Identifying predictive biomarkers in high-dimensional genomic data from randomized clinical trials

Jean-Philippe Vert

8th SFdS International Meeting on Statistical Methods in BIopharmacy, Paris, September 15, 2017

Joint work with

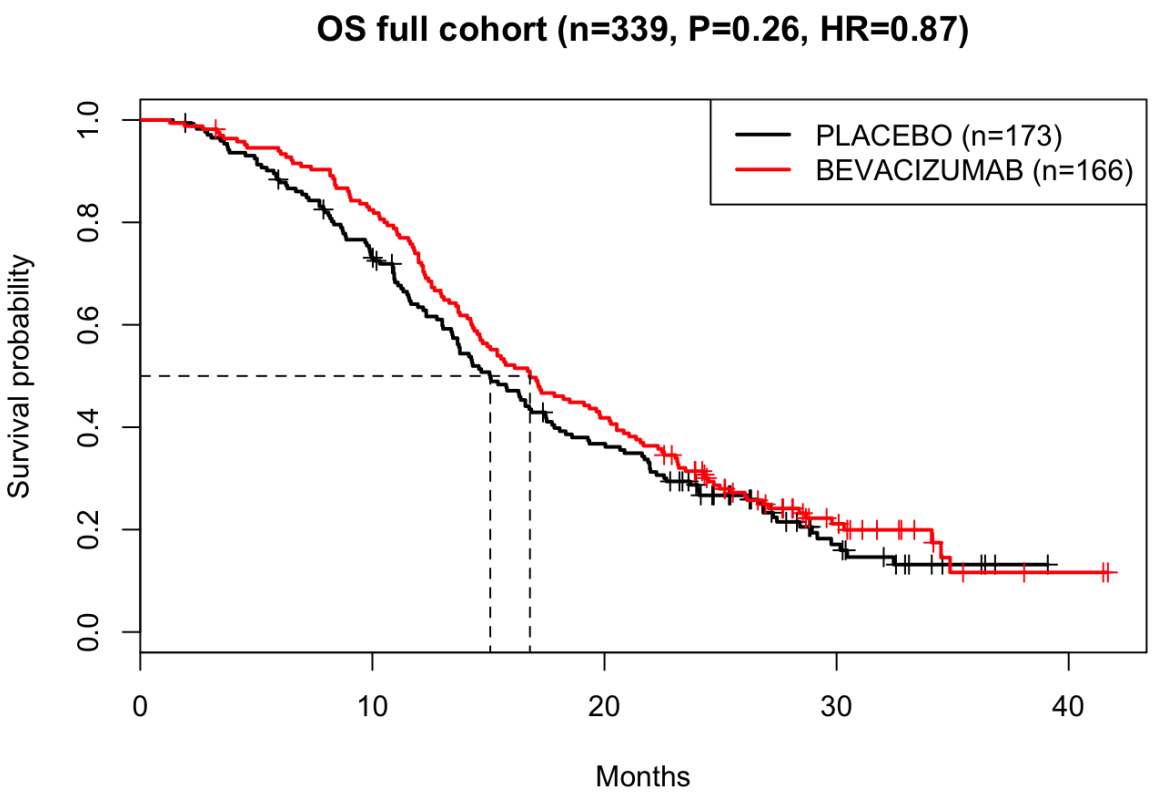
Richard Bourgon @Genentech

Genentech
A Member of the Roche Group

Outline

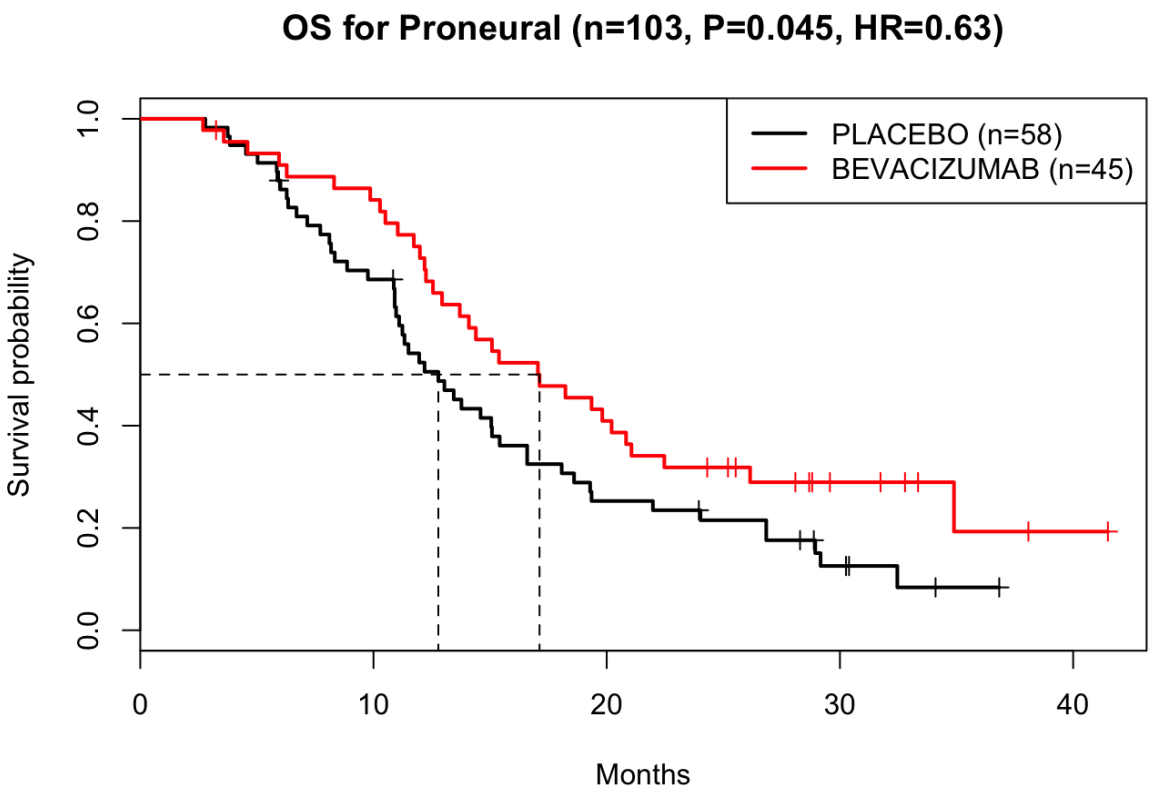
Outline

Motivation



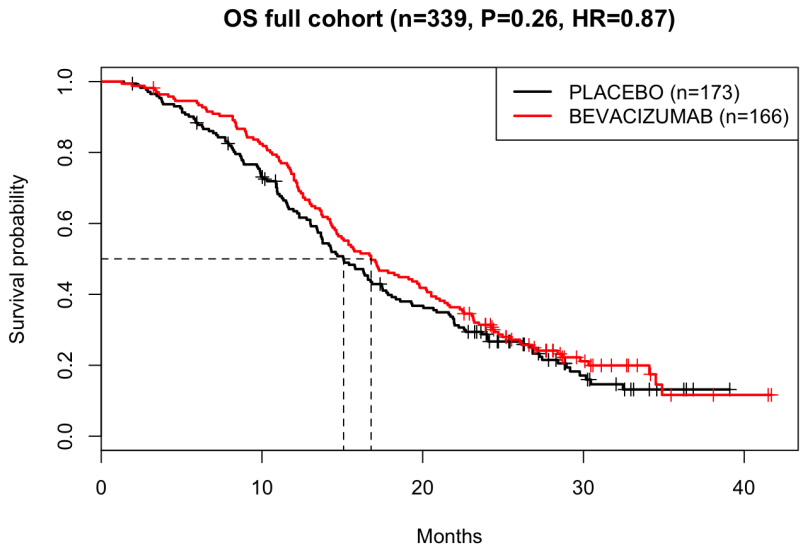
- Avaglio phase III two-arm randomized clinical trial : Bevacizumab (Avestin) vs. placebo + standard-of-care therapy in newly diagnosed glioblastoma (Chinot et al., 2014)
- Improvement in progression-free survival, not in overall survival

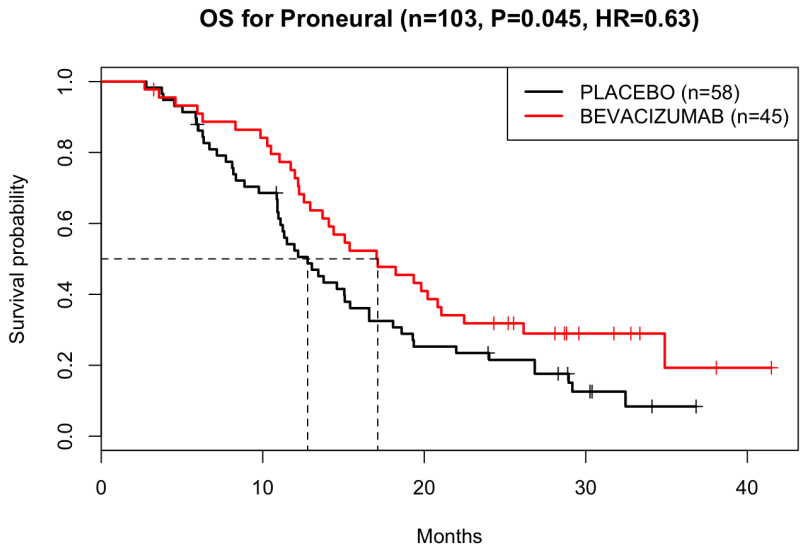
Subgroup analysis



- Post-trial analysis restricted to subgroups based on gene expression data (Sandmann et al., 2015)
- Phillips classification: Mesenchymal / Proliferative / Proneural
- OS benefit in one subgroup (proneural)

Question

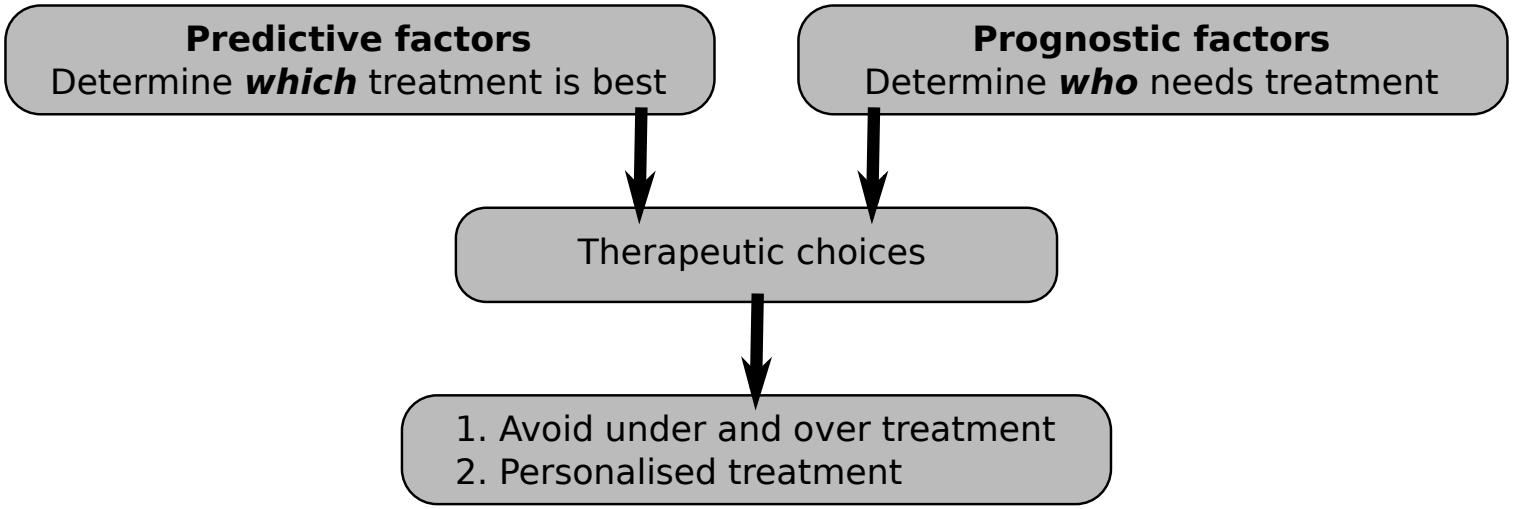




Given the results of a clinical trial, can we automatically learn a decision function to stratify patients based on whether or not they will benefit from the treatment?

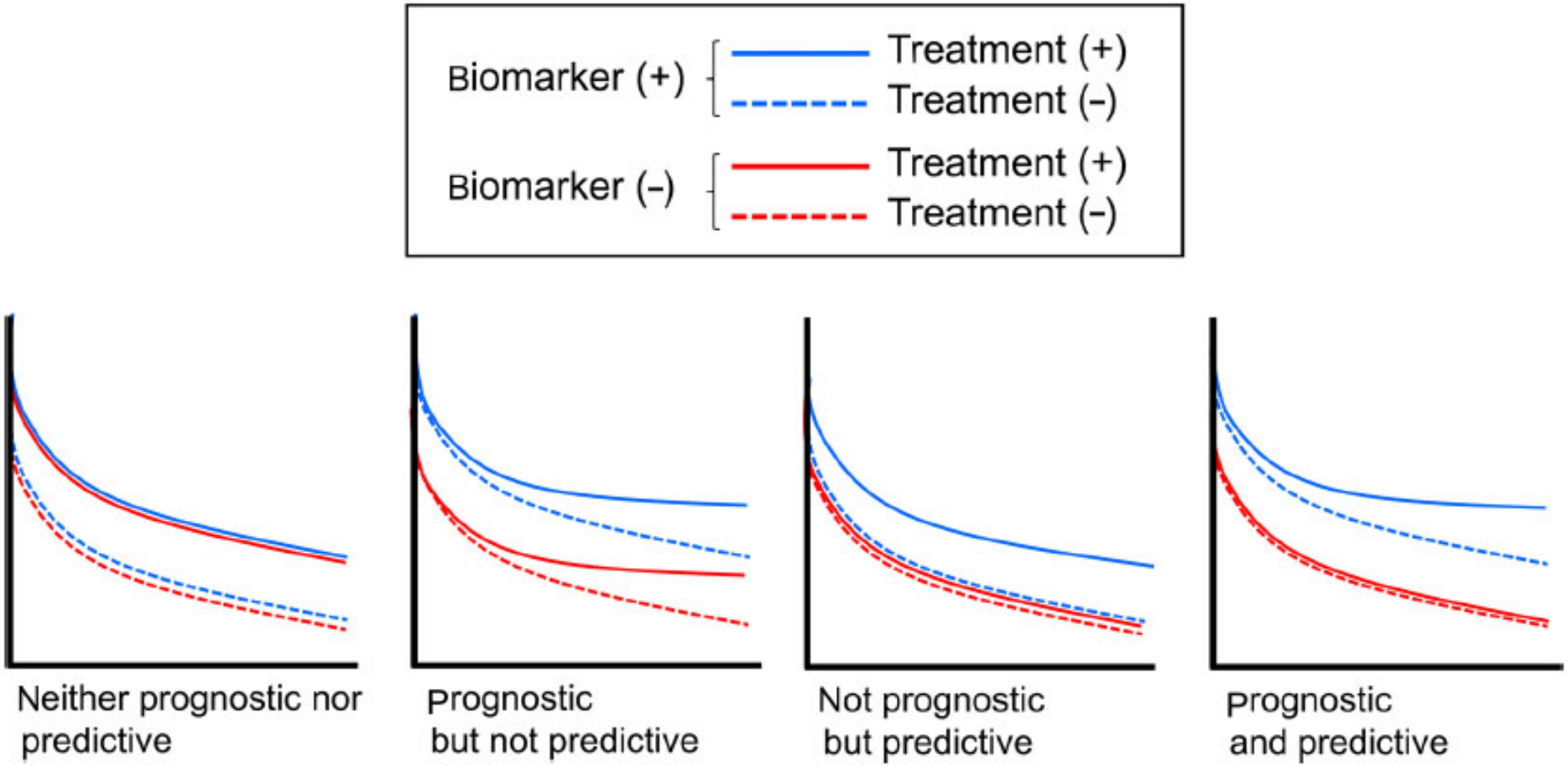
A.k.a. can we learn a predictive marker to identify the optimal treatment for each patient?

Predictive vs Prognostic marker



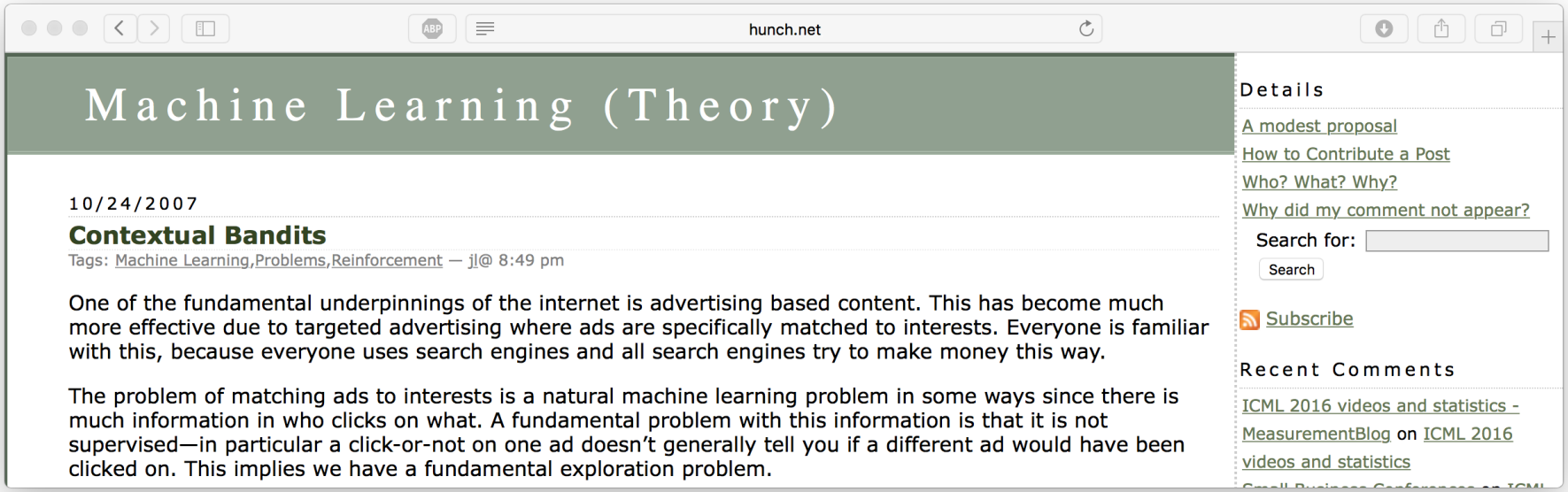
- Prognostic: provides information on the likely outcome of the disease in an untreated individual
- Predictive: provides information on the likely benefit from treatment

Predictive vs Prognostic marker



Difficulty

- For each patient, we only observe the output under one treatment option
- Therefore, it is not possible to simply train a model to discriminate the output with or without treatment.
- Similar to contextual multi-armed bandit problem in e-marketing



More formally

For each patient we have:

- Patient covariates (clinical, transcriptome...): $X \in \mathbb{R}^p$
- Treatment given (randomized arm): $A \in \{-1, 1\}$
- Response (right-censored survival): $R = (Y, \delta) \in \mathbb{R} \times \{0, 1\}$

We want to infer a model for response/hazard of the form

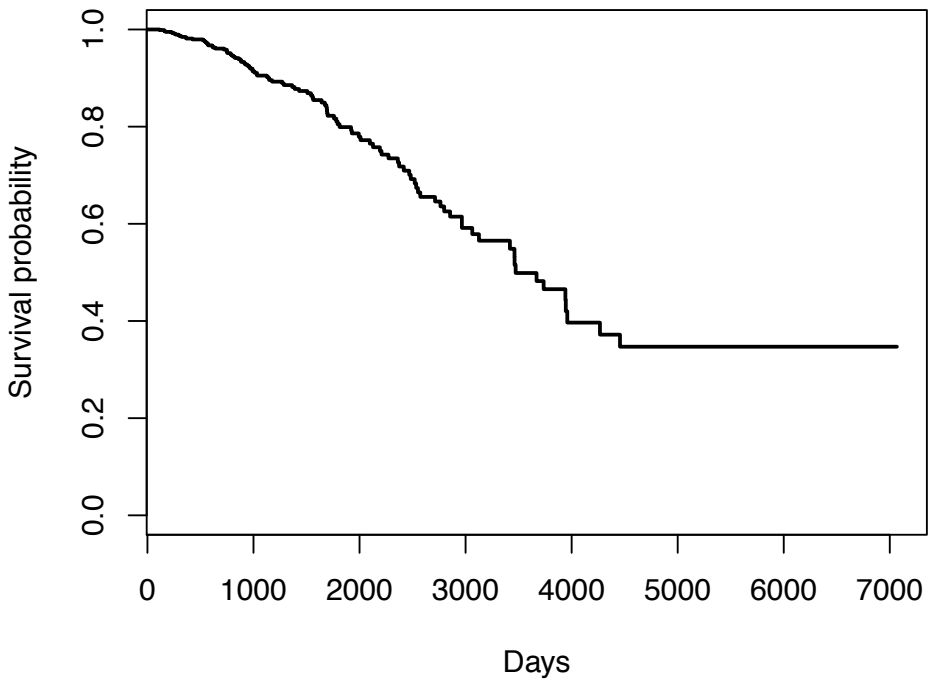
$$\Phi(R(X, A)) = f(X) + g(A) + Ah(X)$$

where

- $f(X)$ is the main patient effect independently of treatment (prognostic)
- $g(A)$ is the main treatment effect, independently of patient (good old drugs)
- $h(X)$ is the patient-specific drug effect (predictive)

Outline

Survival regression



- Patient covariates (clinical, transcriptome...): $X \in \mathbb{R}^p$
- Response (right-censored survival): $R = (Y, \delta) \in \mathbb{R} \times \{0, 1\}$
- Goal: "predict R from X"
- More realistic/useful: predict a score $f(X)$ such as "patient X_1 has a higher risk than patient X_2 is $f(X_1) > f(X_2)$"

Cox proportional hazard model (Cox, 1972)

- Proportional hazard hypothesis: $\lambda(t \mid x) = \lambda_0(t) \exp(\beta^\top x)$
- Model: $f(x) = \beta^\top x := \eta$
- Patient i:
  - $x_i \in \mathbb{R}^p$ covariates
  - $(y_i, \delta_i) \in \mathbb{R} \times \{0, 1\}$ right-censored survival data
  - $R_i = \{j \,:\, y_j \geq y_i\}$ patients at risk at time y_i
- Conditional partial likelihood:

$$L(\beta) = \prod_{i=1}^{n} \left(\frac{e^{\eta_i}}{\sum_{j \in R_i} e^{\eta_i}} \right)^{\delta_i}$$

Cox model estimation

- Maximum conditional partial likelihood:

$$\hat{\beta} \in \arg\max_{\beta} L(\beta)$$

- Equivalently;

$$\hat{\beta} \in \arg\min_{\beta} \ell^{Cox}(X\beta)$$

with

$$\ell^{Cox}(\eta) = \sum_{i=1}^{n} \delta_i \left[-\eta_i + \log\left(\sum_{j \in R_i} e^{\eta_j} \right) \right] ,$$

- Convex optimization problem
- Not good if p is large (overfitting)

Cox model estimation when p is large

- We can regularize the problem, e.g., with a lasso (Tibshirani, 1997) or elastic net penalty (Zou and Hastie, 2005):

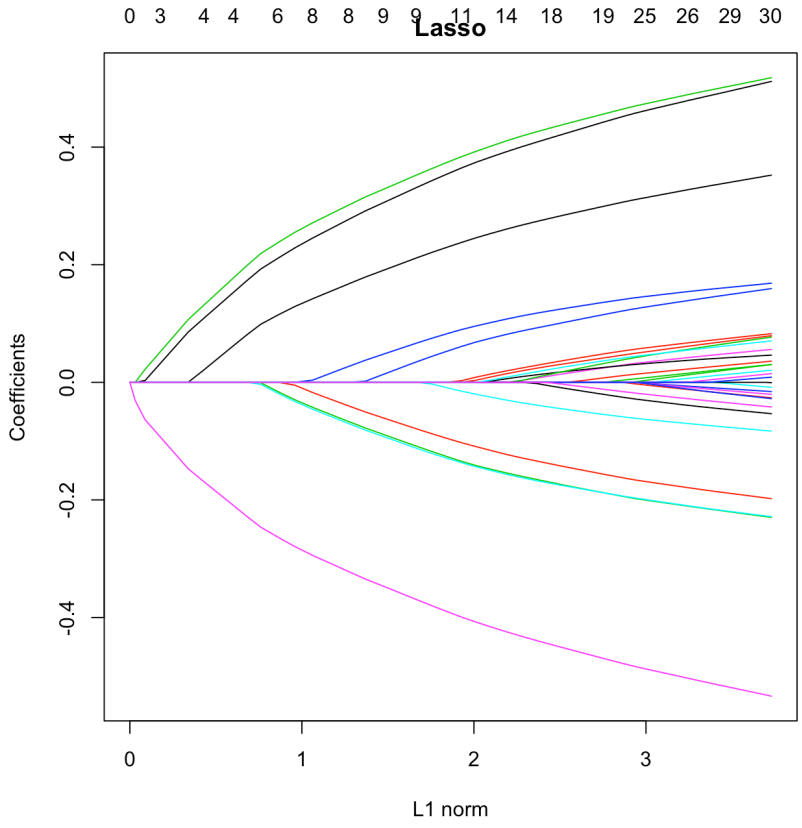
$$\min_{\beta \in \mathbb{R}^p} \frac{1}{n} \ell^{Cox}(X\beta) + \lambda P_\alpha(\beta)\,,$$

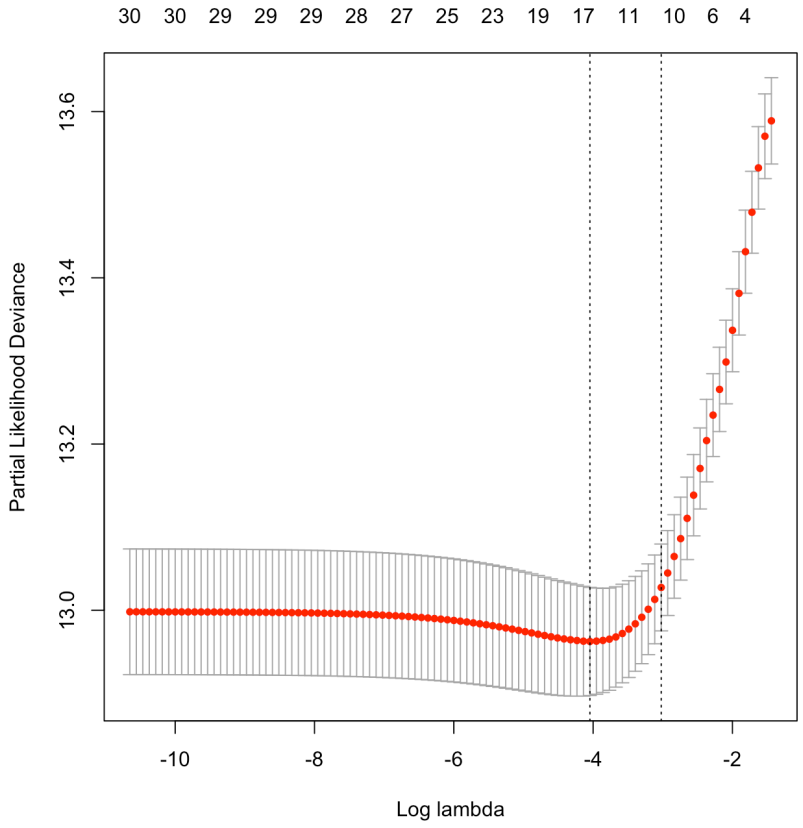
with

$$P_\alpha(\beta) = \alpha \sum_{j=1}^{p} |\beta_j| + \frac{1-\alpha}{2} \sum_{j=1}^{p} \beta_j^2\,.$$

- Regularization allows to learn in high dimension by controlling overfitting
- $\alpha > 0$ shrinks coefficients to 0 and leads to feature selection, leading to a molecular signature

Example


0 3 4 4 6 8 8 9 9 11 14 18 19 25 26 29 30
Lasso
Coefficients
0.4
0.2
0.0
-0.2
-0.4
0
1
2
3
L1 norm


30 30 29 29 29 28 27 25 23 19 17 11 10 6 4
Partial Likelihood Deviance
13.6
13.4
13.2
13.0
-10
-8
-6
-4
-2
Log lambda

Alternatives to Cox regression

- Extensions of machine learning techniques
  - Survival SVM (Van Belle et al., 2007)
  - Random survival forests (Ishwaran et al., 2008)
- Not adapted to learning a molecular signature
- We derive a new variant next, survival logistic regression

Concordance index

- $T_i = \{j : y_j > y_i\}$ patients with strictly longer survival
- Number of discordant pairs

$$n_d(\eta) = \sum_{i=1}^{n} \sum_{j \in T_i} \delta_i 1(\eta_i < \eta_j)$$

- Total number of comparable pairs

$$n_{total} = \sum_{i=1}^{n} \sum_{j \in T_i} \delta_i$$

- Concordance index:

$$CI(\eta) = 1 - \frac{n_d(\eta)}{n_{total}}$$

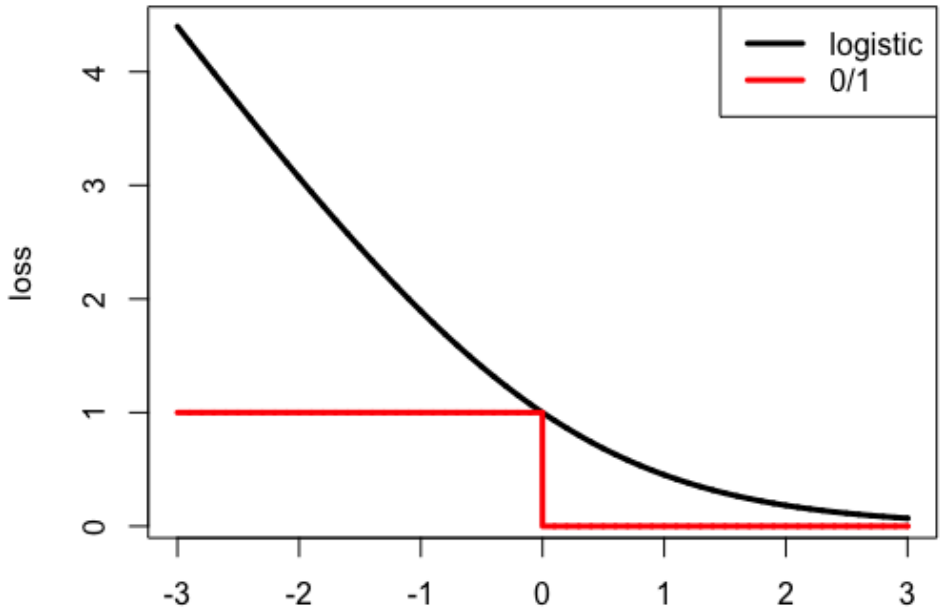
Optimizing the concordance index

- To fit a model β, one could consider:

$$\hat{\beta} = \underset{\beta}{\operatorname{argmax}}\ CI(X\beta) = \underset{\beta}{\operatorname{argmin}}\ n_d(X\beta)\,,$$

  but this is computationally intractable (NP-hard).
- Convex relaxation:

$$1(u < 0) \leq \log_2(1 + e^{-u})$$



Survival logistic regression

$$\min_{\beta \in \mathbb{R}^p} \frac{1}{n} \ell^{Survlog}(X\beta) + \lambda P_\alpha(\beta)\,,$$

with

$$\ell^{Survlog}(\eta) = \sum_{i=1}^{n} \sum_{j \in T_i} \delta_i \log_2 \left(1 + e^{\eta_j - \eta_i}\right)\,.$$

- $\ell^{Survlog}(\eta)$ is a convex upper bound of $n_d(\eta)$
- Convex optimization problem efficiently solved with the algorithm used in `glmnet`
- $\ell^{Survlog}(\eta)$ does not have an obvious likelihood interpretation, but also makes no assumption about the data such as proportional hazard
- Similar trick used, with the hinge loss, in the survival SVM (Van Belle et al., 2007)

Implementation

- C++ implementation in the `optreat` package (soon available..)
- Function `survenet` solves

$$\min_{\beta \in \mathbb{R}^p} \frac{1}{n} \ell(X\beta) + \lambda P_\alpha(\beta)\,,$$

  for $\ell = \ell^{Cox}$ and $\ell = \ell^{Survlog}$
- Syntax similar to `glmnet()`

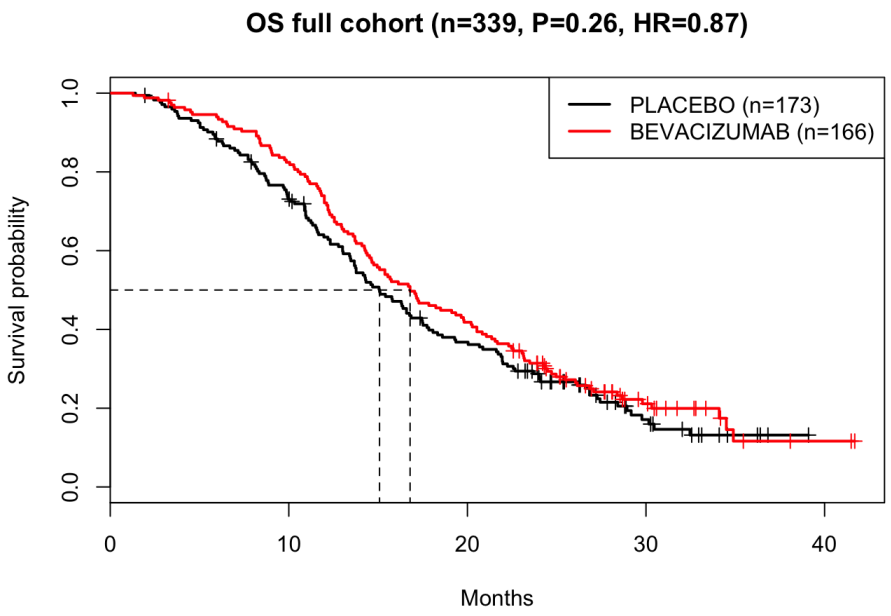
```
library(optreat)
m = survenet(x, y)  # by default, family="cox"
m = survenet(x, y, family="survlog")
m = survenet(x, y, family="survlog", nfolds=5)
plot(m)
predict(m, xtest, s="lambda.min")
```

Cox vs survival logistic regression

- Different objective functions
- Small n large p behaviour?
- Intuitive difference: survenet "uses" more pairs ($O(n^2)$) than Cox ($O(n)$), to be formalized
- Empirical comparison later

Outline

The problem



For each patient i we now have

- $x_i \in \mathbb{R}^p$ covariates
- $(y_i, \delta_i) \in \mathbb{R} \times \{0, 1\}$ survival data
- $a_i \in \{-1, 1\}$ treatment given

How to learn a function $f(x)$ to estimate the benefit of treatment?

Some strategies

Assume we have a model for survival regression (Cox or survival logistic):

1. Learn a survival model for each arm
2. Learn a unique survival model with interactions
3. Learn only the predictive model with the modified covariate trick

Standard model with interaction

- Model to capture treatment/covariate interactions

$$\eta(x, a) = x^\top \beta + \frac{1}{2} a x^\top \gamma \,,$$

where we add to x a constant covariate to account to drug main effect.

- Parameters estimation (e.g., Qian and Murphy, 2011)

$$\min_{\beta,\gamma} \frac{1}{n} \ell(X\beta + \frac{1}{2} AX\gamma) + \lambda P_\alpha(\beta) + \mu P_\alpha(\gamma)$$

- Scoring of a new patient:

$$s(x) = x^\top \gamma = \eta(x, a = 1) - \eta(x, a = -1)$$

is the predicted benefit (in "η" scale) of treating the patient

The "modified covariates" trick

- Tian et al. (2014) propose to replace

$$\ell(X\beta + \frac{1}{2}AX\gamma) = \ell(A * AX\beta + \frac{1}{2}AX\gamma)$$

by

$$\ell(\frac{1}{2}AX\gamma) = \ell(\tilde{X}\gamma)$$

where $\tilde{X} = AX/2$ are modified covariates

- Note that it bypasses the estimation of the main effect β
- In practice:
  1. Modify covariates by inverting columns corresponding to $a = -1$ arm
  2. Estimate a standard model on the modified covariates

Trick 1 justification

- Linear regression: if

$$\gamma_0 = \text{argmin } E(Y - \gamma \tilde{X})^2 \,,$$

i.e.

$$\gamma_0 \tilde{X} = E[Y \,|\, \tilde{X}] \,,$$

then

$$\begin{aligned} E[Y \,|\, X, A = 1] - E[Y \,|\, X, A = -1] \\ &= E[Y \,|\, \tilde{X} = X/2] - E[Y \,|\, \tilde{X} = -X/2] \\ &= \gamma_0 X/2 - (-\gamma_0 X/2) \\ &= \gamma_0 X \end{aligned}$$

- Similar justification for logistic and Cox regression.

Trick 2: Augmented model

- Estimator after covariate modification:

$$\min_{\gamma} \frac{1}{n} \ell(\tilde{X}\gamma) + \lambda P_{\alpha}(\gamma)\,,$$

- The following augmented model estimator is asymptotically the same, for any $r \in \mathbb{R}^n$, because $E[\tilde{X}] = 0$

$$\min_{\gamma} \frac{1}{n} \left[\ell(\tilde{X}\gamma) - r^{\top} \tilde{X}\gamma \right] + \lambda P_{\alpha}(\gamma)\,,$$

- Choose r to minimize the variance of the estimator, which is [...]:

$$r = E[\nabla \ell(0) \,|\, X]$$

- Two step procedure
  - Estimate r
  - Optimize the augmented model

Implementation

- Function `optreat` estimates the drug effect by combining:
  - A survival model: `cox` or `survlog`
  - A method: `interaction` or `modified` or `augmented`
- Syntax similar to `glmnet()`

```
library(optreat)
m = optreat(x, y, a)  # by default, family="cox",
                      # method="interaction"
m = optreat(x, y, a, family="survlog",
                      method="augmented")
m = optreat(x, y, a, family="survlog", nfolds=5)
plot(m)
predict(m, xtest, s="lambda.min")
```

Outline

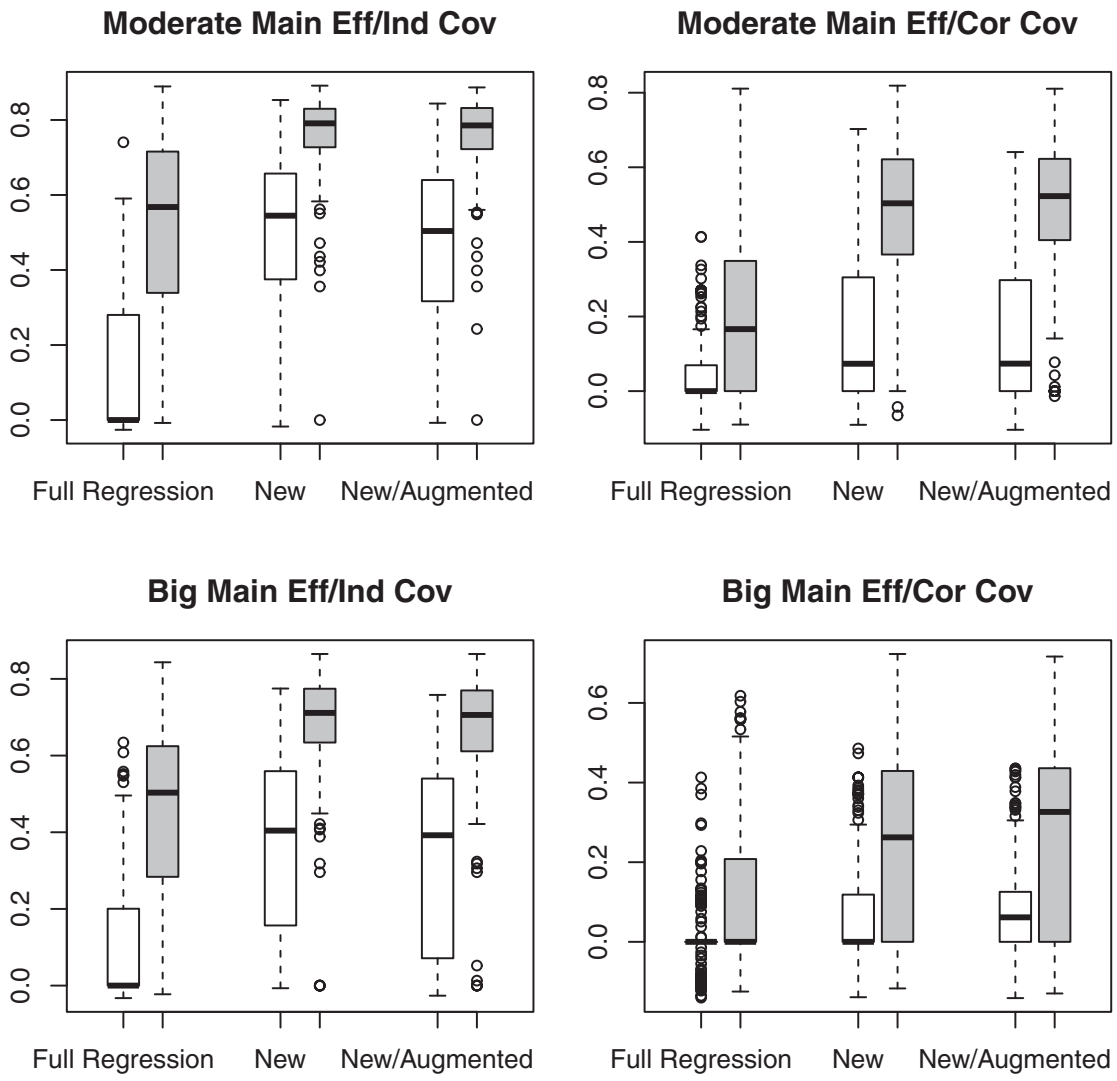
Simulations from Tian et al. (2014)

- Simulate X as a multivariate Gaussian, with or without correlation
- Simulate time according to

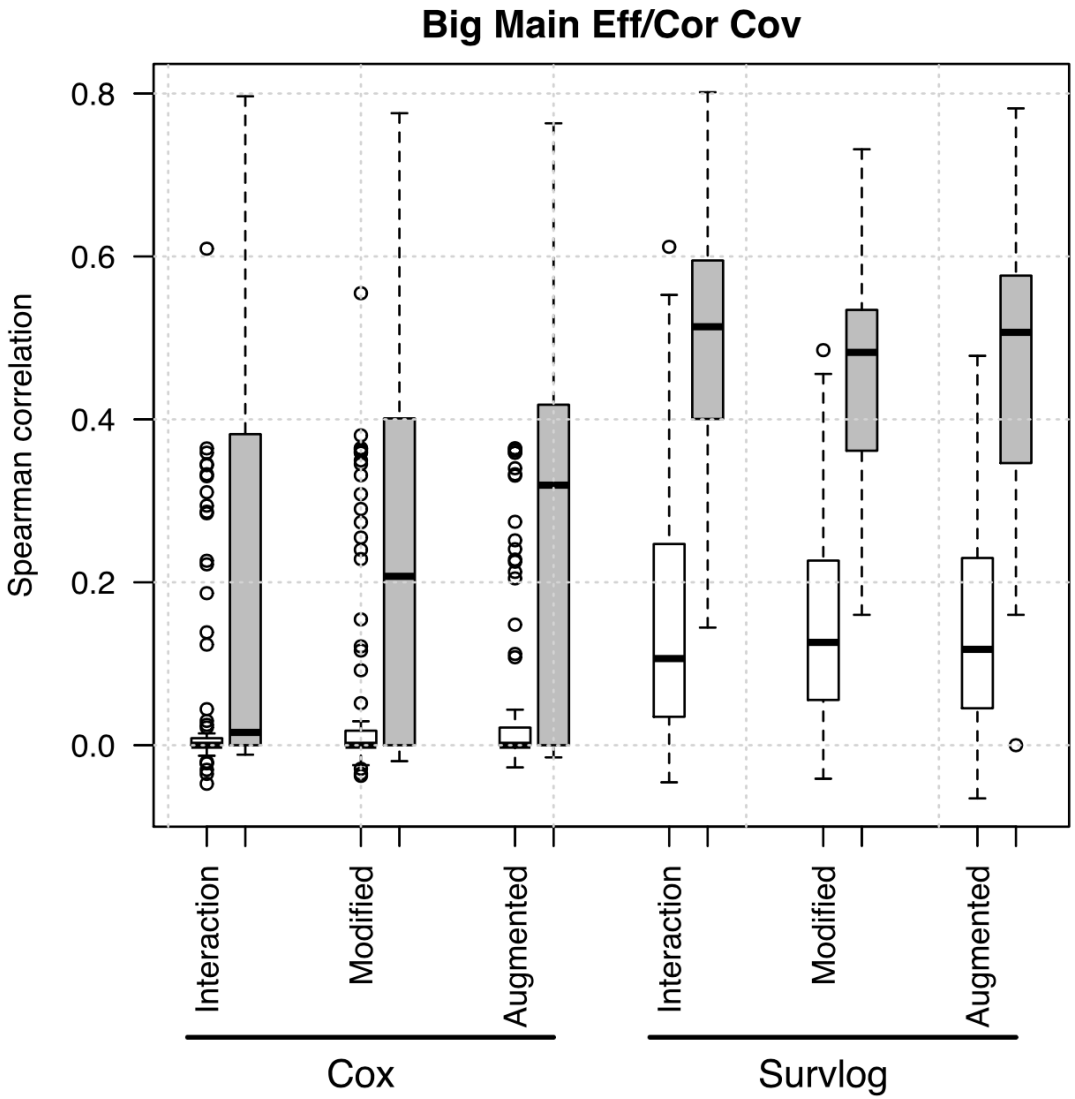
$$Y = \exp\left((\beta^\top x)^2 + A \times \left(\gamma^\top x + x^\top \alpha x\right) + \sigma_0 \epsilon\right)$$

- Censoring time samples uniformly to ensure a 25% censoring proportion
- Consider $2 * 2 * 2 = 8$ scenarios:
  - Small ($p = 30$) or large ($p = 1000$) dimension
  - Small or large main effect (change β)
  - Correlated or independent variables in X
- Assess performance on an independent test set, by Sperman correlation between Y and $\hat{\gamma}(X)$

Results from Tian et al. (2014) with Cox regression



Cox VS survival logistic regression



BEATRICE clinical trial (Cameron et al., 2013)

- Triple-negative operable primary invasive breast cancer.
- Two treatment arms: chemotherapy alone or with bevacizumab (Avastin)
- Gene expression assessed by NanoString (784 genes) in 991 trial participants

Model	Modified covariates	Predictive features	Full data *z*-score	Outer CV mean HR	Outer CV *p*-value
Cox	No	0	NA	NA	NA
Cox	Yes	0	NA	NA	NA
Survival LR	No	47	11.5	1.13	6.6×10^{-5}
Survival LR	Yes	58	13.2	1.16	1.5×10^{-6}

Conclusion

- A new survival regression model for high dimensional data
- Several tricks to learn predictive markers
- Limited theoretical analysis so far
- Quick development of contextual bandit techniques in other fields that could inspire us to:
  - estimate predictive models from randomized trials
  - design new trials for that purpose

THANKS

References

D. Cameron, J. Brown, R. Dent, C. Jackisch, J. Mackey, X. Pivot, G. G. Steger, T. M. Suter, M. Toi, M. Parmar, R. Laeufle, Y.-H. Im, G. Romieu, V. Harvey, O. Lipatov, T. Pienkowski, P. Cottu, A. Chan, S.-A. Im, P. S. Hall, L. Bubuteishvili-Pacaud, V. Henschel, R. J. Deurloo, C. Pallaud, and R. Bell. Adjuvant bevacizumab-containing therapy in triple-negative breast cancer (beatrice): primary results of a randomised, phase 3 trial. *The Lancet. Oncology*, 14: 933–942, Sept. 2013. ISSN 1474-5488. doi: 10.1016/S1470-2045(13)70335-8.

O. L. Chinot, W. Wick, W. Mason, R. Henriksson, F. Saran, R. Nishikawa, A. F. Carpentier, K. Hoang-Xuan, P. Kavan, D. Cernea, A. A. Brandes, M. Hilton, L. Abrey, and T. Cloughesy. Bevacizumab plus radiotherapy-temozolomide for newly diagnosed glioblastoma. *N. Engl. J. Med.*, 370:709–722, Feb. 2014. ISSN 1533-4406. doi: 10.1056/NEJMoa1308345.

D. R. Cox. Regression models and life-tables. *J. R. Stat. Soc. Ser. B*, 34(2):187–220, 1972. doi: 10.2307/2985181. URL `http://www.jstor.org/stable/2985181`.

H. Ishwaran, U. B. Kogalur, E. H. Blackstone, and M. S. Lauer. Random survival forests. *Ann. Appl. Stat.*, 2(3):840–861, 2008. doi: 10.1214/08-AOAS169. URL `http://dx.doi.org/10.1214/08-AOAS169`.

F. Qian and S. A. Murphy. Performance guarantees for individualized treatment rules. *Ann. Stat.*, 39(2):1180–1210, 2011. doi: 10.1214/10-AOS864. URL `http://dx.doi.org/10.1214/10-AOS864`.

References (cont.)

T. Sandmann, R. Bourgon, J. Garcia, C. Li, T. Cloughesy, O. L. Chinot, W. Wick, R. Nishikawa, W. Mason, R. Henriksson, F. Saran, A. Lai, N. Moore, S. Kharbanda, F. Peale, P. Hegde, L. E. Abrey, H. S. Phillips, and C. Bais. Patients with proneural glioblastoma may derive overall survival benefit from the addition of bevacizumab to first-line radiotherapy and temozolomide: Retrospective analysis of the avaglio trial. *J Clin Oncol*, 33(25):2735–2744, Sep 2015. doi: 10.1200/JCO.2015.61.5005. URL `http://dx.doi.org/10.1200/JCO.2015.61.5005`.

L. Tian, A. A. Alizadeh, A. J. Gentles, and R. Tibshirani. A simple method for estimating interactions between a treatment and a large number of covariates. *J. Am. Stat. Assoc.*, 109 (508):1517–1532, Oct 2014. doi: 10.1080/01621459.2014.951443. URL `http://dx.doi.org/10.1080/01621459.2014.951443`.

R. Tibshirani. The lasso method for variable selection in the Cox model. *Stat. Med.*, 16(4): 385–395, Feb 1997.

V. Van Belle, K. Pelckmans, J. A. K. Suykens, and S. Van Huffel. Support vector machines for survival analysis. In E. Ifeachor and A. Anastasiou, editors, *Proceedings of the Third International Congress on Computational Intelligence in Medicine and Healthcare (CIMED 2007)*, Plymouth, UK, 2007.

H. Zou and T. Hastie. Regularization and variable selection via the Elastic Net. *J. R. Stat. Soc. Ser. B*, 67:301–320, 2005. URL `http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.89.1596`.