Optimism-corrected treatment effect estimates in subgroups displayed in forest plots for time to event outcomes

Ke Li, Marcel Wolbers, Kaspar Rufibach

August 27, 2018

Randomized clinical trials (RCT)

- Clinical trials: the experimental approach to evaluating the effectiveness and safety of new interventions for the treatment or prevention of diseases (Cook et al., 2007).

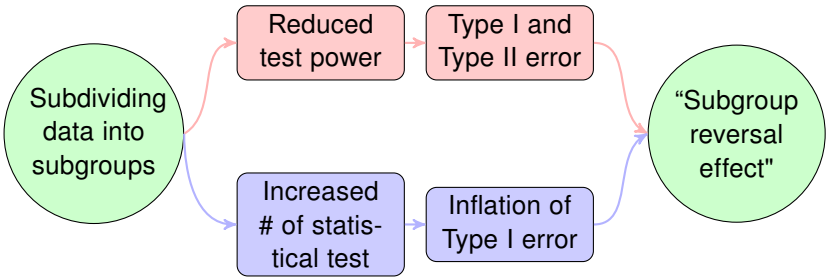
- Randomization: a fundamental part to ensure comparability between subjects receiving the intervention and control.

Why subgroup analysis in RCT?

- to assess the magnitude of the treatment benefit in major subgroups defined by baseline factors
- to investigate consistency of treatment effects across subgroups
- to determine the appropriate patient population for treatment use

Subgroup analysis

- Approach: test of the null hypothesis that there is no interaction between treatment and the investigated subgroup
- Challenges (Lagakos and Stephen W , 2006)

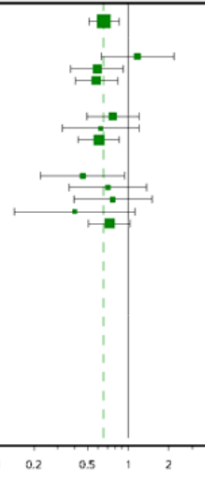


The GALLIUM phase III Trial (Marcus et al., 2017):

- a two-arm, 1:1 randomized, open-label study
- comparison of the control treatment **Rituximab-based chemotherapy** to the new treatment **Gazyva-based chemotherapy**
- 1202 patients with previously untreated, **advanced indolent non-Hodgkin's follicular lymphoma** enrolled between July 6, 2011, and February 4, 2014
- the primary end point was progression-free survival assessed by the investigator

Forest plot of the GALLIUM data

Stratification Factors	Total n	R-chemo (N=601) n	Events	1 Year KM rate	G-chemo (N=601) n	Events	1 Year KM rate	Hazard Ratio	95% Wald CI
All Patients	1202	601	144	89.736	601	101	93.939	0.66	(0.51, 0.85)
IPI/FLIPI									
FLIPI Low	253	125	18	93.059	128	22	94.357	1.17	(0.63, 2.19)
FLIPI Intermediate	447	223	49	90.176	224	31	96.246	0.59	(0.37, 0.92)
FLIPI High	502	253	77	87.785	249	48	91.662	0.58	(0.41, 0.84)
Chemotherapy Regimen									
CHOP	398	203	46	93.841	195	35	93.636	0.77	(0.50, 1.20)
CVP	118	57	20	78.963	61	16	95.000	0.63	(0.32, 1.21)
BENDAMUSTINE	686	341	78	89.021	345	50	93.928	0.61	(0.43, 0.86)
Geographic Region									
Asia	185	93	22	93.049	92	11	94.212	0.46	(0.22, 0.95)
Eastern Europe	157	79	21	85.941	78	15	92.137	0.71	(0.36, 1.37)
North America	152	77	20	92.006	75	15	95.730	0.77	(0.39, 1.50)
Other	128	66	13	92.188	62	5	98.305	0.40	(0.14, 1.12)
Western Europe	580	286	68	88.621	294	55	93.007	0.73	(0.51, 1.04)



Aim and contributions

To develop methods for regularizing treatment effect estimation in **pre-specified** subgroups for **survival outcomes**, we

- propose two methods:
  - penalized composite likelihood approach
  - marginalization of prediction from a penalized Cox model to all data
- examine the properties of these methods in an extensive simulation study
  - multiple realistic clinical trial scenarios
  - overlapping subgroups
- apply the best-performing method to the GALLIUM data.

Penalized regression for subgroup analysis

Naive method for survival outcomes:

$$h_i(t) = h_0(t) \exp(\beta_{\text{tr},k} \cdot z_i), i \in \mathcal{S}_k \tag{1}$$

Problem: lack of coherent regularization across subgroups

Solution:

- a globe model for overall treatment effect, subgroup-specific prognostic effects and predictive effects
- applying ridge-penalty or lasso-penalty to predictive effects

Approach (1): penalized composite likelihood

We denote the treatment effect for kth subgroup by

$$h_i(t) = h_0(t) \cdot \exp(\beta_{\text{tr}} \cdot z_i + \alpha_k + \beta_k \cdot z_i), i \in \mathcal{S}_k. \tag{2}$$

- h_0 is the “overall" baseline hazard function,
- β_{tr} indicates the “overall" treatment effects,
- α_k and β_k indicate subgroup-specific deviations to the “overall" baseline hazard and the “overall" treatment effect.

We assume the independence among different subgroups, the penalized composite likelihood is

$$\sum_{k=1}^{K} l_k(\beta_{\text{tr}}, \alpha_k, \beta_k) - \lambda \sum_{k=1}^{K} \|\beta_k^q\|. \tag{3}$$

lasso-penalty, $q = 1$ and ridge penalty, $q = 2$.

The subgroup treatment effect for $\mathcal{S}_k$ is $\hat{\beta}_{\text{tr}} + \hat{\beta}_k$.

Approach (2): Marginalization of prediction from a penalized Cox model

Without the independence assumption, the globe model is

$$h_i(t) = h_0(t) \exp(\beta_{\text{tr}} z_i + \underbrace{\beta_1 s_{1i} + \cdots + \beta_K s_{Ki}}_{\text{prognostic effects}} + \underbrace{\theta_1 s_{1i} z_i + \cdots \theta_K s_{Ki} z_i}_{\text{predictive effects}}), \quad (4)$$

Problem:

- the PH assumption may not hold.
- the subgroups are overlapping.

$\rightarrow$ subgroup treatment effect for $\mathcal{S}_k$ can not be simply obtained by $\hat{\beta}_{\text{tr}} + \hat{\theta}_K$.

Given the globe model

$$h_i(t) = h_0(t) \exp(\beta_{\text{tr}} z_i + \underbrace{\beta_1 s_{1i} + \cdots + \beta_K s_{Ki}}_{\text{prognostic effects}} + \underbrace{\theta_1 s_{1i} z_i + \cdots \theta_K s_{Ki} z_i}_{\text{predictive effects}}),$$

the penalized partial likelihood is

$$l(\beta_{\text{tr}}, \beta_1, \ldots, \beta_K, \theta_1, \ldots, \theta_K) - \lambda \sum_{k=1}^{K} \|\theta_k^q\|. \tag{5}$$

Marginalization of prediction from a penalized globe Cox model:

- extracting all estimates, including baseline hazard function and every parameters
- constructing **survival probabilities** for each individual
- a summary statistic indicating subgroup-specific treatment effect

Average hazard ratio

$$\text{AHR}_{OC} = \frac{\int S_0(t) f_1(t) \mathrm{d}t}{\int S_1(t) f_0(t) \mathrm{d}t}. \tag{6}$$

The expression can be rewritten as odds of concordance (Schemper et al., 2009):

$$\text{AHR}_{OC} = \frac{\int P(T_0 > t) f_1(t) \mathrm{d}t}{\int P(T_1 > t) f_0(t) \mathrm{d}t} = \frac{P(T_0 > T_1)}{P(T_1 > T_0)} = \frac{P(T_1 < T_0)}{1 - P(T_1 < T_0)} = OC. \tag{7}$$

For subgroup-specific treatment effect estimation, we have

$$\widehat{\text{AHR}}_{\text{OC}}(\mathcal{S}_k) = \frac{\sum_{t \in t_1, \cdots, t_l} \hat{S}_{k,0}(t) \cdot \hat{f}_{k,1}(t)}{\sum_{t \in t_1, \cdots, t_l} \hat{S}_{k,1}(t) \cdot \hat{f}_{k,0}(t)}.$$

- $\hat{S}_{k,0} = 1/|\mathcal{S}_k| \sum_{i \in \mathcal{S}_k} \hat{S}_{i,0}$ and $\hat{S}_{k,1} = 1/|\mathcal{S}_k| \sum_{i \in \mathcal{S}_k} \hat{S}_{i,1}$.
- $\hat{f}_{i,0}(t_1) = 1 - \hat{S}_{k,0}$ at t_1 and $\hat{f}_{i,0}(t_k) = \hat{S}_{k,0}(t_{k-1}) - \hat{S}_{k,0}(t_k)$ for $k = 2, \cdots, l$.

Baseline methods

- **Naive method**

$$h_i(t) = h_0(t)\exp(\beta_{\text{tr},k} \cdot z_i), i \in \mathcal{S}_k \tag{8}$$

The Wald 2-sided (1-α) $\times$ 100% confidence interval:

$$\hat{\beta}_{\text{tr},k} \pm Z_{1-\alpha/2} \cdot \widehat{se}(\hat{\beta}_{\text{tr},k}).$$

- **Naive population-based method**

$$h_i(t) = h_0(t)\exp(\beta_{\text{tr, overall}} \cdot z_i), i \in \mathcal{S}, \tag{9}$$

$$\hat{\beta}_{\text{tr},k} = \hat{\beta}_{\text{tr, overall}}, k \in \{1, 2, \ldots, K\} \tag{10}$$

The Wald 2-sided (1-α) $\times$ 100% confidence interval:

$$\hat{\beta}_{\text{tr,overall}} \pm Z_{1-\alpha/2} \cdot \widehat{se}(\hat{\beta}_{\text{tr,overall}}).$$

Simulation study

- Compare the performance of six statistical methods for estimating subgroup treatment effect
- To resemble real clinical trial data, the parameter setting was inspired by the GALLIUM data.
- 1000 datasets with sample size 1202, and target number of event 245

Estimation method (estimator)	Denoted by
naive	naive
naive overall population-based	naivepop
lasso-penalized average hazard ratio	lassoAHR
ridge-penalized average hazard ratio	ridgeAHR
lasso-penalized composite likelihood	lassocomposite
ridge-penalized composite likelihood	ridgecomposite

Table: Estimators used in simulations

Data simulation

- **Biomarker generation**

$$\mathbf{X} \sim \mathcal{N}_{10}(\mu, \Sigma), \tag{11}$$

where $\boldsymbol{\mu} = [\mathrm{E}(X_1), \cdots, \mathrm{E}(X_{10})]^T = [0, \cdots, 0]^T$,

$$\mathrm{diag}(\boldsymbol{\Sigma}) = 1, \qquad \mathrm{Cov}(X_{i=1,\cdots,5}, X_{j\neq i, j=1,\cdots,5}) = 0,$$
$$\mathrm{Cov}(X_{i=6,\cdots,8}, X_{j\neq i, j=6,\cdots,8}) = \sigma_{\text{moderate}}, \qquad \mathrm{Cov}(X_{i=9,10}, X_{j\neq i, j=9,10}) = \sigma_{\text{high}}.$$

Then, dichotomize the continuous variables

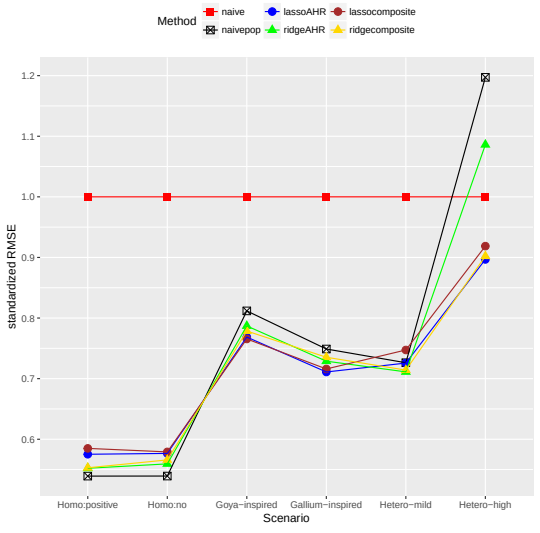
- **Arm**: sampled from a vector with 50% 1's and 50% 0's
- **Survival time**: Weibull distribution, AFT representation
- **Censoring time**: Exponential distribution, censoring rate 2%
- **Progression-free time**: with consideration of non-administrative and administrative censoring

Six simulation scenarios

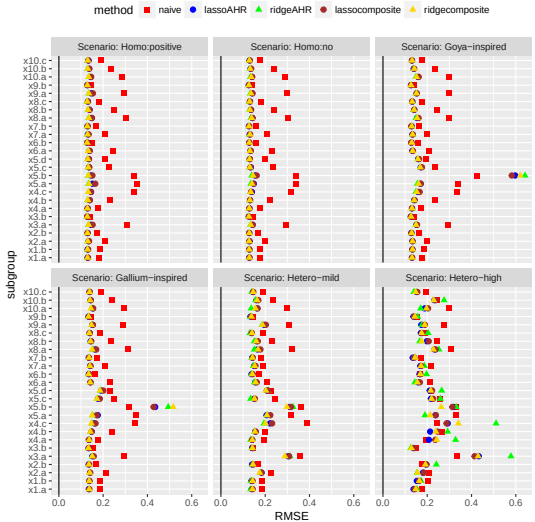
Scenario	Overall treatment effect (HR-scale)	No.subgroup with predictive effect	Differential subgroup treatment effect
Homo:positive	0.67	0	0
Homo:no	1	0	0
GOYA-inspired	1	3	0.5
GALLIUM-inspired	0.67	3	1.2
Hetero-mild	1	15	$\mathcal{N}(0, 0.2)$
Hetero-high	1	15	$\mathcal{N}(0, 0.5)$

Table: Simulation scenarios

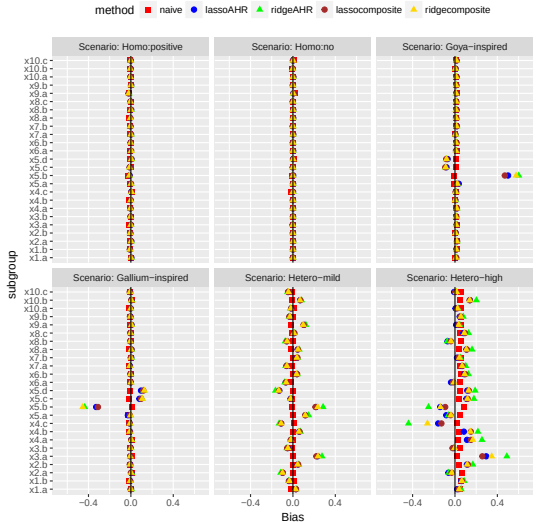
Result: overall RMSE



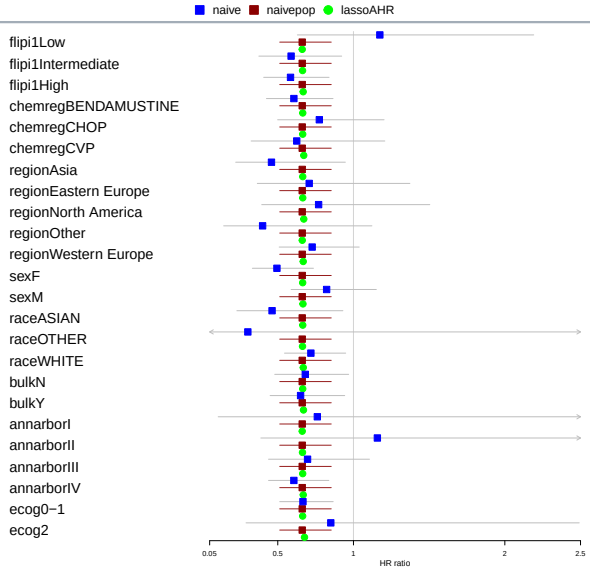
Result: subgroup-specific RMSE

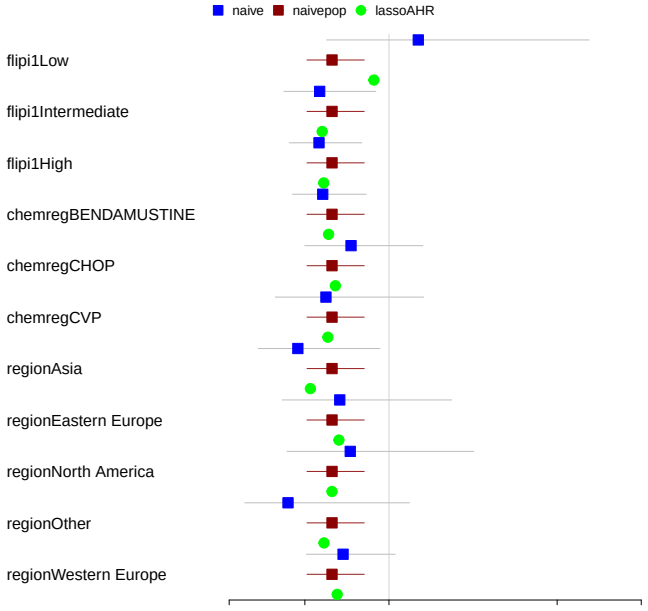


Result: subgroup-specific bias



Application: the Gallium data




naive
naivepop
lassoAHR
flipi1Low
flipi1Intermediate
flipi1High
chemregBENDAMUSTINE
chemregCHOP
chemregCVP
regionAsia
regionEastern Europe
regionNorth America
regionOther
regionWestern Europe
0.05
0.5
1
2
2.5
HR ratio

Conclusion

- Two new methods have been proposed for treatment effect estimation in subgroups for time-to-event data.
- The variants of our methods, generally, outperform the naive method.
- The type of shrinkage (lasso-penalty or ridge-penalty) plays a more influential role, compared to the type of estimation methods.
- No clear differential treatment effects in subgroups has been observed in the GALLIUM data.

Outlook

- Extension of lasso method: elastic net, adaptive lasso, and relaxed lasso
- Confidence interval: developing counterpart models of our methods under the Bayesian framework and leverage the posterior distributions of the parameters to obtain the credible interval.

Reference

Cook, Thomas D and DeMets, David L(2007) Introduction to statistical methods for clinical trials. *CRC Press*

Lagakos and Stephen W (2006) The challenge of subgroup analyses-reporting without distorting. *New England Journal of Medicine*

Thomas, Marius and Bornkamp, Björn(2017) Comparing Approaches to Treatment Effect Estimation for Subgroups in Clinical Trials. *Statistics in Biopharmaceutical Research*

Bayesian models for subgroup analysis in clinical trials. *Clinical Trials*

Bornkamp, Björn and Ohlssen, David and Magnusson, Baldur P and Schmidli, Heinz(2017) Model averaging for treatment effect estimation in subgroups. *Pharmaceutical statistics*

Rosenkranz, Gerd K(2016) Exploratory subgroup analysis in clinical trials by model selection. . *Biometrical Journal*

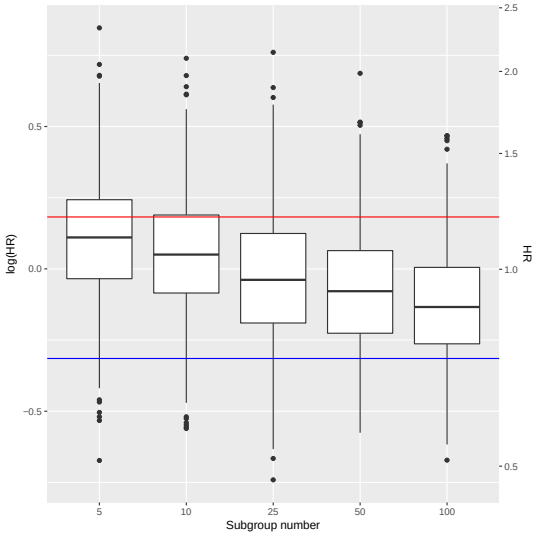
Marcus, Robert and Davies, Andrew and Ando, Kiyoshi and Klapper, Wolfram and Opat, Stephen and Owen, Carolyn and Phillips, Elizabeth and Sangha, Randeep and Schlag, Rudolf and Seymour, John F and others (2017) obinutuzumab for the First-line Treatment of Follicular Lymphoma. *New England Journal of Medicine*

Klein, John P and Moeschberger, Melvin L (2005) Survival analysis: techniques for censored and truncated data.*Springer Science & Business Media*

Schemper, Michael and Wakounig, Samo and Heinze, Georg (2009) The estimation of average hazard ratios by weighted Cox regression. *Statistics in medicine*

Thank you for listening ! Questions?

Result: shrinkage vs. number of subgroups



Result: predictive subgroup estimates under scenarios “Goya-inspired" and “Gallium-inspired"