



Identifying treatment effect heterogeneity in dose-finding trials using Bayesian hierarchical models

Marius Thomas
IDEAS Dissemination Workshop
Basel, Switzerland
26 September 2018

Acknowledgements

- Dr. Björn Bornkamp (Novartis Pharma AG)
- Prof. Dr. Katja Ickstadt (TU Dortmund University)

This work was supported by funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No 633567 and by the Swiss State Secretariat for Education, Research and Innovation (SERI) under contract number 999754557. The opinions expressed and arguments employed herein do not necessarily reflect the official views of the Swiss Government.



Introduction

Investigations of treatment effect heterogeneity routinely part of many trials in Phase II and III

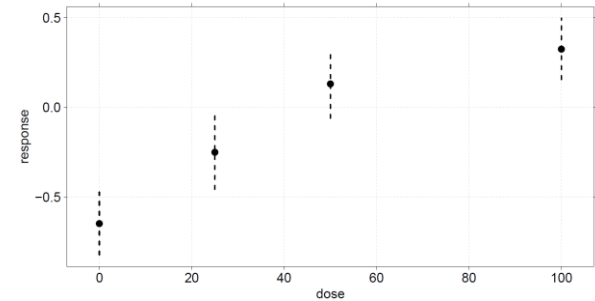
- Common questions:
 - Is the treatment effect the same across the population?
 - Can we define a subgroup of patients with increased treatment effects?
- Usually **exploratory**
- Based on **pre-specified baseline covariates in moderate numbers (< 30)**
- Statistically challenging: **multiplicity, lack of power**

Here we consider these analyses in the context of dose-finding trials

Motivating example

Phase II dose-finding trial

- $N = 270$
- Dose levels: 0, 25, 50, 100
- Continuous endpoint (change from baseline)
- 10 baseline covariates (6 categorical, 4 continuous)



Exploratory analyses investigating treatment effect heterogeneity/ possible subgroups

- Do baseline covariates interact with treatment/dose?
- Are there subgroups with higher treatment effects?
- Are there subgroups requiring different doses?

Available subgroup identification methods usually designed for two-arm trials

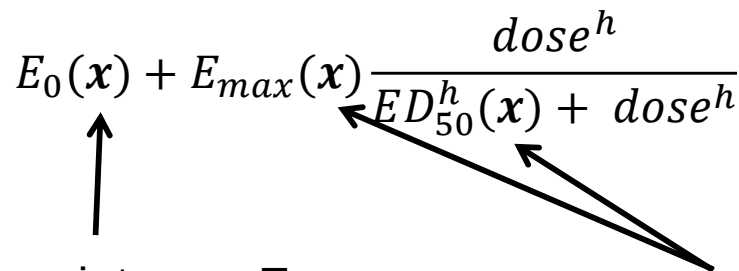
Here we want to take underlying dose-response relationship into account

General idea

- **Standard Emax:**

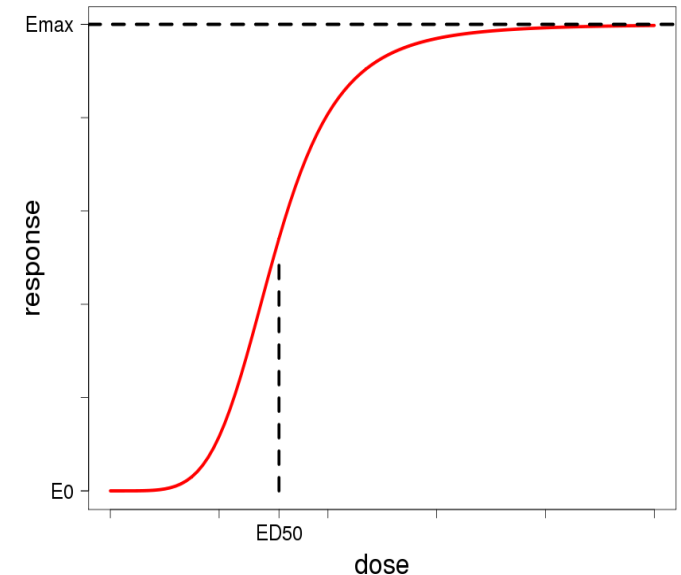
$$E_0 + E_{max} \frac{dose^h}{ED_{50}^h + dose^h}$$

- **In our setting with baseline covariates x :**

$$E_0(x) + E_{max}(x) \frac{dose^h}{ED_{50}^h(x) + dose^h}$$


Covariates on E_0 :
Prognostic covariates
(modify response
independent of treatment)

Covariates on E_{max} or ED_{50} :
Predictive covariates (modify
treatment effects)

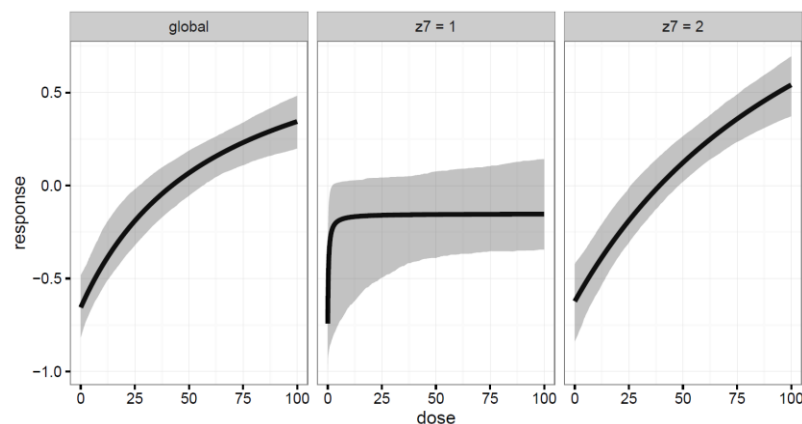
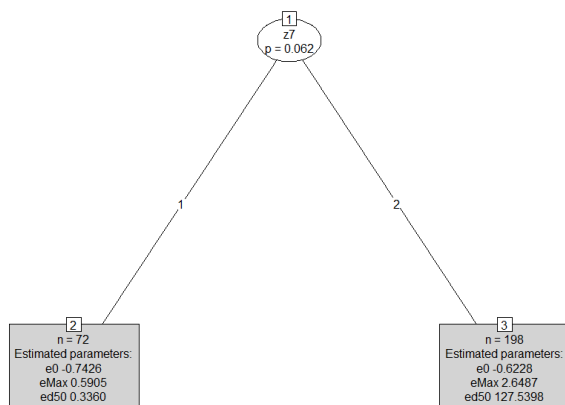


Main aim: Identify **predictive covariates**, which can then be used to **define subgroups**

A possible approach

We previously proposed model-based recursive partitioning (mob) (Seibold et al., 2016, Thomas et al., 2018) for subgroup identification in this setting

- Mob is a tree-based method, which identifies subgroups with differential dose-response
- Uses Bonferroni corrections to control for multiplicity
- Mob applied to the example:



Mob: Pros and Cons

Main advantages of mob

- Quite easy to use
- Finding suitable cut-offs part of the algorithm
- Good performance with regards to variable selection and subgroup identification in simulations
- Can handle covariate-covariate interactions and non-linear covariate effects

Drawback: Dose-response modeling

- Dose-response models are fitted separately in each subgroup, without borrowing information from other subgroups
- Models don't take uncertainty with regards to subgroup selection into account
- Doesn't allow modeling covariate effects on specific dose-response parameters

Can we find a method, that has similar variable selection performance as mob, while improving on the modeling?

Here we propose a Bayesian hierarchical dose-response model

Bayesian dose-response model

For normally distributed data

- Data: $(Y, d, x^{(1)}, \dots, x^{(k)})$
 - Y : response variable
 - d : dose variable
 - $x^{(1)}, \dots, x^{(k)}$: baseline covariates of interest for subgroup analyses

$$Y \sim N(\mu, \sigma^2)$$

$$\mu = E_0 + E_{max} \frac{d^h}{ED_{50}^h + d^h}$$

$$E_0 = \alpha_{E_0} + \beta_1 x^{(1)} + \dots + \beta_k x^{(k)}$$

$$E_{max} = \alpha_{E_{max}} + \gamma_1 x^{(1)} + \dots + \gamma_k x^{(k)}$$

$$\log(ED_{50}) = \alpha_{ED_{50}} + \delta_1 x^{(1)} + \dots + \delta_k x^{(k)}.$$

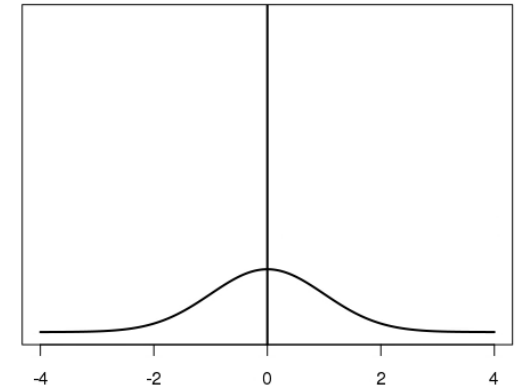
- Choose non-informative priors for $\sigma, h, \alpha_{E_0}, \alpha_{E_{max}}, \alpha_{ED_{50}}$
- Priors for β, γ, δ ?
 - Non-informative priors would lead to overfitting
 - **Instead use shrinkage/variable selection priors**
 - **Here we consider Spike-and-slab and horseshoe**

Considered Shrinkage priors

Spike-and-Slab (Mitchell & Beauchamp, 1988, George & McCulloch, 1993):

$$\theta_j \sim N(0, c^2 \lambda_j), j = 1, \dots, k$$
$$\lambda_j \sim \text{Bern}(p)$$

- Gold standard for Bayesian variable selection
- Mixture between 'spike' at zero and normally-distributed 'slab' with variance c^2
- p represents inclusion probability, e.g. if $p = 0.2$ we expect 20% of the covariates to have a coefficient different from zero

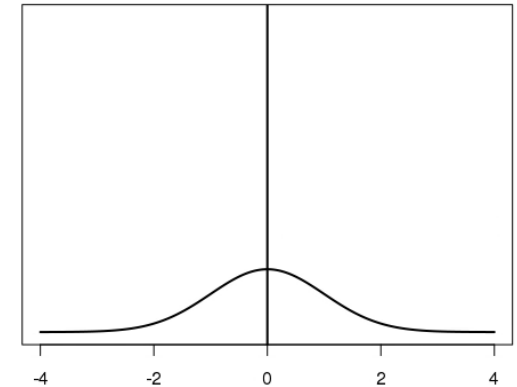


Considered Shrinkage priors

Spike-and-Slab (Mitchell & Beauchamp, 1988, George & McCulloch, 1993):

$$\theta_j \sim N(0, c^2 \lambda_j), j = 1, \dots, k$$
$$\lambda_j \sim \text{Bern}(p)$$

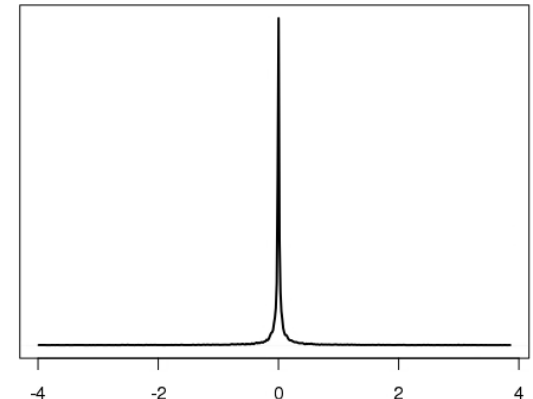
- Gold standard for Bayesian variable selection
- Mixture between 'spike' at zero and normally-distributed 'slab' with variance c^2
- p represents inclusion probability, e.g. if $p = 0.2$ we expect 20% of the covariates to have a coefficient different from zero



Horseshoe (Carvalho et al., 2010, Piironen & Vehtari, 2017):

$$\theta_j \sim N(0, \tau^2 \lambda_j^2), j=1, \dots, k$$
$$\lambda_j \sim C^+(0, 1)$$
$$\tau \sim C^+(0, \eta^2)$$

- Good theoretical properties, clear separation of noise and large effects
- Combination of local (λ_j) and global (τ) shrinkage component
- η determines number of non-zero coefficients a priori
- Wide tails can lead to convergence issues, when using MCMC
- **Regularized horseshoe** with improved MCMC sampling properties proposed by Piironen & Vehtari (2017)



Back to the dose-response model

Model specification using horseshoe priors

$$E_0 = \alpha_{E_0} + \beta_1 x^{(1)} + \dots + \beta_k x^{(k)}$$

$$E_{max} = \alpha_{E_{max}} + \gamma_1 x^{(1)} + \dots + \gamma_k x^{(k)}$$

$$\log(ED_{50}) = \alpha_{ED_{50}} + \delta_1 x^{(1)} + \dots + \delta_k x^{(k)}.$$

Horseshoe priors on coefficients:

- Same local shrinkage components for covariate effects on Emax and ED50
- Reduces model complexity
- Represents focus on distinction *prognostic vs predictive*

$$\beta_j \sim N(0, \tau_\beta^2 \lambda_j^{(prog)^2}), \quad j = 1, \dots, k$$

$$\gamma_j \sim N(0, \tau_\gamma^2 \lambda_j^{(pred)^2}), \quad j = 1, \dots, k$$

$$\delta_j \sim N(0, \tau_\delta^2 \lambda_j^{(pred)^2}), \quad j = 1, \dots, k.$$

Back to the dose-response model

Model specification using horseshoe priors

$$E_0 = \alpha_{E_0} + \beta_1 x^{(1)} + \dots + \beta_k x^{(k)}$$

$$E_{max} = \alpha_{E_{max}} + \gamma_1 x^{(1)} + \dots + \gamma_k x^{(k)}$$

$$\log(ED_{50}) = \alpha_{ED_{50}} + \delta_1 x^{(1)} + \dots + \delta_k x^{(k)}.$$

Horseshoe priors on coefficients:

- Same local shrinkage components for covariate effects on Emax and ED50
- Reduces model complexity
- Represents focus on distinction *prognostic vs predictive*

$$\beta_j \sim N(0, \tau_\beta^2 \lambda_j^{(prog)^2}), \quad j = 1, \dots, k$$

$$\gamma_j \sim N(0, \tau_\gamma^2 \lambda_j^{(pred)^2}), \quad j = 1, \dots, k$$

$$\delta_j \sim N(0, \tau_\delta^2 \lambda_j^{(pred)^2}), \quad j = 1, \dots, k.$$

Independent priors on local shrinkage components (option 1):

- Shrinkage for prognostic and predictive effects of the same covariate independent
- Possible to include interactions (predictive effects) without main effects (prognostic effects)

$$\lambda_j^{(prog)} \sim C^+(0, 1) \quad j = 1, \dots, k$$

$$\lambda_j^{(pred)} \sim C^+(0, 1) \quad j = 1, \dots, k,$$

Back to the dose-response model

Model specification using horseshoe priors

$$E_0 = \alpha_{E_0} + \beta_1 x^{(1)} + \dots + \beta_k x^{(k)}$$

$$E_{max} = \alpha_{E_{max}} + \gamma_1 x^{(1)} + \dots + \gamma_k x^{(k)}$$

$$\log(ED_{50}) = \alpha_{ED_{50}} + \delta_1 x^{(1)} + \dots + \delta_k x^{(k)}.$$

Horseshoe priors on coefficients:

- Same local shrinkage components for covariate effects on Emax and ED50
- Reduces model complexity
- Represents focus on distinction *prognostic vs predictive*

$$\beta_j \sim N(0, \tau_\beta^2 \lambda_j^{(prog)^2}), \quad j = 1, \dots, k$$

$$\gamma_j \sim N(0, \tau_\gamma^2 \lambda_j^{(pred)^2}), \quad j = 1, \dots, k$$

$$\delta_j \sim N(0, \tau_\delta^2 \lambda_j^{(pred)^2}), \quad j = 1, \dots, k.$$

Dependent priors on local shrinkage components (option 2):

- Idea: Don't shrink prognostic effects more than corresponding predictive effects
- Use $\lambda_j^{(pred)}$ as lower bound for $\lambda_j^{(prog)}$
- Reduces probability for 'interaction without main effect' - outcomes

$$\lambda_j^{(*)} \sim C^+(0, 1) \quad j = 1, \dots, k$$

$$\lambda_j^{(prog)} = \max(\lambda_j^{(*)}, \lambda_j^{(pred)}) \quad j = 1, \dots, k$$

$$\lambda_j^{(pred)} \sim C^+(0, 1) \quad j = 1, \dots, k,$$

Simulation study to compare priors

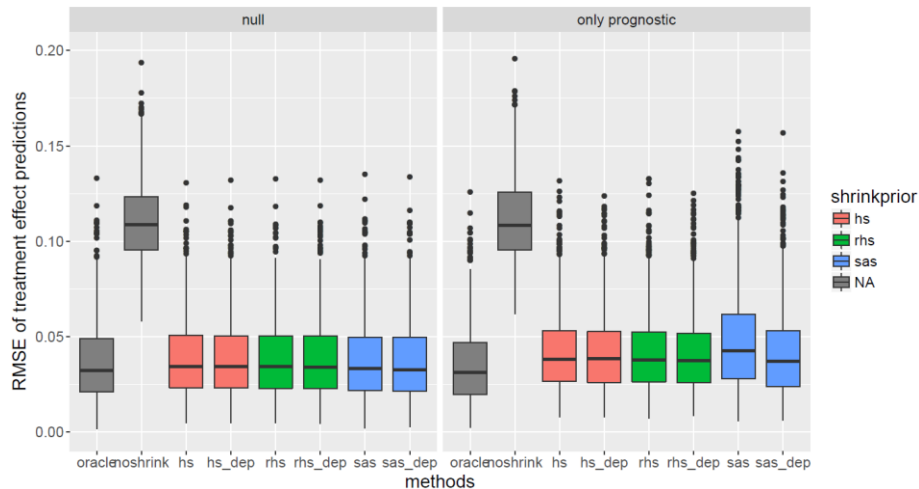
- **Simulation setup (default scenarios):**
 - 5 dose levels: 0, 12.5, 25, 50, 100
 - 50 patients on each dose (250 patients in total)
 - $\sigma = 0.25$
 - $h = 1$
 - 10 independent standard normally distributed covariates $x_1, \dots, x_{10}$

- **Scenarios for Emax model parameters:**

Scenario	$E_0(x)$	$E_{max}(x)$	$ED_{50}(x)$
1: Null	1.2	0.17	20
2: Only prog.	$1.2 + 0.1x_1 + 0.1x_2 + 0.05x_3$	0.17	20
3: Prog. + pred.	$1.2 + 0.1x_1 + 0.1x_2 + 0.05x_3$	$0.17 + 0.1x_2 - 0.1x_3$	$20 * \exp(-0.75x_2 + 0.75x_3)$
4: Only pred.	1.2	$0.17 + 0.1x_2 - 0.1x_3$	$20 * \exp(-0.75x_2 + 0.75x_3)$

- **Comparisons of interest:**
 - Spike-and-Slab vs horseshoe vs regularized horseshoe
 - Independent priors on local shrinkage components for prognostic and predictive effects vs Dependent priors (as on previous slide)
 - Include oracle (true model) and model without shrinkage as general comparators

Simulation results 1: No treatment effect heterogeneity



Estimation of individual treatment effect curves:

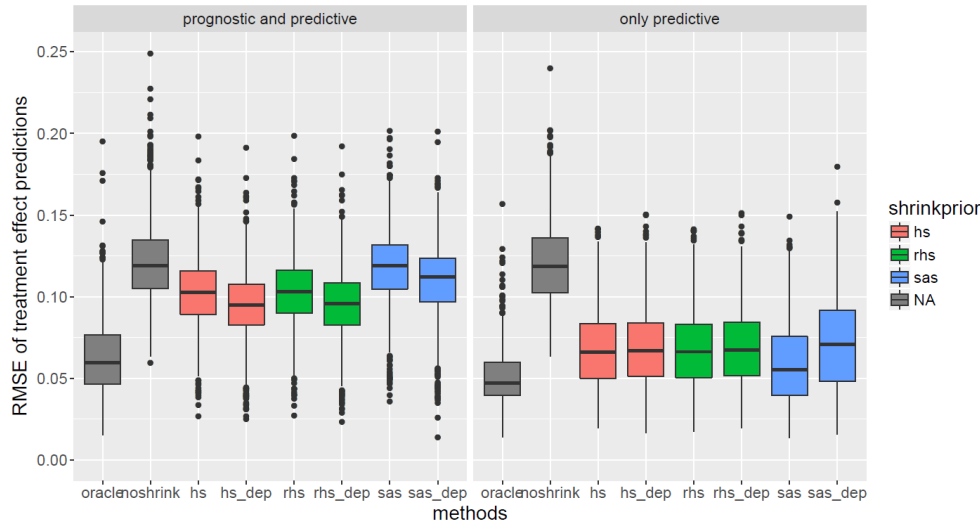
- Shrinkage priors all close to oracle
- Model without shrinkage much worse
- No big differences between shrinkage priors

scenario	method	E_0				ED_{50}				E_{max}			
		x1	x2	x3	x4	x1	x2	x3	x4	x1	x2	x3	x4
null	oracle	0	0	0	0	0	0	0	0	0	0	0	0
	noshrink	0.55	0.51	0.58	0.54	0.53	0.53	0.54	0.51	0.46	0.44	0.45	0.45
	hs	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	hs_con	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	rhs	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	rhs_dep	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	sas	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	sas_dep	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
only progn.	oracle	1	1	1	0	0	0	0	0	0	0	0	0
	noshrink	1.00	1.00	0.95	0.56	0.52	0.52	0.54	0.55	0.46	0.49	0.46	0.47
	hs	0.99	1.00	0.94	0.02	0.01	0.00	0.00	0.00	0.00	0.00	0.01	0.00
	hs_dep	1.00	1.00	0.99	0.09	0.01	0.00	0.00	0.00	0.01	0.01	0.00	0.00
	rhs	0.99	1.00	0.93	0.02	0.00	0.00	0.00	0.00	0.00	0.01	0.01	0.00
	rhs_dep	1.00	1.00	0.99	0.10	0.01	0.00	0.00	0.00	0.00	0.01	0.00	0.00
	sas	0.98	0.99	0.71	0.00	0.01	0.01	0.03	0.00	0.02	0.01	0.05	0.00
	sas_dep	0.99	1.00	0.80	0.00	0.01	0.00	0.01	0.00	0.01	0.01	0.01	0.00

Variable selection:

- Close to zero false positive identifications of predictive covariates
- Negligible differences between different shrinkage priors

Simulation results 2: Existing Treatment effect heterogeneity



scenario	method	E_0				ED_{50}				E_{max}			
		x1	x2	x3	x4	x1	x2	x3	x4	x1	x2	x3	x4
progn. & pred.	oracle	1	1	1	0	0	1	1	0	0	1	1	0
	noshrink	1.00	1.00	0.86	0.54	0.53	0.90	0.89	0.56	0.49	0.96	0.96	0.50
	hs	0.99	0.96	0.14	0.02	0.00	0.18	0.08	0.00	0.01	0.55	0.31	0.00
	hs_dep	1.00	1.00	0.32	0.03	0.01	0.24	0.10	0.00	0.02	0.75	0.34	0.00
	rhs	0.99	0.95	0.14	0.01	0.00	0.19	0.09	0.00	0.01	0.55	0.29	0.00
	rhs_dep	1.00	1.00	0.31	0.02	0.01	0.25	0.11	0.00	0.01	0.77	0.33	0.00
	sas	0.99	0.84	0.09	0.00	0.01	0.29	0.16	0.00	0.01	0.45	0.23	0.00
	sas_dep	0.99	0.96	0.18	0.00	0.00	0.39	0.19	0.00	0.01	0.59	0.28	0.00
only pred.	oracle	0	0	0	0	0	1	1	0	0	1	1	0
	noshrink	0.52	0.63	0.64	0.56	0.54	0.91	0.89	0.55	0.46	0.95	0.95	0.47
	hs	0.00	0.03	0.03	0.00	0.00	0.29	0.30	0.00	0.00	0.91	0.90	0.00
	hs_dep	0.00	0.05	0.05	0.00	0.00	0.27	0.27	0.00	0.00	0.91	0.90	0.00
	rhs	0.00	0.03	0.03	0.00	0.00	0.32	0.32	0.00	0.00	0.91	0.91	0.00
	rhs_dep	0.00	0.04	0.04	0.00	0.00	0.28	0.28	0.00	0.00	0.91	0.91	0.00
	sas	0.00	0.03	0.03	0.00	0.00	0.76	0.77	0.00	0.00	0.96	0.96	0.00
	sas_dep	0.00	0.06	0.07	0.00	0.00	0.67	0.69	0.00	0.00	0.91	0.89	0.00

Estimation of individual treatment effect curves:

- Shrinkage priors generally better than model without shrinkage
- Horseshoe vs Spike-and-slab depends on scenario
- Dependent priors improve estimation in first scenario, are slightly worse in second

Variable selection:

- Only correct covariates are selected often by shrinkage priors
- Distinguishing prognostic from predictive covariates unproblematic

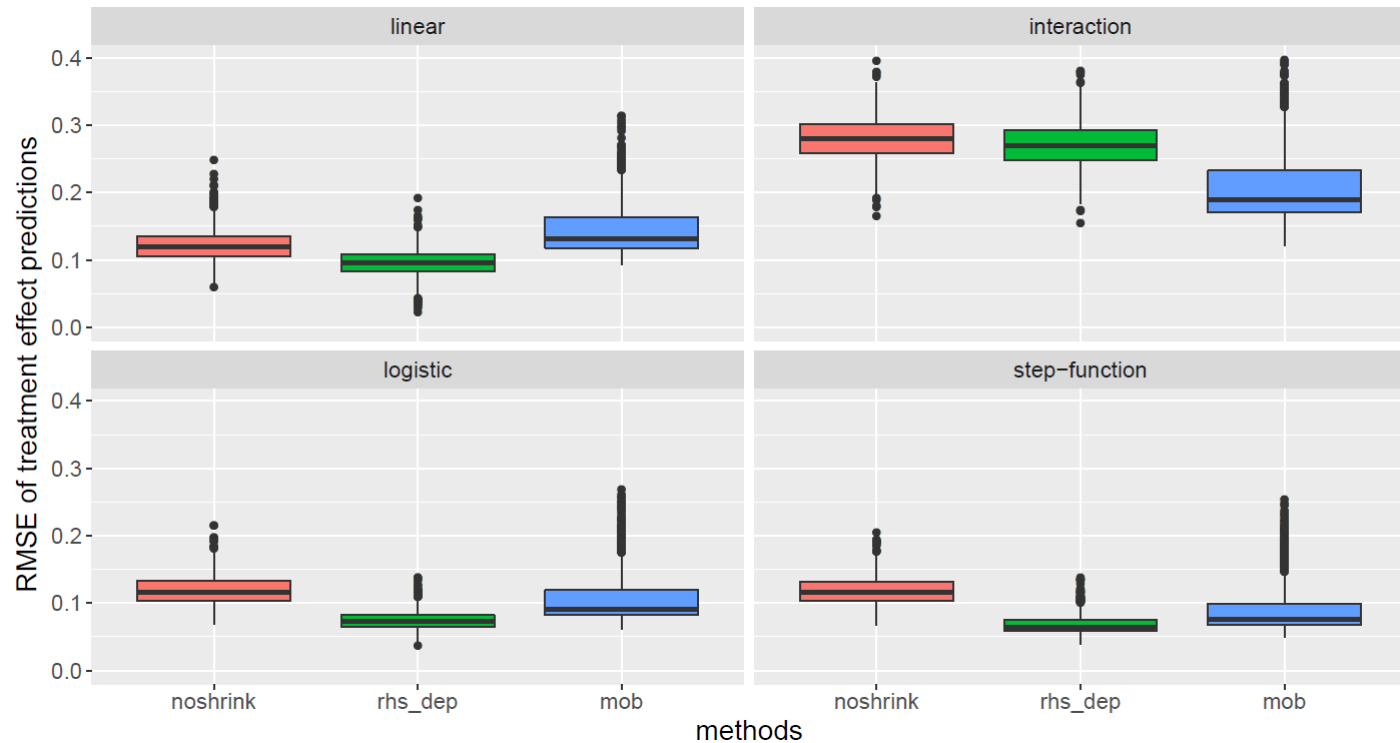
Simulation study: Conclusions

- All shrinkage priors show desired behavior of identifying relevant predictive covariates, while reducing false positives
- Based on our simulations horseshoe gives more consistent results than Spike-and-slab
- Essentially no differences between horseshoe and reg. horseshoe in performance; reg. horseshoe preferred choice, because of better MCMC sampling properties
- Dependent priors increase chance to detect relevant predictive covariates in scenarios with prognostic and predictive effects
- Similar results obtained for larger sample sizes and larger number of covariates

All in all dependent regularized horseshoe seems like a good default choice

Comparison to mob

Compare the Bayesian hierarchical model to *mob* for different types of covariate effects:

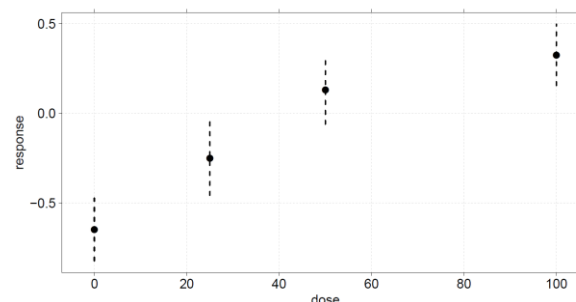


- Variable selection performance relatively similar, *mob* slightly better for non-linear scenarios

Back to the example

Phase II dose-finding trial

- N = 270
- Dose levels: 0, 25, 50, 100
- Continuous endpoint (change from baseline)
- 10 baseline covariates (6 categorical, 4 continuous)

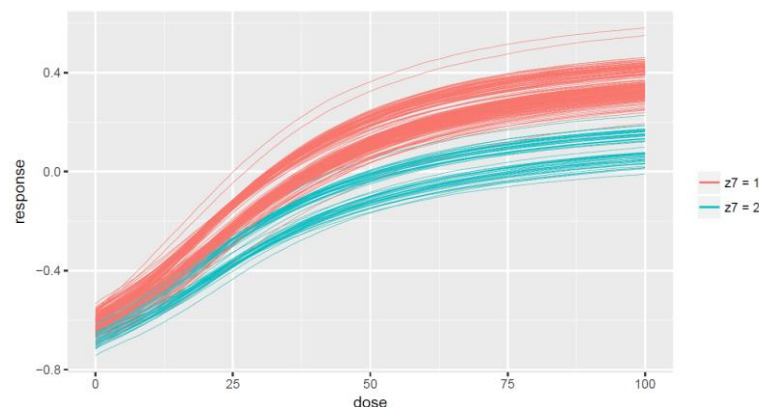


Now analyzed with Bayesian approach (reg. HS)

Posterior summaries for local shrinkage components

parameter	mean	sd	2.5%	25%	50%	75%	97.5%
$\lambda_1^{(pred)}$	1.68	6.88	0.03	0.34	0.78	1.65	8.05
$\lambda_2^{(pred)}$	3.98	77.21	0.03	0.35	0.84	1.79	9.79
$\lambda_3^{(pred)}$	0.91	1.13	0.03	0.27	0.59	1.14	3.83
$\lambda_4^{(pred)}$	1.58	2.80	0.03	0.35	0.82	1.73	7.75
$\lambda_5^{(pred)}$	1.90	6.08	0.04	0.36	0.83	1.81	10.19
$\lambda_6^{(pred)}$	3.35	38.47	0.04	0.41	0.99	2.24	15.47
$\lambda_7^{(pred)}$	8.32	64.87	0.07	0.91	2.54	6.02	39.61
$\lambda_8^{(pred)}$	1.56	3.25	0.03	0.35	0.78	1.65	7.49
$\lambda_9^{(pred)}$	3.06	15.20	0.05	0.49	1.17	2.74	14.16
$\lambda_{10}^{(pred)}$	0.89	1.10	0.03	0.26	0.58	1.13	3.63

Individual dose-response curves



Discussion

The presented approach makes use of Bayesian dose-response models with shrinkage priors to deal with the challenges of subgroup identification

- Reduces rate of false positive findings through shrinkage
- Can handle different types of outcomes and different types of covariates (continuous, categorical, binary)
- Allows estimation of individual dose-response curves
- Limitations: assumes linear function of covariates on DR-parameters, no covariate-covariate interactions, choice of hyperparameters for shrinkage priors non-trivial

How to identify a subgroup with increased treatment effect based on the model? Some possibilities:

- Threshold on posterior individual treatment effect predictions (posterior median or other quantile)
- Use identified predictive covariates to define subgroup
- Fit regression tree with individual treatment effect predictions as target and covariates as features (see Foster et al., 2011)

References

- Carvalho, C. M., Polson, N. G., & Scott, J. G. (2010). The horseshoe estimator for sparse signals. *Biometrika*, 97(2), 465-480.
- Foster, J. C., Taylor, J. M., & Ruberg, S. J. (2011). Subgroup identification from randomized clinical trial data. *Statistics in Medicine*, 30(24), 2867-2880.
- George, E. I., & McCulloch, R. E. (1993). Variable selection via Gibbs sampling. *Journal of the American Statistical Association*, 88(423), 881-889.
- Mitchell, T. J., & Beauchamp, J. J. (1988). Bayesian variable selection in linear regression. *Journal of the American Statistical Association*, 83(404), 1023-1032.
- Piironen, J., & Vehtari, A. (2017). Sparsity information and regularization in the horseshoe and other shrinkage priors. *Electronic Journal of Statistics*, 11(2), 5018-5051.
- Seibold, H., Zeileis, A., & Hothorn, T. (2016). Model-based recursive partitioning for subgroup analyses. *The International Journal of Biostatistics*, 12(1), 45-63.
- Thomas, M., Bornkamp, B., & Seibold, H. (2018). Subgroup identification in dose-finding trials via model-based recursive partitioning. *Statistics in Medicine*, 37(10), 1608-1624.

Thank you