International Biometric Conference 2012/8/31, Kobe JAPAN

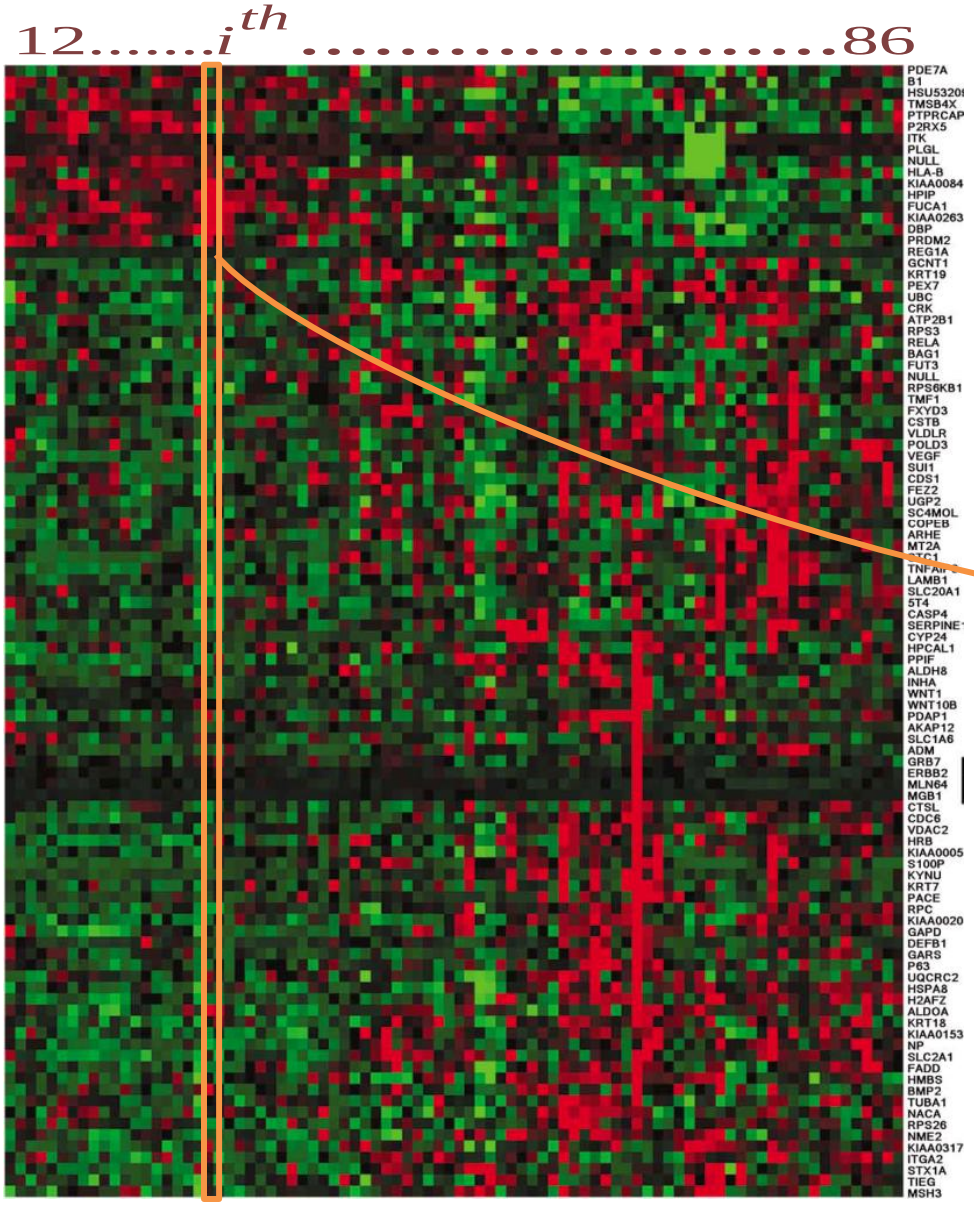
Survival Prediction Based on Compound Covariate under Cox Proportional Hazard Models

PLoS ONE 7(10). doi:10.1371/journal.pone.0047627.
http://dx.plos.org/10.1371/journal.pone.0047627)

Takeshi Emura
Graduate Institute of Statistics,
National Central University, Taiwan

Joint work with Yi-Hau Chen and Dr. Hsuan-Yu Chen
Institute of Statistical Science, Academia Sinica, Taiwan

Survival Data with Microarrays



$p = 100(\text{genes})$

i^{th} patient :

$x_i = (x_{i1}, \ldots, x_{i100})$

t_i = Survival time

δ_i = Censoring indicator

(Lung cancer data from Beer et al., 2002)

Existing methods for high-dimensional survival data

- Lasso (Cox-regression with L_1 penalty)

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

- Ridge regression (Cox-regression with L_2 penalty)

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

- Gene selection via univariate Cox-regression

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

- Others (PC, supervised PC, 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)

Two objectives of our study:

1. Revive *compound covariate prediction method*

   *Previously used in microarrays datasets

   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)

   *But, no theoretical analysis and comparative studies have not yet reported

2. Propose to refine the compound covariate prediction via *shrinkage* technique

Set up

- Survival data :

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

t_i : either time to death or censoring

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

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

Example:

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

  n=125, p=672, Censored proportion = 70%

  ➔ Data analysis (later)

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 $\mathbf{x} = (x_1, \ldots, x_p)'$,

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

Compound covariate method:

- A simple method to resolve the high dimensionality
- Its theoretical justification has not been discussed in the literature

- **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,

$$h(t \mid x_{ij}) = -\frac{\partial}{\partial t}\log S(t \mid x_{ij})$$

$$= -\frac{\partial}{\partial t}\log E[\exp\{-H_0(t)\exp(\boldsymbol{\beta}'\mathbf{x}_i)\} \mid x_{i1}]$$

$$\not\propto \exp(\beta_j x_{ij}).$$

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

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

is a misspecified model (a working model)

Ref:

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

- Univariate partial likelihood equation

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

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

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

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

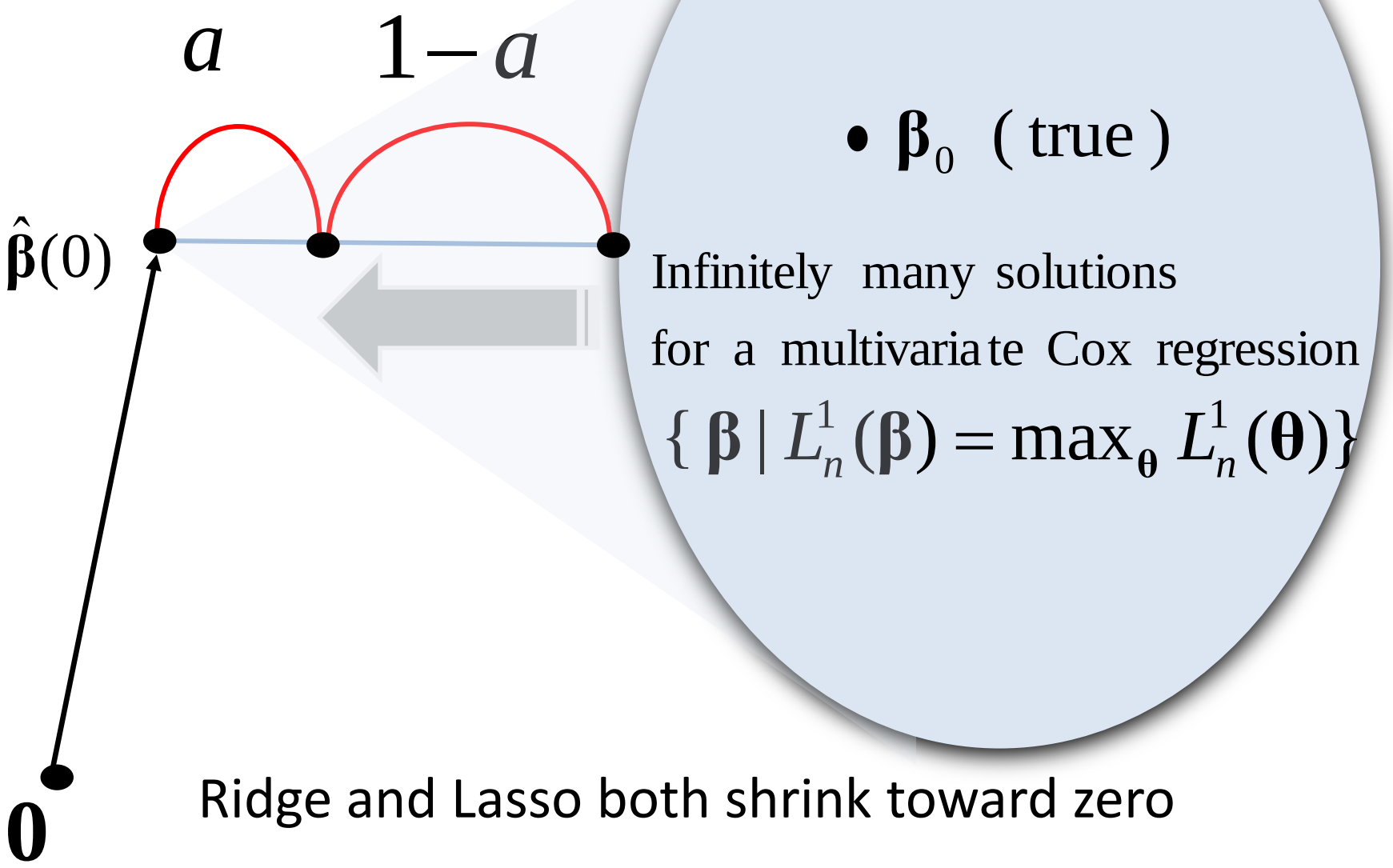
$$\hat{\boldsymbol{\beta}}(a) = \operatorname{argmax} \left\{ a \log L_n^1(\boldsymbol{\beta}) + (1-a) \log L_n^0(\boldsymbol{\beta}) \right\}, \quad a \in [0,1]$$

Special case $a = 0$: $\hat{\boldsymbol{\beta}}(0) =$ a set of p univeriate estimators

(call "compound covariate estimator")

Compound shrinkage estimator :

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



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 $\boldsymbol{\Sigma}_n^{\hat{a}}(\hat{\boldsymbol{\beta}}(\hat{a}))$

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

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

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

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

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

*Reasonable performance even when $p > n$.

Numerical comparison

$\hat{\boldsymbol{\beta}}$ is obtained by 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

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

3. Ridge estimator

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

4. Lasso estimator

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

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

Simulation set up

- Cox model: $h(t \mid \mathbf{x}_i) = \exp(\beta_1 x_{i,1} + \cdots + \beta_{100} x_{i,100})$
- Censoring: $U(0,1)$, moderate censoring (54~63%)
- Training set $\{ (t_i , \delta_i , \mathbf{x}_i); i = 1, \ldots, 100 \}$

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

R "compound.Cox" package
Emura & Chen (2012)

R "penalized" package
Goeman (2010)

- Testing set $\{ (t_i^* , \delta_i^* , \mathbf{x}_i^*); i = 1, \ldots, 100 \}$

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

P-value from a two-sample Log-rank test

(Smaller P-value corresponds to better prediction power)

- Evaluation criterion (Bovelstat et al. 2007 Bioinformatics):

Median P-value among 50 replications

Table 1. Simulation results under **sparse cases**.

CC = compound covariate, CS = compound shrinkage.

LR-test = Log_{10} P-value for discriminating poor / good patients.

Scenario 1: Tag gene / Scenario 2: Gene pathway

		$\boldsymbol{\beta} = (1.5, 1.5, \underbrace{0, ..., 0}_{\times 98})$			
		CC	CS	Ridge	Lasso
Scenario1	LR-test	-5.89	-5.88	-4.99	-10.59
Scenario2	LR-test	-8.88	-9.35	-7.01	-12.39
		$\boldsymbol{\beta} = (\underbrace{0.8, ..., 0.8}_{\times 5}, \underbrace{0, ..., 0}_{\times 95})$			
		CC	CS	Ridge	Lasso
Scenario1	LR-test	-3.88	-4.31	-4.21	-6.64
Scenario2	LR-test	-13.71	-13.69	-11.38	-14.52

Table 2. Simulation results under **Non-sparse cases**.

CC = compound covariate, CS = compound shrinkage.

LR-test = Log_{10} P-value for discriminating poor / good patients.

Scenario 1: Tag gene / Scenario 2: Gene pathway

		$\boldsymbol{\beta} = (\underbrace{0.2, \ldots, 0.2}_{\times 10}, \underbrace{-0.2, \ldots, -0.2}_{\times 10}, \underbrace{0, \ldots, 0}_{\times 80})$			
		CC	CS	Ridge	Lasso
Scenario1	LR-test	-1.22	-1.28	-1.29	-0.39
Scenario2	LR-test	-10.35	-9.49	-9.33	-9.11
		$\boldsymbol{\beta} = (\underbrace{0.1, \ldots, 0.1}_{\times 15}, \underbrace{-0.1, \ldots, -0.1}_{\times 15}, \underbrace{0, \ldots, 0}_{\times 70})$			
		CC	CS	Ridge	Lasso
Scenario1	LR-test	-0.55	-0.61	-0.61	-0.40
Scenario2	LR-test	-7.93	-6.80	-6.67	-6.05

Mostly, $\hat{\boldsymbol{\beta}}' = \mathbf{0}$ for Lasso

Simulation results: Summary

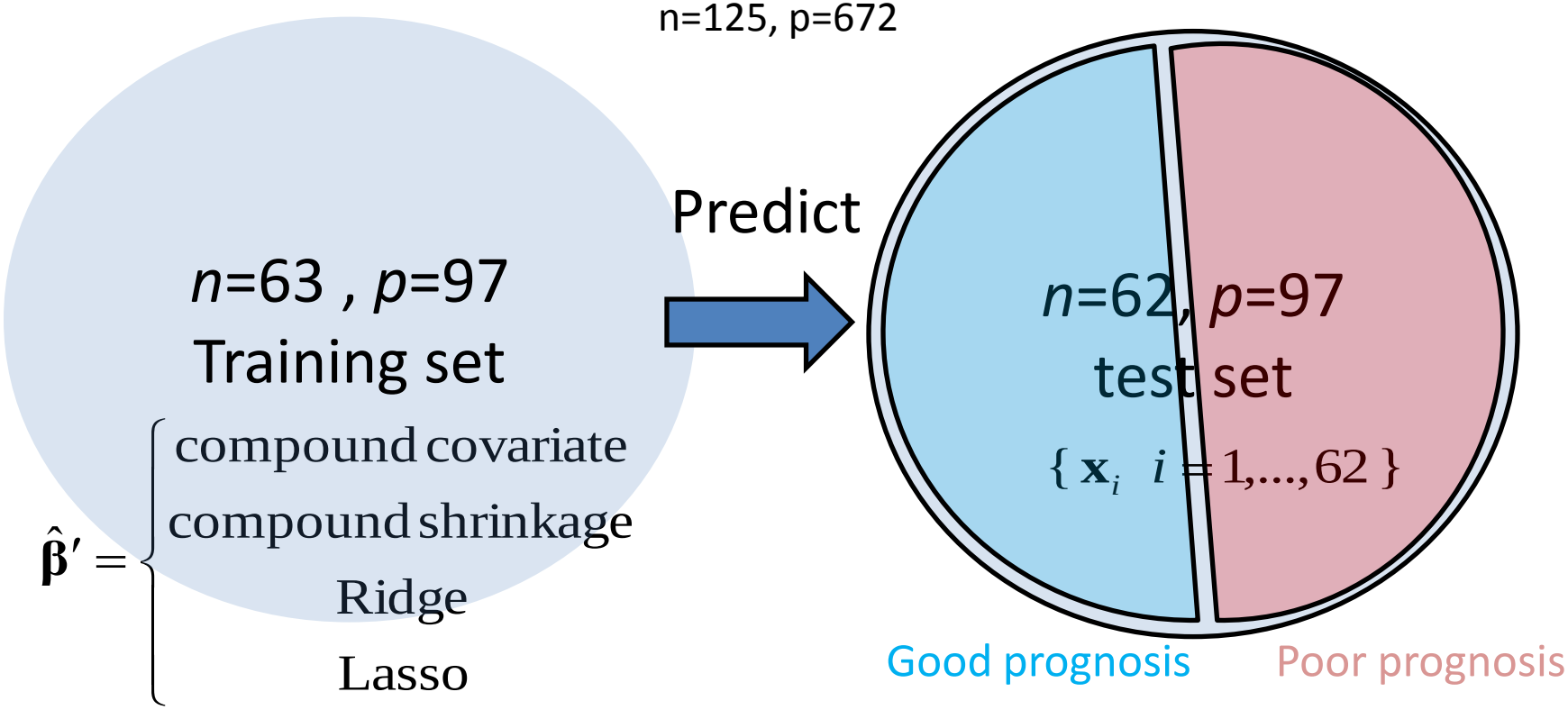
- Ridge is worst in sparse cases
- Lasso is worst in non-sparse cases
- Compound covariate and compound shrinkage performed similar to or slightly better than Ridge
- Since Ridge is reported as the best method in

  Bovelstad et al., 2007; van Weieringen e al., 2009; Bovelstad and Borgan, 2011

  the compound covariate and compound shrinkage are competitive methods

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

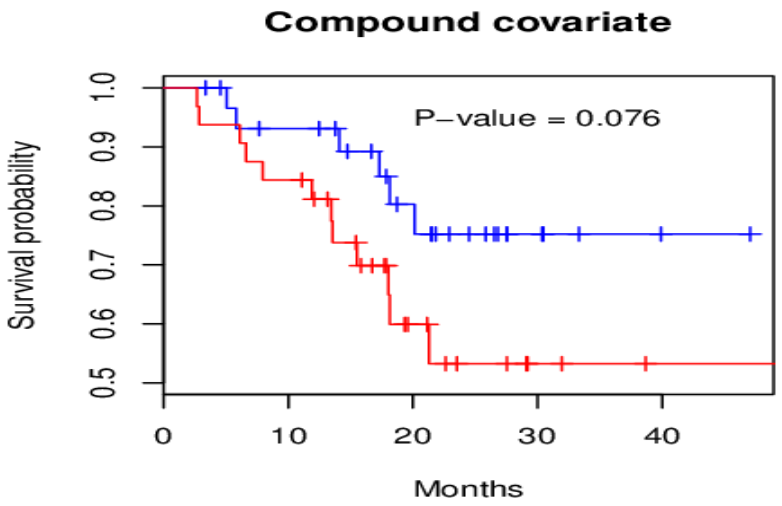
n=125, p=672

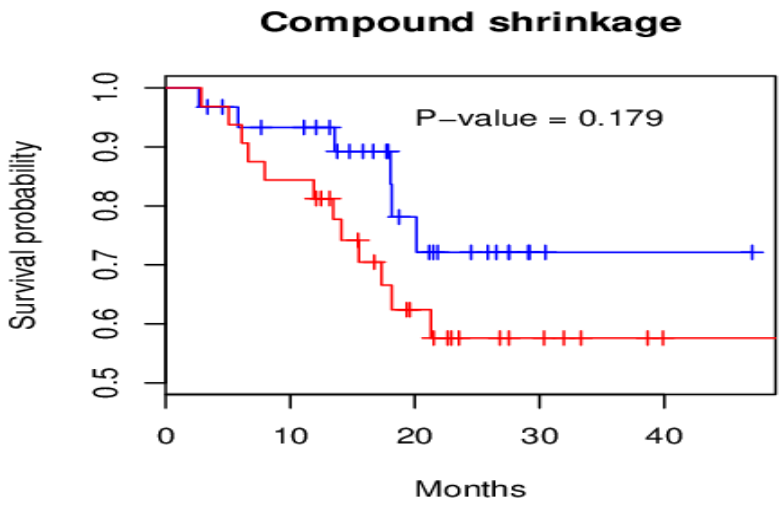


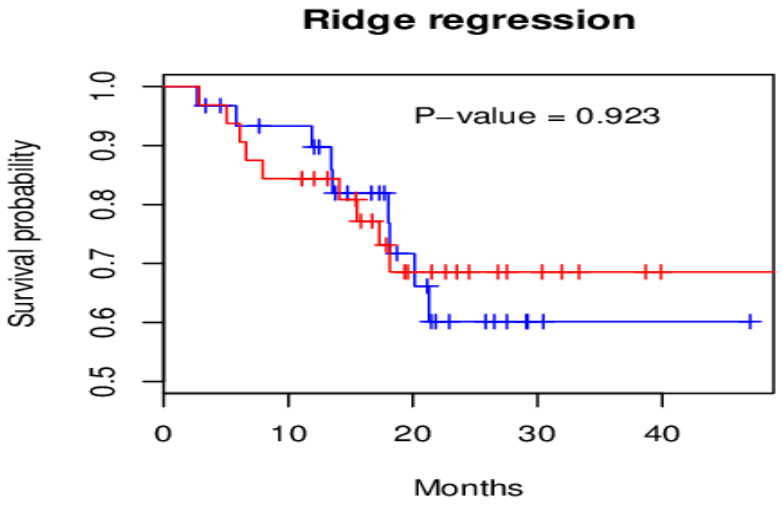
$$\hat{\boldsymbol{\beta}}'\mathbf{x}_i < c \ \text{(Good prognosis)} \ ; \quad \hat{\boldsymbol{\beta}}'\mathbf{x}_i > c \ \text{(Poor prognosis)},$$

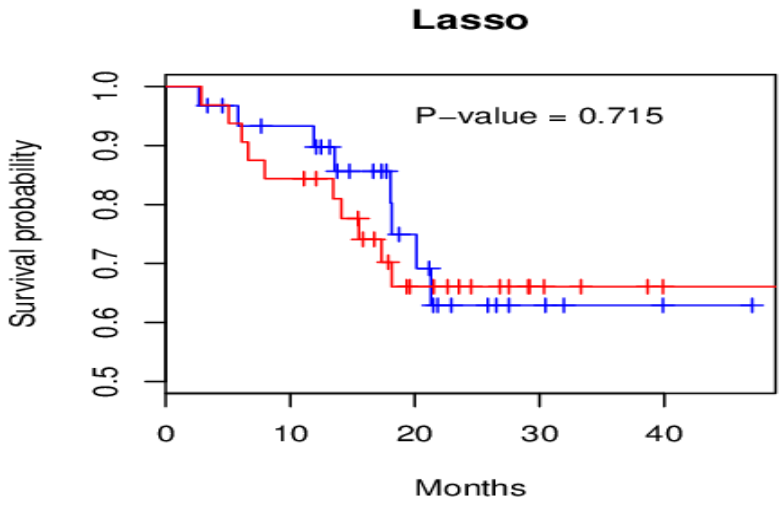
$$\text{where} \, c \ \text{is the median of} \ \{ \hat{\boldsymbol{\beta}}'\mathbf{x}_i, \mathrm{i} = 1,\ldots,\mathrm{n} \}$$

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

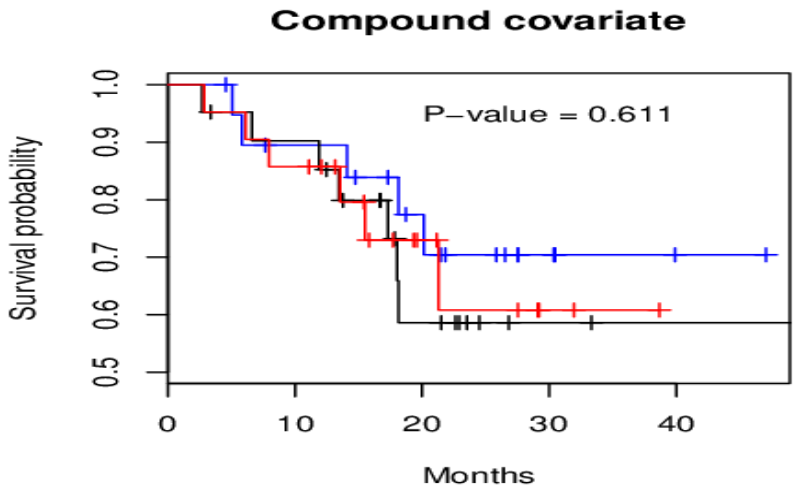


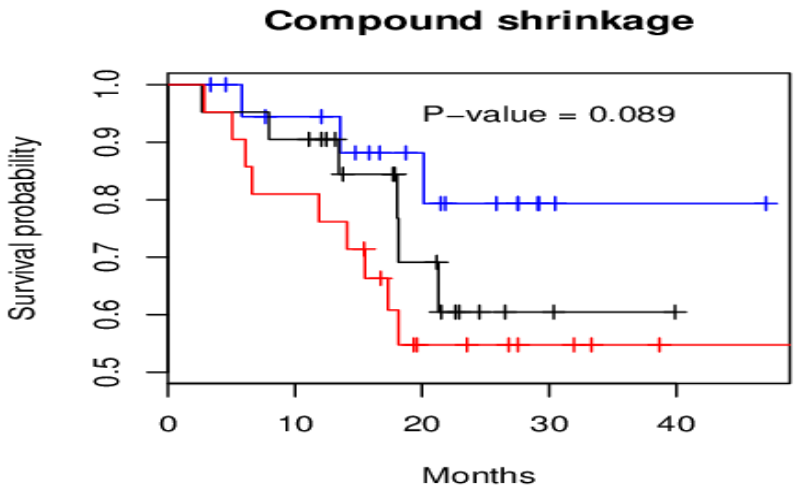


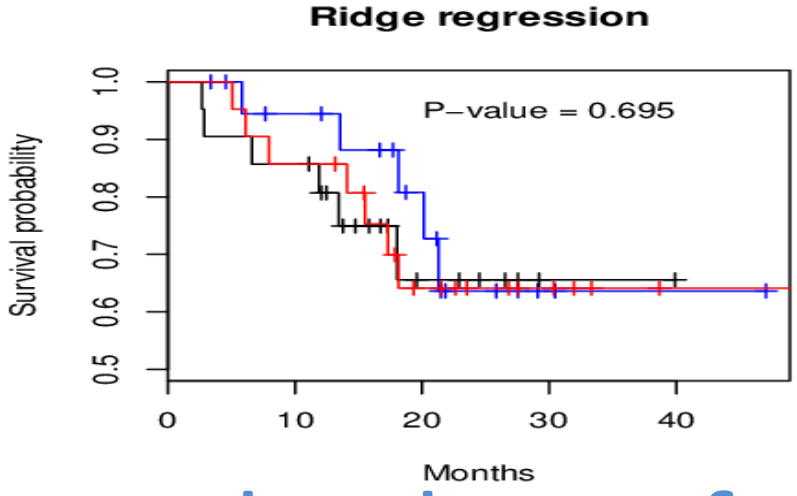


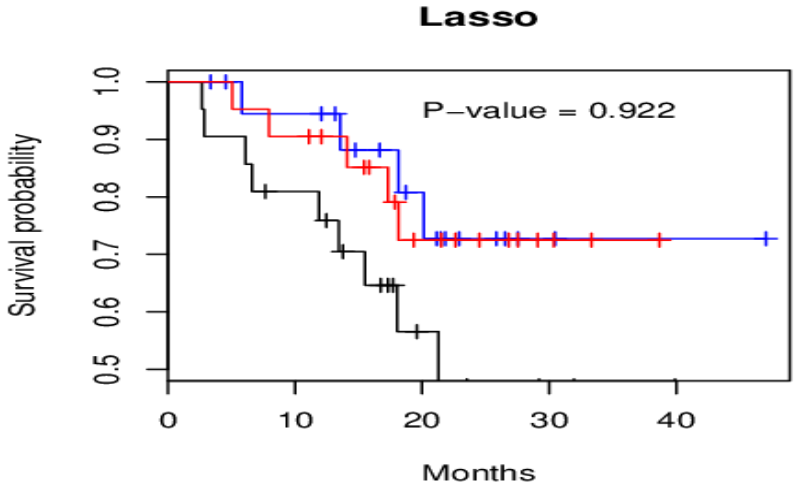


Survival curves for Poor, Medium, Good prognosis groups for n=62 testing data; p-value for Log-rank trend test









Thank you for your attention