Supplementary Material for Prediction of sports injuries in football: a recurrent time-to-event approach using regularized Cox models

Lore Zumeta-Olaskoaga, Maximilian Weigert, Jon Larruskain,
Eder Bikandi, Igor Setuain, Josean Lekue, Helmut Küchenhoff, Dae-Jin Lee

A. Supplementary information on the real case study

Table 1: A detailed information of the 28 variables of the screening tests and lower-limb injury data: type of screening test and number of tests of each type

Type of screening test	Number of tests
Anthropometrics & previous injury	5
Active Straight Leg Raise (ASLR)	2
Cross Over Hop	3
Core Strength Side	1
Drop Jump Kinetics	4
Hand-Held Dynamometry	2
Isokinetics	4
KT1000	1
ROM	4
Star Excursion Balance	2
TOTAL	28

B. Bootstrap .632+ estimates of the Brier score

In this section we briefly describe the " .632 bootstrap approach" used in the calculation of the Brier Score to avoid overfitting.

The bootstrap .632+ method was proposed in Efron (1983) and discussed in Efron and Tibshirani (1997). In the latter they discussed cross-validation and bootstrap estimates of prediction error and showed that the bootstrap .632+ method substantially outperforms cross-validation in the simulation experiments they performed. Besides, the .632+ method for the Brier score prediction error estimate was specifically proposed in Binder and Schumacher (2008) and demonstrated to lead to more accurate estimates.

Namely, the bootstrap .632+ estimate of the Brier score –prediction error curve– is a weighted linear combination of the apparent estimate, the bootstrap cross-validation estimate and the no information estimate. It is implemented in the `pec` package (Gerds, 2020) in **R**, and the following definition is given by them:

$$\text{Boot632plusErr}(t, \hat{S}) = \left(1 - \frac{0.632}{1 - 0.368 \cdot \omega}\right) \text{AppErr}(t, \hat{S}) + \left(\frac{0.632}{1 - 0.368 \cdot \omega}\right) \text{BootCvErr}(t, \hat{S}),$$

where

$$\omega = \frac{\min(\text{BootCvErr}(t,\hat{S}), \text{NoInfErr}(t,\hat{S})) - \text{AppErr}(t,\hat{S})}{\text{NoInfErr}(t,\hat{S}) - \text{AppErr}(t,\hat{S})}.$$

The constant 0.632 is independent of the sample size and corresponds to the probability to draw with replacement subject i into the bootstrap sample: $P(\{(Y_i, X_i)\} \in D_b) = 1 - (1 - 1/N)^N \approx (1 - e^{-1}) \approx 0.632$.

We refer the reader to Mogensen et al. (2012) and the **R** help page of `pec::pec` function for the definitions of the apparent, bootstrap cross-validation and the no information estimate, i.e. $\text{AppErr}(t,\hat{S})$, $\text{BootCvErr}(t,\hat{S})$ and $\text{NoInfErr}(t,\hat{S})$ respectively.

C. Simulation study

In the following, we present further results of the simulation study in the form of tabular and graphical displays: the results of the last setting in scenario 2 and scenario 3 (C.1.), additional performance measures to report more estimates of simulation uncertainty (C.2.) and several graphical presentation of the simulations results for scenario 2 and scenario 3 (C.3.). Furthermore, the source code to reproduce the simulation study conducted can be found at: `https://github.com/lzumeta/TimeToEvent-InjurySim`.

C.1. Results of scenario 2 and scenario 3, last setting, $N_{\text{obs}} = 670$

Table 2: Simulation results for Scenarios 2 and 3, that consider different correlation structure of covariates, $\rho_{ij} = 0$ and $\rho_{ij} = 0.65^{|i-j|}$, for $K = 220$ that give rise to $N_{\text{obs}} = 670$ number of observations. Average model size (AMS), average number of falsely selected variables (ANFS), average number of coefficients incorrectly estimated as zero (ANFNS), mean squared error (MSE) and the median of the integrated Brier scores between [0, 1000] and [0, 3500] time intervals, for all models, are reported.

Sample size (N_{obs})	**Correlation structure** ($i \neq j$)	**Frailty model including vars. that selected**	**AMS** (5)	**ANFS** (0)	**ANFNS** (0)	**MSE** (0)	**IBS** (0)	
							[0,1000]	**[0, 3500]**
$N_{\text{obs}} = 670$	$\rho_{ij} = 0$	BeSS	2.94	0.66	2.72	0.52	0.033	0.109
		Lasso	12.41	8.66	1.25	0.77	0.033	0.107
		Elastic Net	14.18	10.34	1.16	0.79	0.033	0.108
		Ridge	7.53	4.24	1.71	0.65	0.033	0.108
		Group Lasso	21.1	16.1	0	0.80	0.034	0.110
		Cox Boosting	7.36	4.28	1.92	0.67	0.033	0.107
	$\rho_{ij} = 0.65^{\lvert i-j \rvert}$	BeSS	3.03	1.52	2.69	0.88	0.037	0.104
		Lasso	15.41	11.15	0.74	0.90	0.036	0.102
		Elastic Net	17.22	12.70	0.48	0.91	0.036	0.102
		Ridge	9.26	4.91	0.65	0.79	0.036	0.102
		Group Lasso	32.15	27.15	0	1.44	0.037	0.107
		Cox Boosting	10.53	6.89	1.36	0.90	0.036	0.101

C.2. Additional performance measures: MSE, bias and empirical standard errors

We report two additional performance measures, in addition to the mean square error (MSE) reported in the main manuscript in order to better interpret this first measure.

The calculation of MSE is,

$$\text{MSE} = \frac{1}{N_{\text{sim}}} \sum_{i=1}^{N_{\text{sim}}} \sum_{j=1}^{p} (\hat{\beta}_j^{(i)} - \beta_j)^2,$$

bias is calculated as,

$$\text{Bias} = \frac{1}{N_{\text{sim}}} \sum_{i=1}^{N_{\text{sim}}} \sum_{j=1}^{p} (\hat{\beta}_j^{(i)} - \beta_j),$$

and empirical standard error (EmpSE),

$$\text{EmpSE} = \sqrt{\frac{1}{N_{\text{sim}}} \sum_{i=1}^{N_{\text{sim}}} \sum_{j=1}^{p} \left(\hat{\beta}_j^{(i)} - \bar{\hat{\beta}}\right)},$$

which only depends on the estimates $\hat{\beta}_j$.

Scenario 1:

Table 3: Simulation results for the three different settings within Scenario 1, for 66 players with 3 observations each, that results in a sample size of 198. Bias, empirical standard error (EmpSE) and mean squared error (MSE) for all models are reported.

Model	**Bias**	**EmpSE**	**MSE**
True model: frailty (BeSS)			
BeSS	-0.07	1.47	3.85
Lasso	0.22	1.56	4.29
Elastic Net	0.24	1.71	4.84
Ridge	0.04	1.41	3.78
Group Lasso	-0.03	2.69	9.24
Boosting	0.23	1.56	4.29
True model: frailty (Lasso)			
BeSS	-0.36	1.75	5.05
Lasso	-0.42	1.95	5.67
Elastic Net	-0.52	2.09	6.26
Ridge	-0.42	1.64	4.66
Group Lasso	-1.31	3.57	14.57
Boosting	-0.41	1.95	5.65
True model: frailty (Boosting)			
BeSS	0.17	1.69	5.24
Lasso	0.23	1.84	6.06
Elastic Net	0.16	2.11	7.15
Ridge	0.23	1.68	5.50
Group Lasso	0.36	3.09	12.38
Boosting	0.22	1.87	6.17

Scenario 2 and scenario 3:

Table 4: Simulation results for Scenarios 2 and 3, that consider different correlation structures of covariates, $\rho_{ij} = 0$ and $\rho_{ij} = 0.65^{|i-j|}$, for different number of players $K \in \{22, 66, 132, 220\}$ that give rise to $N_{\text{obs}} \in \{60, 191, 391, 670\}$ number of observations. Bias, empirical standard error (EmpSE) and mean squared error (MSE) for all models are reported.

Sample size (N_{obs})	**Correlation structure** ($i \neq j$)	**Frailty model including vars. that selected**	**Bias**	**EmpSE**	**MSE**
$N_{\text{obs}} = 60$	$\rho_{ij} = 0$	BeSS	-0.61	1.61	3.69
		Lasso	-1.08	6.28	56.37
		Elastic Net	1.61	14.64	271.34
		Ridge	-0.22	46.69	57.02
		Group Lasso	-28.34	111.92	$> 10^5$
		Cox Boosting	-0.64	5.43	42.77
	$\rho_{ij} = 0.65^{\lvert i-j \rvert}$	BeSS	-0.14	1.85	4.90
		Lasso	-0.37	1.32	2.72
		Elastic Net	-3.72	45.87	2758.8
		Ridge	1.42	23.40	776.0
		Group Lasso	-21.55	$> 10^5$	$> 10^5$
		Cox Boosting	-0.37	1.36	2.92
$N_{\text{obs}} = 191$	$\rho_{ij} = 0$	BeSS	-0.35	0.71	0.96
		Lasso	-0.37	0.80	1.10
		Elastic Net	-0.33	0.84	1.21
		Ridge	-0.15	0.84	1.21
		Group Lasso	0.36	1.62	3.39
		Cox Boosting	-0.25	0.83	1.18
	$\rho_{ij} = 0.65^{\lvert i-j \rvert}$	BeSS	-0.12	0.88	1.28
		Lasso	-0.48	1.03	1.54
		Elastic Net	-0.61	1.10	1.78
		Ridge	-0.40	1	1.50
		Group Lasso	-182.56	7031.94	$> 10^5$
		Cox Boosting	-0.39	1	1.48
$N_{\text{obs}} = 391$	$\rho_{ij} = 0$	BeSS	0.13	0.57	0.57
		Lasso	0.29	0.72	0.82
		Elastic Net	0.46	0.76	0.89
		Ridge	0.39	0.68	0.74
		Group Lasso	0.74	0.88	1.09
		Cox Boosting	0.33	0.71	0.79
	$\rho_{ij} = 0.65^{\lvert i-j \rvert}$	BeSS	-0.16	0.76	1.02
		Lasso	-0.64	0.89	1.18
		Elastic Net	-0.71	0.93	1.27
		Ridge	-0.36	0.80	0.99
		Group Lasso	-0.35	1.39	2.40
		Cox Boosting	-0.51	0.85	1.12
$N_{\text{obs}} = 670$	$\rho_{ij} = 0$	BeSS	0.10	0.53	0.53
		Lasso	0.17	0.69	0.77
		Elastic Net	0.23	0.70	0.79
		Ridge	0.09	0.62	0.65
		Group Lasso	0.41	0.73	0.80
		Cox Boosting	0.15	0.63	0.67
	$\rho_{ij} = 0.65^{\lvert i-j \rvert}$	BeSS	-0.08	0.72	0.88
		Lasso	-0.76	0.77	0.90
		Elastic Net	-0.75	0.77	0.91
		Ridge	-0.34	0.70	0.79
		Group Lasso	-0.45	1.08	1.44
		Cox Boosting	-0.56	0.76	0.90

C.3. Simulation study: graphical summaries of scenarios 2 and 3

In this section, we present additional graphics regarding scenario 2 and scenario 3 of the simulation study. Source code to reproduce the results is available as a separate Supporting Material in the following GitHub repository: `https://github.com/lzumeta/TimeToEvent-InjurySim`

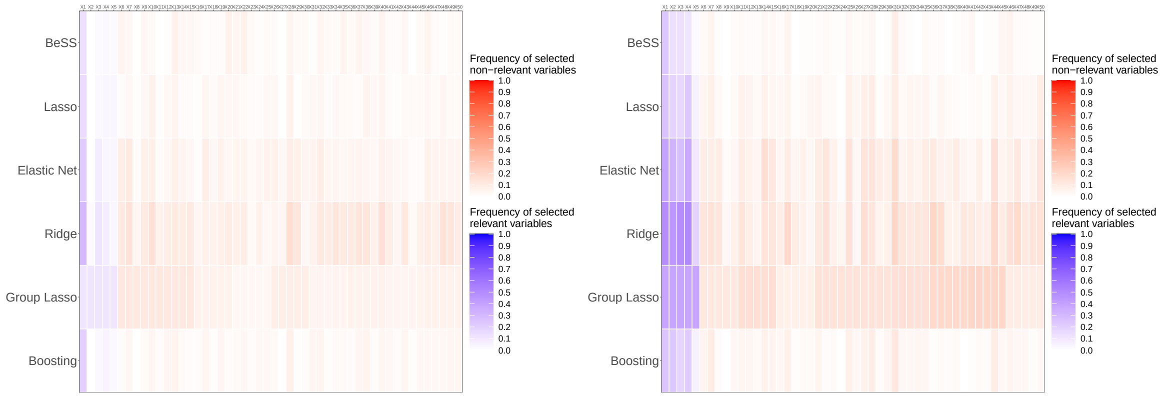

Figure 1: Selected variables within setting 1 in scenario 2 (left) and scenario 3 (right). The frequency of selection, by which each variable has been selected, is shown by color intensity. In addition, blue color represents variables that have a true effect on the outcome, whereas the red refers to variables with no effect

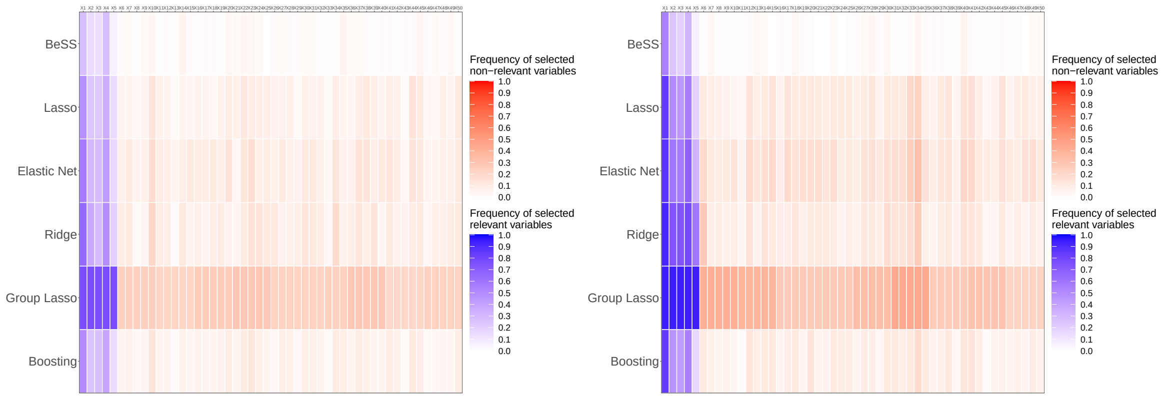

Figure 2: Selected variables within setting 2 in scenario 2 (left) and scenario 3 (right). The frequency of selection, by which each variable has been selected, is shown by color intensity. In addition, blue color represents variables that have a true effect on the outcome, whereas the red refers to variables with no effect

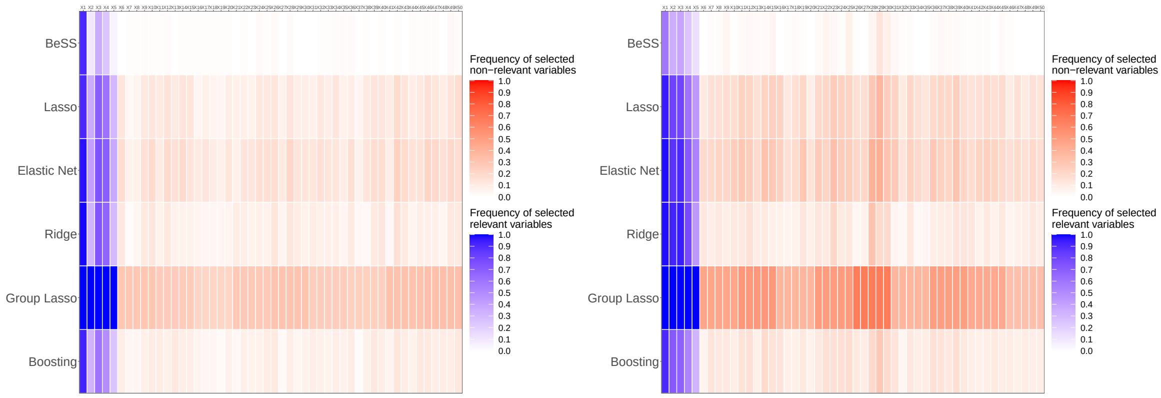

Figure 3: Selected variables within setting 3 in scenario 2 (left) and scenario 3 (right). The frequency of selection, by which each variable has been selected, is shown by color intensity. In addition, blue color represents variables that have a true effect on the outcome, whereas the red refers to variables with no effect

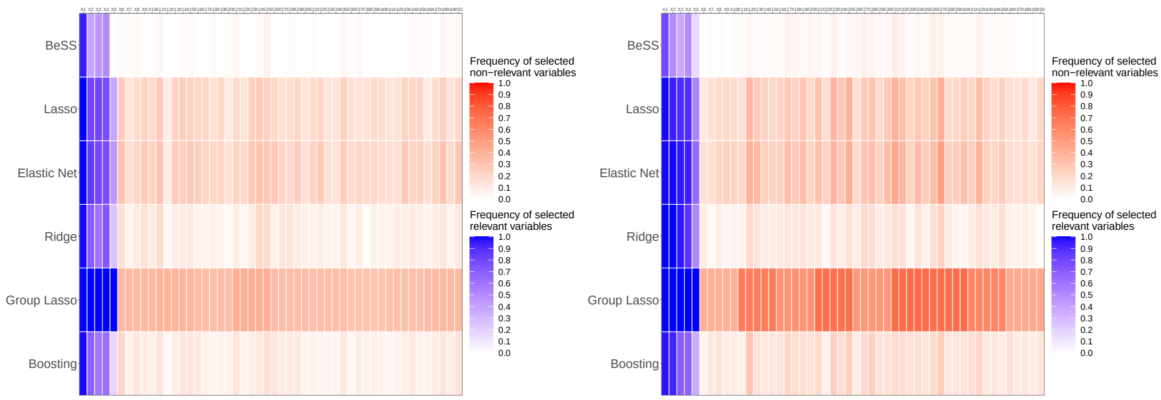

Figure 4: Selected variables within setting 4 in scenario 2 (left) and scenario 3 (right). The frequency of selection, by which each variable has been selected, is shown by color intensity. In addition, blue color represents variables that have a true effect on the outcome, whereas the red refers to variables with no effect

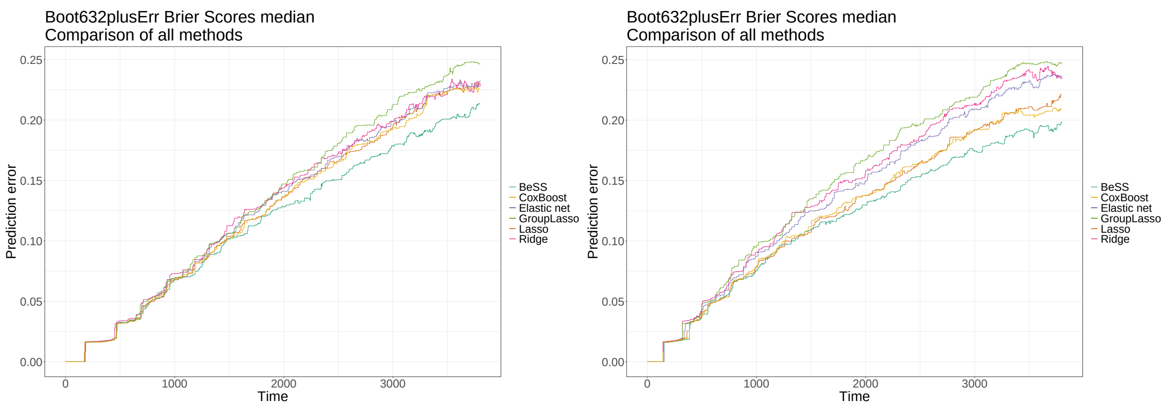

Figure 5: Mean curves of Bootstrap .632+ estimates of Brier Score curves, for frailty models based on BeSS, Lasso, Elastic net, Ridgre regression, Group Lasso and Cox Boosting selected variables, within setting 1 in scenario 2 (left) and scenario 3 (right)

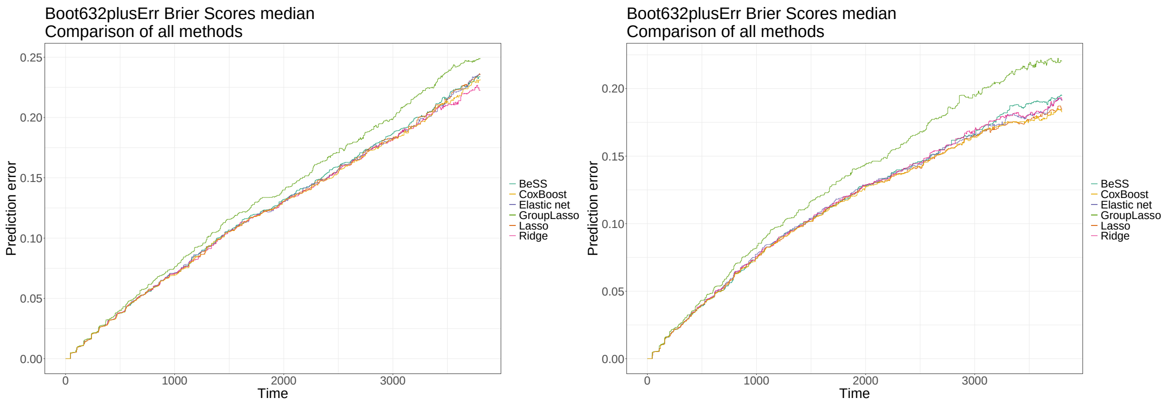

Figure 6: Mean curves of Bootstrap .632+ estimates of Brier Score curves, for frailty models based on BeSS, Lasso, Elastic net, Ridgre regression, Group Lasso and Cox Boosting selected variables, within setting 2 in scenario 2 (left) and scenario 3 (right)

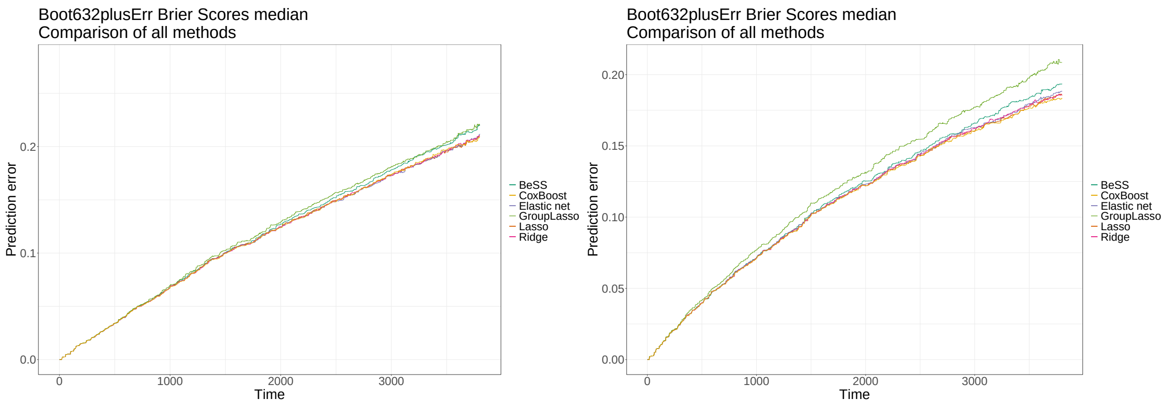

Figure 7: Mean curves of Bootstrap .632+ estimates of Brier Score curves, for frailty models based on BeSS, Lasso, Elastic net, Ridgre regression, Group Lasso and Cox Boosting selected variables, within setting 3 in scenario 2 (left) and scenario 3 (right)

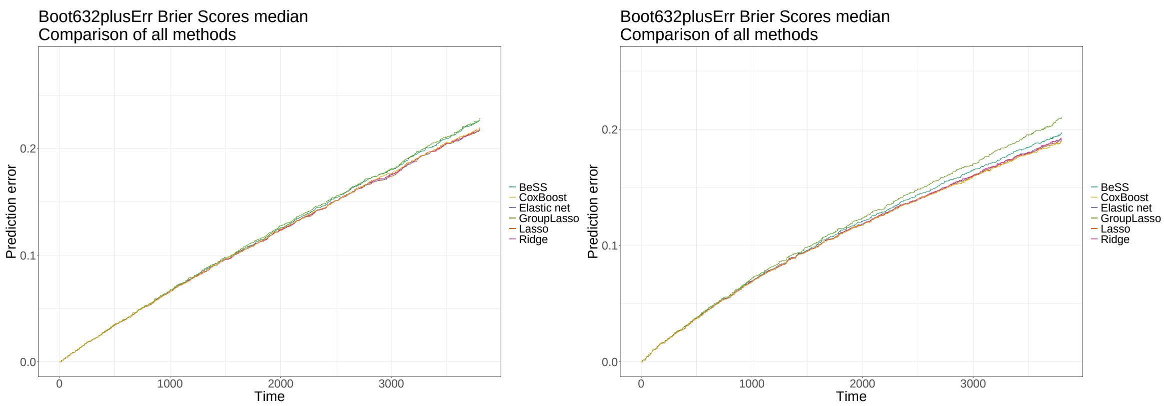

Figure 8: Mean curves of Bootstrap .632+ estimates of Brier Score curves, for frailty models based on BeSS, Lasso, Elastic net, Ridgre regression, Group Lasso and Cox Boosting selected variables, within setting 4 in scenario 2 (left) and scenario 3 (right)

References

Binder, H. and Schumacher, M. (2008). Adapting prediction error estimates for biased complexity selection in high-dimensional bootstrap samples. *Statistical Applications in Genetics and Molecular Biology*, 7(1).

Efron, B. (1983). Estimating the error rate of a prediction rule: improvement on cross-validation. *Journal of the American statistical association*, 78(382):316–331.

Efron, B. and Tibshirani, R. (1997). Improvements on cross-validation: the 632+ bootstrap method. *Journal of the American Statistical Association*, 92(438):548–560.

Gerds, T. A. (2020). *Prediction Error Curves for Risk Prediction Models in Survival Analysis*. R package version 2020.11.17.

Mogensen, U. B., Ishwaran, H., and Gerds, T. A. (2012). Evaluating random forests for survival analysis using prediction error curves. *Journal of statistical software*, 50(11):1.