Principled sure independence screening for Cox models with ultra-high-dimensional covariates

Sihai Dave Zhao and Yi Li

Abstract

It is rather challenging for current variable selectors to handle situations where the number of covariates under consideration is ultra-high. Consider a motivating clinical trial of bortezomib for the treatment of multiple myeloma, where overall survival and expression levels of 9072 probesets were measured for each of 112 patients with the goal of identifying genes that predict survival after treatment. This dataset defies analysis even with regularized regression. Some remedies have been proposed for the linear model and for generalized linear models, but there are few solutions in the survival setting and, to our knowledge, no theoretical support. Furthermore, existing strategies often involve tuning parameters that are difficult to interpret. In this paper we propose and theoretically justify a principled method for reducing dimensionality in the analysis of censored data by selecting only the important covariates. Our procedure involves a tuning parameter that has a simple interpretation as the desired false positive rate. We present simulation results and apply the proposed procedure to analyze the aforementioned myeloma study.

KEYWORDS: Principled sure independence screening; Multiple myeloma; variable selection; sure independence screening; Cox model; ultra-high-dimensional covariates.

RUNNING TITLE: Principled sure independence screening

1 Introduction

An urgent need has emerged in the field of biomedicine for statistical procedures capable of analyzing and interpreting vast quantities of data. Selecting the best predictors of an outcome is a key step in this process, but traditional methods of variable selection, such as best subset selection or backward selection, have been found to be unstable and inaccurate when the dimension of the covariates is close to the number of observations. Furthermore, when there are more covariates than observations, as is often the case in genomic studies, these methods can fail completely.

To address these issues, recent work has focused on regularized regression procedures such as the lasso (Tibshirani, 1996) and adaptive lasso (Zou, 2006), the elastic net (Zou and Hastie, 2005), the smoothly clipped absolute deviation estimator (Fan and Li, 2001), and the Dantzig selector (Candès and Tao, 2007). These methods can handle the high-dimension-low-sample-size paradigm, have superior predictive accuracy, and under certain conditions can achieve the oracle property (Fan and Li, 2001): they are as accurate and efficient as an estimator which knows *a priori* which variables are truly important.

However, these procedures only work well with a moderate number of covariates. When the dimension of the covariates is ultra-high, both traditional and regularization methods have problems with speed, stability, and accuracy (Fan and Lv, 2008). For example, many of the bounds on the accuracy of these methods involve factors of $\log p_n$, where p_n is the dimension of the covariates (Candès and Tao, 2007; Wainwright, 2009). Thus the theoretical performance of these methods degrades as p_n becomes very large, yet this ultra-high dimensionality characterizes many real-world biological datasets. Our work in this paper is motivated by one such dataset, in the area of multiple myeloma.

Multiple myeloma is the world's second-most common hematological cancer and patients often present with bone lesions, immunological disorders, and renal failure. An effective treatment is still being sought, as only about 10% of patients survive 10 years after diagnosis. A deeper understanding of the molecular etiology of this disease would lead to novel

therapeutic targets and more accurate risk classification systems. We studied overall survival for 112 multiple myeloma patients enrolled in a clinical trial of bortezomib (Mulligan et al., 2007). With expression level measurements on 9072 probesets, this dataset defies analysis even with regularized regression.

Without tools to deal with this type of ultra-high dimensionality, many analysts have first employed an ad-hoc initial univariate screening step to reduce the number of covariates under consideration. The remaining covariates could then be fed to one of the more sophisticated regularization techniques in a second stage. But it was only recently that Fan and Lv (2008) placed this ad-hoc practice on firm theoretical ground. They showed that screening could indeed improve the performance of regularization methods. They suggested fitting marginal regression models for each covariate, choosing a threshold, and retaining those covariates for which the magnitudes of the parameter estimates are above the threshold. When the data come from an ordinary linear model with normal errors, Fan and Lv (2008) showed that this pre-screening procedure, which they termed sure independence screening (SIS), has desirable theoretical properties. Fan and Song (2010) later gave theoretical justification for using SIS with generalized linear models.

But two important problems remain. First, one common type of outcome data seen in clinical settings, including in our myeloma dataset, are survival times, which are subject to censoring. Regularized regression methods for censored observations have been studied, as reviewed in Li (2008), but these are subject to the same issues mentioned above when the dimension of the covariates is ultra-high. There is thus a need for a pre-screening procedure in this setting, but the results of Fan and Lv (2008) and Fan and Song (2010) cannot be applied because the issue of censoring is not addressed. Several ad-hoc solutions are available from Tibshirani (2009) and Fan et al. (2010), but none of these proposals has much theoretical support. The extension of the theoretical sure screening results to censored data is not immediate because it turns out that certain conditions on the relationship between the covariates and the censoring distribution are required for screening to have good theoretical

properties, an issue which does not emerge with uncensored data.

The second problem is that existing screening procedures require choosing a threshold to dictate how many variables to retain, but there are no principled methods for making such a choice, making the resulting screened models difficult to evaluate. The threshold can be thought of as a regularization parameter, which in the regression setting is ordinarily chosen by optimizing out-of-sample prediction error using cross-validation or generalized cross-validation. However, this approach is unavailable for screening procedures because no prediction rule is ever generated.

In this paper we provide a screening method for censored survival data with ultra-high-dimensional covariates. We also propose a new, principled method for choosing the number of covariates to retain based on specifying the desired false positive rate. Finally, we give, to our knowledge, the first theoretical justifications of the sure independence screening procedure for censored data. Under the asymptotic framework where the number of covariates can grow with the sample size, we show that with probably going to 1, our procedure will select all of the important variables with a false positive rate close to the prespecified level.

Our paper is organized as follows. We briefly review sure independence screening for generalized linear models in Section 2. In Section 3 we discuss the implementation and the theoretical properties of our principled sure independence screening procedure, and present simulation results in Section 4. Section 5 describes our analysis of the myeloma dataset, and we conclude with a discussion in Section 6. All proofs are given in the Appendix.

2 Sure independence screening in generalized linear models

We first briefly review the sure independence screening formulation of Fan and Song (2010). For subjects $i = 1, \ldots, n$ let $\mathbf{Z}_i = (Z_{i1}, \ldots, Z_{ip_n})$ be the p_n-dimensional covariate vector. Assuming that observations Y_i come from an exponential family, we model $E(Y_i \mid \mathbf{Z}_i)$ as

some function of a linear predictor $\boldsymbol{\alpha}_0^T \mathbf{Z}_i$ with parameter vector $\boldsymbol{\alpha}_0 = (\alpha_{01}, \ldots, \alpha_{0p_n})$. When p_n is much larger than n we are unable to estimate $\boldsymbol{\alpha}_0$ with conventional procedures. To reduce p_n, sure independence screening proceeds by regressing Y_i on each Z_{ij} individually to calculate marginal maximum likelihood estimates $\hat{\beta}_j$. The final screened model retains all covariates $j : |\hat{\beta}_j| \geq \gamma_n$ for some prespecified constant cutoff γ_n.

Fan and Song (2010) showed that under certain conditions, if γ_n follows an ideal rate, this procedure has two desirable properties, namely the sure screening property and the size control property. The former stipulates that the screened model will contain the true model with a probability approaching 1. The latter states that if $\log(p_n) = o(n^{1-2\kappa})$ where $\kappa < 1/2$, the probability that the size of the screened model will be at most $O\{n^{2\kappa}\lambda_{max}(\boldsymbol{\Sigma})\}$ will also go to 1, where $\boldsymbol{\Sigma} = \text{var}(\mathbf{Z}_i)$ and $\lambda_{max}(\boldsymbol{\Sigma})$ is the largest eigenvalue of $\boldsymbol{\Sigma}$.

These results, however, are restricted to non-censored generalized linear models. Furthermore, it is difficult to translate the ideal rate for γ_n into a method for selecting the cutoff in practice. Fan and Lv (2008) suggest $n/log(n)$ or $n-1$ as the number of covariates to retain after screening, but without theoretical justification. To address these issues, we investigate here a reliable pre-screening procedure in a survival setting, where the outcomes are subject to right censoring, and propose a principled method for choosing γ_n based on controlling the false positive rate.

3 Principled Cox sure independence screening

3.1 Method

In the context of survival analysis, we assume that the underlying survival times T_i follow a Cox model (Cox, 1972) with the true hazard function

$$\lambda(x; \mathbf{Z}_i) = \lambda_0(x) \exp(\boldsymbol{\alpha}_0^T \mathbf{Z}_i), \tag{1}$$

where $\lambda_0(x)$ is unspecified. Let $\tilde{C}_i$ be potential censoring times, which are independent of T_i conditional on $\mathbf{Z}_i$. Furthermore let $\tau > 0$ be the finite study duration such that $\mathrm{P}\{\min(\tilde{C}_i, \tau) < T_i\} < 1$, ensuring that enough events will be observed over $[0, \tau]$. The effective censoring times are thus $C_i = \min(\tilde{C}_i, \tau)$. We observe $X_i = \min(T_i, C_i)$, and $\delta_i = I(T_i \leq C_i)$. Without loss of generality, we assume throughout that $E(Z_{ij}) = 0$ for all j.

To perform an initial screening procedure, we propose to fit marginal Cox regressions, possibly misspecified, for each Z_{ij} individually, namely $\lambda_0^*(x)\exp(\beta Z_{ij})$. Let $N_i(t) = I(X_i \leq t, \delta_i = 1)$ be independent counting processes for each subject i and $Y_i(t) = I(X_i \geq t)$ be the at-risk processes. For $k = 0, 1, \ldots$, define

$$S_j^{(k)}(x) = n^{-1}\sum_{i=1}^{n} Z_{ij}^k Y_i(x)\lambda(x; \mathbf{Z}_i), \quad s_j^{(k)}(x) = E\{S^{(k)}(x)\},$$

$$S_j^{(k)}(\beta, x) = n^{-1}\sum_{i=1}^{n} Z_{ij}^k Y_i(x)\exp(\beta Z_{ij}), \quad s_j^{(k)}(\beta, x) = E\{S^{(k)}(\beta, x)\}.$$

Then the maximum marginal partial likelihood estimator $\hat{\beta}_j$ solves the estimating equation

$$U_j(\beta) = \sum_{i=1}^{n}\int_0^\tau \left\{ Z_{ij} - \frac{S_j^{(1)}(\beta, x)}{S_j^{(0)}(\beta, x)} \right\} dN_i(x) = 0. \tag{2}$$

Finally, let β_{0j} be the solution to the limiting estimation equation

$$u_j(\beta) = \int_0^\tau \left\{ s_j^{(1)}(x) - \frac{s_j^{(1)}(\beta, x)}{s_j^{(0)}(\beta, x)} s_j^{(0)}(x) \right\} dx. \tag{3}$$

Define the information matrix to be $I_j(\beta) = -\partial U_j/\partial\beta$ at $\hat{\beta}_j$. We will denote the final screened model by $\hat{\mathcal{M}} = \{j : I_j(\hat{\beta}_j)^{1/2}|\hat{\beta}_j| \geq \gamma_n\}$. We would like a practical way of choosing γ_n such that we can achieve the sure screening property while controlling the false positive rate, or the proportion of unimportant covariates we incorrectly include in $\hat{\mathcal{M}}$. If the true model $\mathcal{M} = \{j : \alpha_{0j} \neq 0\}$ has size $|\mathcal{M}| = s_n$, then the expected false positive rate can be

written as

$$\mathrm{E}\left(\frac{\left|\hat{\mathcal{M}} \cap \mathcal{M}^c\right|}{|\mathcal{M}^c|}\right) = \frac{1}{p_n - s_n} \sum_{j \in \mathcal{M}^c} \mathrm{P}\left\{I_j(\hat{\beta}_j)^{1/2}|\hat{\beta}_j| \geq \gamma_n\right\}. \tag{4}$$

We can show that $I_j(\hat{\beta}_j)^{1/2}\hat{\beta}_j$ has an asymptotically standard normal distribution, so we see that γ_n corresponds to controlling the expected false positive rate at $2\{1 - \Phi(\gamma_n)\}$, where $\Phi(\cdot)$ is the standard normal cumulative distribution function.

However, we would like the false positive rate to decrease to 0 as p_n increases with n, though it can never exactly equal 0 or else $\gamma_n = \infty$. One sensible way to do this would be to first fix the number of false positives f that we are willing to tolerate, which would correspond to a false positive rate of $f/(p_n - s_n)$. Because s_n is unknown, we can be conservative by letting $\gamma_n = \Phi^{-1}\{1 - q_n/2)\}$ where $q_n = f/p_n$, so that the expected false positive rate is $2\{1 - \Phi(\gamma_n)\} = q_n \leq f/(p_n - s_n)$. We can show that this procedure maintains the sure screening property, and more precise arguments will be given later (Theorems 4 and 5).

We term this method a principled Cox sure independence screening procedure (abbreviated PSIS), as the cutoff γ_n is selected to control the false positive rate. Specifically, PSIS is implemented as follows:

1. Fit a marginal Cox model for each of the covariates according to equation (2) to get parameter estimates $\hat{\beta}_j$ and variance estimates $I_j(\hat{\beta}_j)^{-1}$.

2. Fix the false positive rate $q_n = f/p_n$ and let $\gamma_n = \Phi^{-1}(1 - q_n/2)$.

3. Retain covariates $j : I_j(\hat{\beta}_j)^{1/2}|\hat{\beta}_j| \geq \gamma_n$.

Our cutoff selection procedure is related to false discovery rate (FDR) methods (Benjamini and Hochberg, 1995; Benjamini and Yekutieli, 2001). In particular, the FDR is defined as $|\hat{\mathcal{M}} \cap \mathcal{M}^c|/|\hat{\mathcal{M}}|$, which is simply the product of the false positive rate in (4) and $|\mathcal{M}^c|/|\hat{\mathcal{M}}|$, which is less than $p_n/|\hat{\mathcal{M}}|$. Therefore, controlling the false positive rate at $q_n = f/p_n$ is equivalent to controlling the FDR at $f/|\hat{\mathcal{M}}|$, conditional on $|\hat{\mathcal{M}}|$. Bunea et al. (2006) have in fact shown that FDR methods can also have the sure screening property,

though only in the linear regression case.

Our screening procedure resembles the "marginal ranking" methods for censored outcome data proposed by various authors (Fan et al., 2010; Tibshirani, 2009). However, to our knowledge, none of these proposals has much theoretical support. A much more aggressive method of control has been proposed by Fan et al. (2010). We show below that our proposed procedure maintains the sure screening property, and will also control the false positive rate at close to the nominal level. Fan and Lv (2008) also proposed an iterative sure independence screening procedure (ISIS) for linear models, which they showed can perform better than SIS. However, they were unable to offer theoretical support. In this paper we focus on first understanding non-iterative screening for the Cox model.

3.2 Theoretical properties

First, under certain assumptions, we find that we can distinguish α_{0j}, $j \in \mathcal{M}$ from α_{0j}, $j \in \mathcal{M}^c$ in the presence of censoring. It is this guarantee that makes marginal screening approaches even possible.

Theorem 1 *Under regularity conditions 1 and 2 and Assumptions 1–3 in the Appendix,* $\beta_{0j} = 0$ *if and only if* $\alpha_{0j} = 0$*, for all* $j = 1, \ldots, p_n$.

Following Struthers and Kalbfleisch (1986) and under regularity conditions 1 and 2 in the Appendix, we know that the $\hat{\beta}_j$ are consistent for β_{0j}. It is therefore natural to ask how accurate these estimates are, and we can give a tail probability bound for $\hat{\beta}_j$.

Theorem 2 *Under regularity conditions 1–5 and Assumptions 1–3 in the Appendix,*

$$P\left\{\sqrt{n}|\hat{\beta}_j - \beta_{0j}| \geq 4K[1 + C\exp\{2K(A+L)\}](1+t)/H\right\} \leq \exp(-t^2/2)$$

for all $j = 1, \ldots, p_n$*, where* K *is the bound on the covariates* Z_{ij} *for all* j*,* $C = \int_0^\tau \lambda_0(x)dx < \infty$*,* A *is the bound on the parameters* α_{0j} *for all* j*,* $L = \sum_{i=1}^{p_n} |\alpha_{0j}|$ *and is bounded by regularity condition 3, and* H *is defined in regularity condition 5.*

Theorem 2 is important as it suggests that $|\hat{\beta}_j - \beta_{0j}|$ is at most on the order of $n^{-1/2}$ with high probability. Hence in order to detect covariate $j \in \mathcal{M}$, we need $|\beta_{0j}|$ to be at least $O(n^{-1/2})$, which is indeed the case as shown by the following theorem.

Theorem 3 *Under regularity conditions 1 and 2 and Assumptions 1–3 in the Appendix,* $\min_{j\in\mathcal{M}} |\beta_{0j}| \geq c_2 n^{-\kappa}$, *where* $\kappa < 1/2$ *and* c_2 *is a positive constant.*

Because the $|\beta_{0j}|$ are large enough to be detected with our marginal Cox regressions, and because they reflect the importance of the Z_{ij} in the true joint model, we can prove that our procedure maintains the sure screening property and controls the false positive rate at close to the nominal level.

Theorem 4 (Sure screening property) *Under regularity conditions 1–5 and Assumptions 1–3 in the Appendix, if we choose* $\gamma_n = \Phi^{-1}(1 - q_n/2)$, *then for* $\kappa < 1/2$ *and* $\log(p_n) = O(n^{1/2-\kappa})$, *there exists a constant* $c_3 > 0$ *such that*

$$P(\mathcal{M} \subseteq \hat{\mathcal{M}}) \geq 1 - s_n \exp(-c_3 n^{1-2\kappa}).$$

Theorem 5 (False positive control property) *Under regularity conditions 1–5 and Assumptions 1–3 in the Appendix, if we choose* $\gamma_n = \Phi^{-1}(1 - q_n/2)$, *then there exists some* $c_4 > 0$ *such that*

$$E\left(\frac{|\hat{\mathcal{M}} \cap \mathcal{M}^c|}{|\mathcal{M}^c|}\right) \leq q_n + c_4 n^{-1/2},$$

where $|\hat{\mathcal{M}} \cap \mathcal{M}^c|/|\mathcal{M}^c|$ *can be interpreted as the false positive rate.*

It is often assumed that the true model is sparse and s_n is small (Candès and Tao, 2007), in which case Theorem 4 indicates that we will be able to retain all important covariates with high probability. The probability bound will converge to 1 if $\log(p_n) = O(n^{1/2-\kappa})$, which is comparable to the rates allowed in Fan and Lv (2008) and Fan and Song (2010). That p_n is allowed to incease exponentially justifies the use of sure independence screening in the Cox model when p_n is ultra-high-dimensional.

4 Simulations

To evaluate the finite-sample performance of our sure screening and false positive control properties, we performed PSIS on simulated datasets generated from Cox models and examined its average false positive and negative rates. We simulated 200 datasets, each consisting of $p_n = 1000$ covariates and $n = 100$ subjects. We generated the covariates from a multivariate normal distribution where the mean was 0 and the correlation between components Z_{ij} and Z_{ik} was $\rho^{|j-k|}$ for $\rho = 0.5$, and 0.9. Survival times were generated from Cox models with baseline hazards of $\lambda_0(x) = 1$ and linear predictors $\boldsymbol{\alpha}_0^T \mathbf{Z}_i$ for different parameter vectors $\boldsymbol{\alpha}_0$. We let the number of non-zero elements of $\boldsymbol{\alpha}_0$ be either $s_n = 5$ or 15 and set the first s_n components of $\boldsymbol{\alpha}_0$ to be either all equal to 0.35 or all equal to 0.7. Censoring times were generated uniformly between 0 and c, where c was taken such that the rate of censoring was approximately 50%.

To explore how a few popular regularized regression techniques were affected by PSIS with different values of q_n, we followed PSIS by either lasso (Tibshirani, 1997), adaptive lasso (Zhang and Lu, 2007), or SCAD (Fan and Li, 2002). Since the initial parameter estimates required by adaptive lasso do not exist when $p_n > n$, we first apply ordinary lasso to reduce p_n and then apply adaptive lasso using the remaining covariates. To select the tuning parameters for these methods, we used the default procedures provided in the corresponding R packages. For lasso and lasso-adaptive lasso, we used the 5-fold cross-validated partial likelihood (Verweij and van Houwelingern, 1993) using the package `penalized` (Goeman, 2010). For SCAD we implemented the one-step estimator of Zou and Li (2008) with AIC using the package `SIS`. We denote these two-stage procedures by PSIS-L, PSIS-L-A, and PSIS-S, respectively.

Numerical results for our sure screening and false positive control properties, when $\alpha_{0j} = 0.35$, are reported in Table 1. The results when $\alpha_{0j} = 0.7$ are similar and are omitted for the sake of space. We consider $q_n = 10^r$ for $r = -6, \ldots, -2$, and also ranging from 0.1 to 1 (corresponding to no screening) in increments of 0.1. The results support our principled

cutoff procedure: when $\rho = 0.5$, the observed false positive rates closely match the nominal q_n, and when $\rho = 0.9$ the rates match for $q_n \geq 10^{-3}$. For $q_n < 10^{-3}$ the observed rates are higher than the nominal rates, but since $p_n = 1000$ here, $q_n = 10^{-3}$ already corresponds to allowing only one false positive.

Figure 1 plots the average false negative rates for PSIS against q_n, which increase as q_n decreases but don't rise dramatically until $q_n < 0.01$. For a given q_n the false negative rates decrease with larger α_{0j}. The performance of PSIS is not noticeably affected by the amount of correlation. These results suggest that the tolerable false positive rate can be set fairly low and the false negative rate will not suffer much.

The average false negative rates for PSIS-L, PSIS-L-A, and PSIS-S are also plotted in Figure 1. The corresponding false positive rates are all very low and so are not plotted (see Table 1). When $\alpha_{0j} = 0.7$, the false negative rates for most values of q_n are below 20%, though they can rise to 40% when $\rho = 0.9$. When $\alpha_{0j} = 0.35$, however, the false negative rates can go as high as 60% even when q_n is not small. This is most likely due to a low signal-to-noise ratio, which the regularized regressions find difficult to detect. However, we can see from the plots and from Table 1 that when q_n is small and the covariates are not too highly correlated, the PSIS-L, PSIS-L-A, and PSIS-S false negative rates are actually lower than the false negative rates after running lasso, lasso-adaptive lasso, or SCAD alone (i.e. $q_n = 1$). This supports the use of PSIS prior to running regularized regression.

To assess the robustness of our procedure, we also generated 200 datasets from log-normal models. Each dataset had $n = 100$ and $p_n = 1000$, and covariates $\mathbf{Z}_i$ were generated using the same procedure as above, for $\rho = 0.5$ or 0.9. Survival times T_i were generated according to $\log(T_i) = \boldsymbol{\alpha}_0^T \mathbf{Z}_i + \epsilon_i$, where ϵ_i followed a standard normal distribution and $\boldsymbol{\alpha}_0$ had $s_n = 5$ or 15 nonzero elements all equal to either 0.35 or 0.7. Censoring times were generated as before.

Figure 2 and Table 2 show that PSIS can still perform very well when the Cox model is misspecified. Again, the numerical results when $\alpha_{0j} = 0.7$ are omitted from Table 2 to save

space. These results follow the same trends as those of the correctly specified simulations discussed above. In particular, the principled cutoff procedure still shows good performance, and using PSIS when q_n is small can still lead to lower false negative rates than when $q_n = 1$.

5 Analysis of the myeloma study

Recent advances in understanding the biological mechanisms underlying multiple myeloma have offered new possibilities for therapy (Hideshima et al., 2007). Time-to-event outcomes offer information about the progression of the disease, and in this vein several studies have examined the relationship between gene expression levels and survival (Decaux et al., 2008). In one such study conducted by Millennium Pharmaceuticals (Mulligan et al., 2007), mRNA expression levels were collected using Affymetrix U133A/B arrays from myeloma cells of 80 patients enrolled in a clinical trial of bortezomib (accession number GSE9782, trial 039). Median survival time was 684 days after randomization, and 50% of the observations were censored. We apply our methods to this data.

Expression values were measured for 44760 probesets, encompassing more than 22000 genes, and were $\log_2$-transformed. Following Mulligan et al. (2007), we retained only the 9200 probesets with strongest between-sample variance relative to their in-sample replicate variance, and after removing duplicate probesets we were left with 9072 covariates. We perform PSIS and choose $q_n = 0.0001$, for two reasons. First, our simulation results suggest that for large ρ, the true false positive rate could be significantly larger than our nominal level, and genetic datasets are probably highly correlated. Second, gene expression levels are very likely related to the survival outcomes, but only weakly so. Many of our genes are probably not sufficiently important, in the sense of Assumption 2, so allowing even a small false positive rate would result in including a huge number of genes. For these reasons, we want to control the false positive rate to the extent possible, but on the other hand we cannot allow $f = 0$ or else $\gamma_n = \infty$. Thus we considered $f = 1$, which leads to our choice of

$q_n = 0.0001$.

We cannot directly evaluate the performance of PSIS because we do not know which genes are "truly" important. Instead, we run PSIS to get a screened model $\hat{\mathcal{M}}$ and also randomly select $|\hat{\mathcal{M}}|$ probesets. We then compare the prediction accuracies of PSIS-L, PSIS-L-A, and PSIS-S to those obtained by fitting lasso, lasso-adaptive lasso, and SCAD on the randomly selected probesets. Using random genes as negative controls is common in these type of experiments (Hofmann et al., 2002; Aerts et al., 2006; Fan et al., 2010). We also fit the regularized regression procedures on the full dataset, without any screening (i.e. $q_n = 1$).

For each of these methods, we randomly set aside 60 patients in a training set and 20 patients in a testing set. We then use the models fit in the training set to calculate scores for each subject in the testing set, and evaluate the predictive performance using the C-statistic (Uno et al., 2009). We repeat this entire process 200 times. Better performances from the screened methods would provide evidence that PSIS is indeed finding predictively important genes.

Table 3 reports the average C-statistics and model sizes obtained by our different methods. We see that PSIS-L, PSIS-L-A, and PSIS-S perform significantly better than the corresponding regressions fit using randomly selected probesets. When we do not screen the data, SCAD fails, and lasso and lasso-adaptive lasso do not perform as well as the screened versions. Furthermore, screened methods have by far the least variable C-statistics, suggesting that they are more stable estimators.

We next apply PSIS-L-A with $q_n = 0.0001$ to all 80 patients. While Table 3 shows that PSIS-L exhibits higher predictive power, we choose to use PSIS-L-A because of the selection consistency of the adaptive lasso (Zou, 2006; Zhang and Lu, 2007). Table 4 gives the probesets we found to have nonzero parameter estimates, as well as their estimated coefficients. These results indicate that PSIS is an effective way to identify predictively important genes while controlling the false positive rate, and that implementing PSIS before regularized regression can lead to more computationally amenable, interpretable models with

high predictive power.

6 Discussion

This paper advances the field in three distinct ways. First, we have demonstrated that with censored outcomes, sure independence screening using marginal Cox regressions is a theoretically justified, effective way to reduce ultra-high-dimensional data to moderate sizes before applying more sophisticated variable selection procedures. In particular, we have described new, necessary condition on the dependence between the covariates and the censoring distribution. Second, we have provided a simple, principled method to select the number of variables to retain after screening and illustrated its effectiveness with simulated data. Our procedure could be easily extended to other screening methods. Finally, we have demonstrated through the motivating myeloma example that pre-screening may improve risk classification and identify predictive genes. There are a number of ways to broaden the scope of our method. So far we have dealt only with covariates that are constant in time, and we have not considered tied observations. Our method could also be extended to multivariate survival, competing risks, and other extensions of the Cox model.

While our simulations suggest that PSIS performs well even with correlated covariates, it would be interesting to explore other screening methods proposed specifically to deal with this situation. One approach is the ISIS method of Fan and Lv (2008), which starts with an initial model of potentially important covariates, regresses the residuals from the working model on each of the remaining covariates to expand the working model, and iterates this process in order to capture any important covariates that would be missed in univariate screening. Residuals are unavailable with censored observations, but Fan et al. (2010) generalized this iterative idea by working instead with log-likelihood ratios. Their formulation is easily applied to the log-partial likelihood of the Cox model, which they have implemented in the R package `SIS`. However, the theoretical properties of this procedure have not been

investigated.

Finally, our theoretical analysis of sure independence screening touches on some philosophical questions about notions of variable importance. Biological phenomena often arise from the complex interactions of genes and other factors whose individual effects can be fairly weak but still non-zero. Thus merely having a non-zero contribution to the model is not a useful notion of importance, because then nearly every variable would be important. It may be more useful to conceive of importance as a finite sample property, in the sense that covariates whose signals are higher than the noise level of the estimator being used are to be considered important. In our method, for instance, the so-called important covariates satisfy Theorem 3, or else they could not be detected by marginal Cox regressions. Perhaps a good variable selection technique is one that, instead of selecting every variable with a non-zero contribution to the outcome, retains only those variables that, for a given n, meet the finite-sample definition of importance as defined in Theorem 3. The sure screening property of our method indicates that as n increases, we get closer to achieving this goal.

7 Acknowledgements

We thank Professor Jianqing Fan for reading an earlier version of this article and for many helpful suggestions that substantially improved the manuscript.

References

Aerts, S., Lambrechts, D., Maity, S., Van Loo, P., Coessens, B., De Smet, F., Tranchevent, L.-C., De Moor, B., Marynen, P., Hassan, B., Carmeliet, P., and Moreau, Y. (2006). Gene prioritization through genomic data fusion. *Nature Biotechnology* **24,** 537–544.

Benjamini, Y. and Hochberg, Y. (1995). Controlling the False Discovery Rate: a Practical

and Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society, Ser. B* **57,** 289–300.

Benjamini, Y. and Yekutieli, D. (2001). The Control of the False Discovery Rate in Multiple Testing Under Dependency. *The Annals of Statistics* **29,** 1165–1188.

Bunea, F., Wegkamp, M., and Auguste, A. (2006). Consistent Variable Selection in High Dimensional Regression via Multiple Testing. *Journal of Statistical Planning and Inference* **136,** 4349–4364.

Candès, E. and Tao, T. (2007). The Dantzig Selector: Statistical Estimation when p is Much Larger Than n. *The Annals of Statistics* **35,** 2313–2351.

Cox, D. R. (1972). Regression Models and Life Tables (with Discussion). *Journal of the Royal Statistical Society, Ser. B* **34,** 187–220.

Decaux, O., Lodé, L., Magrangeas, F., Charbonnel, C., Gouraud, W., Jézéquel, P., Attal, M., Harousseau, J. L., Moreau, P., Bataille, R., Campion, L., Avet-Loiseau, H., and Minvielle, S. (2008). 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* **26,** 4798–4805.

Fan, J., Feng, Y., and Wu, Y. (2010). Ultrahigh Dimensional Variable Selection for Cox's Proportional Hazards Model. *IMS Collections* page in press.

Fan, J. and Li, R. (2001). Variable Selection via Noncave Penalized Likelihood and Its Oracle Properties. *Journal of the American Statistical Association* **96,** 1348–1360.

Fan, J. and Li, R. (2002). Variable selection for cox's proportional hazards model and frailty model. *The Annals of Statistics* **30,** 74–99.

Fan, J. and Lv, J. (2008). Sure Independence Screening for Ultrahigh Dimensional Feature Space. *Journal of the Royal Statistical Society, Ser. B* **70,** 849–911.

Fan, J. and Song, R. (2010). Sure Independence Screening in Generalized Linear Models and NP-Dimensionality. *The Annals of Statistics* page to appear.

Fleming, T. R. and Harrington, D. P. (2005). *Counting Processes and Survival Analysis*. Wiley, Hoboken.

Goeman, J. J. (2010). L1 Penalized Estimation in the Cox Proportional Hazards Model. *Biometrical Journal* **52,** 70–84.

Hideshima, T., Mitsiades, C., Tonon, G., Richardson, P., and Anderson, K. C. (2007). Understanding Multiple Myeloma Pathogenesis in the Bone Marrow to Identify New Therapeutic Targets. *Nature Reviews Cancer* **7,** 585–598.

Hofmann, W.-K., de Vos, S., Komor, M., Hoelzer, D., Wachsman, W., and Koeffler, H. P. (2002). Characterization of gene expression of $CD34^+$ cells from normal and myelodysplastic bone marrow. *Blood* **100,** 3553–3560.

Li, H. (2008). Censored data regression in high-dimensional and low-sample-size settings for genomic applications. In Biswas, A., Datta, S., Fine, J., and Segal, M., editors, *Statistical Advances in Biomedical Sciences: State of the Art and Future Directions*, pages 384–403. Wiley, Hoboken.

Lin, D. Y. and Wei, L. J. (1989). The robust inference for the cox proportional hazards model. *Journal of the American Statistical Association* **84,** 1074–1078.

Massart, P. (2000). About the constants in talagrand's concentration inequalities for empirical processes. *The Annals of Statistics* **28,** 863–884.

Mulligan, G., Mitsiades, C., Bryant, B., Zhan, F., Chng, W. J., Roels, S., Koenig, E., Fergus, A., Huang, Y., Richardson, P., Trepicchio, W. L., Broyl, A., Sonneveld, P., Shaughnessy,

J. D., Bergsagel, P. L., Schenkein, D., Esseltine, D. L., and Boral, A. (2007). Gene expression profiling and correlation with outcome in clinical trials of the proteasome inhibitor bortezomib. *Blood* **109,** 3177–3188.

Struthers, C. A. and Kalbfleisch, J. D. (1986). Misspecified proportional hazard models. *Biometrika* **73,** 363–369.

Tibshirani, R. J. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society, Ser. B* **58,** 267–288.

Tibshirani, R. J. (1997). The lasso method for variable selection in the cox model. *Statistics in Medicine* **16,** 385–395.

Tibshirani, R. J. (2009). Univariate shrinkage in the cox model for high dimensional data. *Statistical Applications in Genetics and Molecular Biology* **8,** 21.

Uno, H., Cai, T., Pencina, M. J., D'Agostino, R. B., and Wei, L. J. (2009). On the C-statistics for evaluating overall adequacy of risk prediction procedures with censored survival data. *Harvard University Biostatistics Working Paper Series* page Working Paper 101.

van der Vaart, A. W. and Wellner, J. A. (1996). *Weak Convergence and Empirical Processes*. Springer, New York.

Verweij, P. J. M. and van Houwelingern, H. C. (1993). Cross-validation in survival analysis. *Statistics in Medicine* **12,** 2305–2314.

Wainwright, M. J. (2009). Sharp thresholds for high-dimensional and noisy recovery using ℓ_1-constrained quadratic programming (lasso). *IEEE Transactions on Information Theory* **55,** 2183–2202.

Wasserman, L. and Roeder, K. (2009). High dimensional variable selection. *The Annals of Statistics* **37,** 2178–2201.

Zhang, H. H. and Lu, W. (2007). Adaptive lasso for cox's proportional hazards model. *Biometrika* **94,** 691–703.

Zou, H. (2006). The adaptive lasso and its oracle properties. *Journal of the American Statistical Association* **101,** 1418–1429.

Zou, H. and Hastie, T. (2005). Regression shrinkage and selection via the elastic net with application to microarrays. *Journal of the Royal Statistical Society, Ser. B* **67,** 301–320.

Zou, H. and Li, R. (2008). One-step sparse estimates in nonconcave penalized likelihood models (with discussion). *The Annals of Statistics* **36,** 1509–1533.

A Appendix

Let the true hazard function $\lambda(x; \mathbf{Z}_i)$ be given by (1), and denote the true survival functions of T_i and C_i as $S_T(x; \mathbf{Z}_i) = \exp\{-\exp(\alpha_0^T \mathbf{Z}_i)\Lambda_0(x)\}$ and $S_C(x; \mathbf{Z}_i) = \mathrm{P}(C_i > x|\mathbf{Z}_i)$, where $\Lambda_0(x) = \int_0^x \lambda_0(s)ds$. To conserve space we will write these as S_T and S_C. For simplicity we will drop the subject-specific subscripts i, except in the proof of Theorem 2. We will also need the following regularity conditions. We use notation introduced in Section 3.1.

Condition 1 *There exists a neighborhood B of β_{0j} such that for each $t < \infty$,*

$$\sup_{x\in[0,t],\beta\in B} |S_j^{(0)}(\beta, x) - s_j^{(0)}(\beta, x)| \to 0$$

in probability as $n \to \infty$, $s_j^{(0)}(\beta, x)$ is bounded away from zero on $B \times [0, t]$, and $s_j^{(0)}(\beta, x)$ and $s_j^{(1)}(\beta, x)$ are bounded on $B \times [0, t]$.

Condition 2 *For each $t < \infty$ and $j = 1, \ldots, p_n$, $\int_0^t s_j^{(2)}(x)dx < \infty$.*

Condition 3 *The true parameter vector $\boldsymbol{\alpha}_0$ belongs to a compact set such that each component α_{0j} is bounded by a positive constant A. Furthermore, $\|\boldsymbol{\alpha}_0\|_1 = O(1)$.*

Condition 4 *With τ (the study duration) such that $P\{\min(\tilde{C}_i, \tau) < T_i\} < 1$, assume $C = \int_0^\tau \lambda_0(x)dx < \infty$.*

Condition 5 *There is some constant $H > 0$ such that $n^{-1}|U_j(\hat{\beta}_j) - U_j(\beta_{0j})| \geq H|\hat{\beta}_j - \beta_{0j}|$ for all $j = 1, \ldots, p_n$.*

Regularity conditions 1 and 2 are standard in survival analysis. Condition 3 controls the total effect size of the covariates, which intuitively should be bounded and independent of sample size. Condition 4 usually holds with hazard functions commonly encountered in practice. Finally, condition 5 is reasonable because by the mean value theorem, we know that $n^{-1}|U_j(\hat{\beta}_j) - U_j(\beta_{0j})| = |n^{-1}I_j(\beta^*)||\hat{\beta}_j - \beta_{0j}|$ for some β^* between $\hat{\beta}_j$ and β_{0j}. It can be shown that $I_j(\beta^*)$ converges to the absolute value of the limiting information $-\partial u_j(\beta)/\partial\beta$ evaluated at the true β_{0j} (Fleming and Harrington, 2005), and it is reasonable to assume that this limiting information is bounded from below away from zero. Thus for n sufficiently large, we can take $H = \inf_{\beta,j} |\partial u_j(\beta)/\partial\beta|$ such that $H \neq 0$.

A.1 Assumptions

Our PSIS method will have good theoretical properties if the covariates $\mathbf{Z}_i$ satistify the following reasonable assumptions, in addition to the technical conditions above. Versions of these assumptions have been previously proposed (Fan and Lv, 2008; Fan and Song, 2010), but modifications are required when working with censored data.

Assumption 1 *The Z_{ij} are independent of time and bounded by a constant $K > 0$.*

Assumption 2 *If $F_T(x; \mathbf{Z}_i)$ is the cumulative distribution function of T_i given $\mathbf{Z}_i$, then for constants $c_1 > 0$ and $\kappa < 1/2$, $\min_{j\in\mathcal{M}} |cov[Z_{ij}, E\{F_T(C_i; \mathbf{Z}_i) \mid \mathbf{Z}_i\}]| \geq c_1 n^{-\kappa}$.*

Assumption 3 *The Z_{ij}, $j \in \mathcal{M}^c$ are independent of the Z_{ij}, $j \in \mathcal{M}$ and of C_i.*

The validity of our proposed screening procedure hinges on whether the misspecified marginal Cox regressions can reflect the importance of the corresponding covariates in the

joint model. In general it is difficult to directly link the true α_{0j} to the marginal β_{0j} because of the phenomenon of unfaithfulness (Wasserman and Roeder, 2009), where the marginal correlation of Z_{ij} with the outcome can be zero even if α_{0j} is large, due to correlated covariates. Assumption 2 protects against unfaithfulness. Though the outcome is unobservable under censoring, $F_T(C_i; \mathbf{Z}_i)$ is the probability of observing a failure given $\mathbf{Z}_i$ and is a sensible surrogate. Assumption 3 similar to the partial orthogonality condition introduced in Fan and Song (2010).

A.2 Proof of Theorem 1

We first relate β_{0j} to $\text{cov}[Z_j, \text{E}\{F_T(C; Z) \mid \mathbf{Z}\}]$. Assumptions 3 and 2 will then relate the covariance to α_{0j}.

Since β_{0j} is the solution to the estimating equation $u_j(\beta)$ (3), we can write

$$\begin{aligned}
&\int_0^\tau \text{E}\{Z_j \lambda_0(x) \exp(\boldsymbol{\alpha}_0^T \mathbf{Z}) S_T S_C\} dx \\
&= \int_0^\tau \frac{\text{E}\{Z_j \exp(\beta_{0j} Z_j) S_T S_C\}}{\text{E}\{\exp(\beta_{0j} Z_j) S_T S_C\}} \text{E}\{\lambda_0(x) \exp(\boldsymbol{\alpha}_0^T \mathbf{Z}) S_T S_C\} dx.
\end{aligned}$$

Integrating by parts, we can express the left-hand side as $\text{cov}[Z_j, \text{E}\{F_T(C; \mathbf{Z}) \mid \mathbf{Z}\}]$.

Now suppose $\alpha_{0j} = 0$. By Assumption 3, Z_j is independent of $\text{E}\{F_T(C; \mathbf{Z}) \mid \mathbf{Z}\}$ and $\text{cov}[Z_j, \text{E}\{F_T(C; \mathbf{Z}) \mid \mathbf{Z}\}] = 0$. Furthermore if $\beta_{0j} = 0$, then $\text{E}\{Z_j \exp(\beta_{0j} Z_j) S_T S_C\} = \text{E}(Z_j)\text{E}(S_T S_C) = 0$ because Z_j and C are independent for $j \in \mathcal{M}^c$, making the right-hand side zero. Using the Cauchy-Schwarz inequality, we can show that the right-hand side is a monotone function of β_{0j}, so that $\beta_{0j} = 0$ is a unique root. Similarly, suppose that $\alpha_{0j} \neq 0$ so that $j \in \mathcal{M}$. Then by Assumption 2, $|\text{cov}[Z_j, \text{E}\{F_T(C; \mathbf{Z}) \mid \mathbf{Z}\}]| > c_1 n^{-\kappa}$, making the right-hand side nonzero. Therefore $\beta_{0j} \neq 0$ by monotonicity, and we can conclude that $\alpha_{0j} = 0$ if and only if $\beta_{0j} = 0$.

A.3 Proof of Theorem 2

We first bound $|U_j(\hat{\beta}_j) - U_j(\beta_{0j})|$ by the supremum of an empirical process, where $U_j(\beta)$ was defined in (2). We then use the concentration theorem of Massart (2000) to derive a maximal inequality. We will conclude by using regularity condition 5 to extend this inequality to $|\hat{\beta}_j - \beta_{0j}|$.

First, let $\bar{U}_j(\beta) = n^{-1}U_j(\beta)$. Since we still have $\bar{U}_j(\hat{\beta}_j) = 0$, we can write $|\bar{U}_j(\hat{\beta}_j) - \bar{U}_j(\beta_{0j})| = |\bar{U}_j(\beta_{0j})|$. Because $\bar{U}_j(\beta_{0j})$ is not a sum of independent terms, we cannot directly apply empirical process techniques. However, we know from Lin and Wei (1989) that $\bar{U}_j(\beta_{0j}) = n^{-1}\sum_{i=1}^n w_i^{(j)}(\beta_{0j}) + o_p(1)$, where

$$w_i^{(j)}(\beta_{0j}) = \int_0^\tau \left\{ Z_{ij} - \frac{\mathrm{E}\{Z_{ij}\exp(\beta_{0j}Z_{ij})S_T S_C\}}{\mathrm{E}\{\exp(\beta_{0j}Z_{ij})S_T S_C\}} \right\} dN_i(x) - \int_0^\tau \frac{Y_i(x)\exp(\beta_{0j}Z_{ij})}{\mathrm{E}\{\exp(\beta_{0j}Z_{ij})S_T S_C\}} \left\{ Z_{ij} - \frac{\mathrm{E}\{Z_{ij}\exp(\beta_{0j}Z_{ij})S_T S_C\}}{\mathrm{E}\{\exp(\beta_{0j}Z_i)S_T S_C\}} \right\} E\{dN_i(x)\}.$$

and the $w_i^{(j)}(\beta_{0j})$ are independent. Furthermore, it is easy to show that $\mathrm{E}\{w_i^{(j)}(\beta_{0j})\} = 0$. If we let E_n denote the empirical measure, then we can write $|\bar{U}_j(\hat{\beta}_j) - \bar{U}_j(\beta_{0j})| \leq \sup_\beta |(\mathrm{E}_n - \mathrm{E})w_i^{(j)}(\beta)| + o_p(1)$. Thus $|\bar{U}_j(\hat{\beta}_j) - \bar{U}_j(\beta_{0j})|$ is bounded by the sum of the supremum of an empirical process and a term that converges to zero in probability.

To derive a maximal inequality for this process, we first find a bound on $w_i^{(j)}(\beta)$ uniform over β and $j = 1, \ldots, p_n$. Recall that we assumed in Section 3.2 that the covariates are bounded by a constant $K > 0$, and in regularity condition 3 that the α_{0j} are bounded by $A > 0$ and $\|\boldsymbol{\alpha}_0\|_1 = O(1)$. Then $\mathrm{E}\{dN_i(x)\} = \mathrm{E}\{\lambda_0(x)\exp(\boldsymbol{\alpha}_0^T\mathbf{Z})S_T S_C\}$ and $\exp(K\sum_{j=1}^{p_n}|\alpha_{0j}|) = \exp(KL)$ for some $L > 0$. Thus we can $|w_i^{(j)}(\beta)| \leq 2K[1 + C\exp\{2K(A+L)\}]$ for $j = 1, \ldots, p_n$, where C is defined in regularity condition 4.

Next, we must find a bound on the expected value of our supremum. Let ε_i, $i = 1, \ldots, n$ be an independent, identically distributed sequence of random variables taking values ± 1 with probability 1/2. In particular, they are independent of $\mathbf{Z}$. Then $\mathrm{E}\{\sup_\beta |(\mathrm{E}_n - \mathrm{E})w_i^{(j)}(\beta)|\} \leq 2\mathrm{E}[\sup_\beta |\mathrm{E}_n\{\varepsilon_i w_i^{(j)}(\beta)\}|]$, by Lemma 2.3.1 of van der Vaart and Wellner

(1996). But by the Cauchy-Schwarz inequality, independence of ε_i and Z_i, and the bound on $|w_i^{(j)}(\beta)|$ derived above, we can show that the right side is bounded by $4K[1+C\exp\{2K(A+L)\}]\{\text{var}(n^{-1}\sum_{i=1}^n \varepsilon_i)\}^{1/2}$. Then from the concentration theorem of Massart (2000), we know that

$$\text{P}\left[\sup_\beta |(\text{E}_n - \text{E})w_i^{(j)}(\beta)| \geq n^{-1/2}4K[1+C\exp\{2K(A+L)\}](1+t)\right] \leq \exp(-t^2/2). \quad (5)$$

Finally, we can relate this inequality back to $|\hat{\beta}_j - \beta_{0j}|$ with regularity condition 5, though we must also deal with the $o_p(1)$ term. Using a previously proven inequality, $|\hat{\beta}_j - \beta_{0j}| \leq H^{-1}\sup_\beta |(\text{E}_n - \text{E})w_i^{(j)}(\beta)| + o_p(1)$, so we can write

$$\begin{aligned}&\text{P}\left[\sqrt{n}|\hat{\beta}_j - \beta_{0j}| \geq 4K[1+C\exp\{2K(A+L)\}](1+t)/H\right] \\ &\leq \text{P}\left[\sup_\beta |(\text{E}_n - \text{E})w_i^{(j)}(\beta)| + o_p(1) \geq n^{-1/2}4K[1+C\exp\{2K(A+L)\}](1+t)\right]. \quad (6)\end{aligned}$$

But for any $\epsilon > 0$, $\text{P}(A+B \geq c) \leq \text{P}(A \geq c-\epsilon) + \text{P}(B \geq \epsilon)$, where A and B are random variables and c is a constant. We conclude by combining this with (5) and (6) and taking ϵ arbitrarily close to 0.

A.4 Proof of Theorem 3

From Theorem 1, we know that $\beta_{0j} \neq 0$ if $j \in \mathcal{M}$. Then by Theorem 2.1 of Struthers and Kalbfleisch (1986) and the mean value theorem, we know that $|u_j(0)| = |u_j(\beta_{0j}) - u_j(0)| = |u_j'(\beta^*)||\beta_{0j}|$ for some β^* between β_{0j} and 0, where $u_j'(\beta) = du_j(\beta)/d\beta$. We will first bound $u_j'(\beta)$ and then use Assumption 2 to conclude.

Integrating by parts, we can show that $|u_j'(\beta)| \leq 2K^2|\text{E}\{S_T(C;\mathbf{Z}) \mid \mathbf{Z}\}|$. But $\text{E}\{S_T(C;\mathbf{Z}) \mid \mathbf{Z}\}$ is bounded by 1, so

$$|\beta_{0j}| \geq 0.5K^{-2}\left|\text{cov}[Z_j, \text{E}\{F_T(C;\mathbf{Z}) \mid \mathbf{Z}\}] - \int_0^\tau \frac{\text{E}(Z_j S_T S_C)}{\text{E}(S_T S_C)}\text{E}\{\lambda_0(x)\exp(\boldsymbol{\alpha}_0^T\mathbf{Z})S_T S_C\}dx\right|.$$

Because $S_T S_C$ is the probability of being at risk at time x, we can intuitively see, and also prove, that $\mathrm{E}(Z_j S_T S_C) = \mathrm{cov}(Z_j, S_T S_C)$ and $\mathrm{cov}[Z_j, \mathrm{E}\{F_T(C; \mathbf{Z}) \mid \mathbf{Z}\}]$ have opposite signs. This implied that $j \in \mathcal{M}$, $|\beta_{0j}| \geq 0.5K^{-2}|\mathrm{cov}[Z_j, \mathrm{E}\{F_T(C; \mathbf{Z}) \mid \mathbf{Z}\}]|$, and Assumption 2 gives $\min_{j \in \mathcal{M}} |\beta_{0j}| \geq c_2 n^{-\kappa}$ for $c_2 = 0.5K^{-2}c_1$.

A.5 Proof of Theorem 4

We first derive a probability bound for the standardized marginal regression parameters. We can then use this bound to find $\mathrm{P}(\mathcal{M} \subseteq \hat{\mathcal{M}})$.

Let $1 + t = c_2 H n^{1/2-\kappa}/(8K[1 + C\exp\{2K(A + L)\}])$, with c_2 and κ as defined in Theorem 3. Then by Theorem 2 there exists a constant c_3 such that $\mathrm{P}(|\hat{\beta}_j - \beta_{0j}| \geq c_2 n^{-\kappa}/2) \leq \exp(-c_3 n^{1-2\kappa})$.

If we now set our cutoff $\gamma_n = \Phi^{-1}(1 - q_n/2)$, then we can write the probability of retaining the important covariates as $1 - \mathrm{P}\{\min_{j \in \mathcal{M}} I_j(\hat{\beta}_j)^{1/2}|\hat{\beta}_j| < \gamma_n\} \geq 1 - \mathrm{P}\{\min_{j \in \mathcal{M}} |\hat{\beta}_j| \leq \gamma_n (Hn)^{-1/2}\}$. Using Theorem 3 we can show that $c_2 n^{-\kappa} - |\hat{\beta}_j| \leq |\beta_{0j} - \hat{\beta}_j|$, $j \in \mathcal{M}$, so $\mathrm{P}\{\min_{j \in \mathcal{M}} |\hat{\beta}_j| \leq \gamma_n (Hn)^{-1/2}\} \leq \mathrm{P}\{\max_{j \in \mathcal{M}} |\hat{\beta}_j - \beta_{0j}| \geq c_2 n^{-\kappa} - \gamma_n (Hn)^{-1/2}\}$. If we have $\gamma_n \leq c_2 H^{1/2} n^{1/2-\kappa}/2$, then the probability bound above gives $\mathrm{P}(\mathcal{M} \subseteq \hat{\mathcal{M}}) \geq 1 - \exp(-c_3 n^{1-2\kappa})$.

Finally, since $q_n = f/p_n$, the requirement on γ_n can be rewritten as $p_n \leq (f/2)\{1 - \Phi(c_2 H^{1/2} n^{1/2-\kappa}/2)\}^{-1}$. Using the fact that $1 - \Phi(x) \leq x^{-1}\exp(-x^2/2)$, this inequality can be satisfied if $p_n \leq f/2\exp(c_2^2 H n^{1-2\kappa}/8)$. Thus the sure screening property holds as long as $\log(p_n) = O(n^{1-2\kappa})$.

Proof of Theorem 5

We first show that for $j \in \mathcal{M}^c$, $U_j(\beta)$ evaluated at the true β_{0j} can be approximated by a sum of continuous-time martingales, just as it can in a correctly specified Cox regression. We can then appeal to an Edgeworth expansion by Gu (1992) to conclude.

By Theorem 1 we know that $\beta_{0j} = 0$ for $j \in \mathcal{M}^c$. Thus we can rewrite (2) as

$$\begin{aligned}
U_j(\beta_{0j}) =& \sum_{i=1}^{n} \int_0^\tau \left\{ Z_{ij} - \frac{n^{-1}\sum_l Z_{lj} Y_l(x)}{n^{-1}\sum_l Y_l(x)} \right\} dN_i(x) \\
=& \sum_{i=1}^{n} \int_0^\tau \left\{ Z_{ij} - \frac{n^{-1}\sum_l Z_{lj} Y_l(x) e^{\boldsymbol{\alpha}_o^T \mathbf{Z}_l}}{n^{-1}\sum_l Y_l(x) e^{\boldsymbol{\alpha}_o^T \mathbf{Z}_l}} \right\} dN_i(x) + \\
& \sum_{i=1}^{n} \int_0^\tau \left\{ \frac{n^{-1}\sum_l Z_{lj} Y_l(x) e^{\boldsymbol{\alpha}_o^T \mathbf{Z}_l}}{n^{-1}\sum_l Y_l(x) e^{\boldsymbol{\alpha}_o^T \mathbf{Z}_l}} - \frac{n^{-1}\sum_l Z_{lj} Y_l(x)}{n^{-1}\sum_l Y_l(x)} \right\} dN_i(x) \\
=& \sum_{i=1}^{n} \int_0^\tau \left\{ Z_{ij} - \frac{n^{-1}\sum_l Z_{lj} Y_l(x) e^{\boldsymbol{\alpha}_o^T \mathbf{Z}_l}}{n^{-1}\sum_l Y_l(x) e^{\boldsymbol{\alpha}_o^T \mathbf{Z}_l}} \right\} dM_i(x) + \\
& n^{-1} \sum_{l=1}^{n} \int_0^\tau \left\{ \frac{Z_{lj} Y_l(x) e^{\boldsymbol{\alpha}_o^T \mathbf{Z}_l}}{n^{-1}\sum_l Y_l(x) e^{\boldsymbol{\alpha}_o^T \mathbf{Z}_l}} - \frac{Z_{lj} Y_l(x)}{n^{-1}\sum_l Y_l(x)} \right\} \sum_i^n dN_i(x),
\end{aligned}$$

where $M_i(x) = N_i(x) - \int^x Y_i(t)\lambda_0(t) e^{\boldsymbol{\alpha}_0^T \mathbf{Z}_i} dt$ is a continuous martingale in x.

Now let $S_m = \sum_{l=1}^{m} \xi_l$, where

$$\xi_l = \left\{ \frac{Z_{lj} Y_l(x) e^{\boldsymbol{\alpha}_o^T \mathbf{Z}_l}}{n^{-1}\sum_l Y_l(x) e^{\boldsymbol{\alpha}_o^T \mathbf{Z}_l}} - \frac{Z_{lj} Y_l(x)}{n^{-1}\sum_l Y_l(x)} \right\} \sum_i^n dN_i(x).$$

Note that $\mathrm{E}(\xi_m \mid S_{m-1}) = \mathrm{E}\{\mathrm{E}(\xi_m \mid S_{m-1}, \mathbf{Z}_m) \mid S_{m-1}\}$, and

$$\mathrm{E}(\xi_m \mid S_{m-1}, \mathbf{Z}_m) = Z_{mj} \mathrm{E}\left[\left\{ \frac{Y_m(x) e^{\boldsymbol{\alpha}_o^T \mathbf{Z}_m}}{n^{-1}\sum_l Y_l(x) e^{\boldsymbol{\alpha}_o^T \mathbf{Z}_l}} - \frac{Y_m(x)}{n^{-1}\sum_l Y_l(x)} \right\} \sum_i^n dN_i(x) \middle| S_{m-1}, \mathbf{Z}_m \right].$$

Given S_{m-1}, the conditional expectation on the right-hand side above is a random variable in $Z_{mk}, k \in \mathcal{M}$ only, and by Assumption 3 is independent of $Z_{mj}, j \in \mathcal{M}^c$. Since $\mathrm{E}(Z_{mj} \mid S_{m-1}) = \mathrm{E}(Z_{lj}) = 0$ by Assumption 1, we find that $\mathrm{E}(\xi_m \mid S_{m-1}) = 0$, implying that S_m is a discrete martingale in m. Then when $m = n$, by the inequality of Dharmadhikari, Fabian, and Jogdeo (1968) we have that $\mathrm{E}(|n^{-1}S_n|^p) = Cn^{-p/2}$ for $p \geq 2$, where we can show that C does not depend on j.

We have shown that for $j \in \mathcal{M}^c$,

$$U_j(\beta_{0j}) = \sum_{i=1}^{n} \int_0^\tau \left\{ Z_{ij} - \frac{n^{-1} \sum_l Z_{lj} Y_l(x) e^{\boldsymbol{\alpha}_o^T \mathbf{Z}_l}}{n^{-1} \sum_l Y_l(x) e^{\boldsymbol{\alpha}_o^T \mathbf{Z}_l}} \right\} dM_i(x) + n^{-1} S_n,$$

where $n^{-1}S_n$ satisfies the same conditions as $R_{1,n}$ in (4.3) of Gu (1992). We can therefore extend the proof of Theorem 2.1 in Gu (1992) to show that

$$\sup_x \left| \mathrm{P}\{I_j(\hat{\beta}_j)^{1/2} |\hat{\beta}_j| \geq x\} - \Phi(x) \right| \leq c_4 n^{-1/2},$$

where c_4 does not depend on j. Then (4) implies

$$\mathrm{E}\left(\frac{\left| \hat{\mathcal{M}} \cap \mathcal{M}^c \right|}{|\mathcal{M}^c|} \right) \leq \frac{1}{p_n - s_n} \sum_{j \in \mathcal{M}^c} \left[2\{1 - \Phi(\gamma_n)\} + c_4 n^{-1/2} \right].$$

The result follows if we choose $\gamma_n = \Phi^{-1}(1 - q_n/2)$.

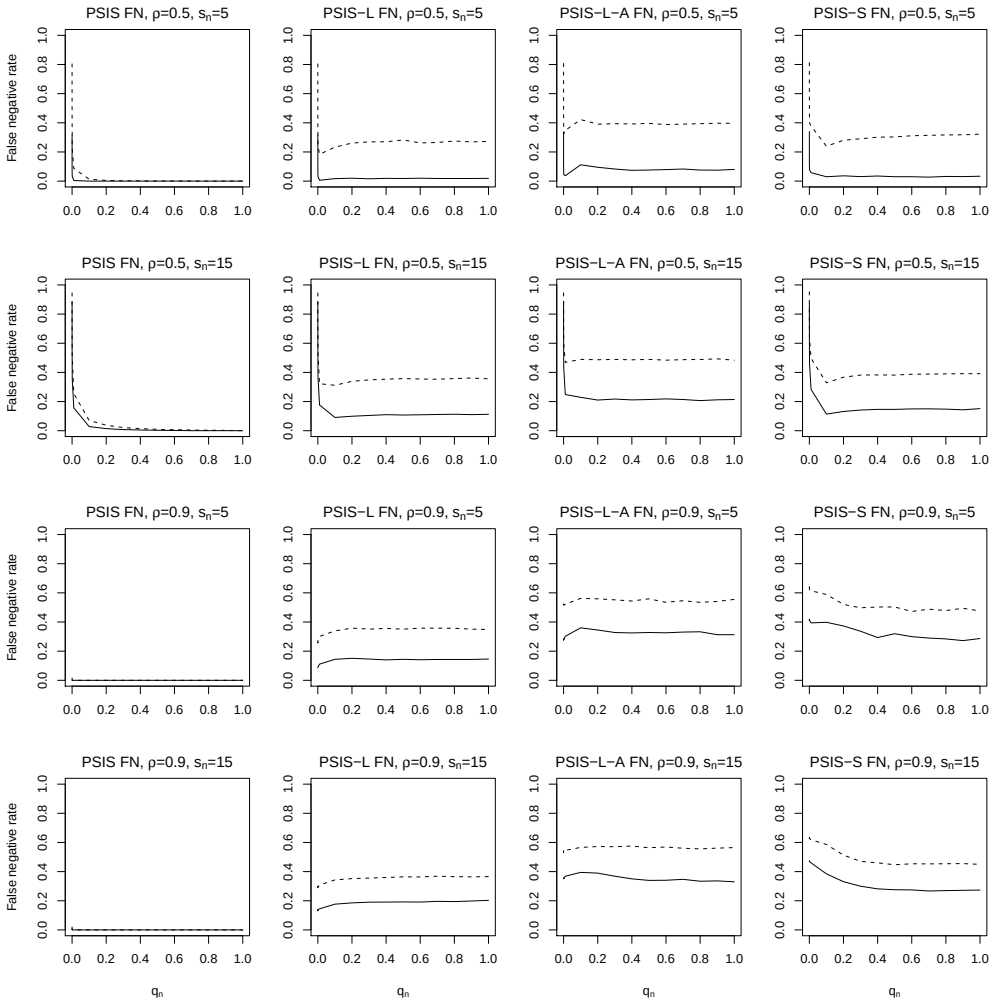

Figure 1: False negative rates for Cox models with $\alpha_{0j} = 0.35$ (dashes) and $\alpha_{0j} = 0.7$ (solid).

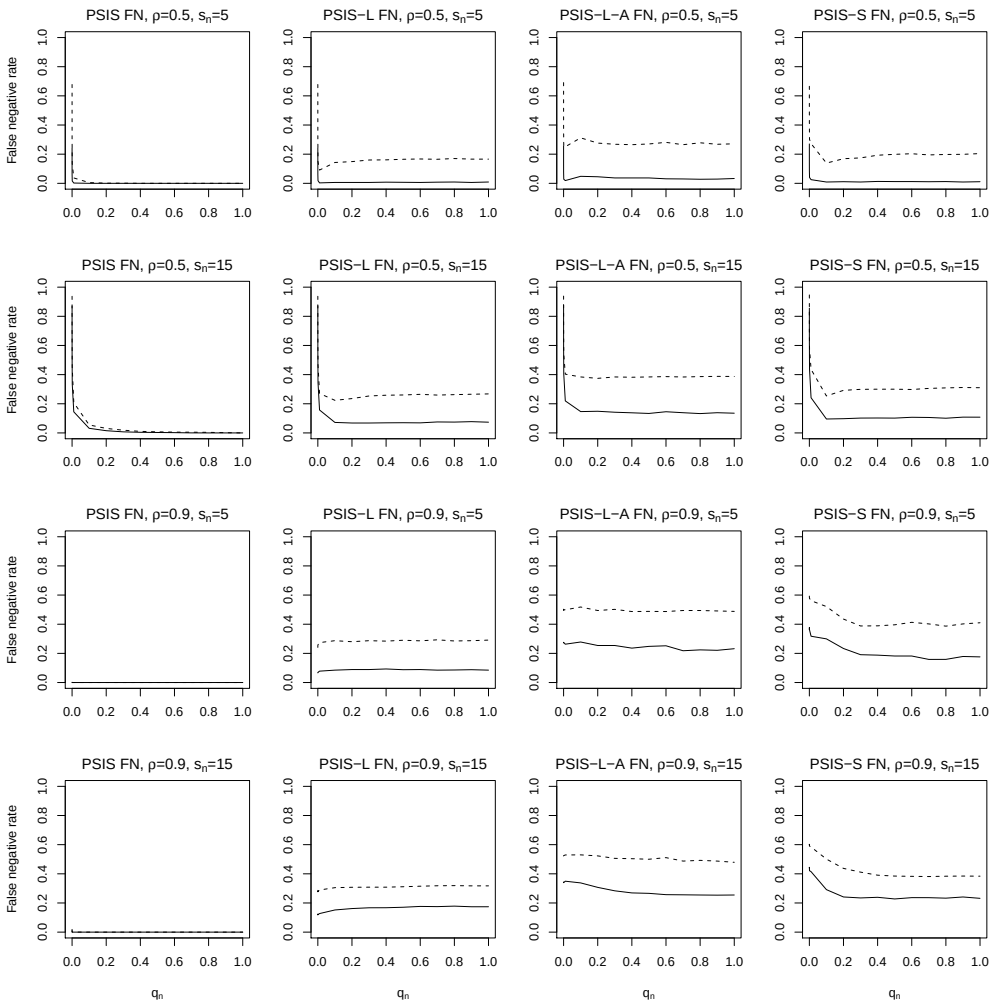

Figure 2: False negative rates for log-normal models $\alpha_{0j} = 0.35$ (dashes) and $\alpha_{0j} = 0.7$ (solid).

Table 1: Simulation results for Cox models with $\alpha_{0j} = 0.35$

				PSIS		PSIS-L		PSIS-A		PSIS-S	
ρ	s_n	q_n	$\lvert\hat{\mathcal{M}}\rvert$	FN	FP	FN	FP	FN	FP	FN	FP
0.5	5	1e-6	0.98	0.80	0.00	0.80	0.00	0.81	0.00	0.81	0.00
0.5	5	1e-5	1.79	0.65	3e-5	0.65	0.00	0.66	0.00	0.63	0.00
0.5	5	1e-4	2.94	0.44	1e-4	0.44	0.00	0.48	0.00	0.49	0.00
0.5	5	1e-3	4.94	0.23	1e-3	0.24	0.00	0.33	0.00	0.40	0.00
0.5	5	0.01	14.79	0.09	0.01	0.18	0.01	0.35	0.01	0.38	0.01
0.5	5	0.10	107.72	0.01	0.10	0.23	0.03	0.42	0.02	0.24	0.05
0.5	5	0.20	207.74	0.01	0.20	0.26	0.02	0.39	0.02	0.28	0.05
0.5	5	0.30	306.95	0.00	0.30	0.27	0.02	0.39	0.01	0.29	0.05
0.5	5	0.40	406.71	0.00	0.40	0.27	0.02	0.39	0.01	0.30	0.06
0.5	5	0.50	506.11	0.00	0.50	0.28	0.02	0.40	0.01	0.30	0.06
0.5	5	0.60	605.37	0.00	0.60	0.26	0.02	0.39	0.01	0.31	0.06
0.5	5	0.70	704.75	0.00	0.70	0.27	0.02	0.39	0.01	0.31	0.06
0.5	5	0.80	803.49	0.00	0.80	0.28	0.02	0.39	0.01	0.32	0.06
0.5	5	0.90	901.71	0.00	0.90	0.27	0.02	0.40	0.01	0.32	0.06
0.5	5	1.00	1000.00	0.00	1.00	0.27	0.02	0.40	0.01	0.32	0.06
0.5	15	1e-6	0.82	0.95	0.00	0.95	0.00	0.95	0.00	0.95	0.00
0.5	15	1e-5	1.81	0.88	5e-6	0.88	0.00	0.89	0.00	0.87	0.00
0.5	15	1e-4	4.08	0.73	7e-5	0.74	0.00	0.76	0.00	0.77	0.00
0.5	15	1e-3	8.34	0.51	1e-3	0.53	0.00	0.59	0.00	0.63	0.00
0.5	15	0.01	21.59	0.26	0.01	0.32	0.01	0.47	0.01	0.50	0.01
0.5	15	0.20	213.49	0.04	0.20	0.34	0.03	0.49	0.02	0.37	0.04
0.5	15	0.30	312.68	0.02	0.30	0.35	0.02	0.49	0.02	0.38	0.04
0.5	15	0.40	410.25	0.01	0.40	0.35	0.02	0.49	0.02	0.38	0.04
0.5	15	0.50	508.06	0.01	0.50	0.36	0.02	0.49	0.02	0.38	0.04
0.5	15	0.60	606.76	0.01	0.60	0.35	0.02	0.48	0.02	0.39	0.05
0.5	15	0.70	705.09	0.00	0.70	0.35	0.02	0.49	0.02	0.39	0.05
0.5	15	0.80	803.67	0.00	0.80	0.36	0.02	0.49	0.01	0.39	0.05
0.5	15	0.90	901.97	0.00	0.90	0.36	0.02	0.49	0.02	0.39	0.05
0.5	15	1.00	1000.00	0.00	1.00	0.36	0.02	0.48	0.02	0.39	0.05
0.9	5	1e-6	6.96	0.02	2e-3	0.26	0.00	0.52	0.00	0.64	0.00
0.9	5	1e-5	8.01	0.00	3e-3	0.26	0.00	0.52	0.00	0.64	0.00
0.9	5	1e-4	9.51	0.00	5e-3	0.26	0.00	0.52	0.00	0.64	0.00
0.9	5	1e-3	12.14	0.00	7e-3	0.28	0.00	0.52	0.00	0.62	0.00
0.9	5	0.01	24.47	0.00	0.02	0.30	0.01	0.52	0.00	0.61	0.00
0.9	5	0.10	119.95	0.00	0.12	0.34	0.02	0.56	0.01	0.59	0.01
0.9	5	0.10	119.95	0.00	0.12	0.34	0.02	0.56	0.01	0.59	0.01
0.9	5	0.20	220.82	0.00	0.22	0.36	0.02	0.56	0.01	0.52	0.03
0.9	5	0.30	318.11	0.00	0.31	0.35	0.02	0.55	0.01	0.50	0.03
0.9	5	0.40	415.01	0.00	0.41	0.36	0.02	0.54	0.01	0.50	0.04
0.9	5	0.50	512.13	0.00	0.51	0.35	0.02	0.56	0.01	0.50	0.04
0.9	5	0.60	609.18	0.00	0.61	0.36	0.02	0.54	0.01	0.47	0.04
0.9	5	0.70	707.00	0.00	0.71	0.36	0.02	0.55	0.01	0.49	0.04
0.9	5	0.80	804.61	0.00	0.80	0.36	0.02	0.54	0.01	0.48	0.04
0.9	5	0.90	901.93	0.00	0.90	0.35	0.01	0.54	0.01	0.49	0.04
0.9	5	1.00	1000.00	0.00	1.00	0.35	0.02	0.56	0.01	0.48	0.04
0.9	15	1e-6	16.12	0.02	1e-3	0.30	0.00	0.54	0.00	0.63	0.00
0.9	15	1e-5	17.12	0.01	2e-3	0.29	0.00	0.53	0.00	0.63	0.00
0.9	15	1e-4	18.49	0.00	4e-3	0.29	0.00	0.54	0.00	0.63	0.00
0.9	15	1e-3	20.95	0.00	6e-3	0.30	0.00	0.55	0.00	0.63	0.00
0.9	15	0.01	32.11	0.00	0.02	0.31	0.00	0.55	0.00	0.62	0.00
0.9	15	0.10	124.98	0.00	0.11	0.34	0.01	0.57	0.01	0.59	0.01
0.9	15	0.20	223.62	0.00	0.21	0.35	0.02	0.57	0.01	0.51	0.02
0.9	15	0.30	321.65	0.00	0.31	0.36	0.02	0.57	0.01	0.47	0.03
0.9	15	0.40	420.10	0.00	0.41	0.36	0.02	0.58	0.01	0.46	0.03
0.9	15	0.50	517.53	0.00	0.51	0.36	0.02	0.56	0.01	0.45	0.04
0.9	15	0.60	613.51	0.00	0.61	0.36	0.02	0.57	0.01	0.45	0.04
0.9	15	0.70	710.24	0.00	0.71	0.37	0.02	0.56	0.01	0.45	0.04
0.9	15	0.80	806.47	0.00	0.80	0.37	0.02	0.56	0.01	0.45	0.04
0.9	15	0.90	903.38	0.00	0.90	0.36	0.02	0.56	0.01	0.45	0.04
0.9	15	1.00	1000.00	0.00	1.00	0.37	0.02	0.56	0.01	0.45	0.04

Table 2: Simulation results for log-normal models with $\alpha_{0j} = 0.35$

				PSIS		PSIS-L		PSIS-A		PSIS-S	
ρ	s_n	q_n	$\lvert\hat{\mathcal{M}}\rvert$	FN	FP	FN	FP	FN	FP	FN	FP
0.5	5	1e-6	1.61	0.68	0.00	0.68	0.00	0.69	0.00	0.67	0.00
0.5	5	1e-5	2.50	0.51	2e-5	0.51	0.00	0.54	0.00	0.53	0.00
0.5	5	1e-4	3.63	0.30	1e-4	0.31	0.00	0.37	0.00	0.40	0.00
0.5	5	1e-3	5.42	0.13	1e-3	0.15	0.00	0.25	0.00	0.30	0.00
0.5	5	0.01	15.45	0.04	0.01	0.09	0.01	0.25	0.01	0.27	0.01
0.5	5	0.10	107.30	0.01	0.10	0.14	0.03	0.31	0.02	0.14	0.05
0.5	5	0.20	209.06	0.00	0.21	0.15	0.03	0.28	0.02	0.17	0.05
0.5	5	0.30	309.48	0.00	0.31	0.16	0.02	0.27	0.02	0.17	0.05
0.5	5	0.40	410.22	0.00	0.41	0.16	0.02	0.27	0.01	0.19	0.05
0.5	5	0.50	509.29	0.00	0.51	0.17	0.02	0.27	0.01	0.20	0.05
0.5	5	0.60	607.66	0.00	0.61	0.17	0.02	0.28	0.01	0.20	0.05
0.5	5	0.70	706.88	0.00	0.71	0.17	0.02	0.27	0.01	0.19	0.05
0.5	5	0.80	803.38	0.00	0.80	0.17	0.02	0.28	0.01	0.20	0.05
0.5	5	0.90	901.01	0.00	0.90	0.17	0.02	0.27	0.01	0.20	0.05
0.5	5	1.00	1000.00	0.00	1.00	0.17	0.02	0.27	0.01	0.20	0.05
0.5	15	1e-6	0.94	0.94	0.00	0.94	0.00	0.94	0.00	0.95	0.00
0.5	15	1e-5	2.35	0.84	5e-6	0.84	0.00	0.85	0.00	0.83	0.00
0.5	15	1e-4	5.03	0.67	1e-4	0.68	0.00	0.70	0.00	0.71	0.00
0.5	15	1e-3	9.29	0.45	1e-3	0.47	0.00	0.54	0.00	0.58	0.00
0.5	15	0.01	22.11	0.21	0.01	0.27	0.01	0.40	0.00	0.44	0.00
0.5	15	0.10	113.98	0.06	0.10	0.22	0.03	0.38	0.02	0.25	0.03
0.5	15	0.20	213.75	0.03	0.20	0.24	0.03	0.37	0.02	0.29	0.03
0.5	15	0.30	312.87	0.02	0.30	0.25	0.03	0.38	0.02	0.30	0.04
0.5	15	0.40	411.37	0.01	0.40	0.26	0.03	0.38	0.02	0.30	0.04
0.5	15	0.50	509.57	0.01	0.50	0.26	0.03	0.38	0.02	0.30	0.04
0.5	15	0.60	608.60	0.01	0.60	0.26	0.03	0.39	0.02	0.30	0.04
0.5	15	0.70	706.09	0.01	0.70	0.26	0.03	0.38	0.02	0.31	0.04
0.5	15	0.80	804.67	0.00	0.80	0.26	0.03	0.39	0.02	0.31	0.04
0.5	15	0.90	902.90	0.00	0.90	0.27	0.02	0.39	0.02	0.31	0.04
0.5	15	1.00	1000.00	0.00	1.00	0.27	0.03	0.39	0.02	0.31	0.04
0.9	5	1e-6	7.59	0.00	3e-3	0.24	0.00	0.49	0.00	0.59	0.00
0.9	5	1e-5	8.49	0.00	4e-3	0.25	0.00	0.50	0.00	0.59	0.00
0.9	5	1e-4	9.79	0.00	7e-3	0.25	0.00	0.50	0.00	0.59	0.00
0.9	5	1e-3	12.39	0.00	0.01	0.26	0.00	0.50	0.00	0.58	0.00
0.9	5	0.01	23.93	0.00	0.02	0.27	0.01	0.50	0.00	0.56	0.00
0.9	5	0.10	119.55	0.00	0.12	0.29	0.02	0.52	0.01	0.52	0.01
0.9	5	0.20	220.84	0.00	0.22	0.28	0.02	0.49	0.01	0.44	0.03
0.9	5	0.30	320.67	0.00	0.32	0.29	0.02	0.50	0.01	0.39	0.04
0.9	5	0.40	419.29	0.00	0.42	0.28	0.02	0.49	0.01	0.39	0.04
0.9	5	0.50	517.38	0.00	0.51	0.29	0.02	0.49	0.01	0.40	0.04
0.9	5	0.60	615.42	0.00	0.61	0.29	0.02	0.49	0.01	0.41	0.04
0.9	5	0.70	711.63	0.00	0.71	0.29	0.02	0.49	0.01	0.40	0.04
0.9	5	0.80	807.91	0.00	0.81	0.29	0.02	0.49	0.01	0.39	0.04
0.9	5	0.90	904.24	0.00	0.90	0.29	0.02	0.49	0.01	0.40	0.04
0.9	5	1.00	1000.00	0.00	1.00	0.29	0.02	0.49	0.01	0.41	0.04
0.9	15	1e-6	16.17	0.02	1e-3	0.29	0.00	0.52	0.00	0.60	0.00
0.9	15	1e-5	17.09	0.01	2e-3	0.28	0.00	0.52	0.00	0.60	0.00
0.9	15	1e-4	18.44	0.00	4e-3	0.28	0.00	0.52	0.00	0.60	0.00
0.9	15	1e-3	21.27	0.00	0.01	0.28	0.00	0.53	0.00	0.59	0.00
0.9	15	0.01	33.07	0.00	0.02	0.29	0.00	0.53	0.00	0.58	0.00
0.9	15	0.10	125.19	0.00	0.11	0.30	0.01	0.53	0.01	0.50	0.01
0.9	15	0.20	225.52	0.00	0.21	0.31	0.02	0.52	0.01	0.44	0.02
0.9	15	0.30	321.33	0.00	0.31	0.31	0.02	0.51	0.01	0.41	0.03
0.9	15	0.40	419.49	0.00	0.41	0.31	0.02	0.50	0.01	0.39	0.03
0.9	15	0.50	516.70	0.00	0.51	0.31	0.02	0.50	0.01	0.38	0.03
0.9	15	0.60	613.00	0.00	0.61	0.32	0.02	0.51	0.01	0.38	0.03
0.9	15	0.70	709.83	0.00	0.71	0.32	0.02	0.49	0.01	0.38	0.03
0.9	15	0.80	807.23	0.00	0.80	0.32	0.02	0.49	0.01	0.38	0.03
0.9	15	0.90	903.17	0.00	0.90	0.32	0.02	0.49	0.01	0.38	0.03
0.9	15	1.00	1000.00	0.00	1.00	0.32	0.02	0.48	0.01	0.38	0.03

Table 3: Predictive accuracies using myeloma data)

	PSIS, $q_n = 0.0001$		Random probesets		All probesets	
Method	C-stat (SD)	Size (SD)	C-stat (SD)	Size (SD)	C-stat (SD)	Size (SD)
Lasso	0.64 (0.10)	8.27 (3.39)	0.42 (0.28)	3.51 (3.60)	0.58 (0.19)	10.64 (6.17)
Lasso-adaptive lasso	0.63 (0.10)	5.61 (1.94)	0.40 (0.29)	2.60 (2.82)	0.55 (0.18)	7.43 (3.74)
SCAD	0.62 (0.09)	7.89 (7.69)	0.43 (0.26)	7.29 (13.13)	—	—

Table 4: Genes found using PSIS-L-A, $q_n = 0.0001$

Probeset	Gene	Description	Coefficient
207677_s_at	NCF4	neutrophil cytosolic factor 4, 40kDa	-0.38
216510_x_at	IGHV3-23	immunoglobulin heavy variable 3-23	-0.30
204319_s_at	RGS10	regulator of G-protein signaling 10	-0.24
220057_at	XAGE1A/B/C/D/E	X antigen family, member 1A/B/C/D/E	0.05
214414_x_at	HBA1/2	hemoglobin, alpha 1/2	-0.27
217889_s_at	CYBRD1	cytochrome b reductase 1	-0.24