November 17th, 2017

Multimarker-based subgroups in time-to-event randomized clinical trials

Federico Rotolo

joint work with

Nils Ternès *and* **Stefan Michiels**

UNIVERSITÉ PARIS SUD
université PARIS-SACLAY
Inserm
GUSTAVE ROUSSY CANCER CAMPUS GRAND PARIS
CESP

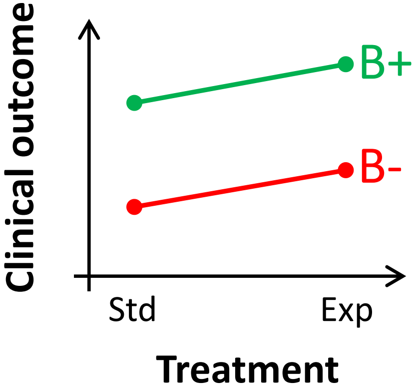
Stratified medicine

Individualized treatment based on tumor characteristics

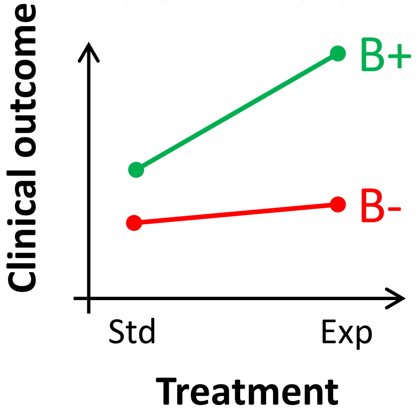
IDENTIFICATION OF BIOMARKERS

PREDICTION OF THE TREATMENT BENEFIT

Prognostic biomarkers



Predictive biomarkers



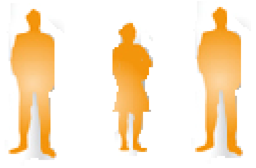
Prediction model combining clinical and genomic data

✓

Responders

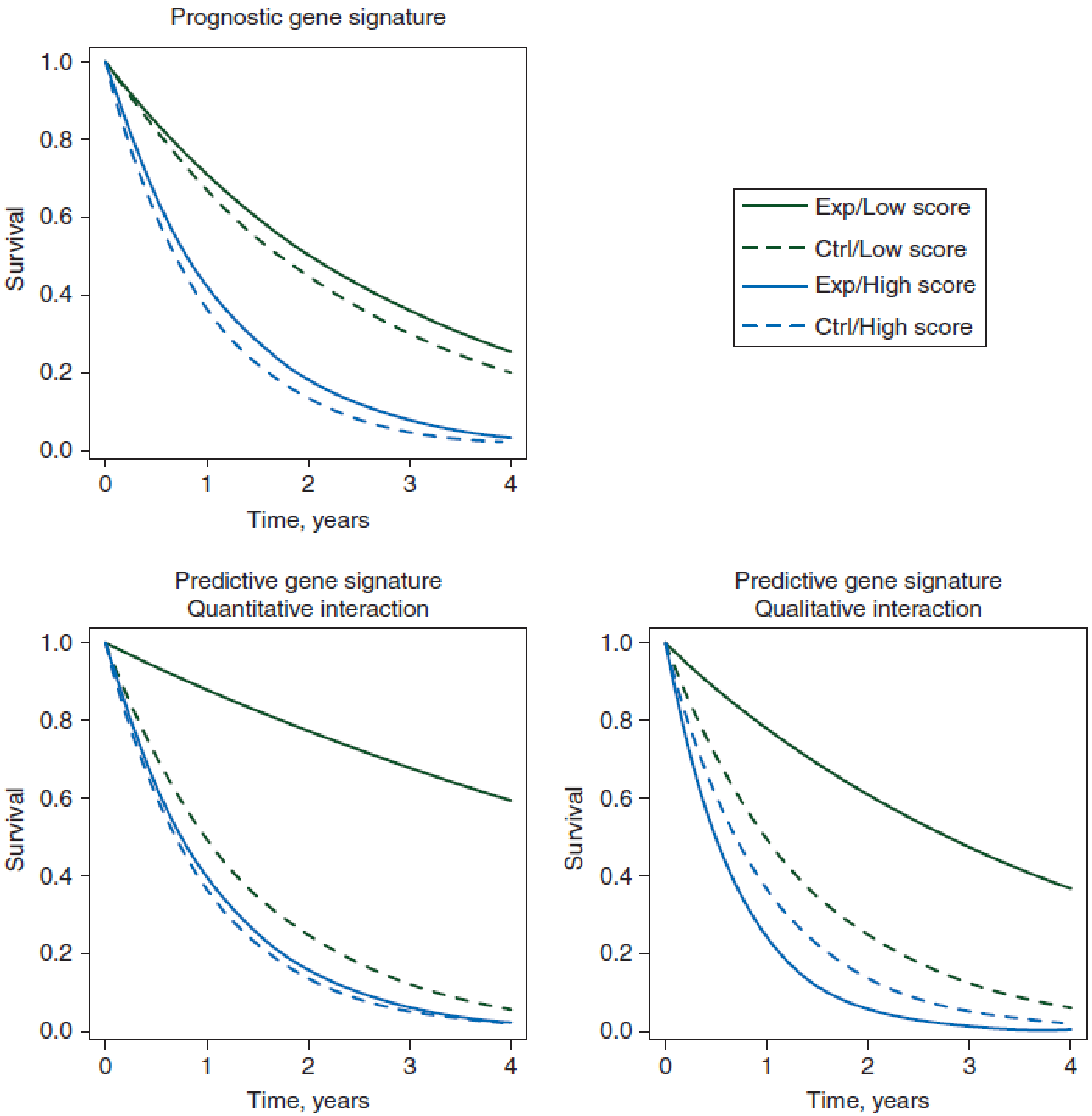
✗

Non responders



Stratified medicine – survival outcome

Michiels, Ternès, Rotolo (2016)



Statistical framework

Proportional hazard model

$$h(t; \boldsymbol{X_i}) = h_0(t) \exp\{\boldsymbol{\beta}^\intercal \boldsymbol{X_i}\}$$

with

- $h_0(t)$ the baseline hazard function at times $t > 0$
- $\boldsymbol{X_i}$ the *p*-dimensional vector of biomarkers

Partial log-likelihood

$$l(\boldsymbol{\beta}; \boldsymbol{X}) = \sum_{t_{(i)}=t_{(1)}}^{t_{(n_e)}} \left\{ \boldsymbol{\beta}^\intercal \boldsymbol{X}_{(i)} - \log \sum_{j \geq i} \boldsymbol{\beta}^\intercal \boldsymbol{X}_{(i)} \right\}$$

Statistical issues

More and more (high-dimensional) data available:

mutations, gene expression, copy number aberrations, etc.

High-dimensional data

Number of biomarkers (**p**) ≫ Sample size (**n**)

Issues raised

- Non-identifiability of the models
- Collinearity
- Instability of the estimated coefficients

- Sparse model selection
- Multiple testing

Alternative estimation strategies

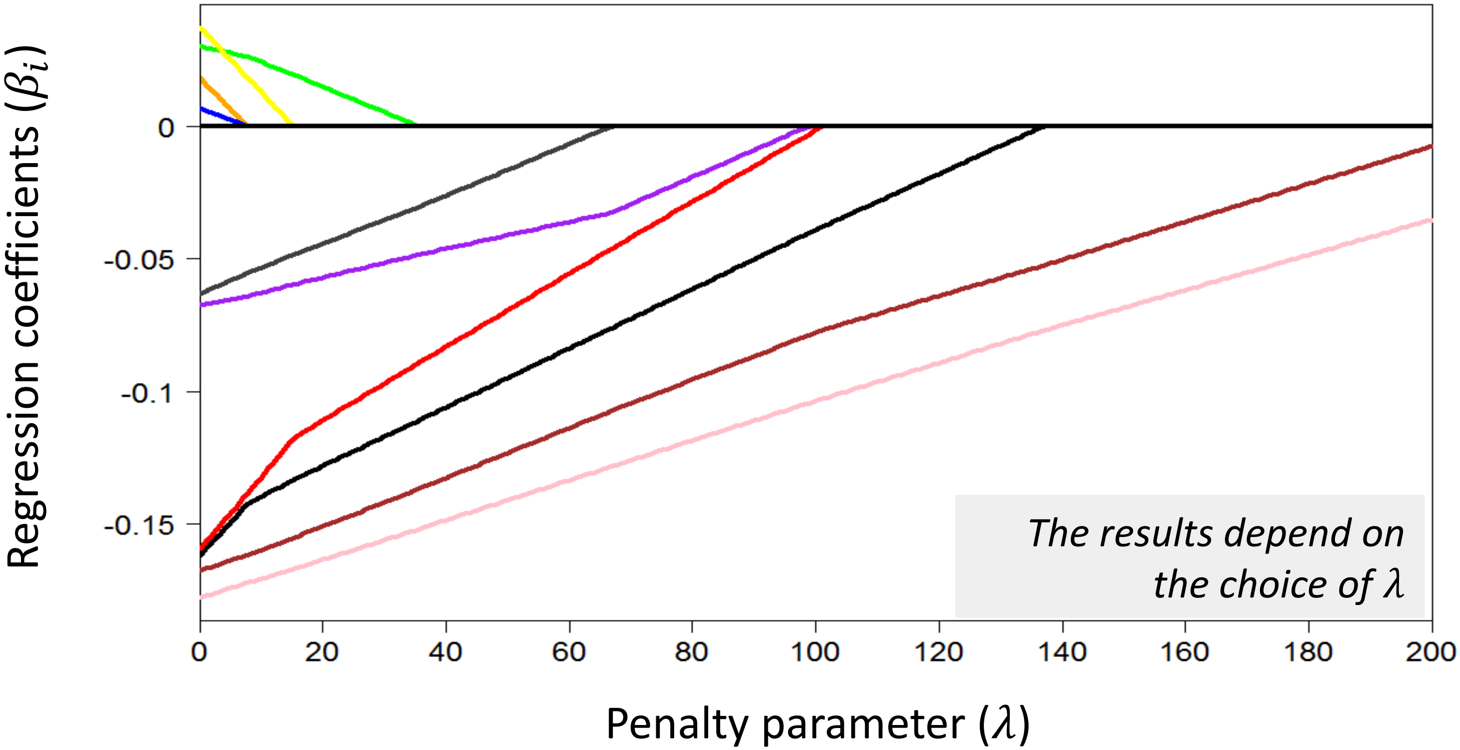
Univariate and stepwise multivariate regressions, dimension reduction, machine learning, penalized regression, …

Penalized regression: the LASSO

Aim: Introduction of a penalty term into the likelihood to shrink regression coefficients

$$l_p(\boldsymbol{\beta}) = l(\boldsymbol{\beta}; \boldsymbol{X}) - p_\lambda(\boldsymbol{\beta})$$

The LASSO penalty (Tibshirani, 1996): $p_\lambda(\boldsymbol{\beta}) = \lambda \mathbf{1}^\top |\boldsymbol{\beta}|$



The results depend on the choice of λ

The adaptive LASSO

$$l_p(\boldsymbol{\beta}) = l(\boldsymbol{\beta}; \boldsymbol{X}) - p_\lambda(\boldsymbol{\beta})$$

LASSO penalty (Tibshirani, 1996)

$$p_\lambda(\boldsymbol{\beta}) = \lambda \mathbf{1}^\top |\boldsymbol{\beta}|$$

Adaptive LASSO penalty (Zou, 2006 ; Zhang and Lu, 2007)

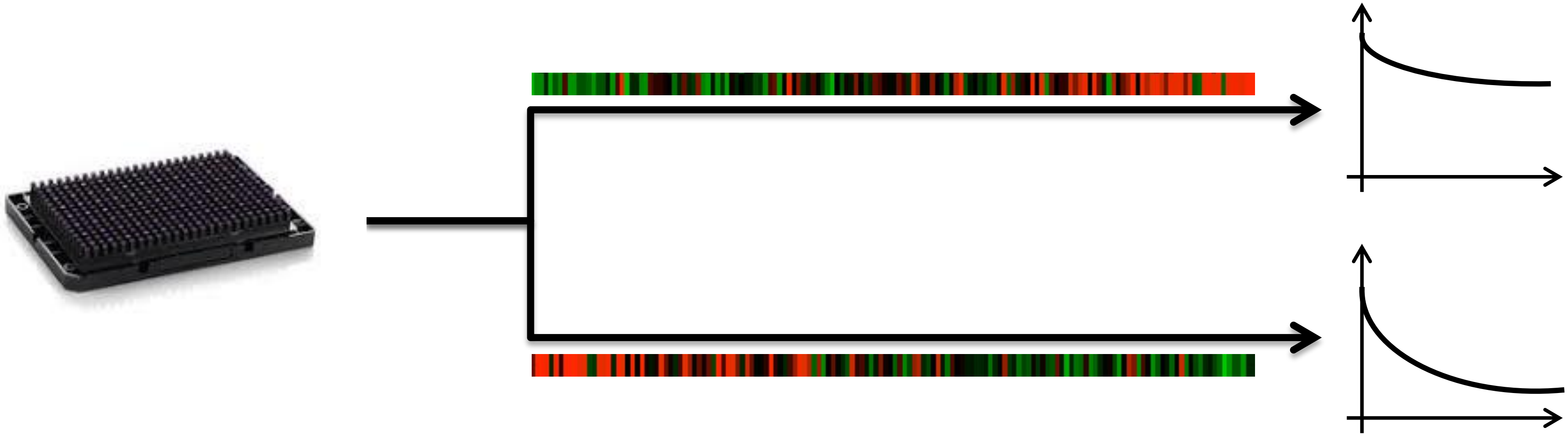
$$p_\lambda(\boldsymbol{\beta}) = \lambda \boldsymbol{w}^\top |\boldsymbol{\beta}|$$

with weights $w_j = 1/\left|\widehat{\beta}_j\right|$ based on

- univariate estimates
- simple lasso
- ridge regression: $l_p(\boldsymbol{\beta}) = l(\boldsymbol{\beta}; \boldsymbol{X}) - \lambda_2 \boldsymbol{\beta}^\top \boldsymbol{\beta}$
- full model (when identifiable)
- …

How to reliably identify prognostic biomarkers?

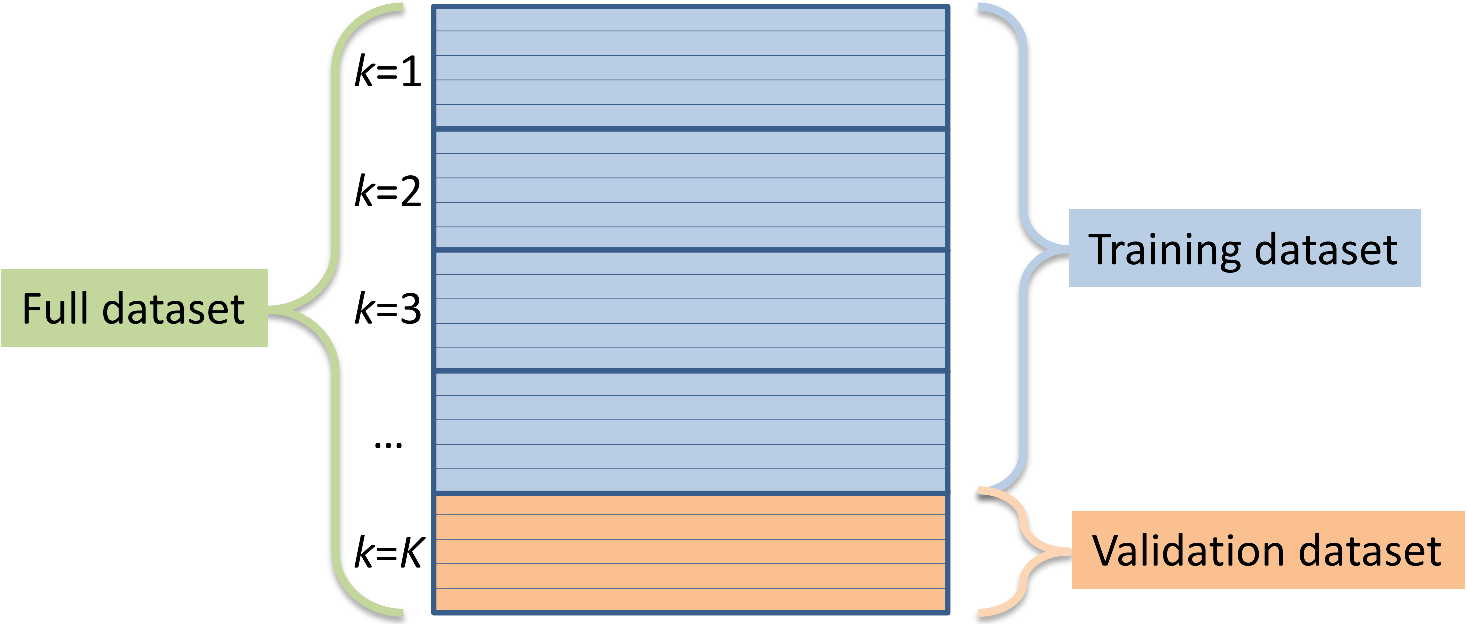
(i.e. associated to patient survival)



The choice of the penalty λ

The usual choice of λ is the value maximizing the **cross-validated penalized likelihood** (Tibshirani, 1997)

$$cvl(\lambda) = \sum_{k=1}^{K}\left(l_p\left(\hat{\beta}_{tr,k}, X_{full}\right) - l_p\left(\hat{\beta}_{tr,k}, X_{tr,k}\right)\right)$$



Full dataset

k=1

k=2

k=3

…

k=*K*

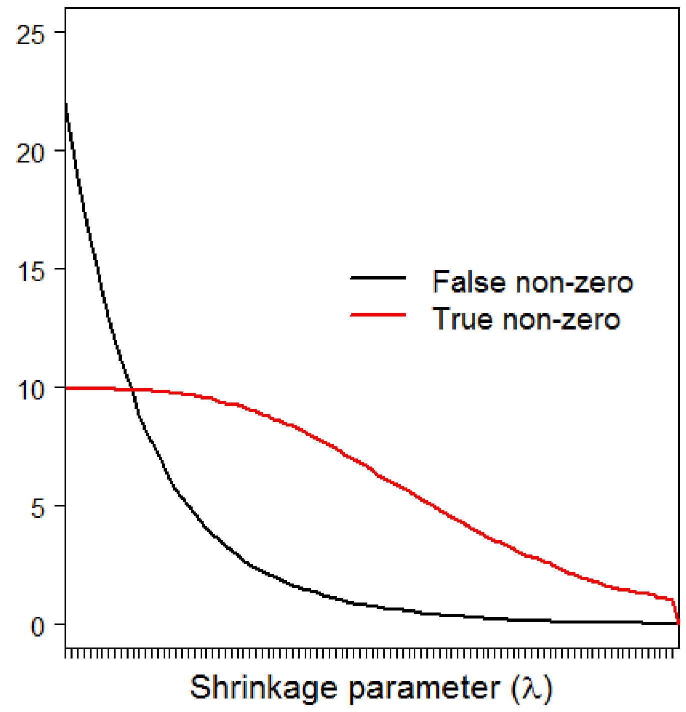
Training dataset

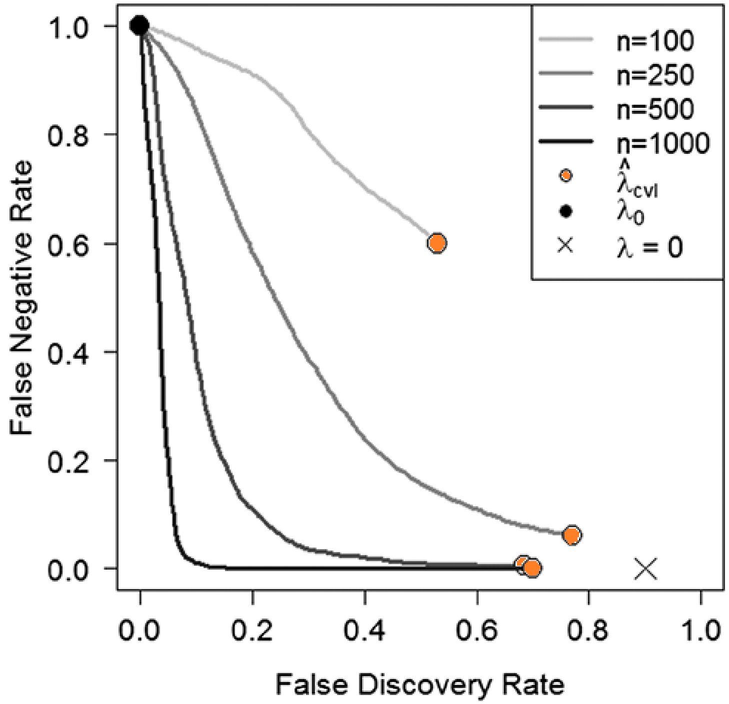
Validation dataset

The choice of the penalty λ

Ternès, Rotolo, Michiels (2016)

Selection performed with λ chosen by max-cvl
ends up with a high proportion of false positives
(Meinshausen and Bühlmann 2006, Benner et al. 2010, Roberts and Nowak 2014)

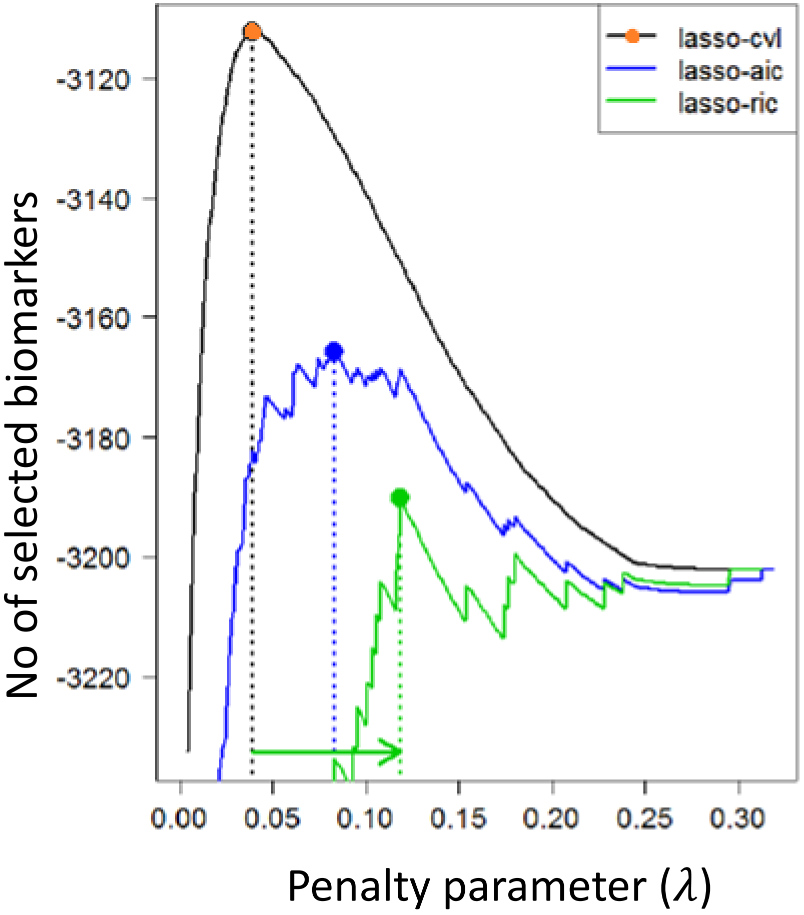




⇒ Need to increase the penalty λ to select fewer biomarkers

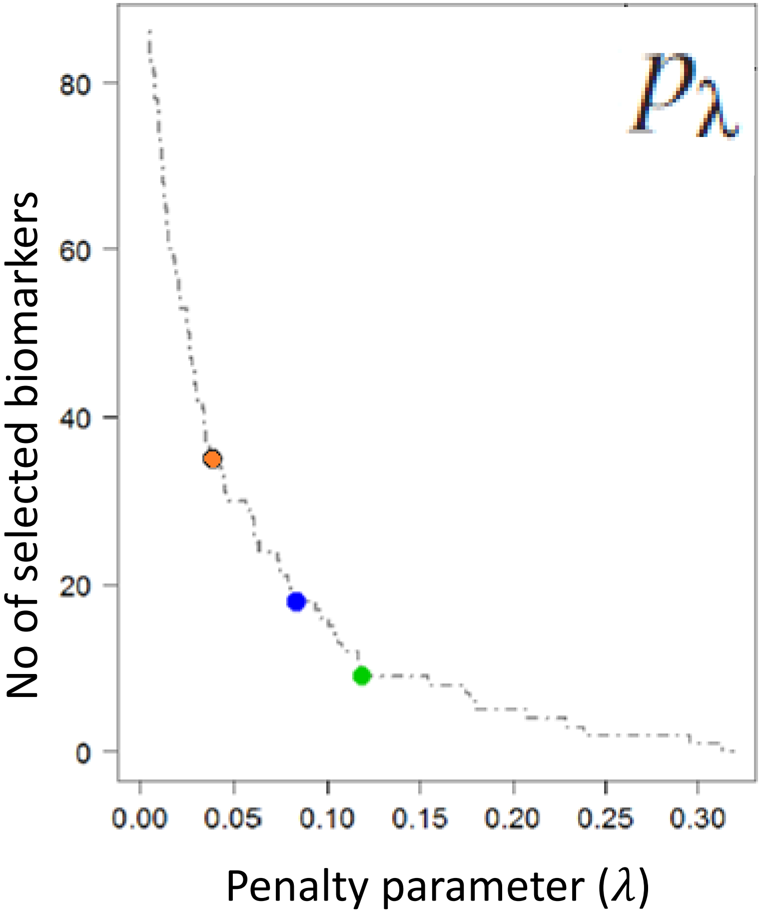
Penalized cross-validated likelihood

$\text{argmax}_\lambda(cvl \; - pen)$



$pen_{\text{AIC}}(\lambda) = 2 \; \times \; p_\lambda$

$pen_{\text{RIC}}(\lambda) = 2 \; \times \log(p_\lambda) \times p_\lambda$

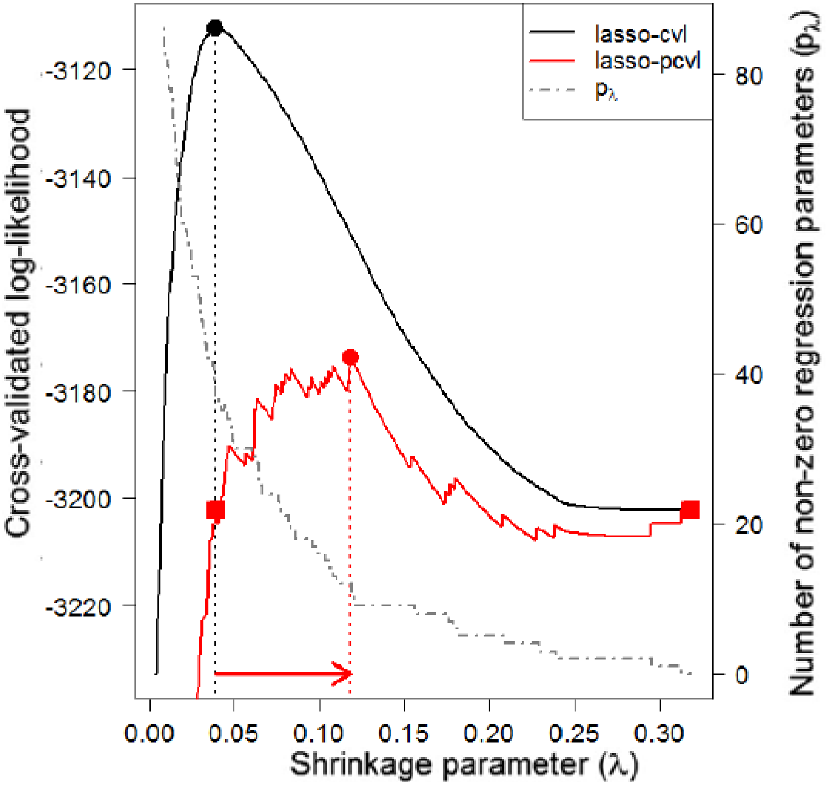


Penalized cross-validated likelihood

Ternès, Rotolo, Michiels (2016)

Lasso with penalized cvl

$$pcvl(\lambda) = cvl(\lambda) - \left(cvl_{\lambda_{cvl}} - cvl_{\lambda_0}\right)\frac{p_\lambda}{p_{\lambda_{cvl}}}$$

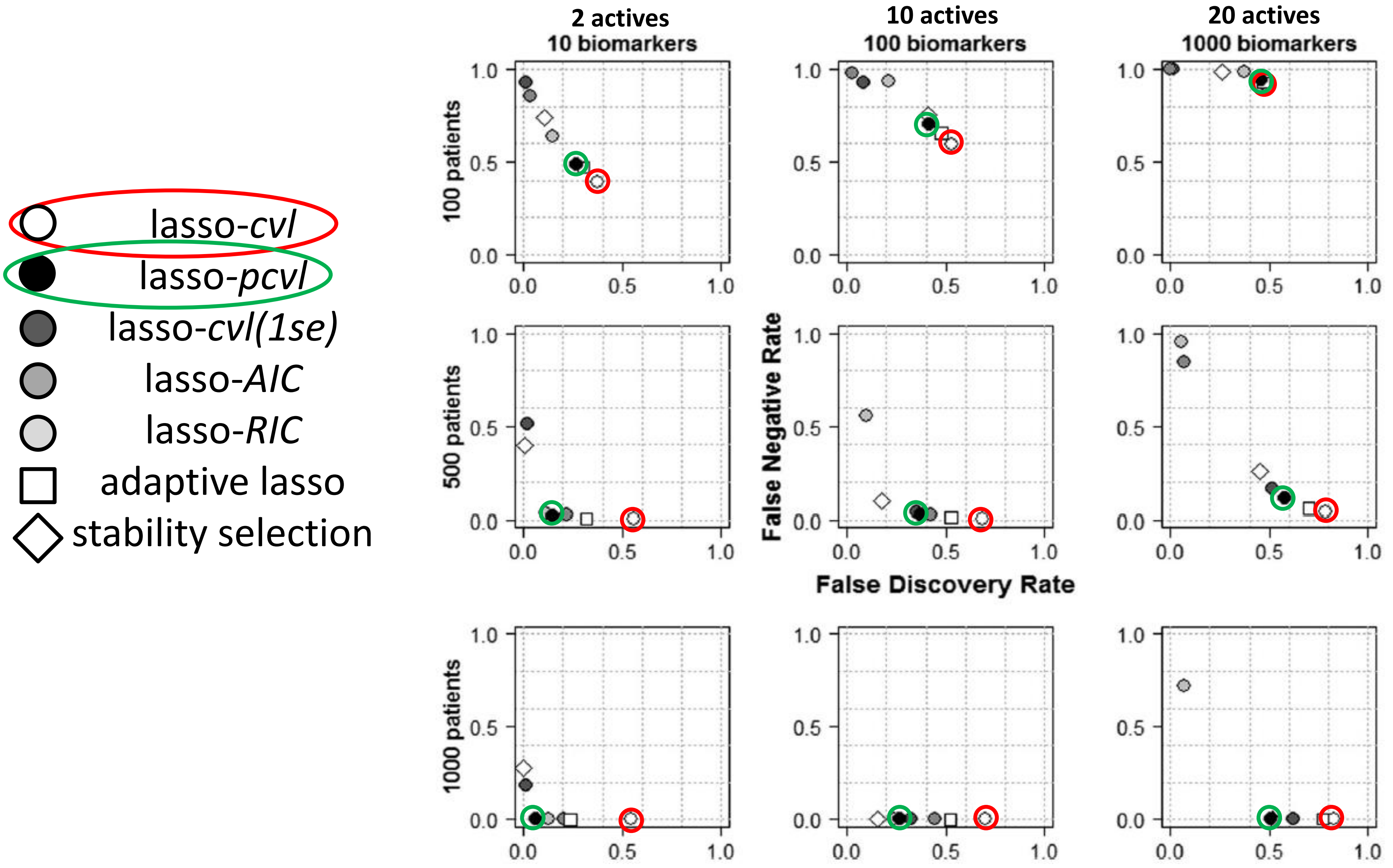


$\lambda \geq \lambda_{cvl} \rightarrow p_\lambda \leq p_{\lambda_{cvl}}$
i.e. selection of fewer biomarkers

$pcvl(\lambda_{cvl}) = pcvl(\lambda_0)$
i.e. the model selected by cvl is penalized as much as the empty model
→ trade-off between goodness of fit (max cvl) and parcimony (small p_λ)

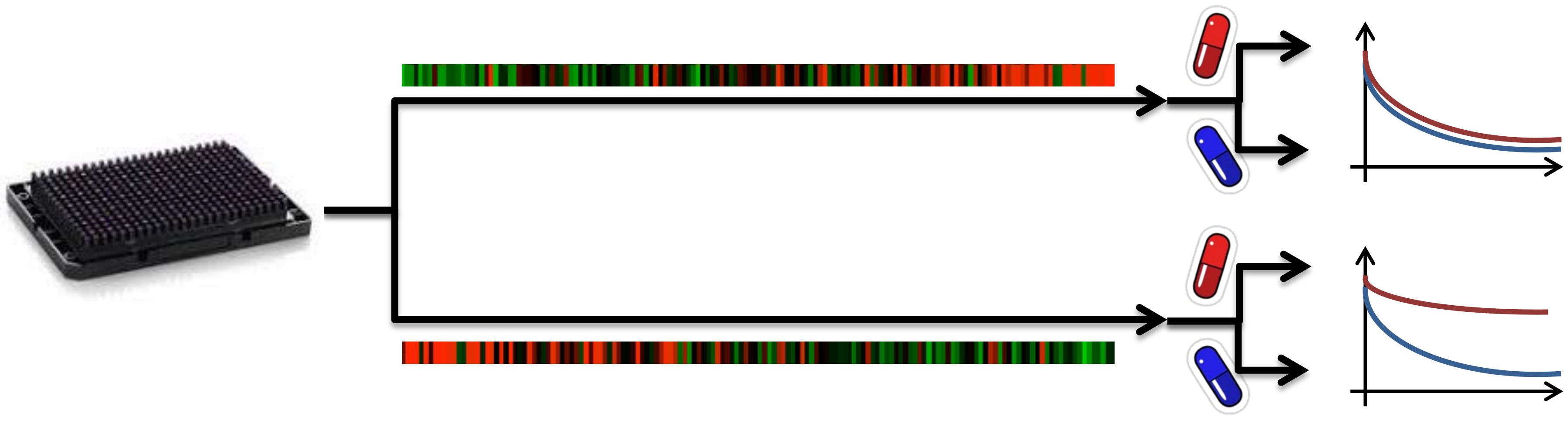
Simulation results

Ternès, Rotolo, Michiels (2016)



How to reliably identify predictive biomarkers?

(i.e. associated to the magnitude of the treatment effect)



Statistical framework

Proportional hazard model

$$h(t; T_i, \boldsymbol{X_i}) = h_0(t)\exp\{\alpha T_i + \boldsymbol{\beta}^\top \boldsymbol{X_i} + \boldsymbol{\gamma}^\top T_i \boldsymbol{X_i}\}$$

with

- $h_0(t)$ the baseline hazard function at times $t > 0$
- T_i the treatment arm
- $\boldsymbol{X_i}$ the *p*-dimensional vector of biomarkers

and with $\boldsymbol{\gamma}^\top T_i \boldsymbol{X_i}$ accounting for the

treatment-by-biomarkers interaction

Statistical issues

$$h(t; T_i, \boldsymbol{X_i}) = h_0(t) \exp\{\alpha T_i + \boldsymbol{\beta}^\intercal \boldsymbol{X_i} + \boldsymbol{\gamma}^\intercal T_i \boldsymbol{X_i}\}$$

Aim: selection of the relevant interactions $T_i \boldsymbol{X_i}$

Issue: The model with *all* the main effects $\boldsymbol{\beta}^\intercal \boldsymbol{X_i}$ is not identifiable
or at least very DoF-consuming

**→ How to select the relevant interactions
while properly accounting for the main effects?**

(A)LASSO

Full LASSO.

$$l_p(\alpha, \boldsymbol{\beta}, \boldsymbol{\gamma}) = l(\alpha, \boldsymbol{\beta}, \boldsymbol{\gamma}; \boldsymbol{T}, \boldsymbol{X}) - \lambda(\mathbf{1}^\top|\boldsymbol{\beta}| + \mathbf{1}^\top|\boldsymbol{\gamma}|)$$

Adaptive LASSO.

$$l_p(\alpha, \boldsymbol{\beta}, \boldsymbol{\gamma}) = l(\alpha, \boldsymbol{\beta}, \boldsymbol{\gamma}; \boldsymbol{T}, \boldsymbol{X}) - \lambda\left(\boldsymbol{w}_{\boldsymbol{\beta}}^\top|\boldsymbol{\beta}| + \boldsymbol{w}_{\boldsymbol{\gamma}}^\top|\boldsymbol{\gamma}|\right)$$

Pros. Simple. Sparse models, easy interpretation

Cons. Selection with same penalty, irrespective of parameter meaning → No hierarchy constraint

Group LASSO

Yuan and Lin (2006)

$$l_p(\alpha,\boldsymbol{\beta},\boldsymbol{\gamma}) = l(\alpha,\boldsymbol{\beta},\boldsymbol{\gamma};\boldsymbol{T},\boldsymbol{X}) - \lambda\sum_{j=1}^{p}\left\|\left(\beta_j,\gamma_j\right)^{\intercal}\right\|_2$$

Either selects both β_j and γ_j or none of them,

Pros. Hierarchy constraint fulfilled:
no interaction (with treatment) without main effect

Cons. Interactions (with treatment)
of strongly prognostic biomarker are likely selected

Ridge + LASSO

$$l_p(\alpha, \boldsymbol{\beta}, \boldsymbol{\gamma}) = l(\alpha, \boldsymbol{\beta}, \boldsymbol{\gamma}; \boldsymbol{T}, \boldsymbol{X}) - \lambda_2 \boldsymbol{\beta}^\top \boldsymbol{\beta} - \lambda \mathbf{1}^\top |\boldsymbol{\gamma}|$$

All the **main effects** are selected, with **ridge** penalization

Lasso selection on the **interactions**

Pros. Hierarchy constraint fulfilled.

Adjustment for prognostic factors, predictive or not

Cons. Model not sparse at all.

Two shrinkage parameters λ and λ_2

Modified covariate model

Tian et al (2014)

$$h(t;T_i,\boldsymbol{X_i}) = h_0(t)\exp\{\boldsymbol{\gamma}^\intercal \boldsymbol{M_i}\},\boldsymbol{M}_i = T_i\boldsymbol{X_i}/2$$

$$l_p(\alpha,\boldsymbol{\beta},\boldsymbol{\gamma}) = l(\alpha,\boldsymbol{\beta},\boldsymbol{\gamma};\boldsymbol{T},\boldsymbol{X}) - \lambda\mathbf{1}^\intercal|\boldsymbol{\gamma}|$$

No **main effects** in the model. **Lasso** selection on the **modified covariates** $\boldsymbol{M}_i$

Pros. Only one set of covariates to make selection on. Shown to perform well in linear model

Cons. No explicit prognostic effects

Other methods

Other methods (among many possible)

- **Dimension reduction** (PCA or PLS) for main effects + LASSO on interactions
- **Univariate**: Wald test for interaction in the biomarker-specific model with FDR control
- **Boosting**
- …

Simulation results

Ternès, Rotolo, Heinze, Michiels (2017)

Data generation.

- Gaussian biomarkers with autoregressive correlation ($\sigma_{ij} = 0.7^{|i-j|}$) within 50-biomarker blocks, respectively
- Randomly assignment (1:1) to experimental or control arm
- Exponential survival times
- Independent censoring from a *U(*2, 5*)* distribution (i.e. 3-y accrual and 2-y follow-up)

Treatment effect: $\alpha = \log(0.5)$
Prognostic biomarker effect: $\beta_j = \log(0.5)$ for $X_j = 1$ vs $X_j = 0$
Predictive biomarker effect: $\gamma_j = \log(0.5)$ for $X_j = 1$ vs $X_j = 0$

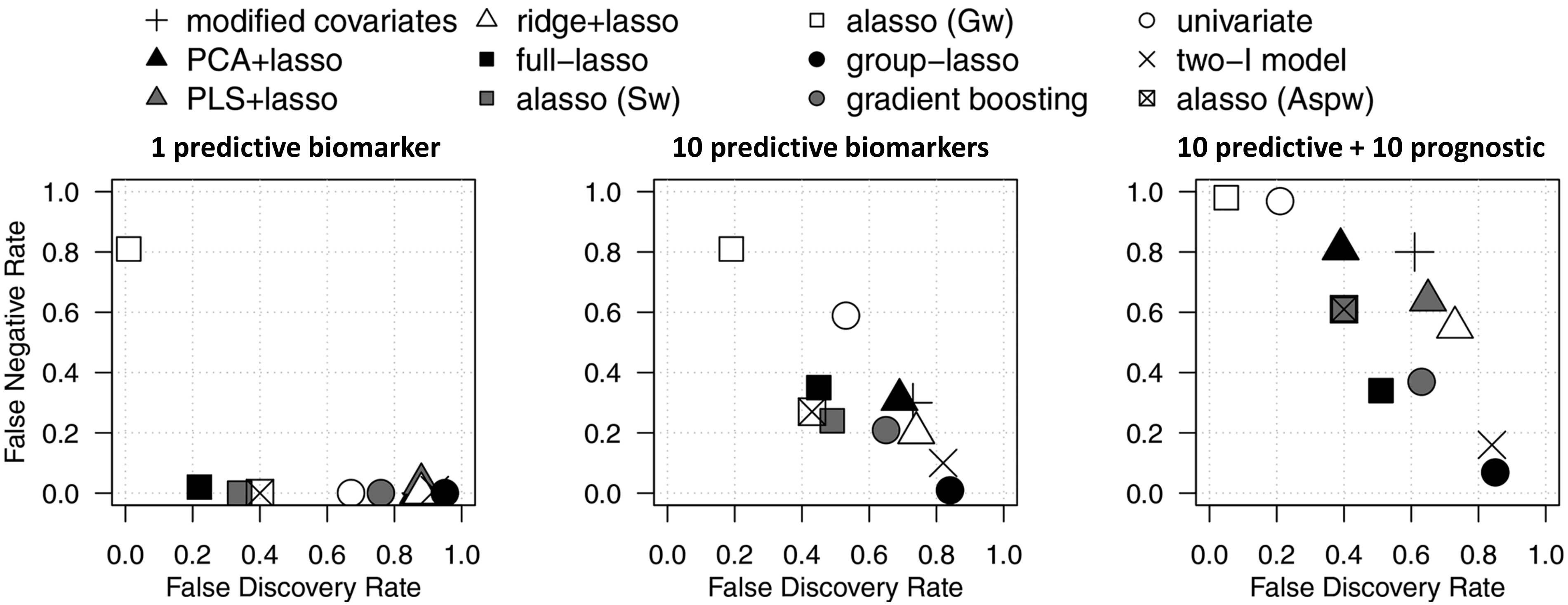
Results in the null scenarios (no predictive biomarker)

Type-1 error rate	**Univariate + FDR**	**Modified Covariates**	**PCA + LASSO**	**Ridge + LASSO**	**Group LASSO**	**Full LASSO**	**Adaptive LASSO**	**Gradient Boosting**
n=500, p=1000								
No effect at all	0.06	0.38	0.35	0.38	0.52	0.01	0.12	0.68
Treatment effect	0.04	0.41	0.43	0.38	0.52	0.02	0.16	0.69
Prognostic effects	0.08	0.45	0.27	0.58	1.00	0.98	0.32	1.00

Simulation results

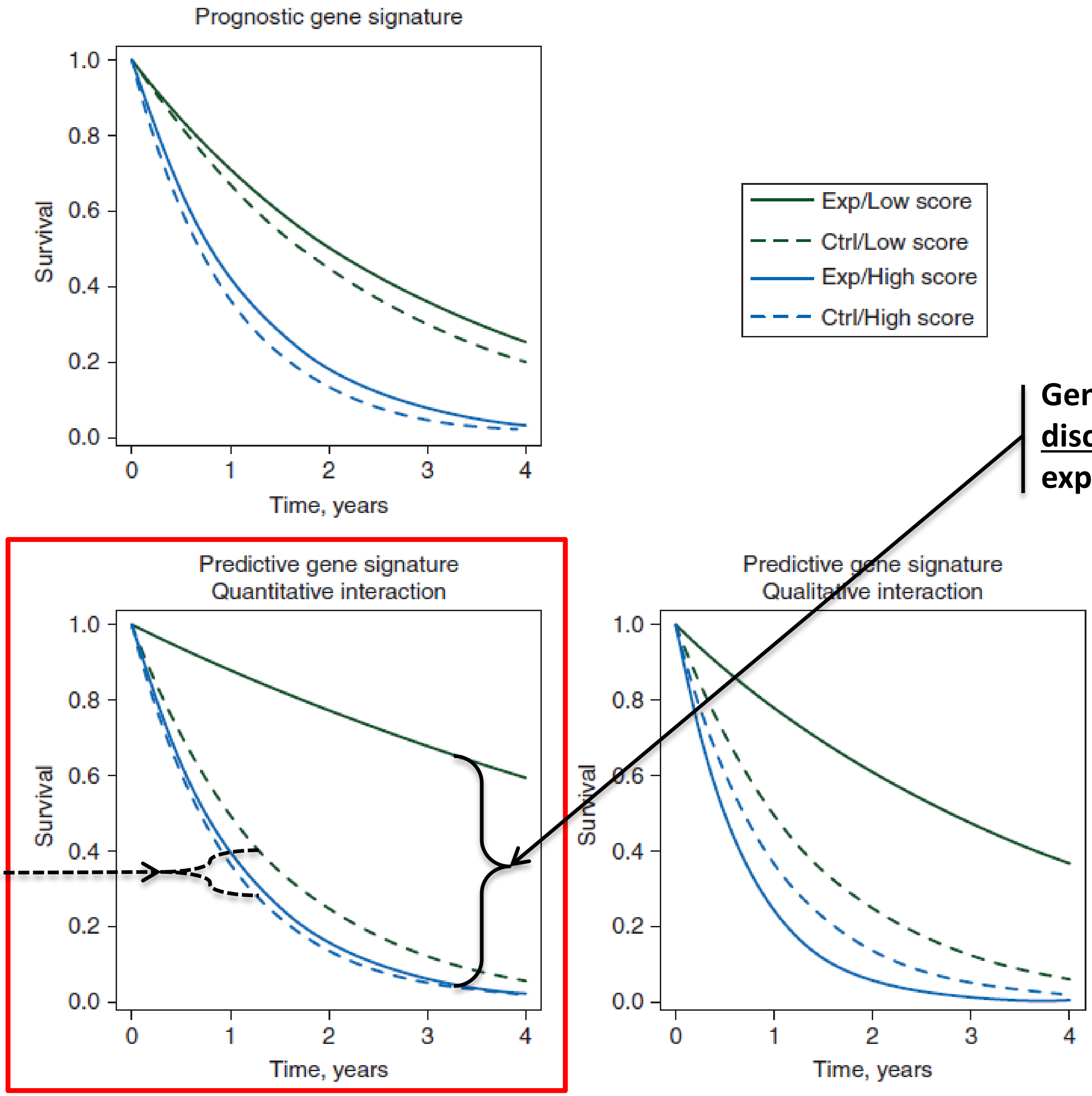
Ternès, Rotolo, Heinze, Michiels (2017)

Results in the alternative scenarios (1+ predictive biomarkers)



Stratified medicine – survival outcome

Michiels, Ternès, Rotolo (2016)



Gene signature is very discriminant in the experimental arm

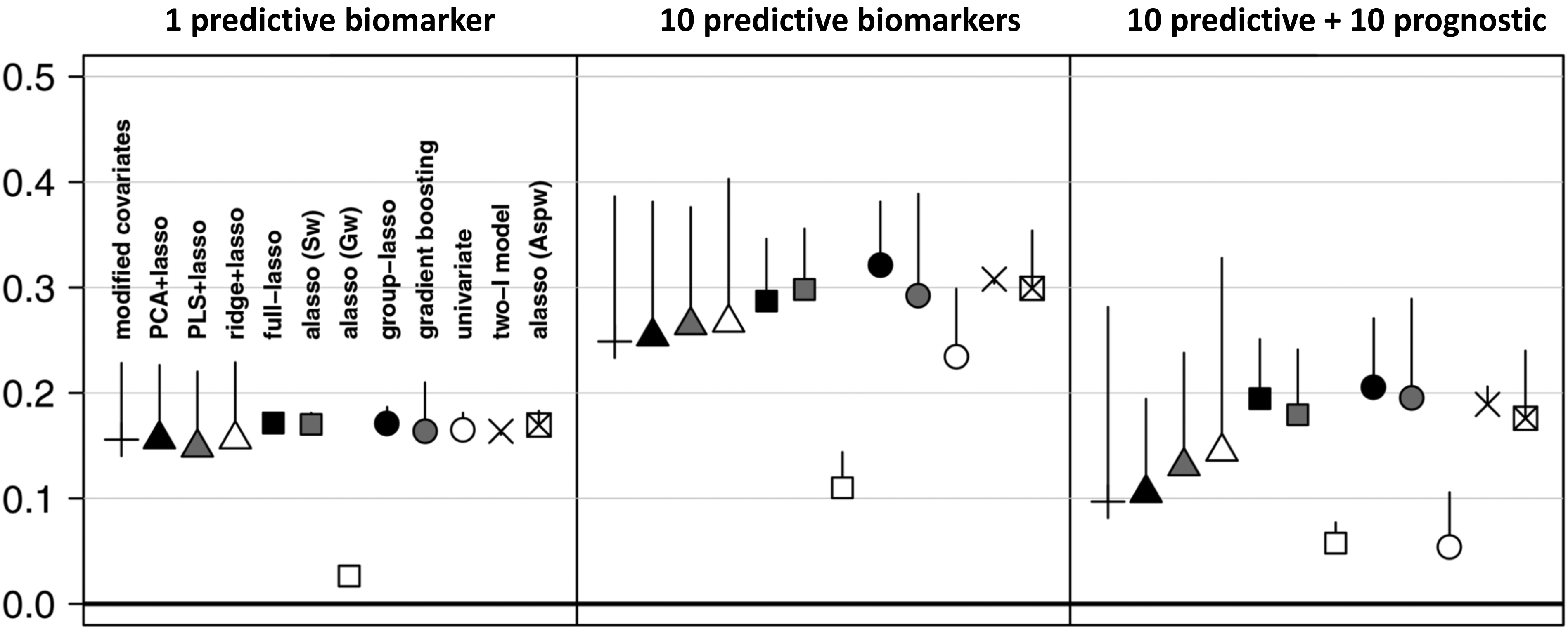
Gene signature is little discriminant in the control arm

Simulation results

Ternès, Rotolo, Heinze, Michiels (2017)

Results in the alternative scenarios (1+ predictive biomarkers)

Between-arms difference in Uno *C*-statistics (ΔC)



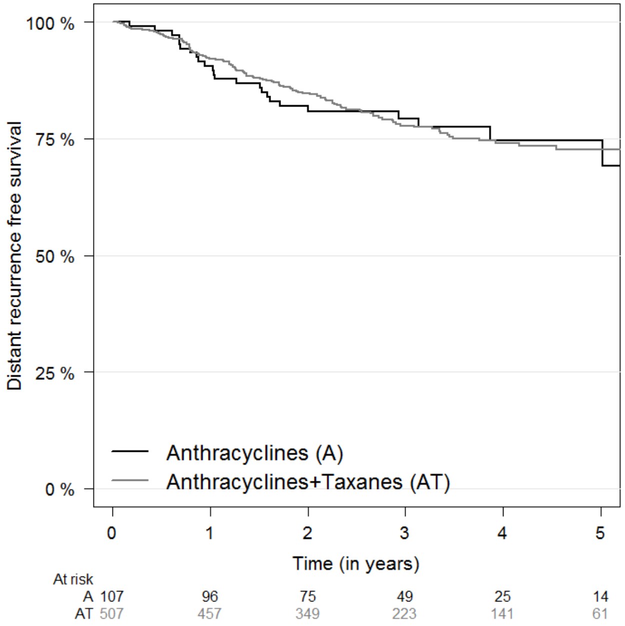
Points: ΔC computed in an independent validation dataset

Vertical segments go up to the values of the ΔC computed in the training dataset

GEO early breast cancer data

Publicly available data in breast cancer patients (GEO Database)

(Desmedt et al, 2011; Hatzis et al, 2011)



Clinical data

n = 614 patients

Gr. A: anthracycline only (**n** = **107**)

Gr. AT: anthracycline + taxane-based (**n** =**507**)

Gene expression data

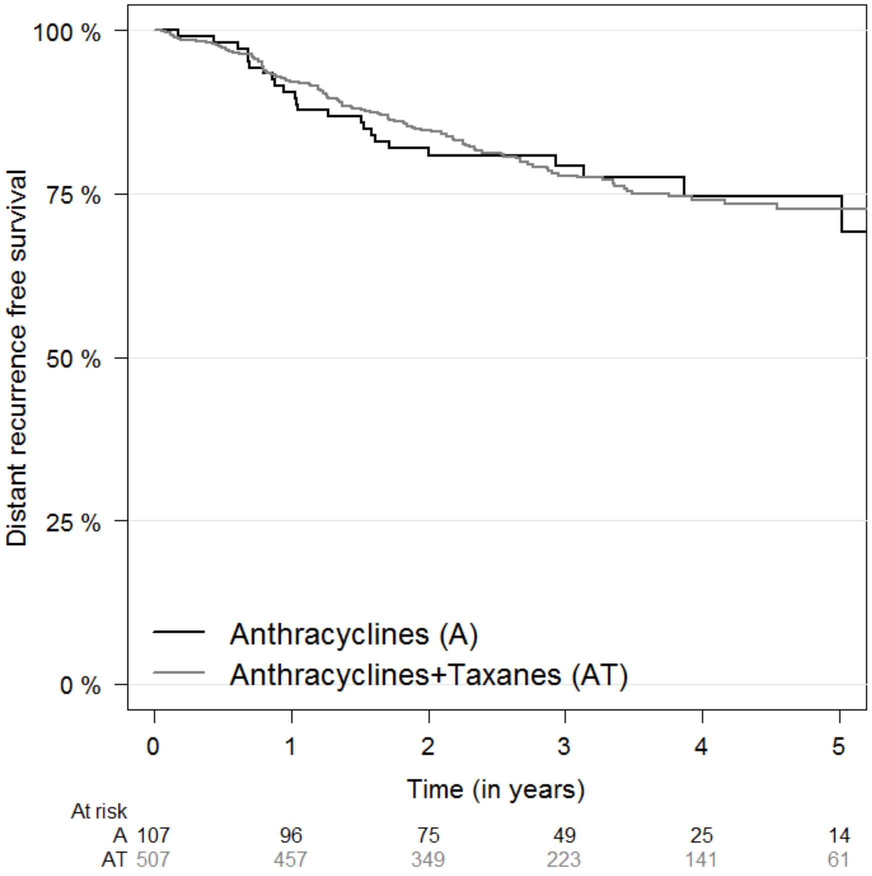
Initially, 22,277 normalized genes

Filtering step: Only genes with the largest variability (IQR > 1)

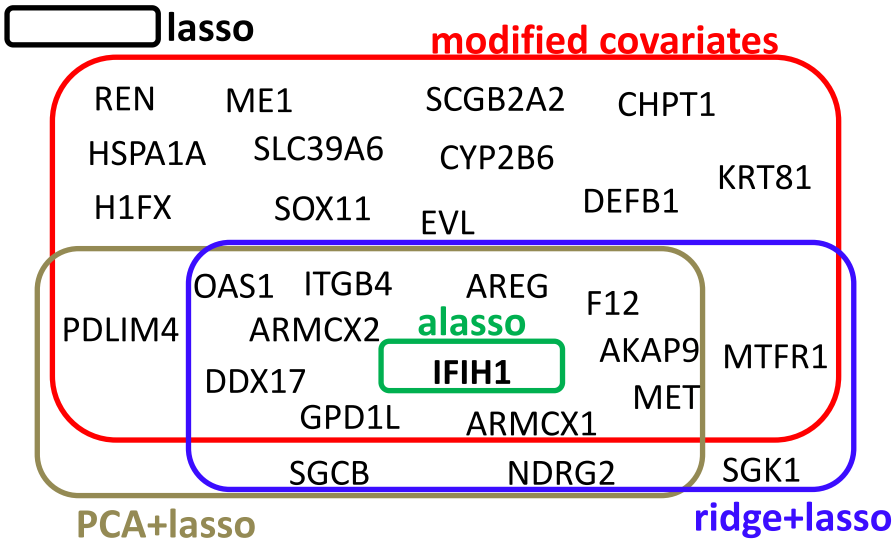
p = 1689 standardized genes

Aim: Identify treatment-modifiers for taxane in tumors biopsies from breast cancer

GEO early breast cancer data



Training set (315 patients)


lasso

modified covariates

REN ME1 SCGB2A2 CHPT1

HSPA1A SLC39A6 CYP2B6 KRT81

H1FX SOX11 EVL DEFB1

PDLIM4 OAS1 ITGB4 AREG F12

ARMCX2 alasso AKAP9 MTFR1

DDX17 **IFIH1** MET

GPD1L ARMCX1

SGCB NDRG2 SGK1

PCA+lasso

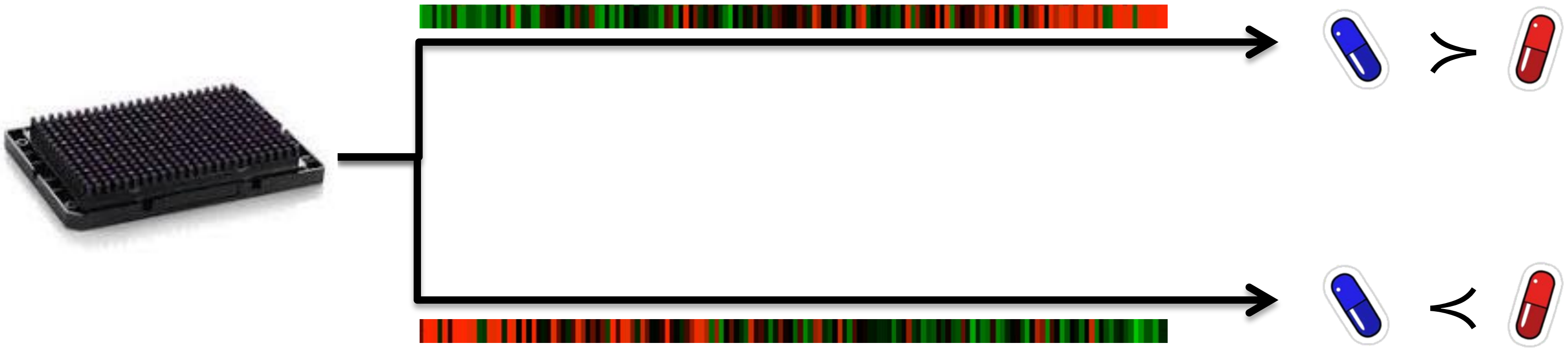
ridge+lasso

Validation set (299 patients)

	# selected biomarkers	ΔC-statistics
modified covariates	25	0.11
PCA+lasso	14	0.07
ridge+lasso	15	0.06
lasso	0	0
adaptive lasso	1	0.11

Knowledge on IFIH1 gene: associated with recurrence in non-responders to taxane CT in early Breast K; is part of 2 patents for predicting the benefit of taxanes in K; its expression is associated with resistance to taxane in prostate K cell lines

How to translate a predictive signature into survival probabilities to support the therapeutic decision?



From a statistical viewpoint

- How can we **estimate** accurately
  the **survival probability** for a new patient
  starting **from** a **penalized regression** model?

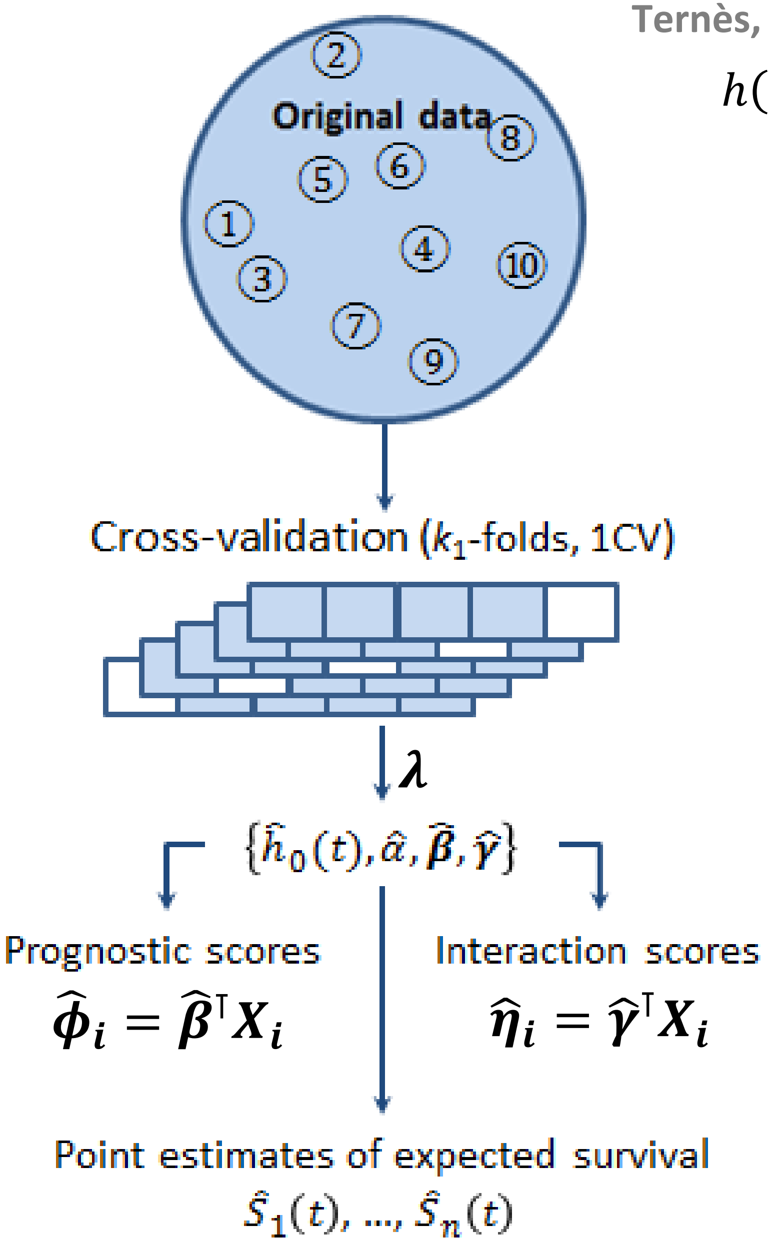
- How can we **compute confidence intervals**
  for such estimation?

- How can we **control for overfitting**
  in the case no external validation data is available?

Estimation of the survival probability

Ternès, Rotolo, Michiels (2017a)

$$h(t; T_i, \boldsymbol{X_i}) = h_0(t)\exp\{\alpha T_i + \boldsymbol{\beta}^\top \boldsymbol{X_i} + \boldsymbol{\gamma}^\top T_i \boldsymbol{X_i}\}$$
$$l_p(\boldsymbol{\beta}) = l(\boldsymbol{\beta}; \boldsymbol{X}) - p_\lambda(\boldsymbol{\beta})$$



Original data

Cross-validation (k_1-folds, 1CV)

$\boldsymbol{\lambda}$

$\{\hat{h}_0(t), \hat{\alpha}, \hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\gamma}}\}$

Prognostic scores

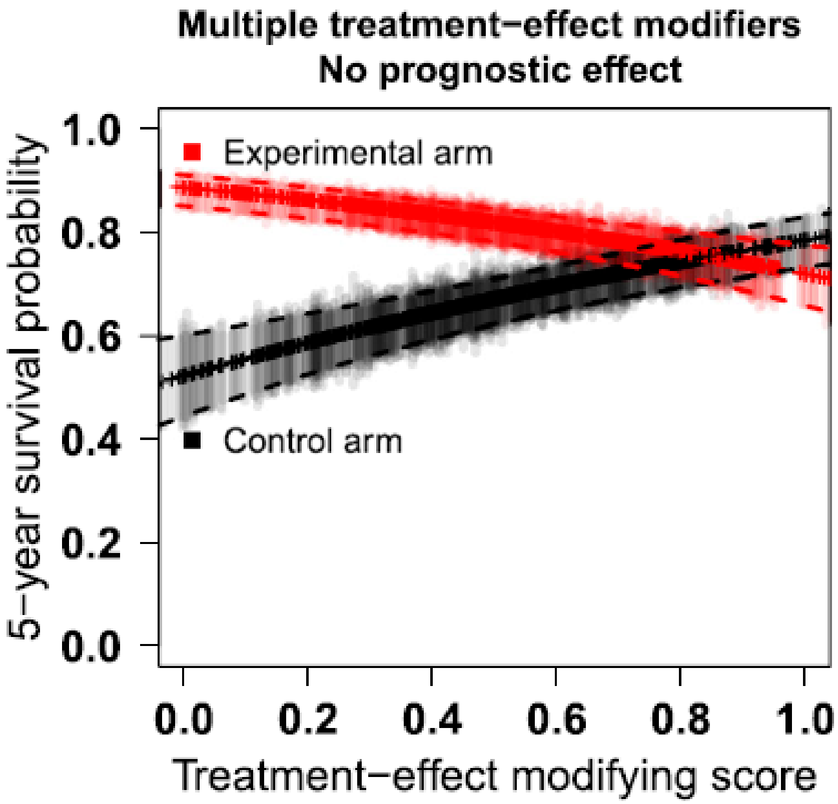
$$\hat{\boldsymbol{\phi}}_i = \hat{\boldsymbol{\beta}}^\top \boldsymbol{X}_i$$

Interaction scores

$$\hat{\boldsymbol{\eta}}_i = \hat{\boldsymbol{\gamma}}^\top \boldsymbol{X}_i$$

Point estimates of expected survival

$\hat{S}_1(t), \ldots, \hat{S}_n(t)$



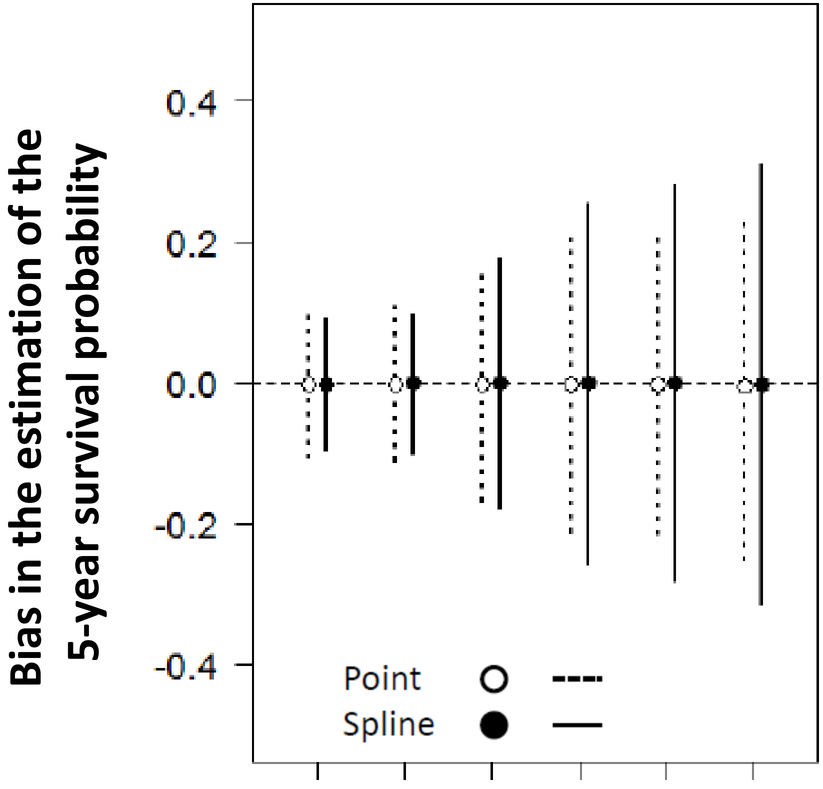
$$\hat{\boldsymbol{\eta}}_i = \hat{\boldsymbol{\gamma}}^\top \boldsymbol{X}_i$$

Estimation of the survival probability

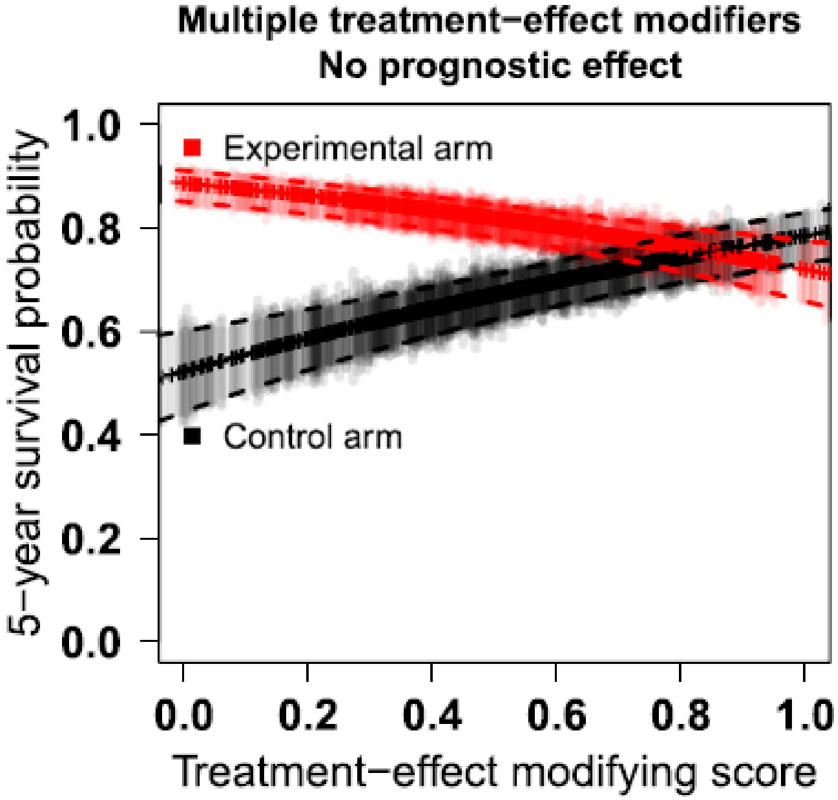
Ternès, Rotolo, Michiels (2017a)

Confidence interval estimation:

- Analytic $\mathrm{IC}_{1-\theta}\left(\hat{S}_k(t)\right) = \exp\left(-\widehat{H}_k(t) \pm z_{1-\frac{\theta}{2}}\sqrt{\widehat{\mathrm{var}}\left(\widehat{H}_k(t)\right)}\right)$
- Smoothed by B-splines



	(1)	(2)	(3)	(4)	(5)	(6)
Treatment effect	N	Y	N	N	Y	Y
Prognostic bmks	0	0	20	0	0	20
Predictive bmks	0	0	0	15	15	15

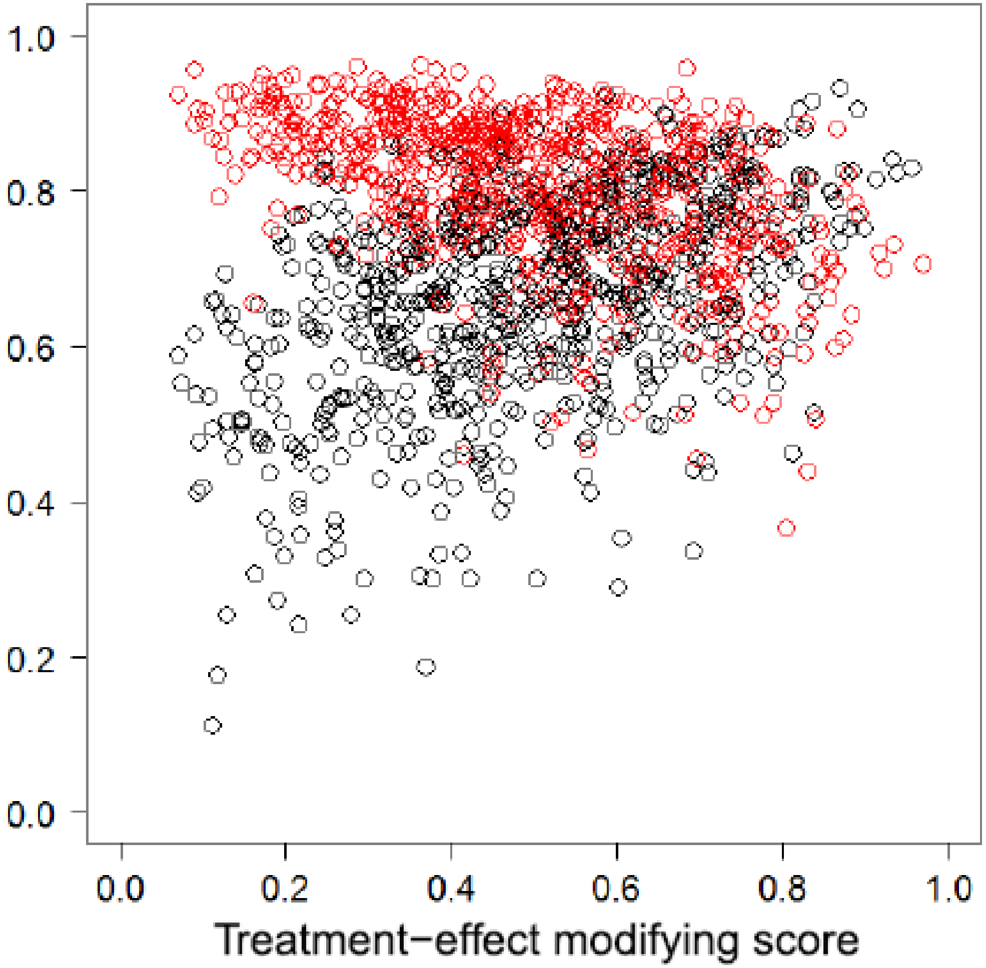

Multiple treatment−effect modifiers
No prognostic effect

$$\hat{\eta}_i = \hat{\gamma}^\top X_i$$

Accounting for prognostic biomarkers

Ternès, Rotolo, Michiels (2017a)

$$h(t; T_i, \boldsymbol{X_i}) = h_0(t) \exp\{\alpha T_i + \boldsymbol{\beta}^\intercal \boldsymbol{X_i} + \boldsymbol{\gamma}^\intercal T_i \boldsymbol{X_i}\}$$

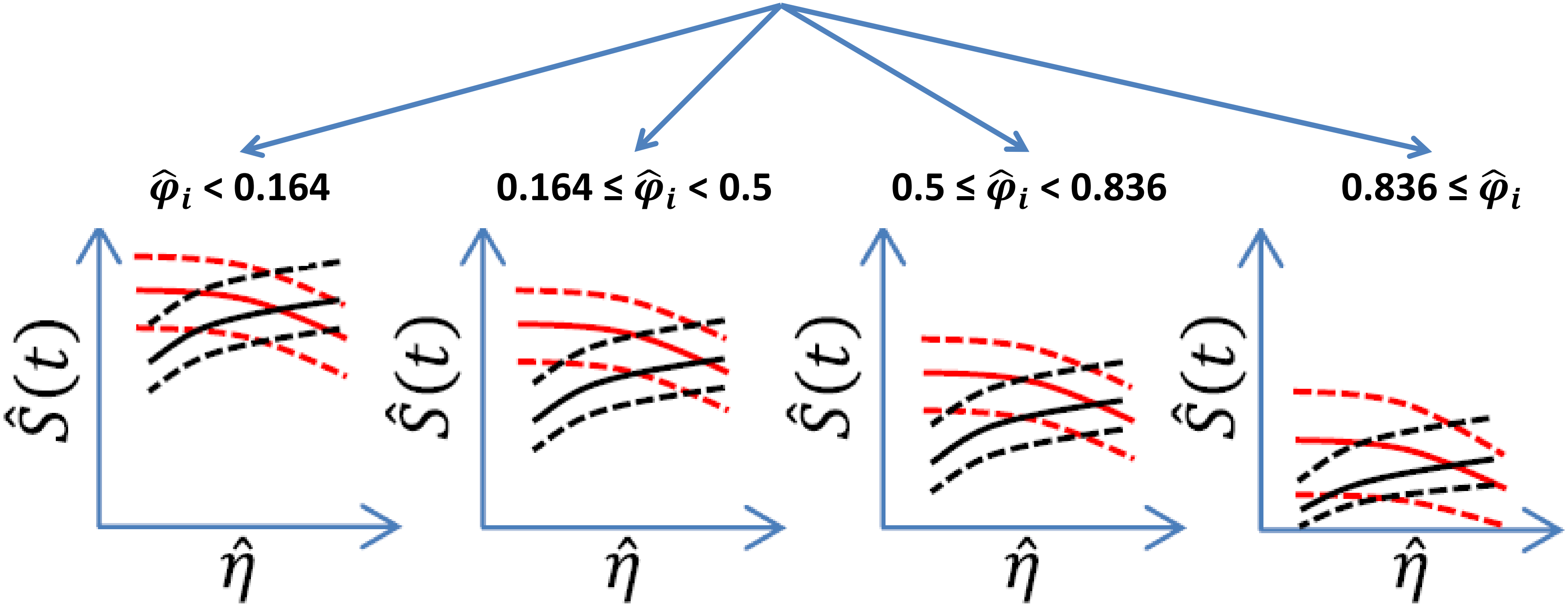


$$\widehat{\boldsymbol{\eta}}_{\boldsymbol{i}} = \widehat{\boldsymbol{\gamma}}^\intercal \boldsymbol{X_i}$$

Accounting for prognostic biomarkers

$$h(t; T_i, \boldsymbol{X_i}) = h_0(t)\exp\{\alpha T_i + \boldsymbol{\beta}^\intercal \boldsymbol{X_i} + \boldsymbol{\gamma}^\intercal T_i \boldsymbol{X_i}\}$$

Prognostic classes based on $\widehat{\varphi}_i = \widehat{\boldsymbol{\beta}}^\intercal \boldsymbol{X_i}$
using percentiles 16.4%, 33.6%, 33.6%, 16.4% (Cox 1957)



$\widehat{\varphi}_i < 0.164$

$0.164 \leq \widehat{\varphi}_i < 0.5$

$0.5 \leq \widehat{\varphi}_i < 0.836$

$0.836 \leq \widehat{\varphi}_i$

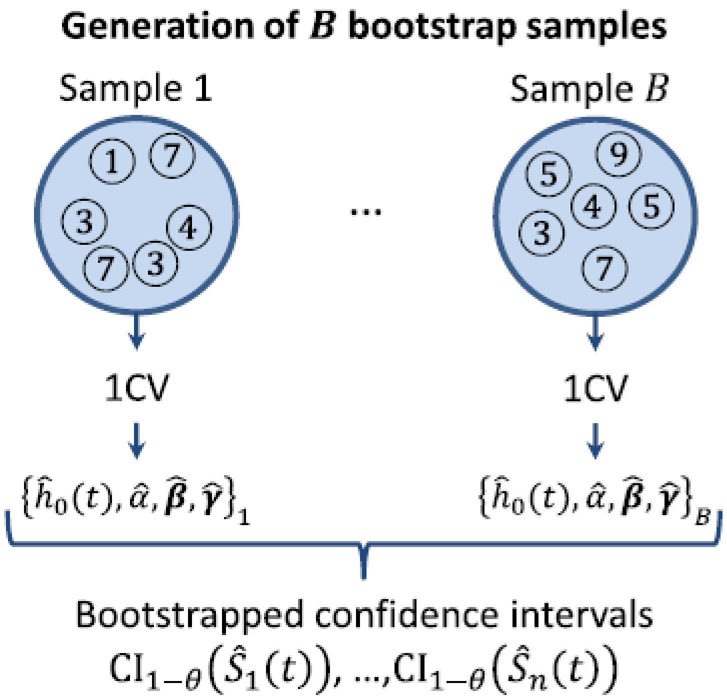
Estimation of the survival probability

Ternès, Rotolo, Michiels (2017a)

Confidence interval estimation:

- Analytic $\mathrm{IC}_{1-\theta}\left(\hat{S}_k(t)\right) = \exp\left(-\widehat{H}_k(t) \pm z_{1-\frac{\theta}{2}}\sqrt{\widehat{\mathrm{var}}\left(\widehat{H}_k(t)\right)}\right)$
- Smoothed by splines

either in the original data or
in bootstrap samples



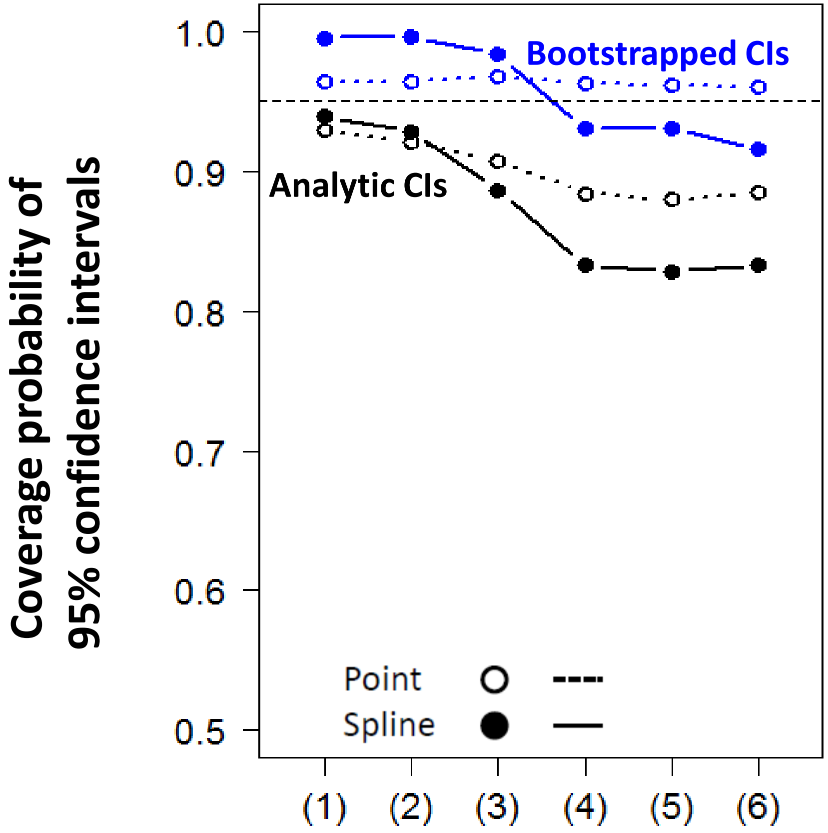
Generation of B bootstrap samples

Sample 1 … Sample B

1CV → $\{\hat{h}_0(t), \hat{\alpha}, \hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\gamma}}\}_1$

1CV → $\{\hat{h}_0(t), \hat{\alpha}, \hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\gamma}}\}_B$

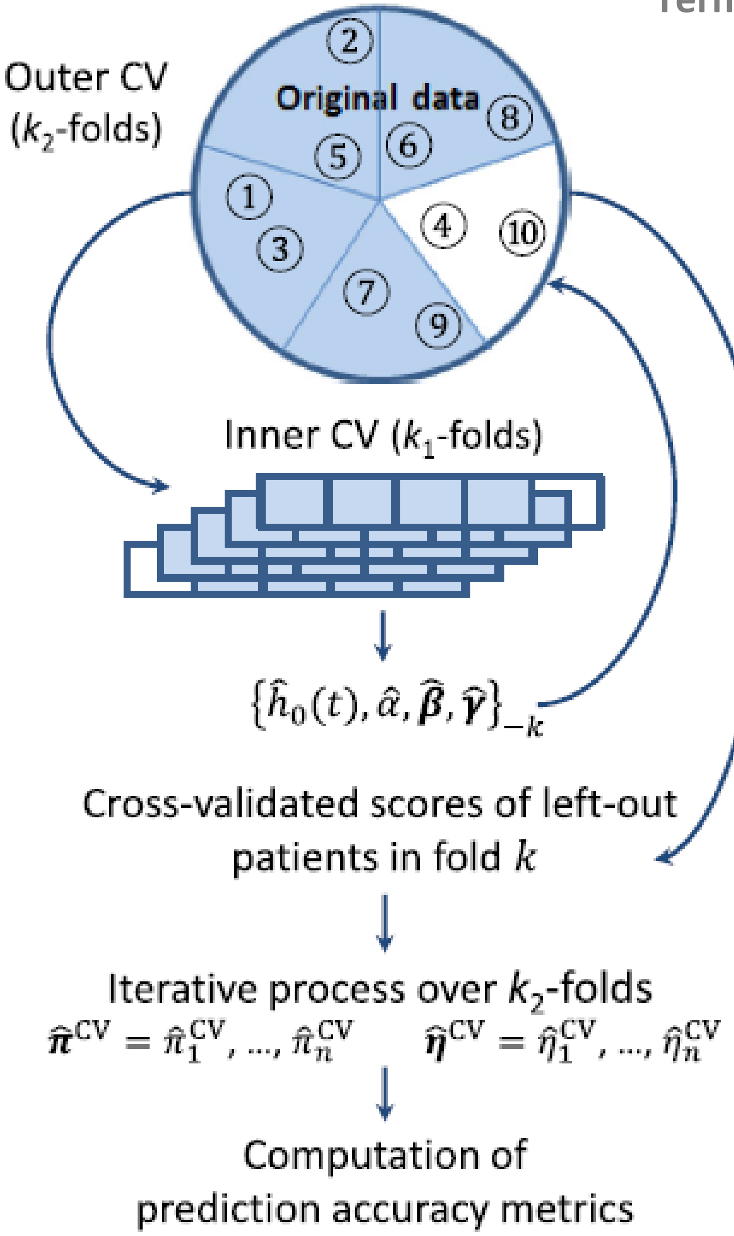
Bootstrapped confidence intervals
$\mathrm{CI}_{1-\theta}\left(\hat{S}_1(t)\right), \ldots, \mathrm{CI}_{1-\theta}\left(\hat{S}_n(t)\right)$



	(1)	(2)	(3)	(4)	(5)	(6)
Treatment effect	N	Y	N	N	Y	Y
Prognostic bmks	0	0	20	0	0	20
Predictive bmks	0	0	0	15	15	15

Controlling for overfitting

Ternès, Rotolo, Michiels (2017a)



Outer CV (k_2-folds)

Inner CV (k_1-folds)

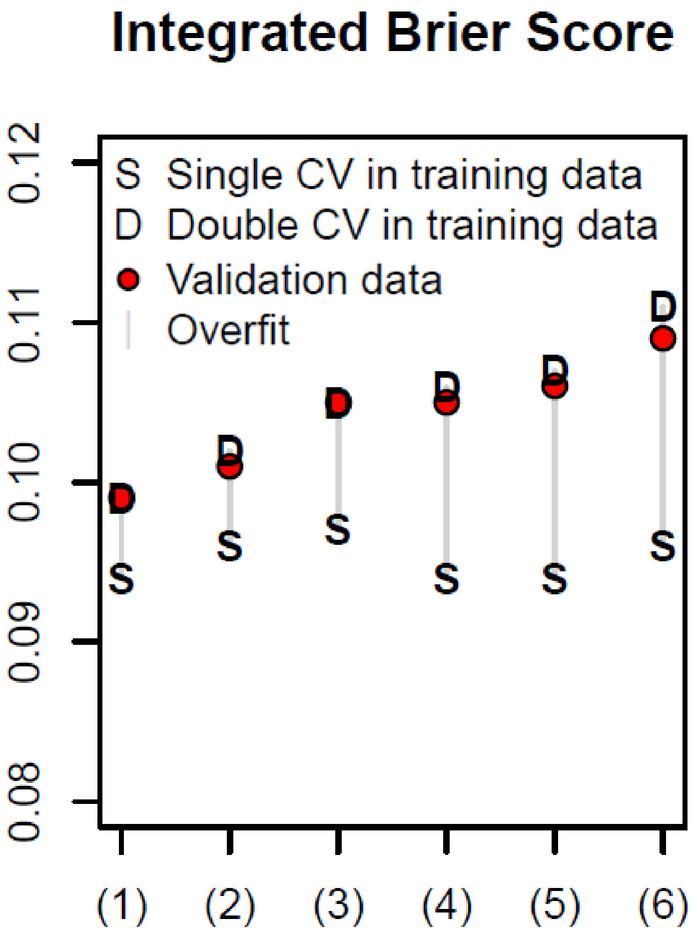
$$\{\hat{h}_0(t), \hat{\alpha}, \hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\gamma}}\}_{-k}$$

Cross-validated scores of left-out patients in fold k

Iterative process over k_2-folds

$$\hat{\boldsymbol{\pi}}^{\mathrm{CV}} = \hat{\pi}_1^{\mathrm{CV}}, \ldots, \hat{\pi}_n^{\mathrm{CV}} \qquad \hat{\boldsymbol{\eta}}^{\mathrm{CV}} = \hat{\eta}_1^{\mathrm{CV}}, \ldots, \hat{\eta}_n^{\mathrm{CV}}$$

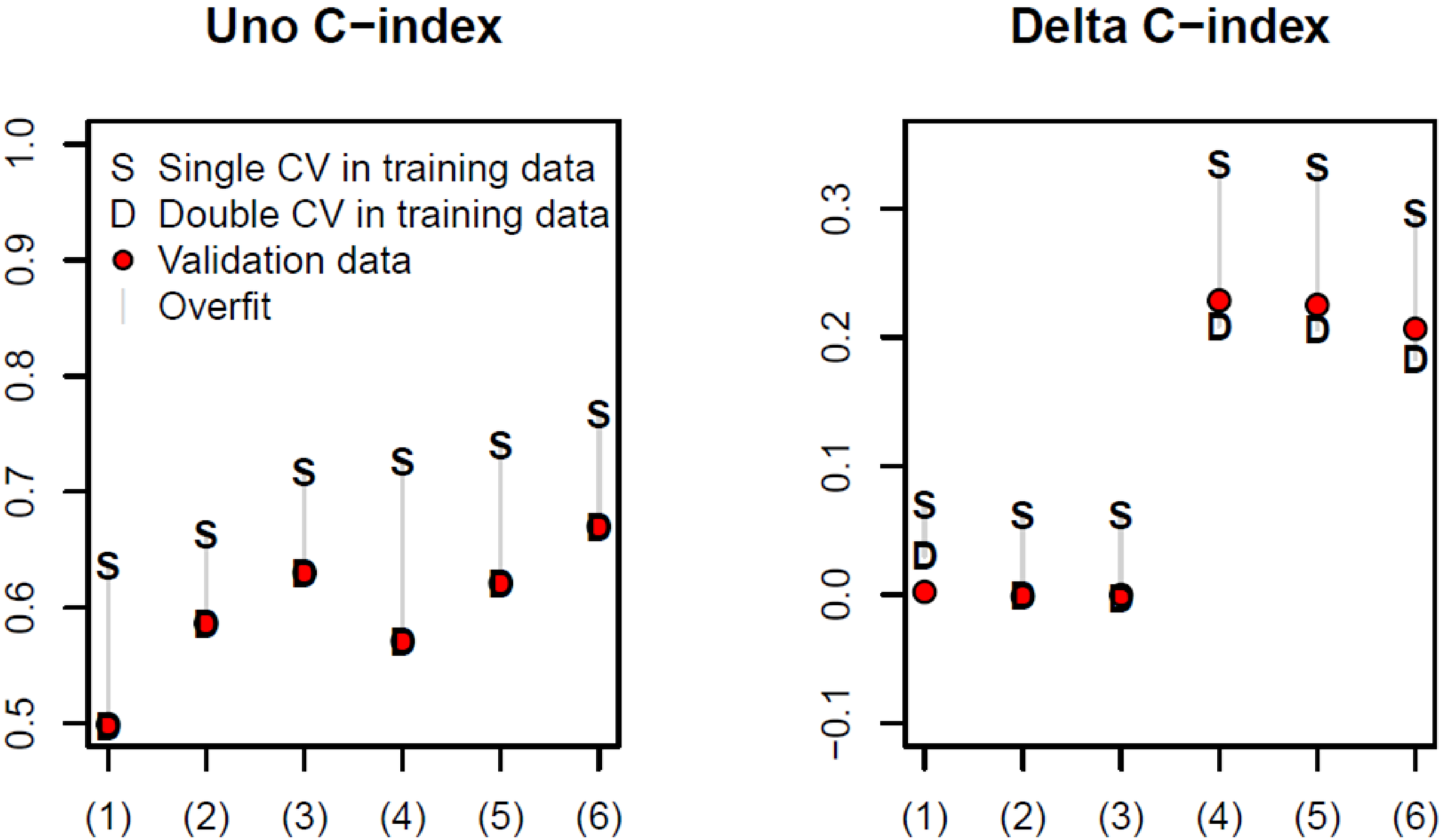
Computation of prediction accuracy metrics



Treatment effect	N	Y	N	N	Y	Y
Prognostic bmks	0	0	20	0	0	20
Predictive bmks	0	0	0	15	15	15

Controlling for overfitting

Ternès, Rotolo, Michiels (2017a)



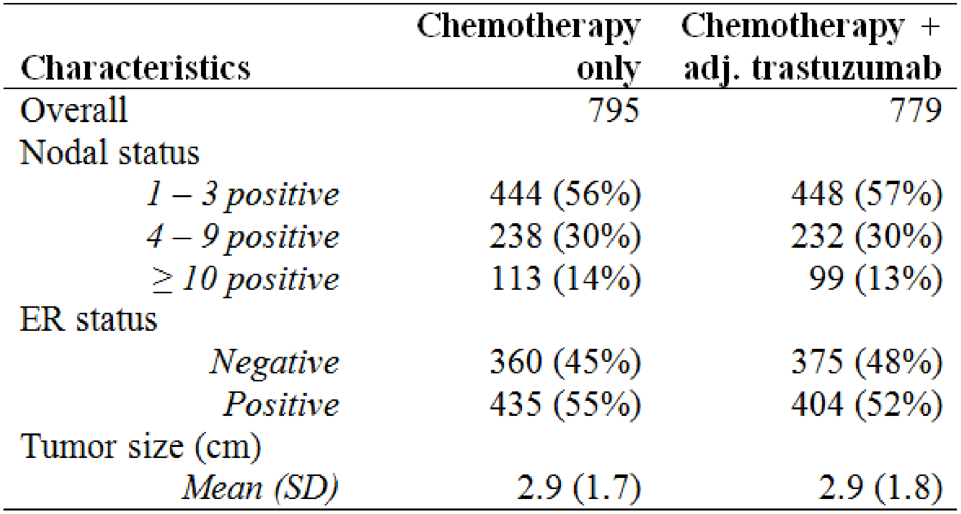
N	Y	N	N	Y	Y	**Treatment effect**	N	Y	N	N	Y	Y
0	0	20	0	0	20	**Prognostic bmks**	0	0	20	0	0	20
0	0	0	15	15	15	**Predictive bmks**	0	0	0	15	15	15

RCT in early breast cancer

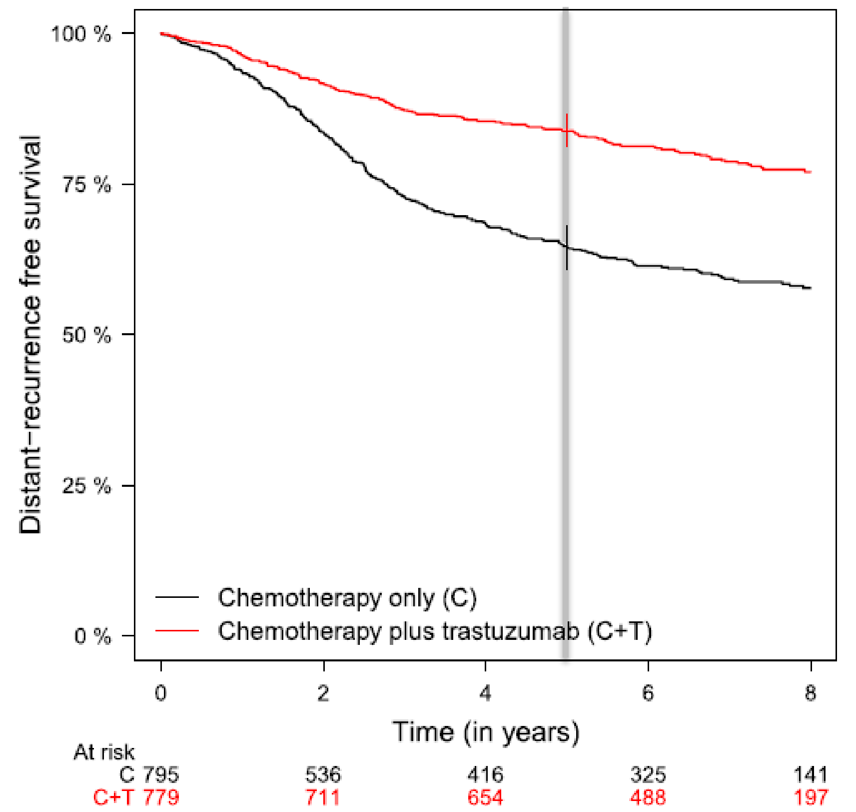
Pogue-Geile et al (2013)

Retrospective biomarker study in RCT of early breast K patients

⇒ Randomized clinical trial (**n** = 1574 patients, **p** = 462 genes)



Characteristics	Chemotherapy only	Chemotherapy + adj. trastuzumab
Overall	795	779
Nodal status		
1 – 3 positive	444 (56%)	448 (57%)
4 – 9 positive	238 (30%)	232 (30%)
≥ 10 positive	113 (14%)	99 (13%)
ER status		
Negative	360 (45%)	375 (48%)
Positive	435 (55%)	404 (52%)
Tumor size (cm)		
Mean (SD)	2.9 (1.7)	2.9 (1.8)



At risk					
C	795	536	416	325	141
C+T	779	711	654	488	197

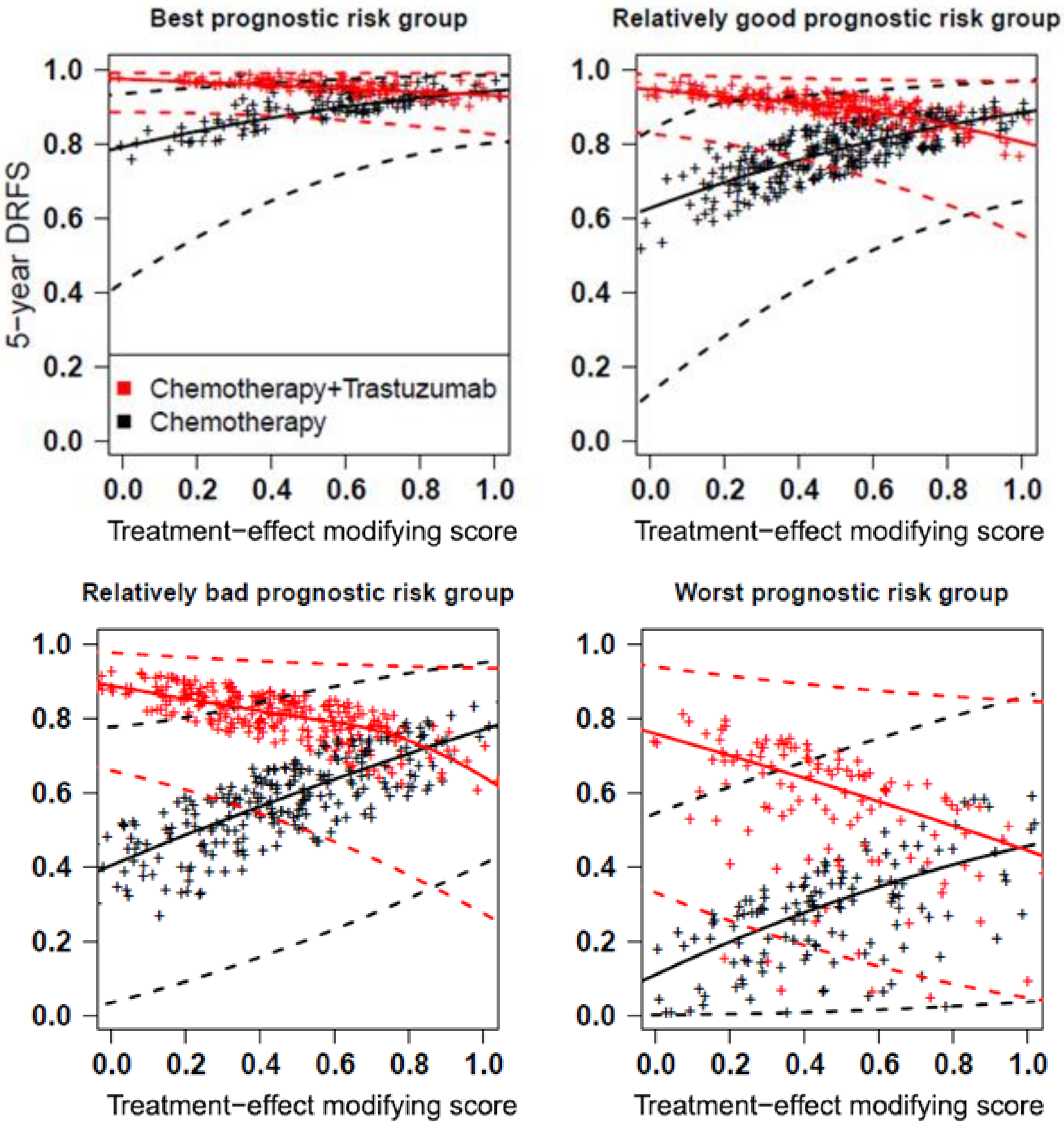
Fig. 3 Arm-specific distant-recurrence free survival in the illustrated breast cancer trial. Vertical lines: 95% confidence interval at 5 years

Developed signature

with the ALASSO penalty

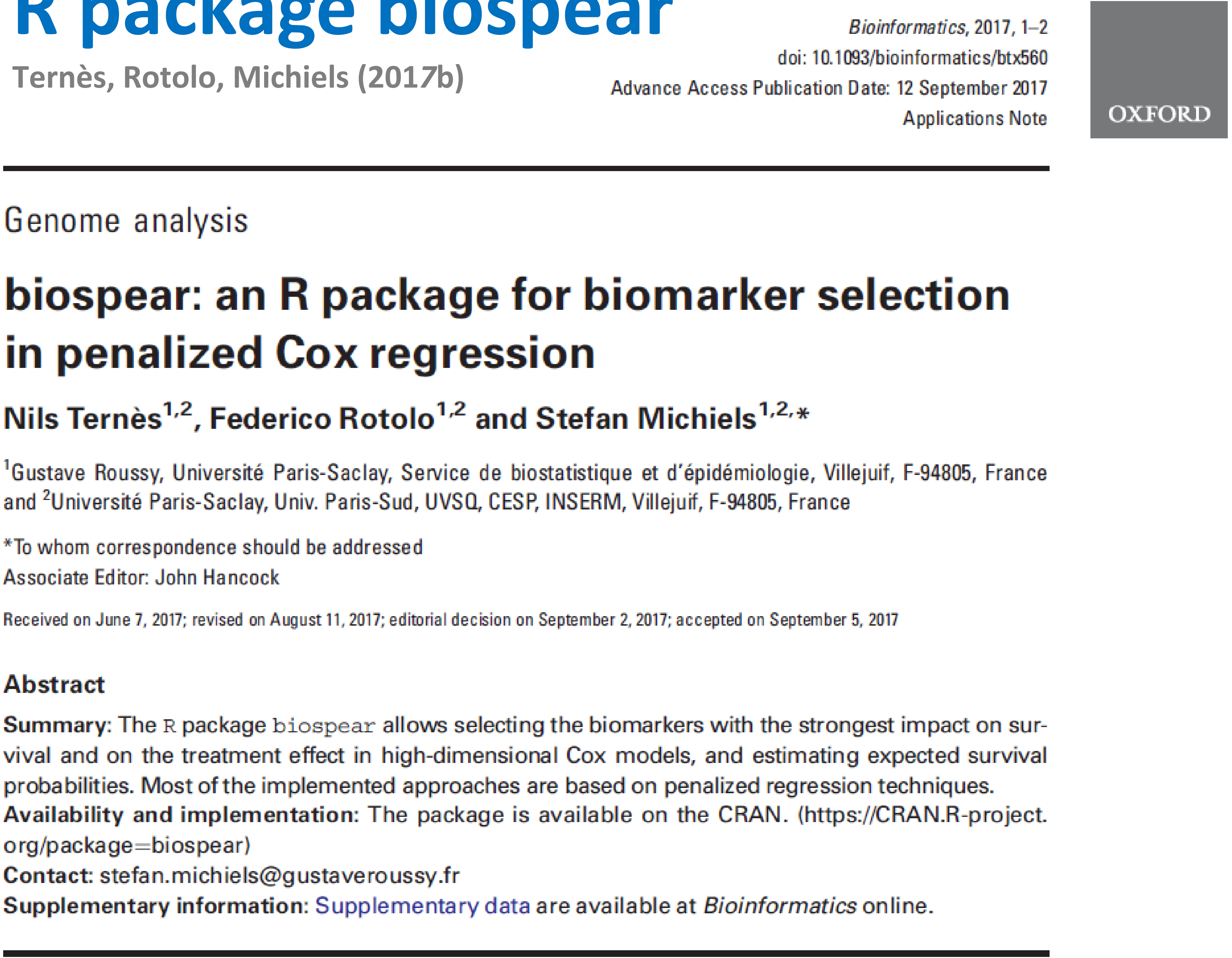
Prognostic component	
Clinical variables (4)	Treatment, ER status, Tumor size, Nodal status
Genomic variables (p = 98)	ACTB, ADCYAP1, ANGPTL4, ARL8A, BBC3, BDH2, CAPS, CASC3,CCDC74A, CDC6,CDH3, CFLP1, CSNK1A1, *CSNK1D*, CXXC5,DHPS, DNAJC4, DPY19L4, ELAVL4, ELN, ENO1, ERBB4, FABP5,FAM84B, FBXW11,FKSG30, FLJ22659, FLJ22795, FLJ35390, FRAG1,FRMD4A, GHR, GPRIN1, GSN, HIST1H2AA,HIST2H2BE, IDUA, IGJ,IGKV2.24, ILF2,KCNE4, KIAA1920, KIF2C, KRT81, L3MBTL2,LCE3E, LOC400590, MAD2L2, MAP3K13,MBOAT2, **MED13L**, METTL3, MSI2, MTCH2, MVP, NAT1, NAT10, NDC80, NECAB3,NXPH3, OGFR, PCK2,PGM5, PHGDH,PITPNC1, PRPF40A, PTTG1, RBM14, RELB, RHBDD1, RND3, RPL34, RPS2, SFRP1,SLC25A28, SLC25A31,SLC25A5, SLC30A10,SLC6A19, SMCP, *SOX4,SPDEF*, SPP1, ST6GALNAC4, STEAP3, STK11IP,SULT1A2,TBXAS1, TCEB2, TFRC,TMSB10, TRABD, TUBB2C, UBE2W, UGDH, XYLT1, ZNF592, ZNF609
Treatment-effect modifying component	
Genomic variables (p = 24)	ATAD3A, C16orf14, C1orf93, *CCL21*, *CD9,CIAPIN1*, CLIC1, DKFZP434A0131, FAM148A,FNDC4, FURIN, KRTAP2.4, **MED13L**, MIA, MMD, ORMDL3, RPLP0,SIAH2, SLC39A14, SSBP2, THOP1, THRAP1,TMEM45B, UNC119
Prediction measures	
C-statistic (C)	0.80 (1CV), 0.67 (2CV)
ΔC-statistic (ΔC)	0.23 (1CV), 0.02 (2CV)

Graphical illustration



R package biospear

Ternès, Rotolo, Michiels (2017b)



Bioinformatics, 2017, 1–2
doi: 10.1093/bioinformatics/btx560
Advance Access Publication Date: 12 September 2017
Applications Note

OXFORD

Genome analysis

biospear: an R package for biomarker selection in penalized Cox regression

Nils Ternès[1,2], Federico Rotolo[1,2] and Stefan Michiels[1,2,*]

[1]Gustave Roussy, Université Paris-Saclay, Service de biostatistique et d'épidémiologie, Villejuif, F-94805, France and [2]Université Paris-Saclay, Univ. Paris-Sud, UVSQ, CESP, INSERM, Villejuif, F-94805, France

*To whom correspondence should be addressed

Associate Editor: John Hancock

Received on June 7, 2017; revised on August 11, 2017; editorial decision on September 2, 2017; accepted on September 5, 2017

Abstract

Summary: The `R` package `biospear` allows selecting the biomarkers with the strongest impact on survival and on the treatment effect in high-dimensional Cox models, and estimating expected survival probabilities. Most of the implemented approaches are based on penalized regression techniques.
Availability and implementation: The package is available on the CRAN. (https://CRAN.R-project.org/package=biospear)
Contact: stefan.michiels@gustaveroussy.fr
Supplementary information: Supplementary data are available at *Bioinformatics* online.

Acknowledgments

- This work was supported (PhD scholarship) by the *Foundation Philanthropia Lombard-Odier*

- We thank the
  NSABP investigators of the B-31 trial
  who submitted data from the study to dbGaP
  (Study Accession:phs000826.v1.p1)
  and the NIH data repository
  The B-31 trial was supported by
  - NCI (Grants U10-CA-12027, U10-CA-69651, U10-CA-37377, and U10-CA-69974)
  - a grant from the Pennsylvania Department of Health.

Scientific output of the project

- Michiels, Ternès, Rotolo (2016). Statistical Controversies in Clinical Research: Prognostic gene signatures are not (yet) useful in clinical practice. *Ann Oncol.* DOI: 10.1093/annonc/mdw307

- Ternès, Rotolo, Michiels (2016). Empirical extensions of the lasso penalty to reduce the false discovery rate in high-dimensional Cox regression models. *Stat Med*. DOI: 10.1002/sim.6927

- Ternès, Rotolo, Heinze, Michiels (2017). Identification of biomarker-by-treatment interactions in randomized clinical trials with survival outcomes and high-dimensional spaces. *Biom J*. DOI: 10.1002/bimj.201500234

- Ternès, Rotolo, Michiels (2017a). Robust estimation of the expected survival probabilities from high-dimensional Cox models with biomarker-by-treatment interactions in randomized clinical trials, *BMC Med Res Methodol*. DOI: 10.1186/s12874-017-0354-0

- Ternès, Rotolo, Michiels (2017b). biospear: an R package for biomarker selection in penalized Cox regression, *Bioinformatics*. DOI: 10.1093/bioinformatics/btx560

Bibliography

- Benner, Zucknick, Hielscher, Ittrich, Mansmann (2010). High-dimensional Cox models: the choice of penalty as part of the model building process. *Biom J*
- Cox (1957). Note on grouping. *J Am Stat Assoc*
- Desmedt, Di Leo, de Azambuja et al (2011). Multifactorial approach to predicting resistance to anthracyclines. *J Clin Oncol*
- Hatzis, Pusztai, Valero et al (2011). A genomic predictor of response and survival following taxane-anthracycline chemotherapy for invasive breast cancer. *JAMA*
- Meinshausen, Bühlmann (2006). High-dimensional graphs and variable selection with the Lasso. *Ann Stat*
- Pogue-Geile, Kim, Jeong, Tanaka et al (2013). Predicting degree of benefit from adjuvant trastuzumab in NSABP trial B-31. *J Natl Cancer Inst*
- Tian, Alizadeh, Gentles, Tibshirani (2014). A simple method for estimating interactions between a treatment and a large number of covariates. *J Am Stat Assoc*
- Tibshirani (1996). Regression shrinkage and selection via the lasso. *J R Stat Soc B*
- Tibshirani (1997). The lasso method for variable selection in the Cox model. *Stat Med*
- Yuan, Lin (2006). Model selection and estimation in regression with grouped variables. *J R Stat Soc B*
- Zou (2006). The adaptive lasso and its oracle properties. *J Am Stat Assoc*
- Zhang, Lu (2007). Adaptive lasso for Cox's proportional hazards model. *Biometrika*