7^{th} IASC-ARS, Joint 2011 Taipei Symposium

Predicting Survival Outcomes Based on Compound Covariate Method under Cox Proportional Hazard Models with Microarrays

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

Outlines

1. Intro: Survival analysis with microarrays
2. Existing methods (Ridge regression & Lasso)
3. A method known as *compound covariate prediction*

   (this method is not studied in sufficient details in literature)
4. Refinement of compound covariate method

   (Proposed method)
5. Comparison with existing methods (by real data)

Survival prediction

Clinical characteristics
(Age / Stage / Tumor type, etc.

Breast cancer patient

- 5 year survival probability
- Classification (High-risk / Low-risk)
  (criteria for chemotherapy)

Cox proportional hazard model: $h(t \mid \mathbf{x}_i) = h_0(t)\exp(\boldsymbol{\beta}'\mathbf{x}_i)$,

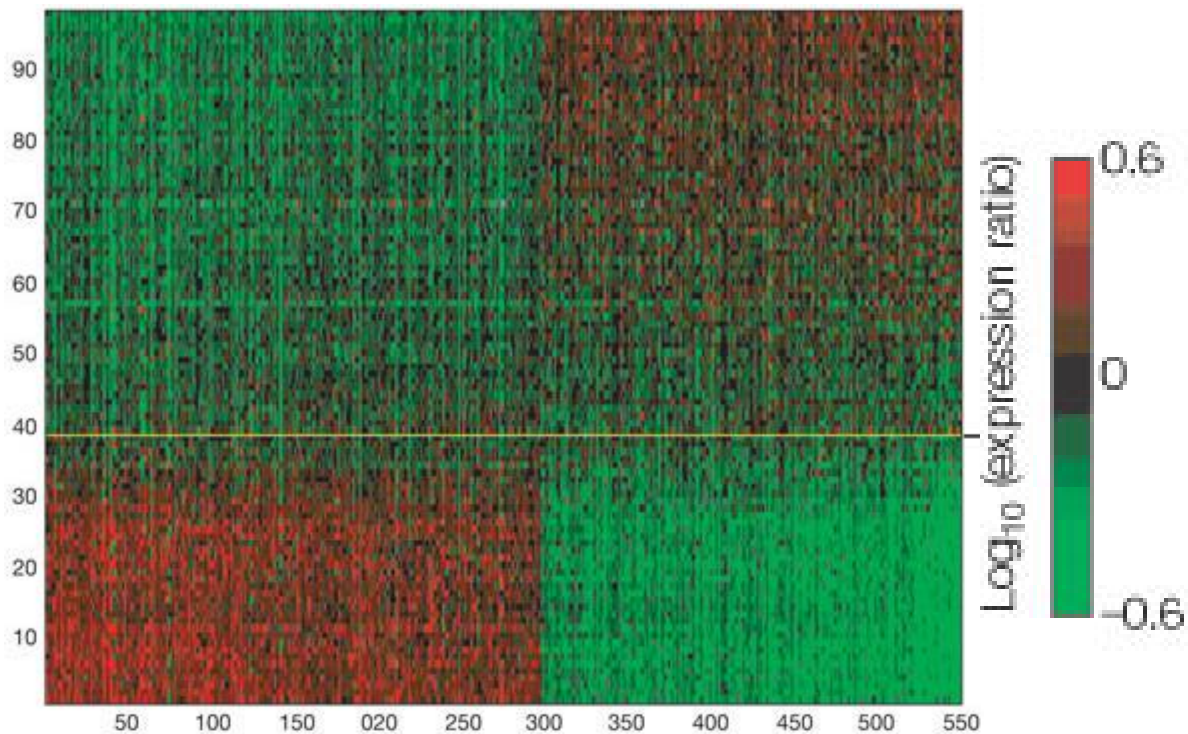
$\mathbf{x}_i = (\text{Age, Stage, Tumor type, etc.})$

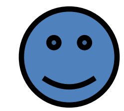
$\boldsymbol{\beta} = \hat{\boldsymbol{\beta}}$: Partial likelihood estimates from data

Classification : $\hat{\boldsymbol{\beta}}'\mathbf{x}_i < c$ (Low-risk) ; $\hat{\boldsymbol{\beta}}'\mathbf{x}_i > c$ (High-risk)

Survival prediction with microarrays

Microarrays (van't Veer et al., 2002 Nature)





Breast cancer patient

$$\mathbf{x}_i = (x_{i1}, \ldots, x_{i550}),$$
$$i = 1, 2, \ldots, 98$$

Microarrays is useful for predicting breast cancer patients

(Jensen et al., 2002; van't Veer et al., 2002; Vijiver et al., 2002; Zhao et al., 2011)

$$\text{Cox proportional hazard model:}\quad h(t \mid \mathbf{x}_i) = h_0(t)\exp(\boldsymbol{\beta}'\mathbf{x}_i),$$

$$\Rightarrow \text{ Difficult to get } \boldsymbol{\beta} = \hat{\boldsymbol{\beta}} \text{ due to high-dimensionality } (p >> n)$$

Available methods with microarray

- Lasso (Cox-regression with L_1 panalty)

  Tibshirani (1997), Gui & Li (2005), Segal (2006)

- Ridge regression (Cox-regression with L_2 penalty)

  Verveij & Howelingen(1994), Zhao et al. (2011)

- Univariate selection via Cox-regression

  Jenssen et al. (2002), Chen et al. (2007)

- Cluster analysis van't Veer et al., 2002; medical studies
- Others (PC, supervised PC, partial lease square, etc.)

Among many methods, ridge regression has the overall-best prediction power

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

Two objectives of our study:

1. Study ***compound covariate prediction*** (Tukey 1993)

   In survival data, compound covariate is empirically used:
   (Beer et al., 2002; Chen et al., 2007;
   Radamacher et al, 2002; Matsui, 2006)
   * But, less studied in the statistical literature
   * So, its comparative performance is unknown

2. Propose to refine a compound covariate prediction via ***Shrinkage*** technique

Set up

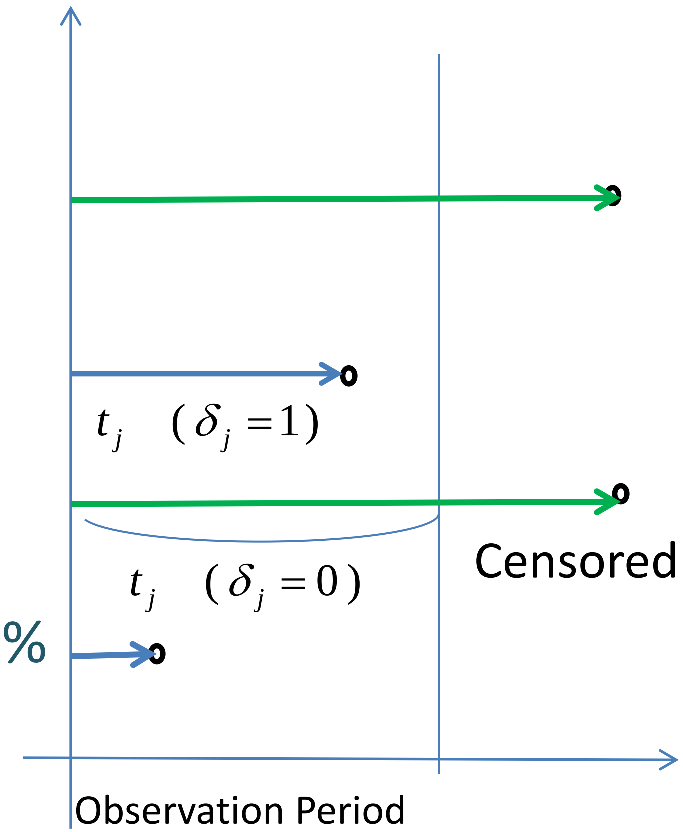
- <u>Survival data :</u>

$$\{(t_j, \delta_j, \mathbf{x}_i);\ j = 1, \ldots, n\}$$

t_j : either time to death or time to censoring

$\delta_j = 1$ if death, $\delta_j = 1$ if censoring

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



Example:

- Breast cancer data
  (van Houwlingen et al. 2006)
  n=295, p=4919, Censored proportion = 73%
- Lung cancer data (Chen et al., 2007)
  n=125, p=672, Censored proportion = 70%

➔➔ Data analysis (later)

Cox regression with *p>n*

Cox proportional hazard model:

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

•Partial likelihood

$$L_n^1(\boldsymbol{\beta}) = \prod_{i=1}^{n}\left(\frac{\exp(\boldsymbol{\beta}'\mathbf{x}_i)}{\sum_{l \in R_i}\exp(\boldsymbol{\beta}'\mathbf{x}_l)}\right)^{\delta_i}, \quad \text{where} \quad R_i = \{l : t_l \geq t_i\}$$

If $p > n$, the maximum is not unique

•Penalized partial likelihood (well-known methods)

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

(Lasso) (Cox - Ridge regression)

Even if $p > n$, the maximum is unique.

($\lambda > 0$ is determined by cross-validation, Verveij & Houwelingen, 1993)

Compound covariate prediction

- Univariate Cox regression

$$\Pr(t \le t_i \le t + dt \mid t_i \ge t, x_{ij}) / dt = h_0(t) \exp(\beta_j x_{ij}) \quad \text{for} \quad j = 1, \ldots, p$$

- *A collection of p* univariate likelihood estimators

$$\hat{\boldsymbol{\beta}}(0) = (\hat{\beta}_1, \ldots, \hat{\beta}_p)' \quad \text{where} \quad \hat{\beta}_j = \arg\max_{\beta_j} L^0_{n,j}(\beta_j),$$

$$\text{and where} \quad L^0_{n,j}(\beta_j) = \prod_{i=1}^{n} \left(\frac{\exp(\beta_j x_{ij})}{\sum_{l \in R_i} \exp(\beta_j x_{lj})} \right)^{\delta_i} \quad \text{and} \quad R_i = \{ l : t_l \ge t_i \}$$

- *Compound covariate prediction*

$$\hat{\boldsymbol{\beta}}'(0) \mathbf{x}_i < c \ \text{(Low - risk)} \ ; \ \hat{\boldsymbol{\beta}}'(0) \mathbf{x}_i > c \ \text{(High - risk)}$$

Refining Compound covariate prediction

- Compound covariate prediction uses marginal (univariate) likelihood only:

$$\hat{\boldsymbol{\beta}}(0) = (\hat{\beta}_1, \dots, \hat{\beta}_p)' \quad \text{where} \quad \hat{\boldsymbol{\beta}}(0) = \arg\max_{\boldsymbol{\beta}} L_n^0(\boldsymbol{\beta}) = \arg\max_{\boldsymbol{\beta}} \prod_{j=1}^{p} L_{n,j}^0(\beta_j)$$

- We try to enhance prediction power by incorporating multivariate likelihood information

$$L_n^1(\boldsymbol{\beta}) = \prod_{i=1}^{n} \left(\frac{\exp(\boldsymbol{\beta}'\mathbf{x}_i)}{\sum_{l \in R_i} \exp(\boldsymbol{\beta}'\mathbf{x}_l)} \right)^{\delta_i}$$

- Idea: Mixture of Univariate and multivariate likelihood

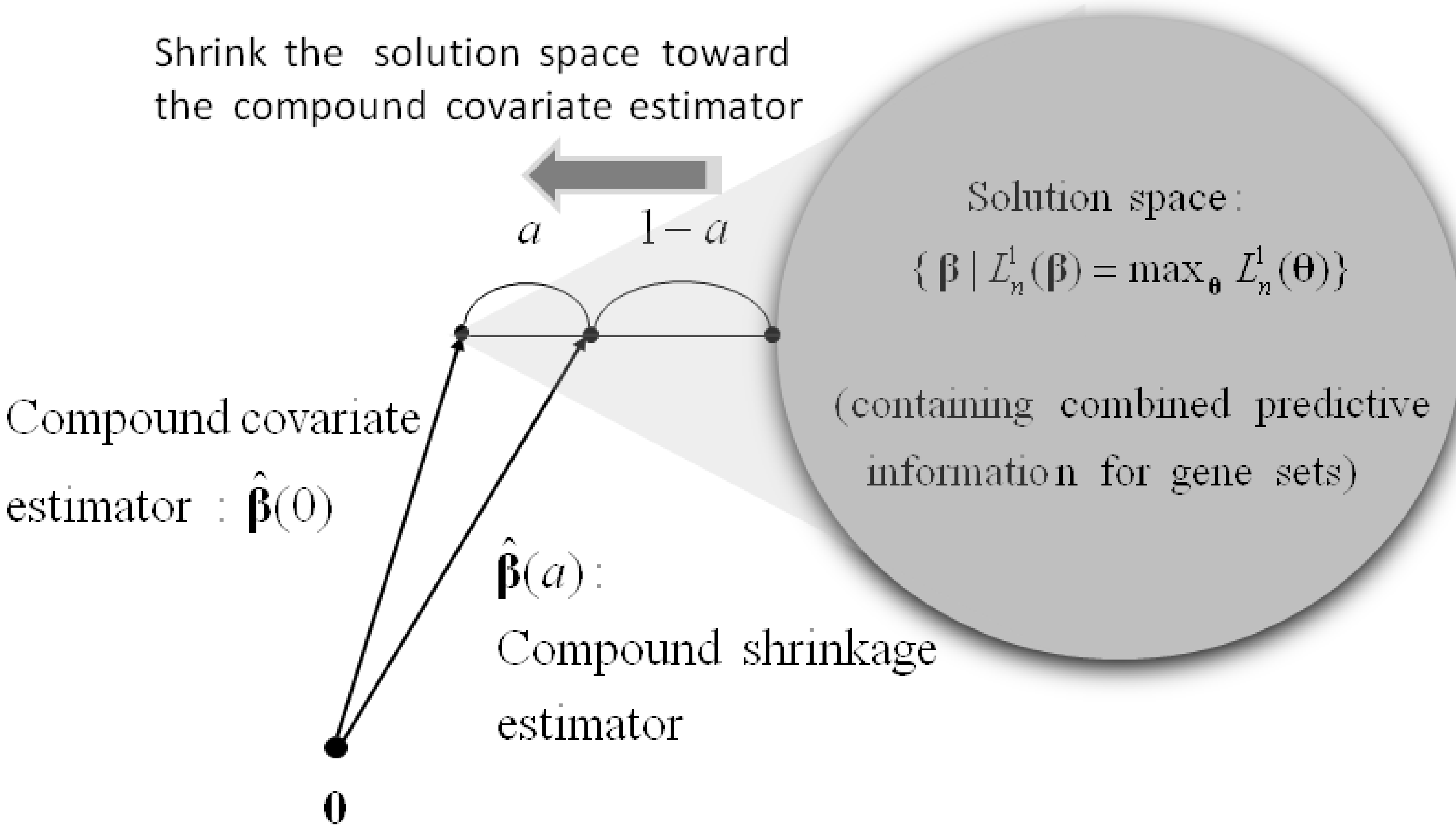
$$l_n^a(\boldsymbol{\beta}) = a \log L_n^1(\boldsymbol{\beta}) + (1-a) \log L_n^0(\boldsymbol{\beta})$$

where $a \in [0,1]$ is prespecified

$a = \hat{a}$ is determined by cross validation (Verveij & Houwelingen, 1993)

$$\text{Compound Shrinkage estimator: } \hat{\boldsymbol{\beta}}(a) = \arg\max l_n^a(\boldsymbol{\beta}),$$

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



Theoretical results

- Asymptotic normality

$$\sqrt{n}(\hat{\boldsymbol{\beta}}(\hat{a}) - \boldsymbol{\beta}_0) \to N(\mathbf{0}, \boldsymbol{\Sigma}(\boldsymbol{\beta}_0))$$

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

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

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

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

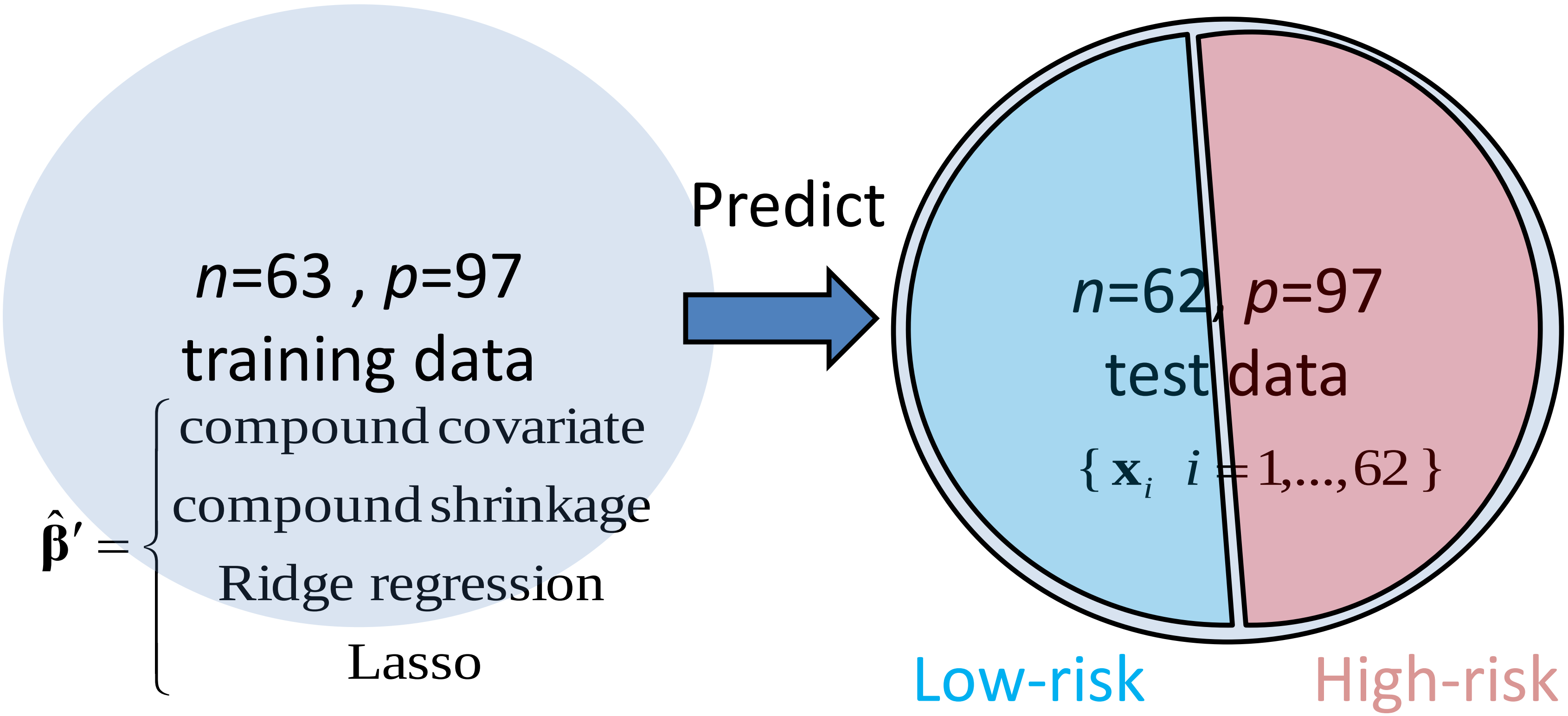
$$CV(a) = \text{Estimating function of } a,$$

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

*Reasonable performance even when *p* > *n*.

Comparison with real data

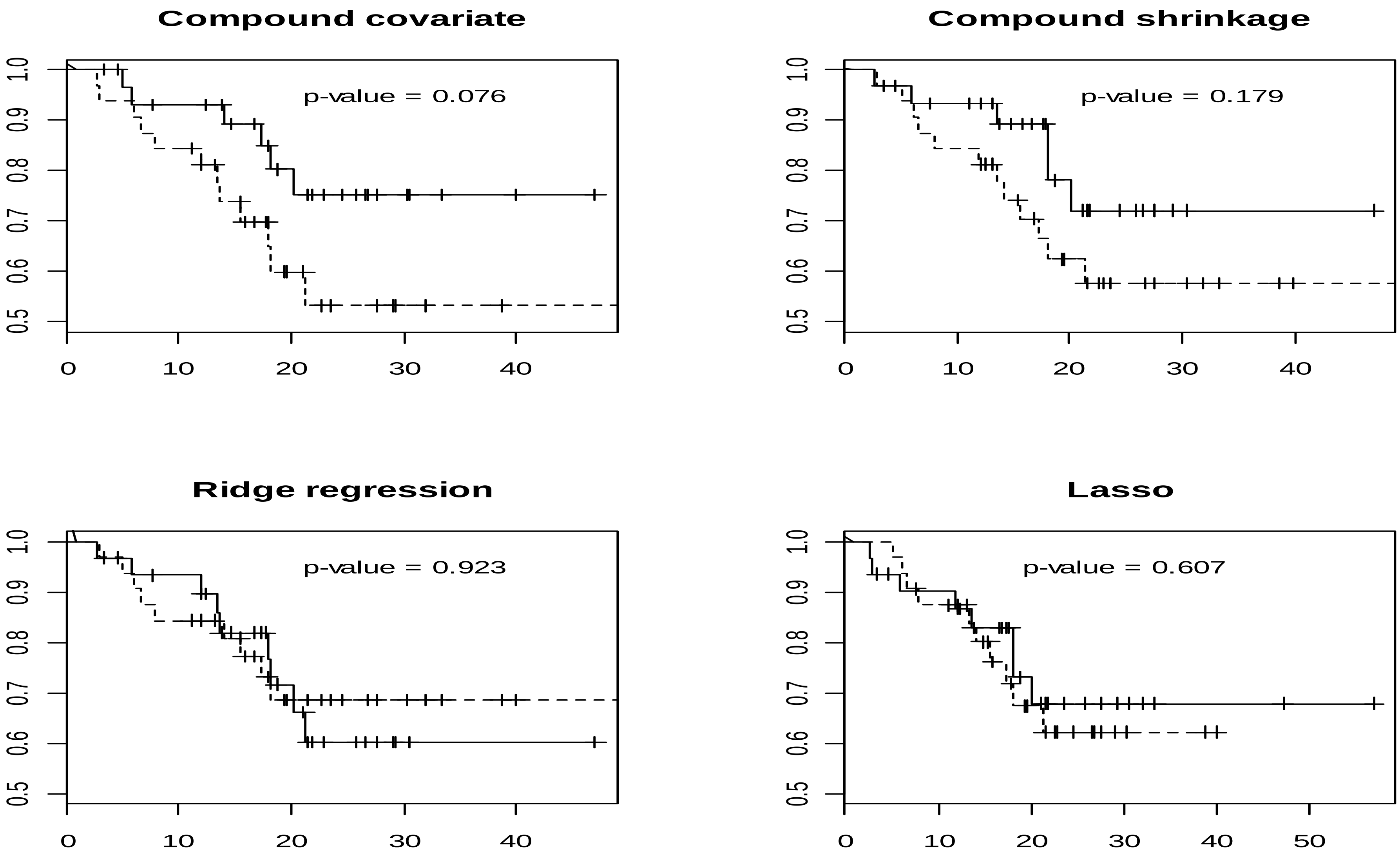
- Data: Lung cancer data (Chen et al., 2007 NEJM)



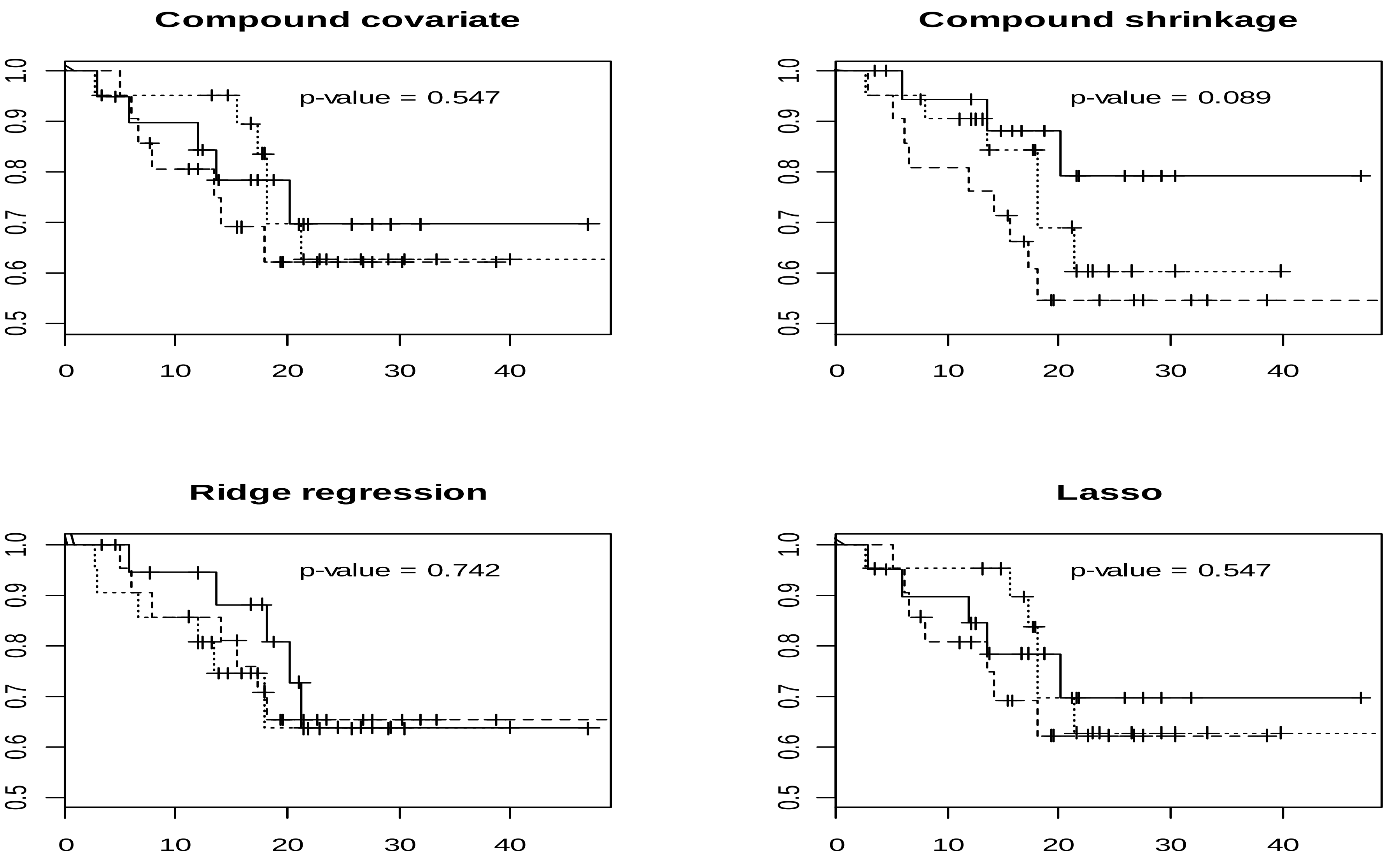
- Prediction: $\hat{\boldsymbol{\beta}}'\mathbf{x}_i < c$ (Low-risk) ; $\hat{\boldsymbol{\beta}}'\mathbf{x}_i > c$ (High-risk),

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

Survival curves for High vs. Low risk groups for n=62 testing data; p-value for testing the equality of two groups



Survival curves for High, Medium, Low risk groups for n=62 testing data; p-value for testing the equality of two groups



Summary of data analysis

- The compound covariate method is best in terms of the binary (good/poor) classification of patients' survival prospect.
- On the other hand, the three survival curves are best-separated by the proposed (compound shrinkage) method
- Overall ranking of patients' risk may be best predicted by the proposed method

Thank you for your attention