國立東華大學　應用數學系

An extension of the compound covariate prediction under the Cox proportional hazard models

Emura , Chen & Chen [2012, PLoS ONE 7(10)]

Takeshi Emura (NCU)

Joint work with Dr. Yi-Hau Chen and Dr. Hsuan-Yu Chen (Sinica)

Outline:

1) Review: Survival data with high-dimensional

        covariates

2) Review: Ridge regression

3) Review: Compound covariate (CC) method

4) Some theory on CC method

5) Extension of the CC method

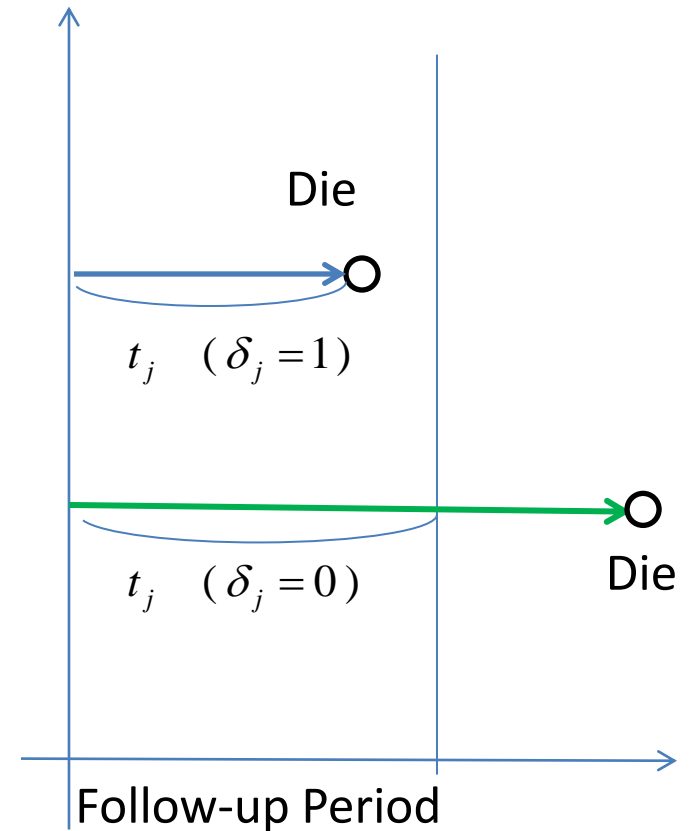 ➔ Propose a new method

6) Data example

- Survival data :

$\{ (t_i, \delta_i, \mathbf{x}_i); i = 1, ..., n \}$

t_i : either time to death or censoring

$$\delta_i = \begin{cases} 1 & \text{if death} \\ 0 & \text{if censoring} \end{cases}$$

$\mathbf{x}_i = (x_{i1}, ..., x_{ip})'$, possibly $p > n$

(Gene $\Leftrightarrow$ Covariate)

Die

$t_j \quad (\delta_j = 1)$

Die

$t_j \quad (\delta_j = 0)$

Follow-up Period

Lung cancer data from Chen et al. 2007 NEJM

```
Patient_ID              Survival_Status
100                     alive , 47 months   (censored)      $t_i = 47, \ \delta_i = 0$
101                     alive, 49 months    (censored)
102                     death, 20 months                    $t_i = 20, \ \delta_i = 1$
109                     death, 26
110                     alive, 39          (censored)
113                     alive, 35         (censored)
115                     alive, 45         (censored)
116                     death, 9
128                     death, 21
.
.
.
.
.
365                     alive, 5 months       (censored)
```

n=125 samples

1ˢᵗ Patient (ID = 100)

- Gene : $\mathbf{x}_i = (x_{i1}, \ldots, x_{i672})'$

P=672 covariates >> n = 125

ID_REF	SLOG TRANFORMED VALUE
1	15.27004532
2	13.17203115
3	14.21802644
4	15.12513123
5	13.20893358
6	14.8388795
7	13.8996511
8	13.93310453
9	14.4358955
10	13.94191912
11	14.80745797
12	13.73624082
13	13.07752608
666	14.63251884
667	14.53994587
668	14.60524106
669	14.48299068
670	11.55074679
671	11.55074679
672	11.55074679

- Genetic information is useful in survival prediction:

Breast cancer:

(Jenssen et al., 2002; van de Vijver et al., 2002;
  van't Veer et al., 2002; Zhao et al., 2011)

Lung cancer:

(Beer et al., 2002; Chen et al., 2007; Shedden et al., 2008)

- Demand for statistical prediction methods adapted to high-dimensionality, especially on recent 10 years

Hazard function:

$$h(t \mid \mathbf{x}_i) = \Pr(t \leq t_i \leq t + dt \mid t_i \geq t, \mathbf{x}_i) / dt$$

- Cox proportional hazard model (Cox 1972)

$$h(t \mid \mathbf{x}_i) = h_0(t) \exp(\boldsymbol{\beta}' \mathbf{x}_i), \quad \boldsymbol{\beta} \in \mathbf{R}^p$$

•Partial likelihood estimator:

$$\hat{\boldsymbol{\beta}} \in \mathbf{R}^p : \quad \text{maximize} \quad L_n^1(\boldsymbol{\beta}) = \prod_{i=1}^{n} \left(\frac{\exp(\boldsymbol{\beta}' \mathbf{x}_i)}{\sum_{t_l \leq t_i} \exp(\boldsymbol{\beta}' \mathbf{x}_l)} \right)^{\delta_i}$$

If $n \to \infty$, then $\hat{\boldsymbol{\beta}} \to_P \boldsymbol{\beta}_0$

If $p > n$, $\hat{\boldsymbol{\beta}} \in \mathbf{R}^p$ is not unique (infinitely many maxima)

Idea of *Ridge* estimator (Hoerl & Kennard 1970)

Infinitely many solution in OLS

$$\hat{\boldsymbol{\beta}} = \text{argmin} \|y - X\boldsymbol{\beta}\|^2$$

Unique nearest point to 0

0

Bayesian Interpretation :
Hastile, Tibshirani, Friedman (2009)
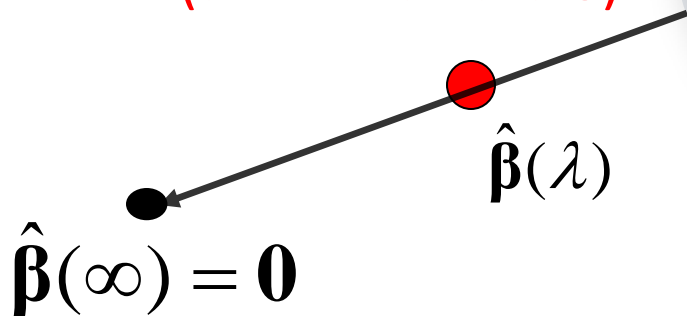The Element of Statistical Learning

- Penalized partial likelihood
(Verweij & Houwelingen 1994)

$$\hat{\boldsymbol{\beta}}(\lambda): \quad L_{\lambda}^{Ridge}(\boldsymbol{\beta}) = \prod_{i=1}^{n} \left(\frac{\exp(\boldsymbol{\beta}'\mathbf{x}_i)}{\sum_{l \in R_i} \exp(\boldsymbol{\beta}'\mathbf{x}_l)} \right)^{\delta_i} - \lambda \parallel \boldsymbol{\beta} \parallel^2$$

(Ridge estimator)

Ridge estimator
(shrink toward **0**)

$$\hat{\boldsymbol{\beta}}(0): \quad \mathbf{U}(\boldsymbol{\beta}) = \frac{\partial}{\partial \boldsymbol{\beta}} \log L_n^1(\boldsymbol{\beta}) = \mathbf{0}$$

$$\hat{\boldsymbol{\beta}}(\lambda)$$

$$\hat{\boldsymbol{\beta}}(\infty) = \mathbf{0}$$

- Prediction by the Ridge estimator

$\mathbf{x}$:   Covariates for a future patient

$\hat{\boldsymbol{\beta}}'(\lambda)\mathbf{x} < c$ (Low hazard $-->$ Good prognosis)

$\hat{\boldsymbol{\beta}}'(\lambda)\mathbf{x} > c$ (High hazard $-->$ Poor prognosis)

where c is the cut-off value

Existing methods for high-dimensional survival data

- Ridge regression (Cox-regression with L_2 penalty)

 Verve & van Howelingen(1994 Stat. Med.), Zhao et al. (2011 PLoS ONE)

- Lasso (Cox-regression with L_1 penalty)

 Gui & Li (2005 Bioinformatics), Segal (2006 Biostatistics)

- Gene selection via univariate Cox-regression

 Jenssen et al. (2002 Nature Med.), Chen et al. (2007 NEJM), name but a few

- Others (Principal component, partial lease square, etc.)

Among above methods, ridge regression has the best performance in terms of survival prediction
(Bovelstad et al., 2007; van Weieringen e al., 2009; Bovelstad and Borgan, 2011)

Objectives of our study:

Propose a new prediction method that generalizes the so-called *compound covariate prediction*

Compound covariate prediction method

*Previously used in medical context

Tukey (1993 Controlled Clinical Trial), Beer et al. (2002 Nature Med.)
Chen et al. (2007 NEJM), Radamacher et al (2002 J. of Theoretical Bio.)
Matsui (2006 BMC Bioinformatics)

* Less known in statistical literature

Compound covariate prediction

Step1: For each gene $j(=1,\ldots,p)$, fit a univariate Cox model

$$\Pr(t \le t_i \le t + dt \mid t_i \ge t, x_{ij})/dt = h_{0j}(t)\exp(\beta_j x_{ij})$$

Step2: A set of p regression coefficients

$$\hat{\boldsymbol{\beta}}(0) = (\hat{\beta}_1,\ldots,\hat{\beta}_p)', \quad \text{where } \hat{\beta}_j = \arg\max \prod_{i=1}^{n}\left(\frac{\exp(\beta_j x_{ij})}{\sum_{t_l \ge t_i}\exp(\beta_j x_{lj})}\right)^{\delta_i}$$

Remark: This is possible even when *p > n*

Step 3: *Compound covariate prediction*

For a future patient with genes $\mathbf{x} = (x_1,\ldots,x_p)'$,

$\hat{\boldsymbol{\beta}}'(0)\mathbf{x} < c$ (Good prognosis) ; $\hat{\boldsymbol{\beta}}'(0)\mathbf{x} > c$ (Poor prognosis)

Compound covariate method

- is a univariate method to resolve the high dimensionality

- empirically perform well in microarray studies

To motivate our proposal, we first give some theoretical results on the compound covariate method

- **Assumption**: The Cox model holds with

$$h(t \mid \mathbf{x}_i) = h_0(t) \exp(\boldsymbol{\beta}' \mathbf{x}_i) = h_0(t) \exp(\beta_1 x_{i1} + \cdots + \beta_p x_{ip})$$

at the true parameter $\boldsymbol{\beta}' = \boldsymbol{\beta}_0' = (\beta_{0,1}, \ldots, \beta_{0,p}) \neq \mathbf{0}$

- **Remark:** Under the multivariate Cox model assumption, the *univariate Cox model does not hold*, i.e,

$$h(t \mid x_{ij}) = -\frac{\partial}{\partial t} \log E[\exp\{-H_0(t) \exp(\boldsymbol{\beta}' \mathbf{x}_i)\} \mid x_{i1}]$$
$$\neq h_{0j}(t) \exp(\beta_j x_{ij}).$$

- Univariate Cox model for each gene $j(=1,\ldots,p)$

$$\Pr(t \leq t_i \leq t + dt \mid t_i \geq t, x_{ij}) / dt = h_{0j}(t)\exp(\beta_j x_{ij})$$

is a misspecified model (a working model)

Ref:

Struthers & Kalbfleisch (1986) Misspecified proportional hazard models, Biometrika 73 pp.363-9.

- Univariate partial likelihood equation

$$\hat{\beta}_j: \quad \text{Solution to} \quad 0 = U_j(\beta_j) = \frac{1}{n}\sum_{i=1}^{n}\delta_i\left\{ x_{ij} - \frac{\sum_{\ell=1}^{n} I(t_\ell \geq t_i) x_{\ell j} \exp(\beta_j x_{\ell j})}{\sum_{\ell=1}^{n} I(t_\ell \geq t_i)\exp(\beta_j x_{\ell j})} \right\}$$

$$\beta_j^* \quad \text{Solution to} \quad 0 = u_j(\beta_j) \xleftarrow{\ P\ } U_j(\beta_j)$$

$$\hat{\beta}_j \xrightarrow{\ P\ } \beta_j^* \neq \beta_{0j} \quad \text{(true value in the Assumption)}$$

Remark I: If all genes $\mathbf{x} = (x_1, \ldots, x_p)'$ **are independent**

$$\text{sign}(\beta_j^*) = \text{sign}(\beta_{0j}), \qquad |\beta_j^*| \leq |\beta_{0j}|$$

Remark II:

Let $\boldsymbol{\beta}^*(0) = (\beta_1^*, \ldots, \beta_p^*)'$ and $\mathbf{0} = (0, \ldots, 0)'$.

Then, $\boldsymbol{\beta}^*(0)$ is between $\boldsymbol{\beta}_0$ and $\mathbf{0}$.

Above results deduced from :

Struthers & Kalbfleisch (1986 Biometrika) ; Bretagnolle & Huber-Carol(1988 Scand. JS)

Proposed estimator

- Univariate *compound* likelihood (unique maxima)

$$L_n^0(\boldsymbol{\beta}) = \prod_{j=1}^{p} \prod_{i=1}^{n} \left(\frac{\exp(\beta_j x_{ij})}{\sum_{t_l \geq t_i} \exp(\beta_j x_{lj})} \right)^{\delta_i}$$

- Multivariate likelihood (infinitely many maxima when $p > n$)

$$L_n^1(\boldsymbol{\beta}) = \prod_{i=1}^{n} \left(\frac{\exp(\boldsymbol{\beta}' \mathbf{x}_i)}{\sum_{l \in R_i} \exp(\boldsymbol{\beta}' \mathbf{x}_l)} \right)^{\delta_i}$$
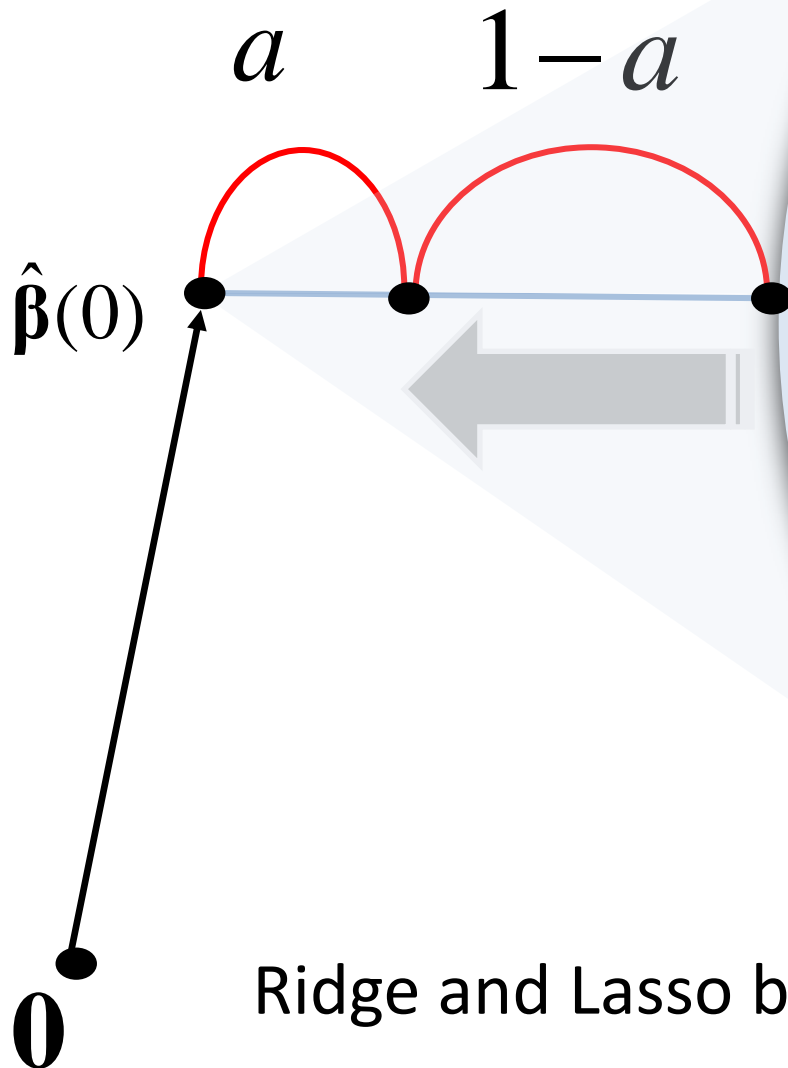
- Idea: Mixture of univariate and multivariate likelihood

$$\hat{\boldsymbol{\beta}}(a) = \operatorname{argmax} \left\{ a \log L_n^1(\boldsymbol{\beta}) + (1-a) \log L_n^0(\boldsymbol{\beta}) \right\}, \quad a \in [0, 1]$$

We call it "compound shrinkage estimator"

Compound shrinkage estimator :

$$\hat{\boldsymbol{\beta}}(a) = \operatorname{argmax}\left\{ a \log L_n^1(\boldsymbol{\beta}) + (1-a)\log L_n^0(\boldsymbol{\beta}) \right\}$$

a $1-a$

$\hat{\boldsymbol{\beta}}(0)$

$\bullet\ \boldsymbol{\beta}_0$ （true）

Infinitely many solutions

for a multivariate Cox regression

$$\hat{\boldsymbol{\beta}}:\quad \mathbf{U}(\boldsymbol{\beta}) = \frac{\partial}{\partial \boldsymbol{\beta}}\log L_n^1(\boldsymbol{\beta}) = \mathbf{0}$$

$\mathbf{0}$

Ridge and Lasso both shrink toward zero

- Proposition 2: (in our paper)

$$\sqrt{n}\,(\hat{\boldsymbol{\beta}}(\hat{a}) - \boldsymbol{\beta}_0) \to N(\mathbf{0}, \boldsymbol{\Sigma}(\boldsymbol{\beta}_0)) \text{ with } \hat{a} = \operatorname{argmax} CV(a).$$

(CV = Cross-Validated likelihood of Verveij & Houwelingen 1993)

- Plug-in variance estimator $\boldsymbol{\Sigma}_n^{\hat{a}}(\hat{\boldsymbol{\beta}}(\hat{a}))$

$$\boldsymbol{\Sigma}_n^a(\boldsymbol{\beta}) = \mathbf{A}_n^a(\boldsymbol{\beta})\{\mathbf{V}_n^a(\boldsymbol{\beta})/n\}^{-1}\mathbf{A}_n^a(\boldsymbol{\beta})'$$

$$\mathbf{A}_n^a(\boldsymbol{\beta}) = \mathbf{V}_n^a(\boldsymbol{\beta})^{-1}\dot{\mathbf{h}}_n(\boldsymbol{\beta})\{-d^2 CV(a)/da^2\}^{-1}\dot{\mathbf{h}}_n(\boldsymbol{\beta})' + \mathbf{I}_p$$

$$\dot{h}_n(\boldsymbol{\beta}) = \partial \mathbf{U}_n^a(\boldsymbol{\beta})/\partial a, \text{ where } \mathbf{U}_n^a(\boldsymbol{\beta}) = \text{Score function}$$

$$\frac{d}{da}CV(a) = \text{Estimating function of } a,$$

$$\mathbf{V}_n^a(\boldsymbol{\beta}) = \text{observed Fisher information}$$

*Reasonable performance even when $p > n$.

Numerical comparison

Compare 4 methods

1. Compound covariate (CC) estimator

$$\hat{\boldsymbol{\beta}} = (\hat{\beta}_1, ..., \hat{\beta}_p)', \quad \text{where} \quad \hat{\beta}_j = \text{univariate Cox regression estimators}$$

2. Compound shrinkage (CS) estimator

$$\hat{\boldsymbol{\beta}}(\hat{a}): \quad a \log L_n^1(\boldsymbol{\beta}) + (1-a) \log L_n^0(\boldsymbol{\beta})$$

←R compound.Cox package
  (Emura and Chen, 2012)

3. Ridge estimator

$$\hat{\boldsymbol{\beta}}(\hat{\lambda}): \quad \log L_n^1(\boldsymbol{\beta}) - (\lambda/2) \sum_{j=1}^{p} \beta_j^2$$

←R penalized package
  (Goeman et al., 2012)

4. Lasso estimator

$$\hat{\boldsymbol{\beta}}(\hat{\lambda}): \quad \log L_n^1(\boldsymbol{\beta}) - \lambda \sum_{j=1}^{p} | \beta_j |$$

← R penalized package

* $\hat{a}$ or $\hat{\lambda}$ is obtained by cross-validation (Verveij & Houwelingen 1993 Stat.Med.)

- Data: Lung cancer data (Chen et al., 2007 NEJM)

n=125, p=672

n=63 , *p*=97
Training set

$$\hat{\boldsymbol{\beta}}' = \begin{cases} \text{compound covariate} \\ \text{compound shrinkage} \\ \text{Ridge} \\ \text{Lasso} \end{cases}$$

Predict

n=62 , *p*=97
test set

$\{\ \mathbf{x}_i \quad i = 1,...,62\ \}$

Good prognosis   Poor prognosis
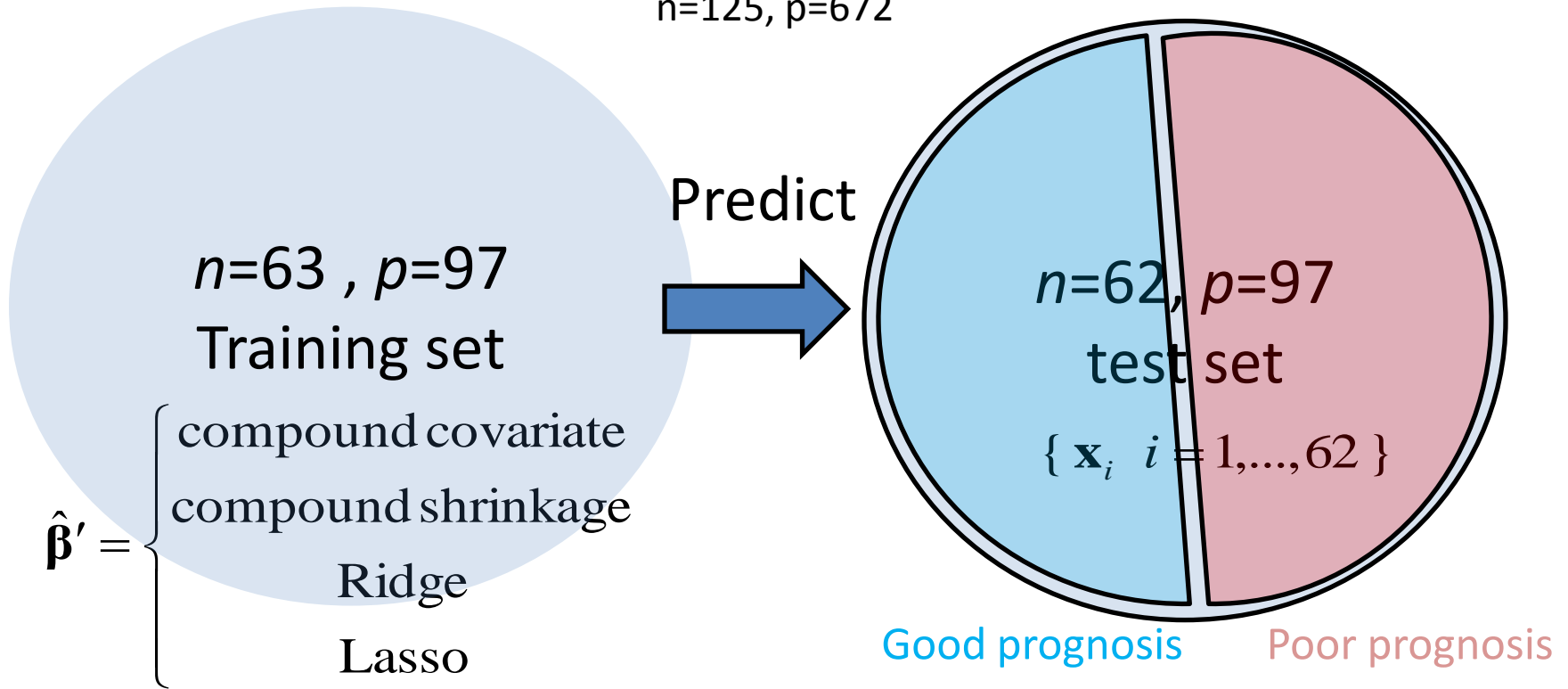
$$\hat{\boldsymbol{\beta}}'\mathbf{x}_i < c \ \text{(Good prognosis)} \ ; \ \ \hat{\boldsymbol{\beta}}'\mathbf{x}_i > c \ \text{(Poor prognosis)},$$

$$\text{where } c \ \text{is the median of } \{\ \hat{\boldsymbol{\beta}}'\mathbf{x}_i, i = 1,...,n\ \}$$

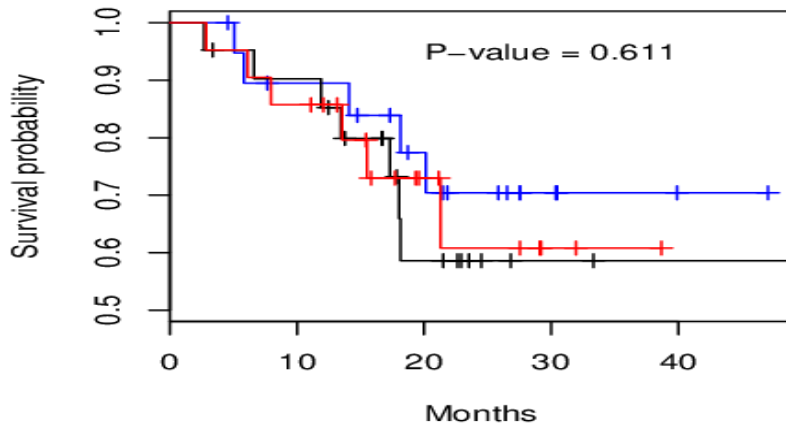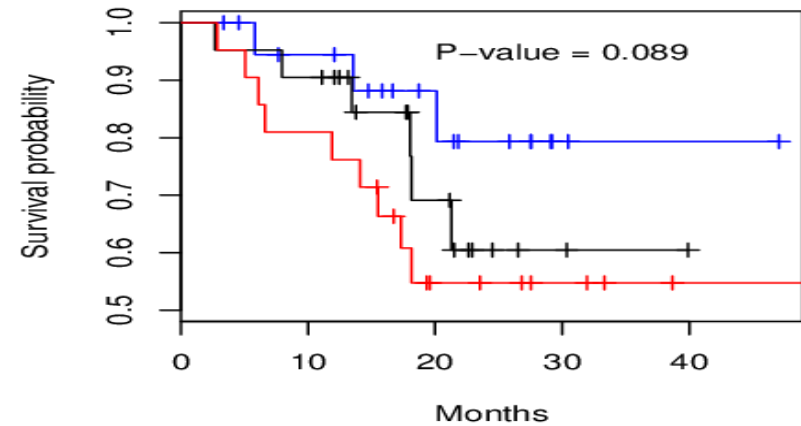# Survival curves for Poor vs. Good prognosis groups in n=62 testing data; p-value for Log-rank test
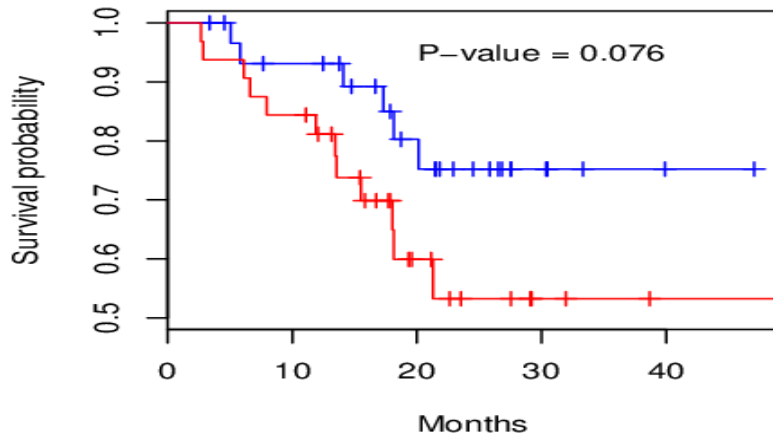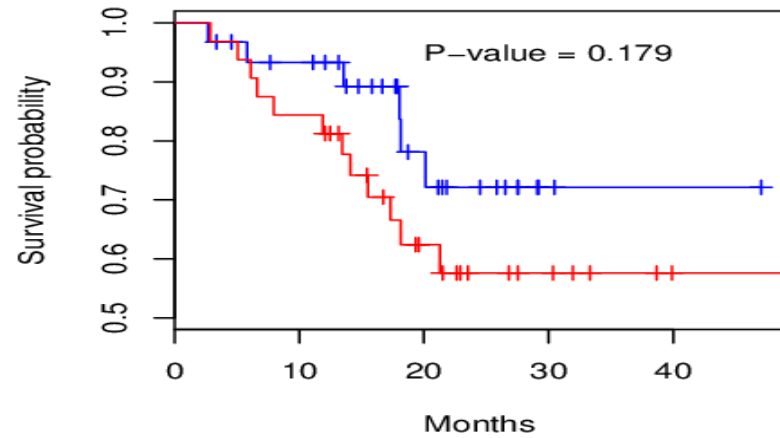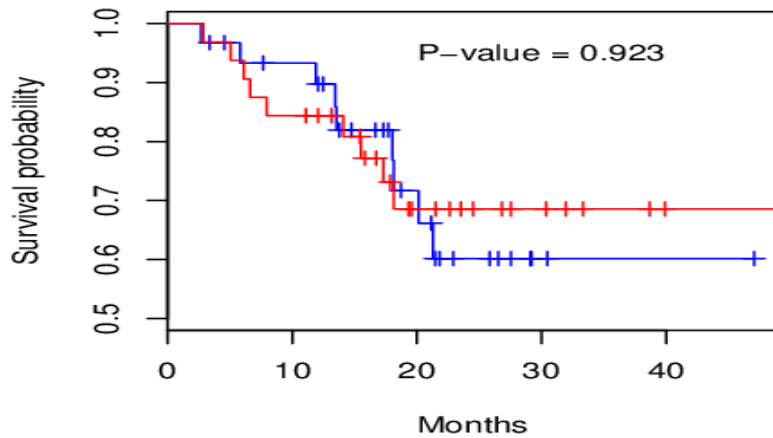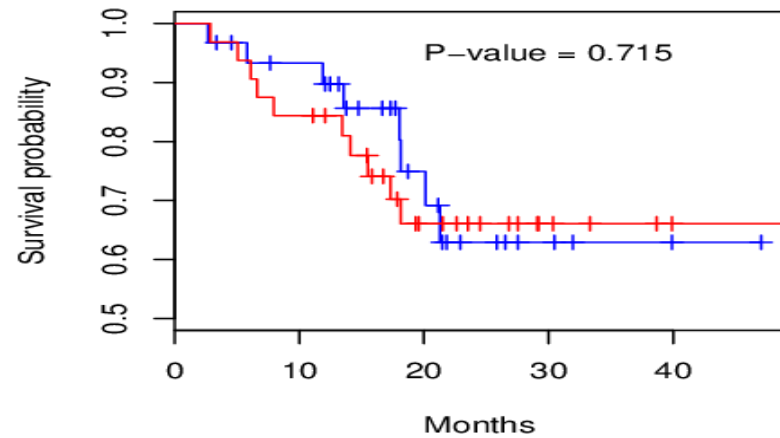
Survival curves for Poor, Medium, Good prognosis groups for n=62 testing data; p-value for Log-rank trend test

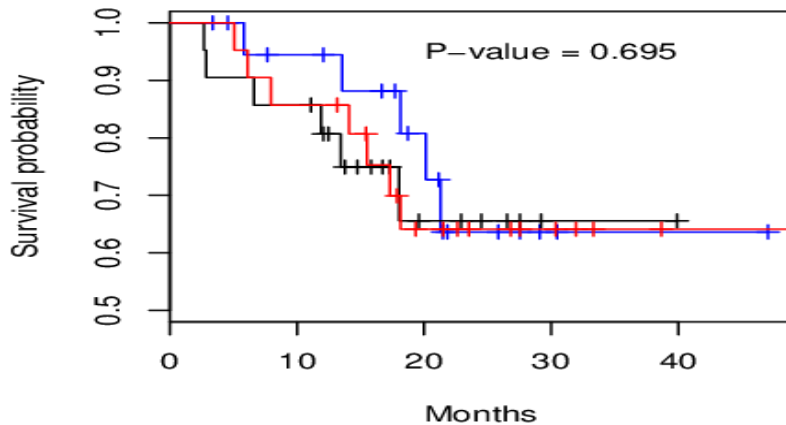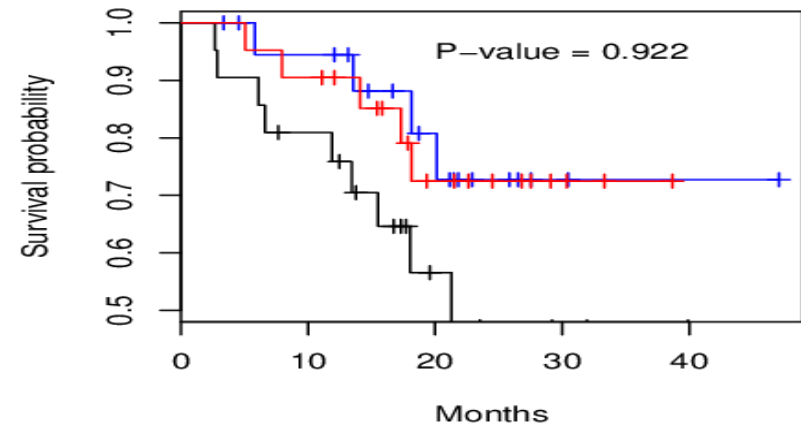**Thank you for your attention**