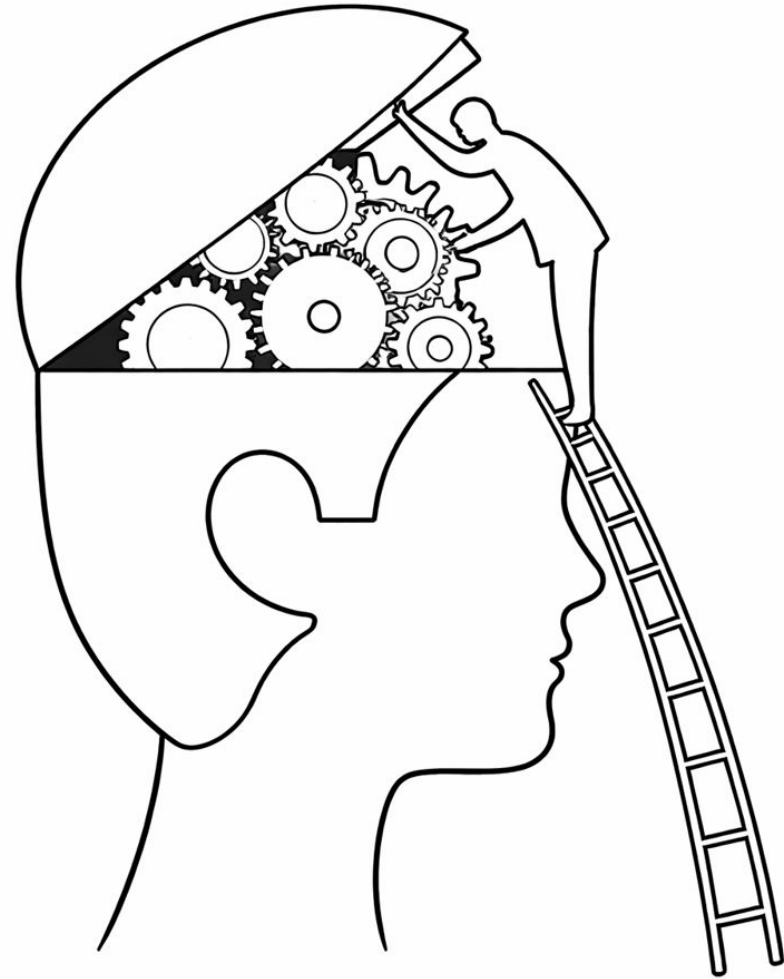


Shapley values with Bayesian uncertainty quantification for rule ensembles

Marjolein Fokkema

Behavioral science



Behavioral science



Journal of Affective Disorders

Volume 227, February 2018, Pages 164-182

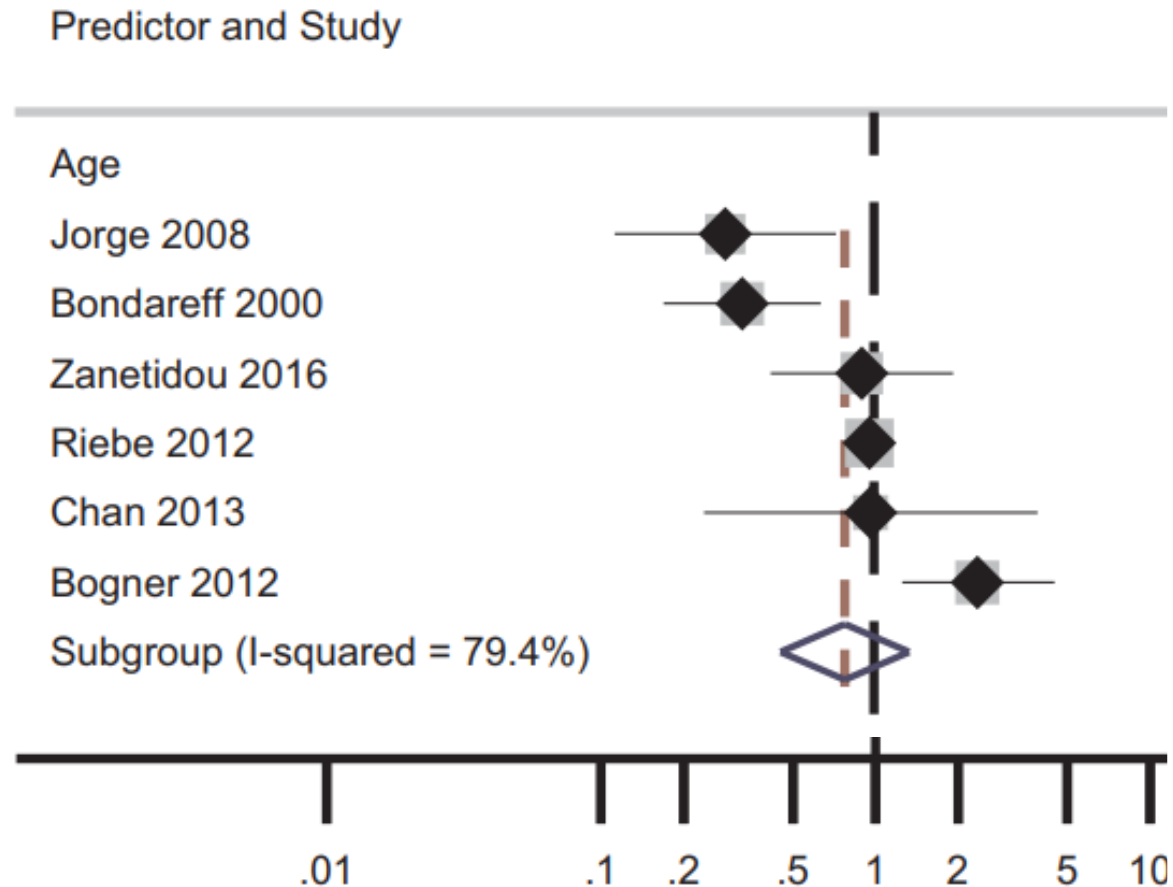


Review article

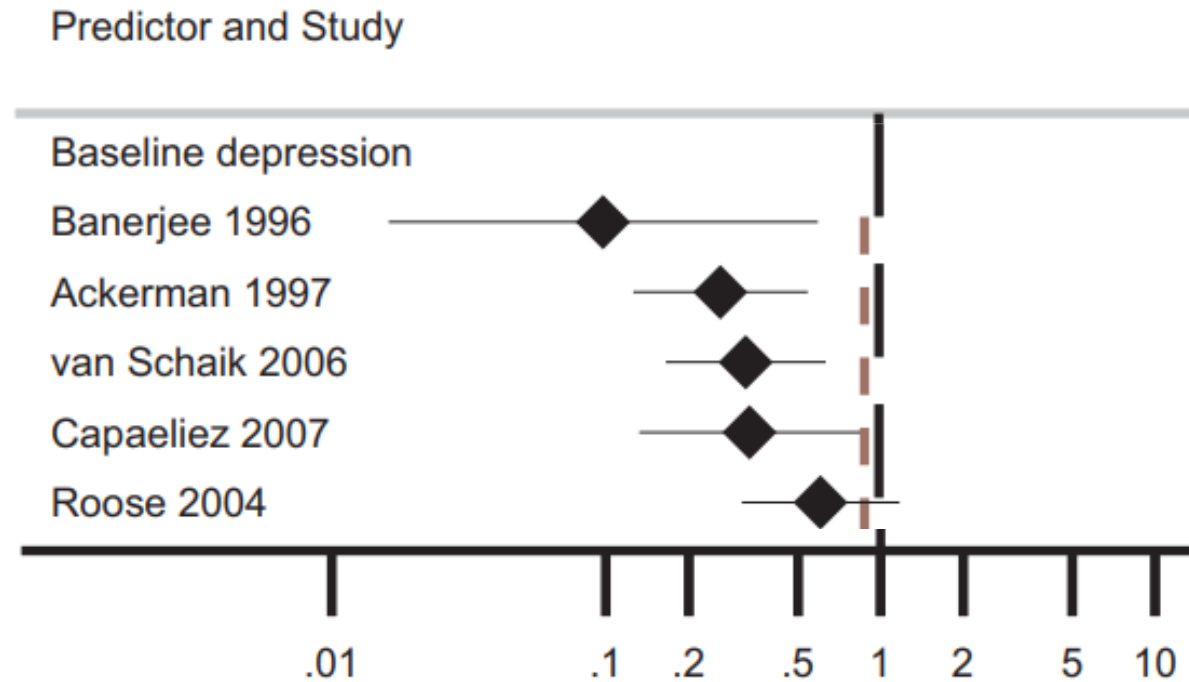
Predictors of treatment outcome in depression in later life: A systematic review and meta-analysis

Chavit Tunvirachaisakul ^{a,d}  , Rebecca L. Gould ^{a,b}, Mark C. Coulson ^c, Emma V. Ward ^c,
Gemma Reynolds ^c, Rebecca L. Gathercole ^a, Hannah Grocott ^a, Thitiporn Supasitthumrong ^d,
Athicha Tunvirachaisakul ^e, Kate Kimona ^b, Robert J. Howard ^{a,b}

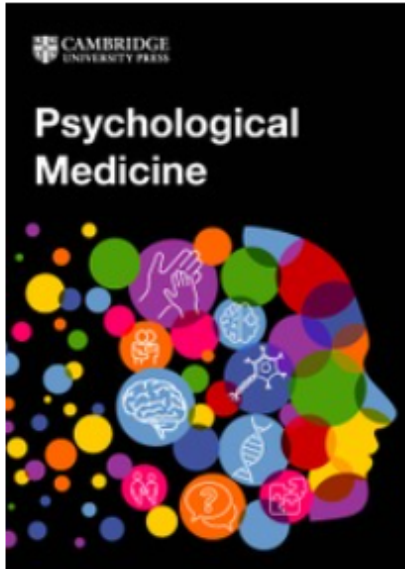
Traditional: Effect sizes + uncertainty



Traditional: Effect sizes + uncertainty



Machine learning in behavioral science



Machine learning in the prediction of depression treatment outcomes: a systematic review and meta-analysis

Published online by Cambridge University Press: 12 October 2021

Mehri Sajjadian, Raymond W. Lam, Roumen Milev, Susan Rotzinger, Benicio N. Frey, Claudio N. Soares, Sagar V. Parikh, Jane A. Foster, Gustavo Turecki and Daniel J. Müller

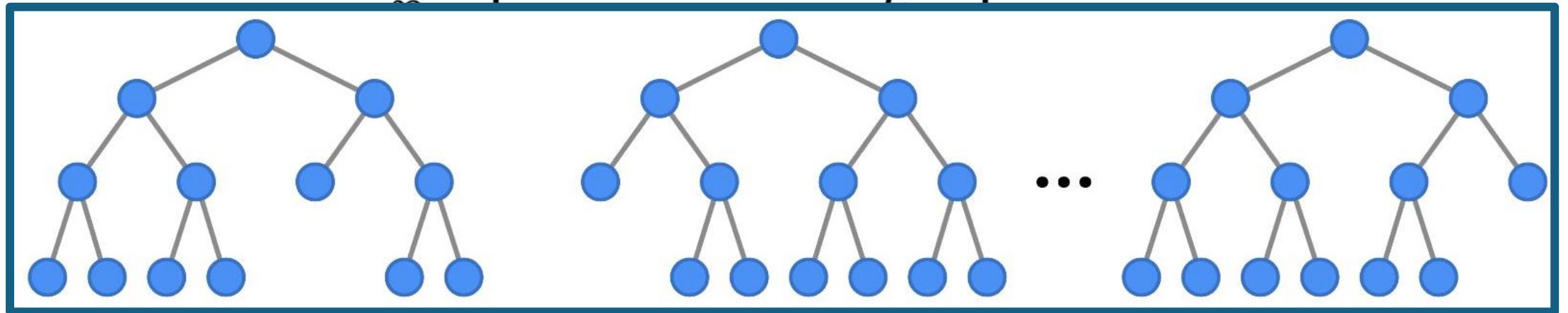
“studies reported a mean accuracy of 0.63
[95% CI 0.56-0.71]”

RuleFit (Friedman & Popescu, 2008)

$$\hat{y} = \hat{b}_0 + \sum_{p=1}^P \hat{b}_p x_p + \sum_{k=1}^K \hat{a}_k r_k(x)$$

RuleFit (Friedman & Popescu, 2008)

$$\hat{y} = \hat{b}_0 + \sum_{p=1}^P \hat{b}_p x_p + \sum_{k=1}^K \hat{a}_k r_k(x)$$



Example: Depression treatment outcomes

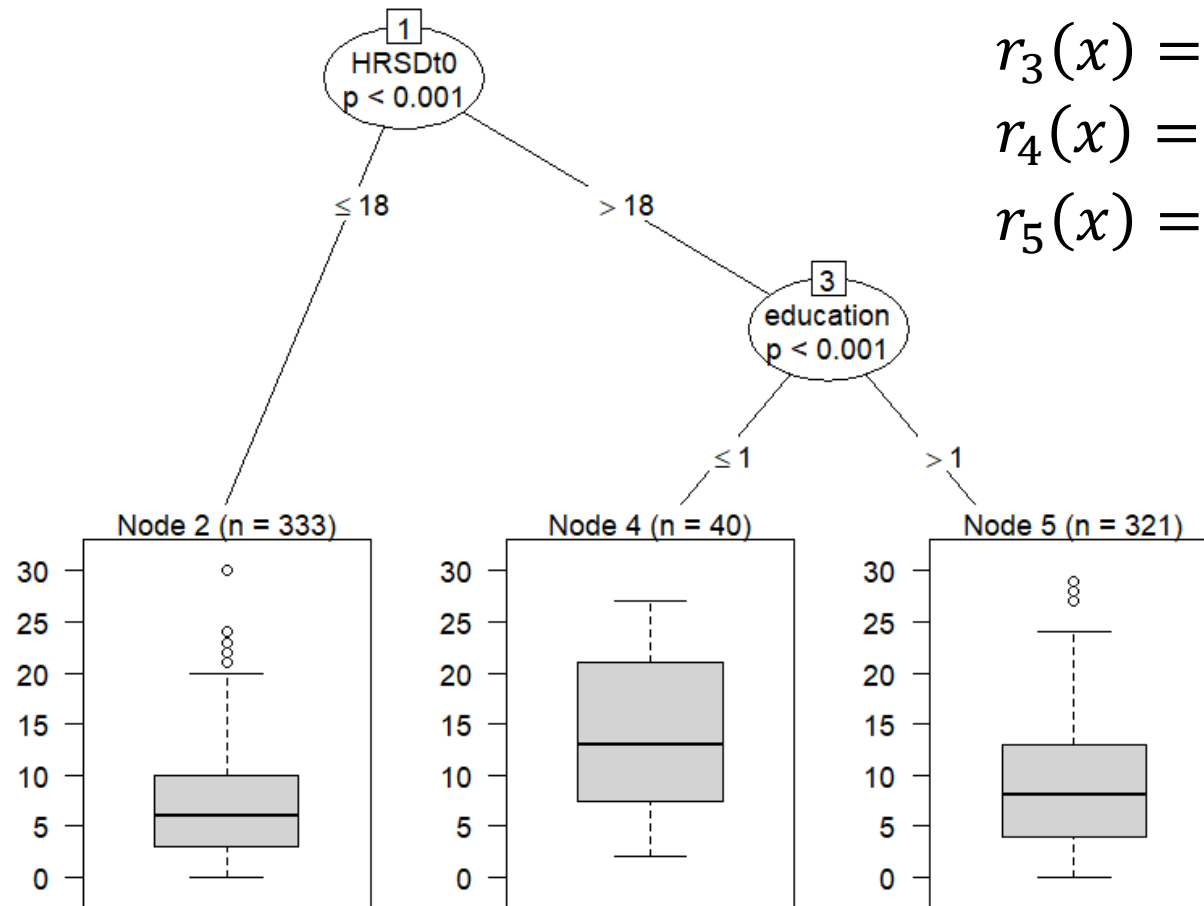
- Psychotherapy (CBT) vs. Pharmacotherapy (PHA)
- Treatment outcome: Depression rating scale score
- *N*=694 observations (from 7 studies)

```
library("pre")
set.seed(42)
pre_mod <- pre(HRSDt1 ~ Tx_group + Age + Gender + education +
               ComorbidAnxiety + HRSDt0, data = metadata)
```

```
print(pre_mod)
```

rule	coefficient		description
(Intercept)	14.0434584		1
rule41	1.9168924	Tx_group %in% c("CBT") & ComorbidAnxiety %in% c("yes")	
education	-1.2840478		1 <= education <= 4
rule18	-1.1955925		HRSDt0 <= 24
rule61	1.1180350		HRSDt0 > 20 & Tx_group %in% c("CBT")
rule38	1.1021041		ComorbidAnxiety %in% c("yes") & HRSDt0 > 20
rule65	-1.0429025	HRSDt0 <= 24 & HRSDt0 <= 16 & Gender %in% c("female")	
rule24	-0.8392146		HRSDt0 <= 18 Gender %in% c("female")
rule1	-0.1621209		HRSDt0 <= 18

From trees to rules



$$r_1(x) = I(\text{HRSDt0} > 18)$$

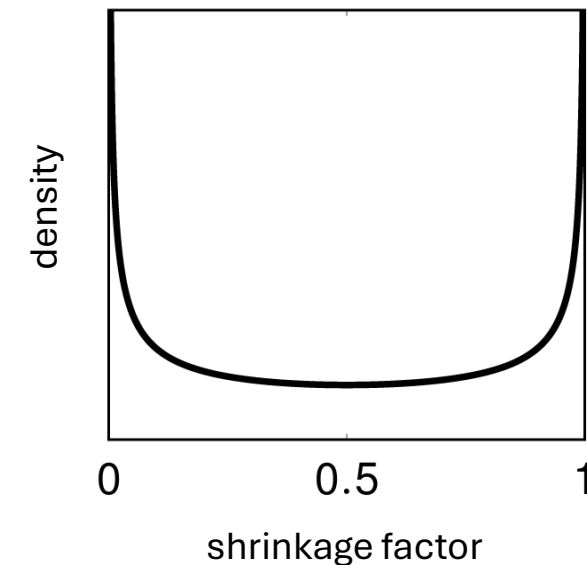
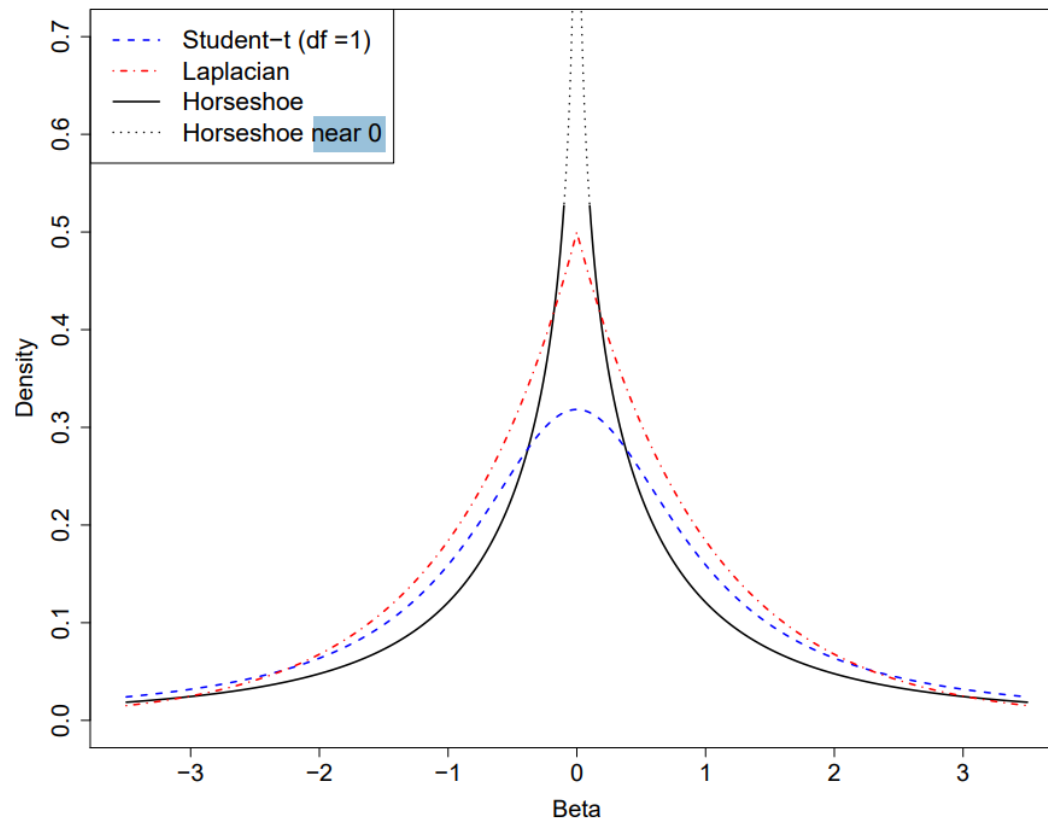
$$r_3(x) = I(\text{education} > 1)$$

$$r_4(x) = I(\text{HRSDt0} > 18) \cdot I(\text{education} \leq 1)$$

$$r_5(x) = I(\text{HRSDt0} > 18) \cdot I(\text{education} > 1)$$

RuleSHAP (Spadaccini et al., 2025)

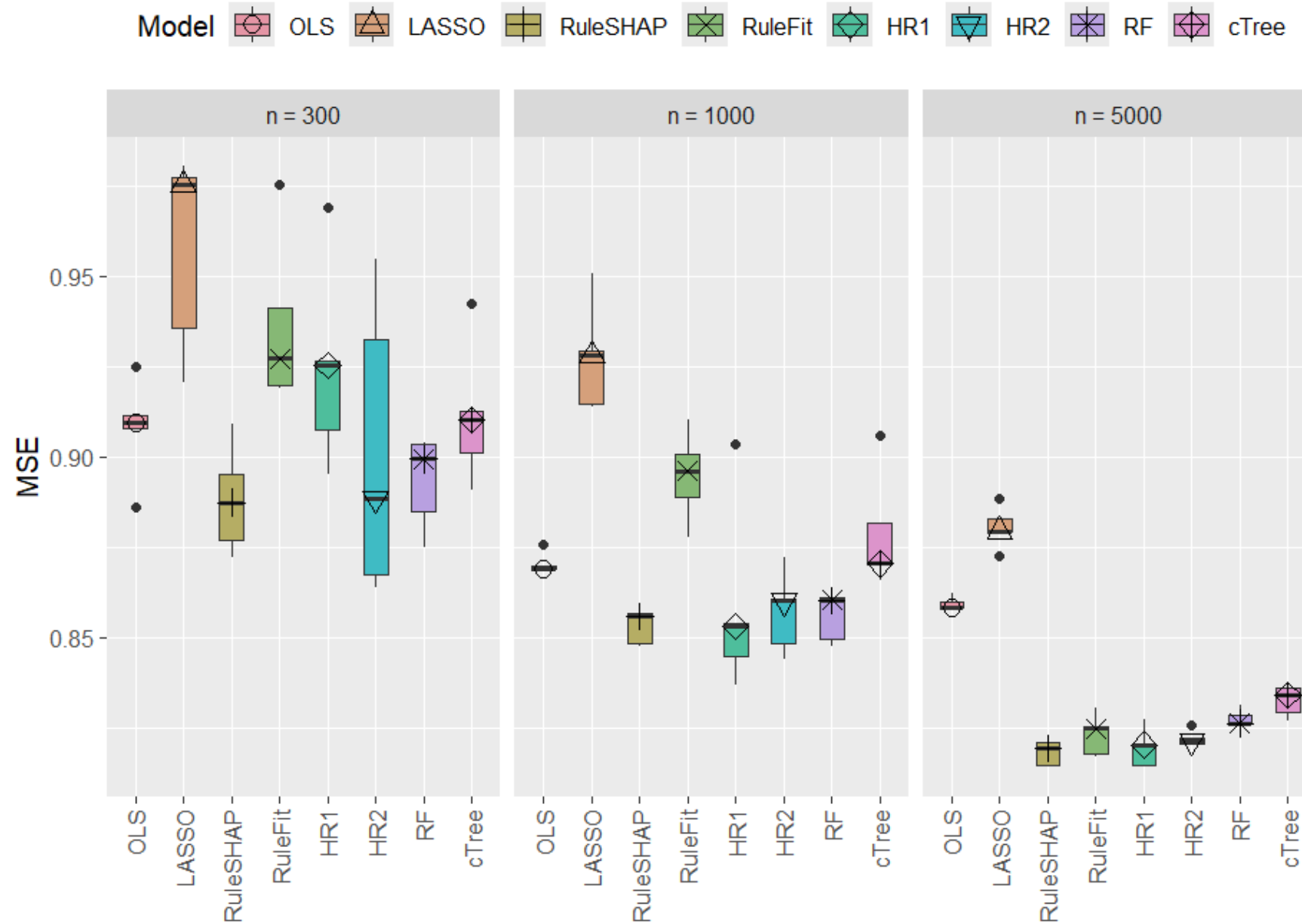
Estimation and uncertainty with hierarchical Horseshoe prior
(Busatto & van de Wiel, 2025)



Example: Cholesterol levels

- Merged data from epidemiological cohort studies ($N \approx 20,000$)
- 8 predictors, 4 artificial noise features

(b) Cholesterol



Estimating effects

Notoriously difficult in presence of multicollinearity and interactions!

SHAP methods (e.g., Lundberg & Lee, 2017) use *Shapley* values

Estimating effects: Shapley values

$$\varphi_j(v) = \frac{1}{p} \sum_{S \subseteq P \setminus \{j\}} \binom{p-1}{s}^{-1} (v(S \cup \{j\}) - v(S))$$

v is a model's output (e.g., prediction)

j a predictor of interest

P the full set of p predictors

S a subset of s predictors

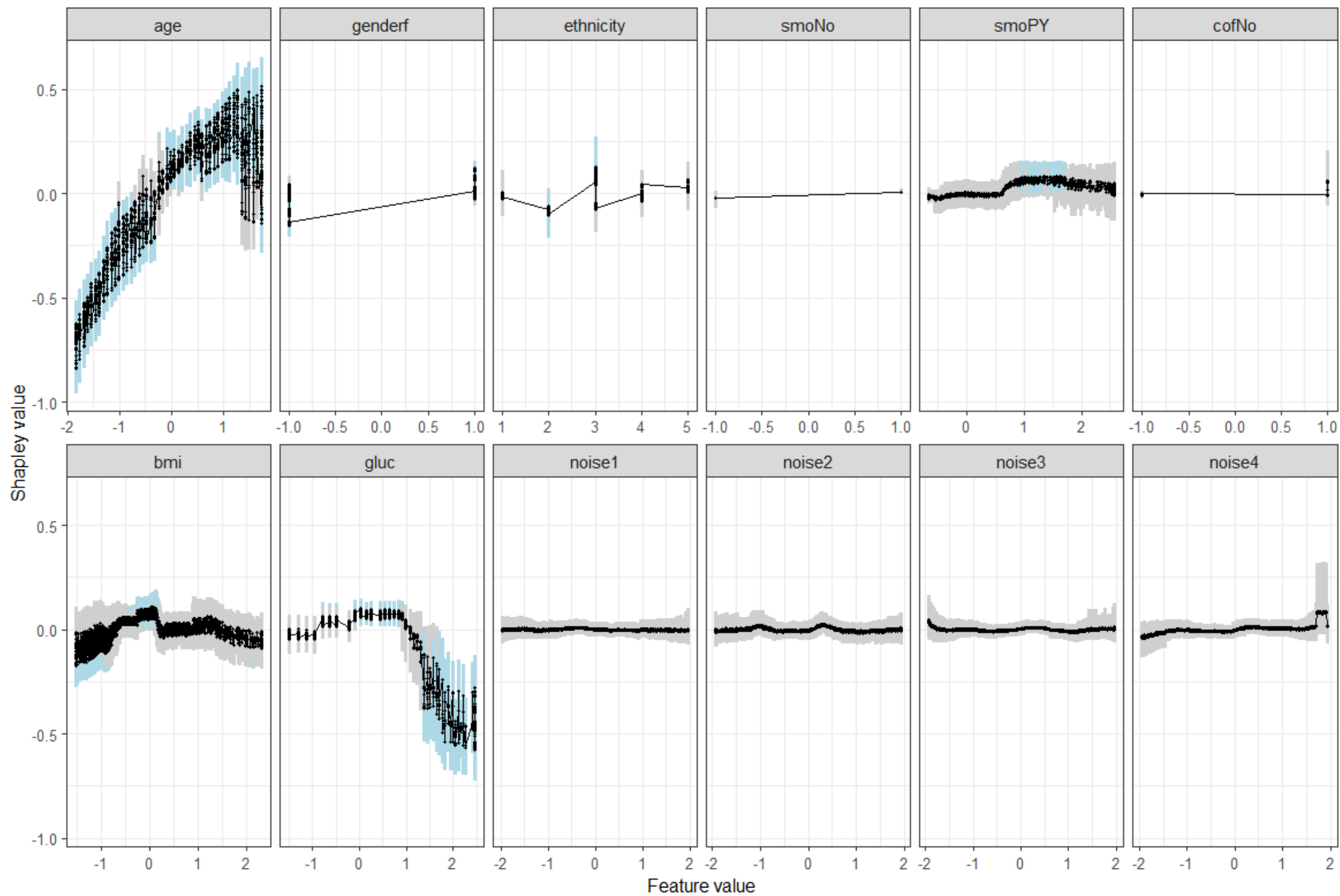
Estimating effects

Notoriously difficult in presence of multicollinearity and interactions

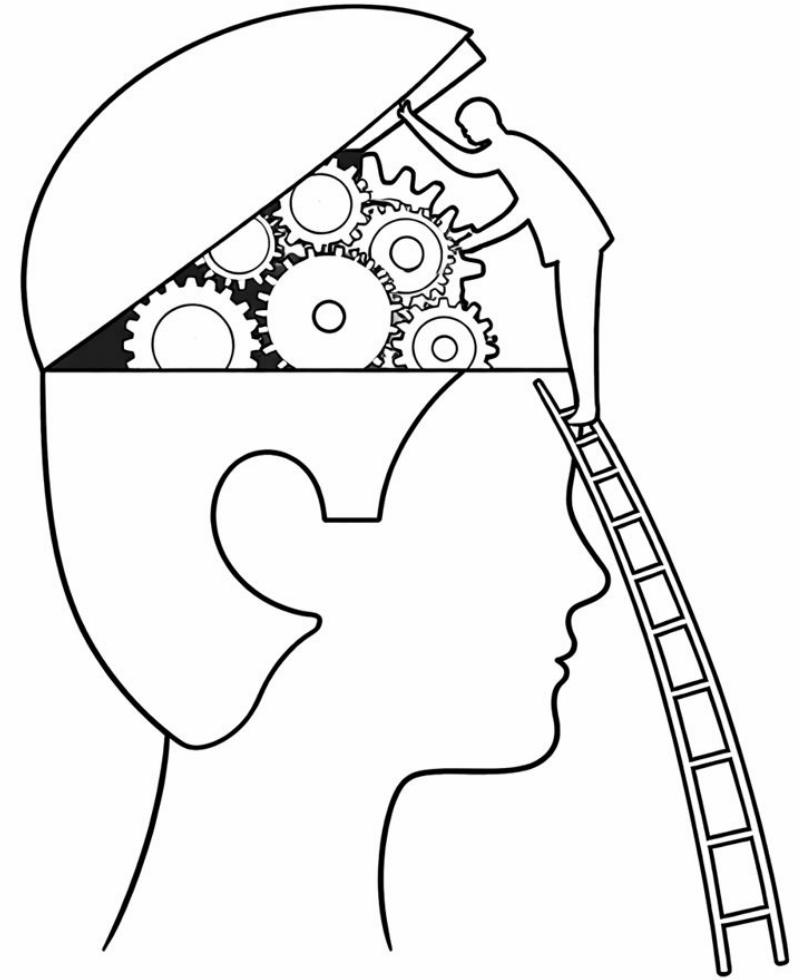
SHAP methods (e.g., Lundberg & Lee, 2017) use *Shapley* values

Notoriously heavy computation!

RuleSHAP (Spadaccini et al., 2025): Simplified expression for interventional SHAP values for rule ensembles

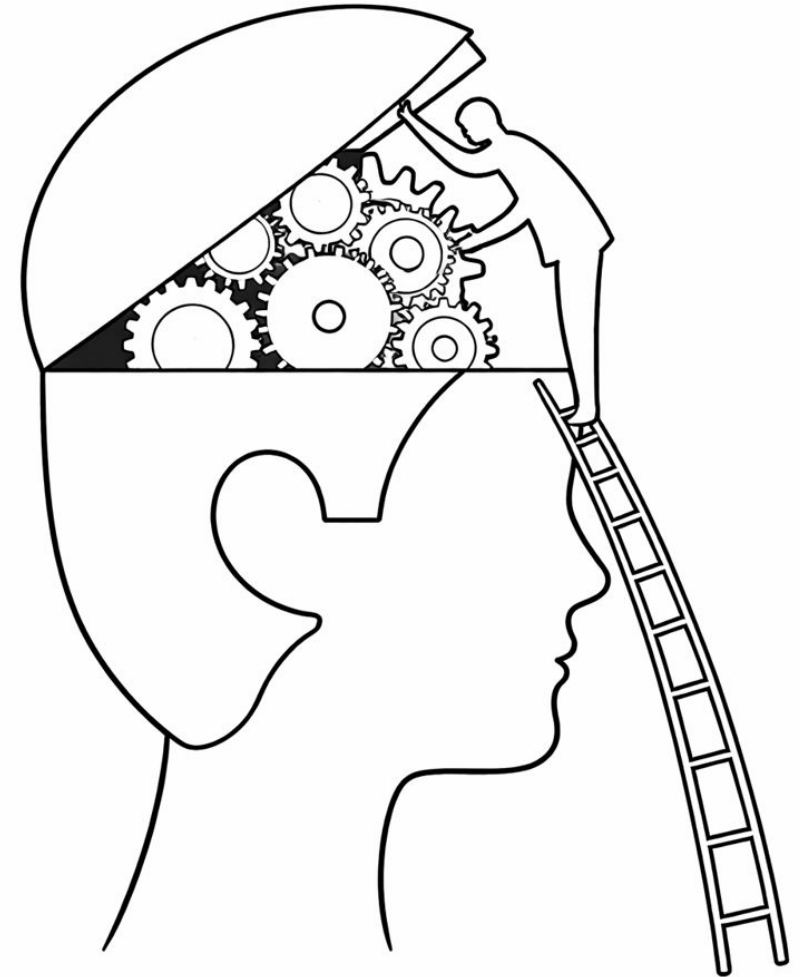






Vacancy for a PhD student!

m.fokkema@fsw.leidenuniv.nl



References

- Busatto, C., & van de Wiel, M. (2023). Informative co-data learning for high-dimensional horseshoe regression. *Biometrical Journal*, 68, e70105.
- Fokkema, M. (2020). Fitting prediction rule ensembles with R package pre. *Journal of Statistical Software*, 92, 1-30.
- Fokkema, M., & Strobl, C. (2020). Fitting prediction rule ensembles to psychological research data: An introduction and tutorial. *Psychological Methods*, 25(5), 636.
- Friedman, J. H., & Popescu, B. E. (2008). Predictive Learning Via Rule Ensembles. *Annals of Applied Statistics*, 2(3), 916-954.
- Lundberg, S. M., & Lee, S.-I. (2017). A unified approach to interpreting model predictions. *Advances in Neural Information Processing Systems*, 30.
- Spadaccini, G., Fokkema, M., & van de Wiel, M. A. (2025). Hypothesis-free discovery from epidemiological data by automatic detection and local inference for tree-based nonlinearities and interactions. *arXiv preprint arXiv:2505.00571*.