score screening 20% of the correlations were at least 0.58 and reached 0.99. In other words, the different screening methods most likely selected blocks of correlated covariates together, and the same covariates could be ranked very differently by different methods if they were in different blocks. This highlights the importance of reproducibility.

To choose between the four models, we used the fitted regression models to predict the 0.1 conditional quantiles in the TT3 arm and calculated validation metrics in two ways. First, to estimate the quantile prediction error we used the loss function

$$n^{-1} \sum_i \frac{\delta_i}{\hat{S}_{h(C)}\{h(Y_i)\}} \{\tau - I(Y_i - \hat{Y}_i < 0)\} Y_i, \quad (1)$$

where  $\delta_i$  is the censoring indicator,  $Y_i$  is the observed follow-up time,  $\tau = 0.1$  is the target quantile, and  $\hat{Y}_i$  is the predicted  $\tau$  conditional quantile. A similar loss function was described by Honore et al. (2002). Second, we used the censored quantile regression approach of Peng and Huang (2008) to estimate the associations between the predicted quantiles and the true 0.1 quantile. We report the  $t$ -statistics of association. Table 5 shows that the models selected after score screening performed the best under both validation metrics, followed by semiparametric screening. In contrast, the quantiles predicted after Wald and semiparametric screening were actually negatively associated with the true quantile. This suggests that the true relationship between the genes and the quantile may be significantly nonlinear. This nonlinearity can still be detected by the score screening methods.

## 7 Discussion

Motivated by our analysis of genomic factors influencing the high risk multiple myeloma patients, we introduced a new framework for variable screening based on score tests. Score screening is widely applicable to parametric, semiparametric, and nonparametric models, relatively easy to theoretically justify, and computationally efficient. Using score test screening in our MM analysis resulted in a predictive model for the conditional 10% quantile (high risk group) which was more accurate the models obtained using other screening methods.

We introduced a method for selecting the number of covariates to retain based on the principle of reproducible screening. It would be interesting to investigate the sure screening and false positive control properties of this procedure, in the context of Theorems 1 and 2. Our score testing framework also suggested a new approach to iterative screening based on projected subgradient methods, which can be applied even to nonsmooth estimating equations. It is related to sparse regression techniques and it is possible that this connection can lead to better a theoretical understanding of iterative screening, which is still elusive.

## Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

## Acknowledgments

We are grateful to the editor, the associate editor, and the anonymous referee for their helpful comments. We also thank Professors Lee Dicker and Julian Wolfson for reading an earlier version of this manuscript. This research is partially supported by an NIH grant. We'd like to credit NIH grant R01 HL107240-01.

## References

Agarwal A, Negahban S, Wainwright MJ. Fast global convergence of gradient methods for high-dimensional statistical recovery. *The Annals of Statistics*. 2012; 40(5):2452–2482.

- Avet-Loiseau H, Li C, Magrangeas F, Gouraud W, Charbonnel C, Harousseau JL, Attal M, Marit G, Mathiot C, Facon T, et al. Prognostic significance of copy-number alterations in multiple myeloma. *Journal of Clinical Oncology*. 2009; 27(27):4585–4590. [PubMed: 19687334]
- Bühlmann, PL.; van de Geer, SA.; Van de Geer, S. *Statistics for high-dimensional data*. Springer; 2011.
- Cai T, Huang J, Tian L. Regularized estimation for the accelerated failure time model. *Biometrics*. 2009; 65:394–404. [PubMed: 18573133]
- Cheng S, Wei L, Ying Z. Analysis of transformation models with censored data. *Biometrika*. 1995; 82(4):835–845.
- Daubechies I, Fornasier M, Loris I. Accelerated projected gradient method for linear inverse problems with sparsity constraints. *Journal of Fourier Analysis and Applications*. 2008; 14(5–6):764–792.
- Decaux O, Lodé L, Magrangeas F, Charbonnel C, Gouraud W, Jézéquel P, Attal M, Harousseau JL, Moreau P, Bataille R, Campion L, Avet-Loiseau H, Minvielle S. Prediction of survival in multiple myeloma based on gene expression profiles reveals cell cycle and chromosome instability signatures in high-risk patients and hyperdiploid signatures in low-risk patients: a study of the Intergroupe Francophone du Myélome. *Journal of Clinical Oncology*. 2008; 26(29):4798–4805. [PubMed: 18591550]
- Donoho D, Jin J. Higher criticism thresholding: Optimal feature selection when useful features are rare and weak. *Proceedings of the National Academy of Sciences*. 2008; 105(39):14790–14795.
- Duchi, J.; Shalev-Shwartz, S.; Singer, Y.; Chandra, T. *Proceedings of the 25th international conference on Machine learning*. ACM; 2008. Efficient projections onto the  $\ell_1$ -ball for learning in high dimensions; p. 272-279.
- Fan J, Lv J. Sure independence screening for ultrahigh dimensional feature space. *Journal of the Royal Statistical Society, Ser B*. 2008; 70(5):849–911.
- Fan J, Song R. Sure independence screening in generalized linear models and NP-dimensionality. *The Annals of Statistics*. 2010; 38(6):3567–3604.
- Fan J, Samworth R, Wu Y. Ultrahigh dimensional feature selection: beyond the linear model. *The Journal of Machine Learning Research*. 2009; 10:2013–2038. [PubMed: 21603590]
- Fan J, Feng Y, Song R. Nonparametric independence screening in sparse ultra-high-dimensional additive models. *Journal of the American Statistical Association*. 2011; 106:544–557. [PubMed: 22279246]
- Gorst-Rasmussen A, Scheike TH. Independent screening for single-index hazard rate models with ultra-high dimensional features. *Journal of the Royal Statistical Society, Ser B*. 2013; 75:217–245.
- He Q, Lin DY. A variable selection method for genome-wide association studies. *Bioinformatics*. 2011; 27(1):1–8. [PubMed: 21036813]
- He X, Wang L, Hong HG. Quantile-adaptive model-free variable screening for high-dimensional heterogeneous data. *The Annals of Statistics*. 2013; 41(1):342–369.
- Hoeffding W. Probability inequalities for sums of bounded random variables. *Journal of the American Statistical Association*. 1963; 58:13–30.
- Honore B, Khan S, Powell JL. Quantile regression under random censoring. *Journal of Econometrics*. 2002; 109(1):67–105.
- Jin Z, Lin DY, Wei LJ, Ying Z. Rank-based inference for the accelerated failure time model. *Biometrika*. 2003; 90(2):341–353.
- Li G, Peng H, Zhang J, Zhu L. Robust rank correlation based screening. *The Annals of Statistics*. 2012a; 40(3):1846–1877.
- Li R, Zhong W, Zhu L. Feature screening via distance correlation learning. *Journal of the American Statistical Association*. 2012b; 107(499):1129–1139. [PubMed: 25249709]
- Meinshausen N, Bühlmann P. Stability selection. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*. 2010; 72(4):417–473.
- Peng L, Huang Y. Survival analysis with quantile regression models. *Journal of the American Statistical Association*. 2008; 103(482):637–649.
- Portnoy S. Censored regression quantiles. *Journal of the American Statistical Association*. 2003; 98(464):1001–1012.

- Shaughnessy J, Zhan F, Burington BE, Huang Y, Colla S, Hanamura I, Stewart JP, Kordsmeier B, Randolph C, Williams DR, et al. A validated gene expression model of high-risk multiple myeloma is defined by deregulated expression of genes mapping to chromosome 1. *Blood*. 2007; 109(6):2276–2284. [PubMed: 17105813]
- Shi L, Campbell G, Jones WD, et al. The MAQC-II Project: A comprehensive study of common practices for the development and validation of microarray-based predictive models. *Nature Biotechnology*. 2010; 28:827–838.
- Shor, NZ.; Kiwiel, KC.; Ruszcayski, A. *Minimization methods for non-differentiable functions*. Springer-Verlag New York, Inc; 1985.
- Song X, Huang Y. On corrected score approach for proportional hazards model with covariate measurement error. *Biometrics*. 2005:702–714. [PubMed: 16135021]
- Tibshirani RJ. Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society, Ser B*. 1996; 58:267–288.
- Tsiatis AA. Estimating regression parameters using linear rank tests for censored data. *The Annals of Statistics*. 1996; 18:354–372.
- Uno H, Cai T, Pencina MJ, D'Agostino RB, Wei LJ. On the C-statistics for evaluating overall adequacy of risk prediction procedures with censored survival data. *Statistics in Medicine*. 2011; 30:1105–1117. [PubMed: 21484848]
- Wang HJ, Wang L. Locally weighted censored quantile regression. *Journal of the American Statistical Association*. 2009; 104(487):1117–1128.
- Wei LJ. The accelerated failure time model: a useful alternative to the cox regression model in survival analysis. *Statistics in Medicine*. 1992; 11:14–15. 1871–1879.
- Zhan F, Huang Y, Colla S, Stewart J, Hanamura I, Gupta S, Epstein J, Yaccoby S, Sawyer J, Burington B, et al. The molecular classification of multiple myeloma. *Blood*. 2006; 108(6):2020. [PubMed: 16728703]
- Zhao SD, Li Y. Principled sure independence screening for Cox models with ultra-high-dimensional covariates. *Journal of Multivariate Analysis*. 2012; 105:397–411. [PubMed: 22408278]
- Zhu LP, Li L, Li R, Zhu LX. Model-free feature screening for ultrahigh-dimensional data. *Journal of the American Statistical Association*. 2011; 106:1464–1475. [PubMed: 22754050]



**Table 1**

Average runtime (seconds) of different screening methods.

Example	Wald	Score	Zhu et al. (2011)	He et al. (2013)
1	16533.06	1.91	8.20	
2	1206.89	2.42	7.16	123.27
3		0.21	0.88	5.41
4		2.19	5.76	

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

**Table 2**

Medians (interquartile ranges) of minimum model sizes required to retain the covariates in the second column. In Example 2,  $\beta_{0,5}$  is relevant only when  $\tau = 0.5$  and  $\beta_{0,55}$  is relevant only when  $\tau = 0.25$ . Similarly, in Example 3  $\beta_{0,5}$  is relevant only when  $\tau = 0.5$  and  $\beta_{0,25}$  is relevant only when  $\tau = 0.25$ .

Covariates	Wald	Score	Zhu et al. (2011)	He et al. (2013)
<i>Example 1</i>				
All	5699.5 (4231)	5500 (4091)	5645 (5022.25)	
<i>Example 2, <math>\tau = 0.5</math></i>				
All	6070.5 (4655.75)	6168.5 (3832.5)	5742 (4609.25)	9166 (2057)
$\beta_{0,5}$	30.5 (62.75)	36.5 (94.25)	23 (28.25)	3451 (5856.5)
$\beta_{0,55}$	2393 (4184.5)	1882.5 (5565.75)	547.5 (1028.25)	3708.5 (5094.5)
<i>Example 2, <math>\tau = 0.25</math></i>				
All	5111 (4736.5)	5094 (4104.5)	5742 (4376)	9720 (554)
$\beta_{0,5}$	1724.5 (4758.25)	1763 (5112.75)	23 (28.25)	5516.5 (5380.5)
$\beta_{0,55}$	67.5 (309)	88.5 (273.5)	547.5 (1028.25)	2381 (6252.5)
<i>Example 3, <math>\tau = 0.5</math></i>				
All	9620 (548.25)	9729 (337.75)	9277.5 (361.75)	9945 (243)
<i>Example 4</i>				
All	5495.5 (4365.5)	5584.5 (4367.5)	5483 (4237.25)	

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

**Table 3**

Performance of different methods for choosing the screening threshold. Methods: RS = reproducible screening, described in Section 3; Auxiliary = auxiliary variables method of Zhu et al. (2011). Average performance metrics (standard deviation): TP = true positive rate, TD = true discovery rate. Median size is reported (interquartile range).

Screening	Threshold	TP	TD	Size
<i>Example 1</i>				
Wald	$n/\log n$	90 (0)	13.64 (0)	66 (0)
Score	RS	86.2 (7.49)	22.84 (1.52)	40 (0)
Zhu et al. (2011)	Auxiliary	89.4 (2.39)	21.02 (1.46)	43 (4)
<i>Example 2, <math>\tau = 0.5</math></i>				
Wald	$n/\log n$	68.9 (13.25)	10.44 (2.01)	66 (0)
Score	RS	62.9 (27.35)	13.26 (13)	30 (1602.5)
Zhu et al. (2011)	Auxiliary	40.7 (17.25)	31.18 (15.09)	15 (10.25)
He et al. (2013)	$n/\log n$	3.9 (5.84)	0.59 (0.89)	66 (0)
<i>Example 2, <math>\tau = 0.25</math></i>				
Wald	$n/\log n$	62.8 (12.56)	9.52 (1.9)	66 (0)
Score	RS	72 (25.74)	7.73 (12.06)	1540 (1642.5)
Zhu et al. (2011)	Auxiliary	36.9 (15.87)	28.66 (15.87)	15 (10.25)
He et al. (2013)	$n/\log n$	5.5 (8.57)	0.83 (1.3)	66 (0)
<i>Example 3, <math>\tau = 0.5</math></i>				
Score	RS	62.38 (15.02)	29.76 (22.56)	10 (130)
Zhu et al. (2011)	Auxiliary	16.38 (17.65)	49.49 (44.36)	2 (2)
He et al. (2013)	$n/\log n$	56.25 (22.86)	6.82 (2.77)	66 (0)
<i>Example 4</i>				
Score	RS	68 (15.04)	26.26 (6.76)	30 (10)
Zhu et al. (2011)	Auxiliary	72.4 (14.36)	25.15 (4.23)	30.5 (12)

**Table 4**

Performance of iterative screening. The second column reports the average percentage of times (SD) the marginally unimportant variables (see Section 5.1) were capture by iterative screening. Average performance metrics (standard deviation): TP = true positive rate, TD = true discovery rate. Median size is reported (interquartile range).

<b>Hidden</b>	<b>TP</b>	<b>TD</b>	<b>Size</b>
<i>Example 1</i>			
91 (28.76)	95.9 (6.21)	1.66 (0.42)	652 (259.5)
<i>Example 2, <math>\tau = 0.5</math></i>			
1 (10)	79.8 (11.97)	2.71 (0.73)	282 (44.25)
<i>Example 2, <math>\tau = 0.25</math></i>			
1 (10)	71.7 (10.83)	3.29 (0.93)	210 (64.5)

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

**Table 5**

Final regression models for the 0.1 conditional quantile of MM patient survival. Validation metrics: PE = prediction error (1);  $t$ -statistic =  $t$ -statistic of the regression of the true 0.1 quantile on the predicted quantile, using Peng and Huang (2008).

Validation	Wald	Zhu et al. (2011)	He et al. (2013)	Iterative	Score
	CDK13	ADAR	ATP6	hmRNPK	hmRNPK
	MAPKAP1	ATP6V0E1	CARD8	hmRNPKP4	hmRNPK4
	PEX11B	DPY30	CTCF	MATR3	MATR3
	VCP	HNRNPU	DDX3X	OAZ1	OAZ1
		NOLC1		RAB10	RAP10
				RPS3A	RPS3A
				SPCS1	SPCS1
				SPCS2	SPCS2
				SUMO2	SUMO2
				TMBIM4	TMBIM4
				UBC	UBC
				SERP1	SERP1
PE	0.660	1.158	0.430	<b>0.079</b>	0.083
$t$ -statistic	-1.133	-0.515	1.319	1.690	<b>1.731</b>