Supplemental Appendix

Lexin Li (2005). Survival prediction of diffuse large-B-cell lymphoma based on both clinical and gene expression information. *Bioinformatics*.

Algorithm of partial sliced inverse regression for survival data

For the i-th patient, $i = 1, \ldots, n$, let X_i denote the p-dimensional predictor vector, Y_i denote the observed survival time, δ_i denote the censoring indicator, and W_i denote the observed categorical variable. W_i corresponds to the IPI group in DLBCL data, and is assumed to take value in $\{1, \ldots, C\}$. The proposed algorithm of partial sliced inverse regression for survival data operates on the data $\{(X_i, Y_i, \delta_i, W_i), i = 1, \ldots, n\}$ in the following way:

1. Divide the data into C subsets according to the value of W_i, yielding $\{(X_{wi}, Y_{wi}, \delta_{wi}), w = 1, \ldots, C, i = 1, \ldots, n_w\}$, and $\sum_{w=1}^{C} n_w = n$.

2. Within each subset $\{(X_{wi}, Y_{wi}, \delta_{wi})\}$, the covariance matrix $\mathrm{Cov}(X_w)$ (on the right hand side of Equation (4) of the paper) is estimated by

$$\hat{\Sigma}_{X_w} = \frac{1}{n_w} \sum_{i=1}^{n_w} (X_{wi} - \bar{X}_w)(X_{wi} - \bar{X}_w)^{\mathsf{T}},$$

   where $\bar{X}_w = 1/n_w \sum_{i=1}^{n_w} X_{wi}$.

3. Within each subset $\{(X_{wi}, Y_{wi}, \delta_{wi})\}$,

   (1) Partition Y_{wi} into 2 subsets according to $\delta_{wi} = 1$ or 0. Then partition the range of Y_{wi} within each subset to h non-overlapping intervals. This yields $2h$ intervals (slices), and each Y_{wi} belongs to one and only of those slices.

   (2) Compute the sample estimate of $\mathrm{E}(X_w \,|\, Y_w)$ within the s-th slice, $s = 1, \ldots, 2h$, as

$$\hat{\mathrm{E}}_s = \frac{1}{n_{sw}} \sum_{i|s} X_{wi},$$

   where $\sum_{i|s}$ denotes that the sum is over indexes i of the corresponding response observations Y_{wi} that fall into slice s, and n_{sw} denotes the number of observations in slice s.

   (3) The covariance matrix $\mathrm{Cov}(\mathrm{E}(X_w \,|\, Y_w))$ (on the left hand side of Equation (4) of the paper) is estimated by

$$\hat{\Sigma}_{X_w \,|\, Y_w} = \sum_{s=1}^{2h} \frac{n_{sw}}{n_w} (\hat{\mathrm{E}}_s - \bar{\hat{\mathrm{E}}})(\hat{\mathrm{E}}_s - \bar{\hat{\mathrm{E}}})^{\mathsf{T}},$$

   where $\bar{\hat{\mathrm{E}}} = \sum_{s=1}^{2h} \frac{n_{sw}}{n_w} \hat{\mathrm{E}}_s$.

4. The pooled sample covariance matrix of predictors, $\sum_{w=1}^{C} P(W = w)\mathrm{Cov}(X_w)$ (on the right hand side of Equation (4) of the paper), is estimated by

$$\hat{\Sigma}_X^{pool} = \sum_{w=1}^{C} \frac{n_w}{n} \hat{\Sigma}_{X_w},$$

and the pooled sample covariance matrix of inverse mean, $\sum_{w=1}^{C} P(W = w)\mathrm{Cov}(\mathrm{E}(X_w \mid Y_w))$, (on the left hand side of Equation (4) of the paper) is estimated by

$$\hat{\Sigma}_{X|Y}^{pool} = \sum_{w=1}^{C} \frac{n_w}{n} \hat{\Sigma}_{X_w \mid Y_w}.$$

5. Perform eigen-decomposition as in Equation (4) of the paper, with the substituted sample estimators $\hat{\Sigma}_X^{pool}$ and $\hat{\Sigma}_{X|Y}^{pool}$. This gives the sample partial sliced inverse regression estimators $\hat{\eta}_1, \ldots, \hat{\eta}_d$.

Supplemental figure

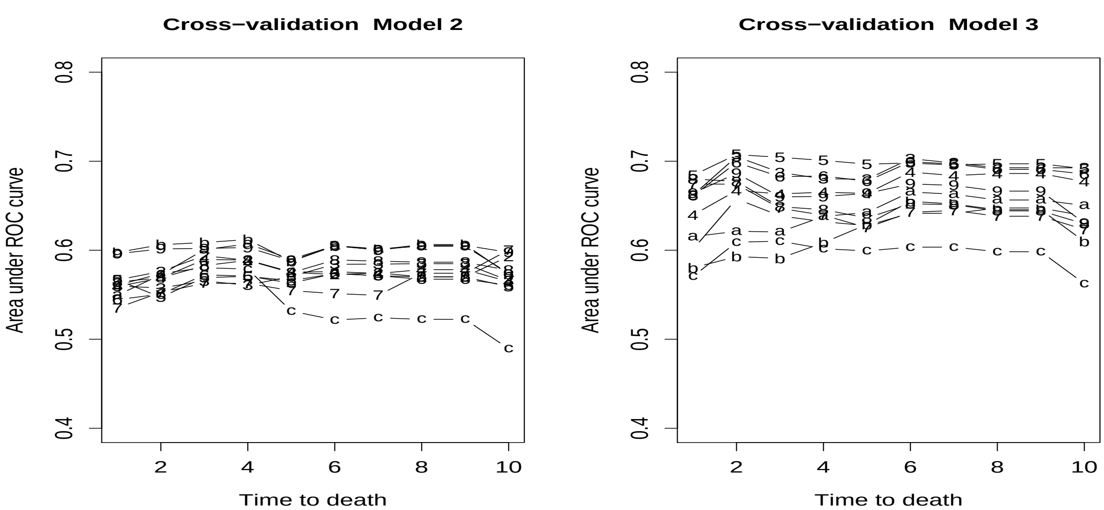

Figure S1: Area under ROC curves at time 1 year to 10 years for model 2 (left panel) and model 3 (right panel) based on a range of number of principal components q. Numbers 2 through 9, plus characters a through c denote $q = 20, 30, \ldots, 120$, respectively.

Supplemental tables

Table S1: Three fitted Cox proportional hazards models based on the training data.

	Term	Coefficient	S.E.	P-value
Model 1	IPI-Intermediate	1.06	0.28	1.9e-04
	IPI-High	1.77	0.32	4.4e-08
Model 2	SIR	-0.15	0.02	1.3e-13
Model 3	IPI-Intermediate	1.24	0.29	1.6e-05
	IPI-High	2.22	0.34	4.1e-11
	PSIR	0.14	0.02	5.4e-13

Table S2: Area under ROC curves at time 1, 3, and 6 (years) for four designs and four models in Monte Carlo study. Shown are the median and the median absolute deviation (in parenthesis) of AUCs based on 100 replications. All results are evaluated on the testing data.

	Time	Model 1	Model 2	Model 3	Model 4
Design 1	1	0.785 (0.029)	0.497 (0.034)	0.824 (0.025)	0.824 (0.023)
	3	0.834 (0.017)	0.490 (0.029)	0.855 (0.017)	0.856 (0.018)
	6	0.730 (0.023)	0.474 (0.078)	0.730 (0.019)	0.731 (0.018)
Design 2	1	0.501 (0.026)	0.686 (0.026)	0.682 (0.032)	0.682 (0.030)
	3	0.509 (0.041)	0.671 (0.031)	0.670 (0.027)	0.664 (0.027)
	6	0.500 (0.131)	0.676 (0.039)	0.674 (0.042)	0.672 (0.042)
Design 3	1	0.635 (0.046)	0.648 (0.028)	0.703 (0.026)	0.704 (0.025)
	3	0.670 (0.093)	0.663 (0.035)	0.726 (0.050)	0.727 (0.045)
	6	0.676 (0.193)	0.666 (0.056)	0.723 (0.082)	0.716 (0.084)
Design 4	1	0.606 (0.038)	0.643 (0.030)	0.692 (0.030)	0.710 (0.029)
	3	0.589 (0.071)	0.667 (0.036)	0.689 (0.043)	0.715 (0.045)
	6	0.554 (0.137)	0.667 (0.058)	0.678 (0.078)	0.718 (0.091)