# MetaForest

Using random forests to explore heterogeneity in meta-analysis

Caspar J. van Lissa, Utrecht University NL
c.j.vanlissa@uu.nl

# Applied meta-analysis

* Considered "golden standard" of evidence Crocetti, 2016
* "Superstitions" that it is somehow immune to small-sample problems because each data point is based on an entire study
* Often small N, but many moderators (either measured or ignored)

# Dealing with heterogeneity

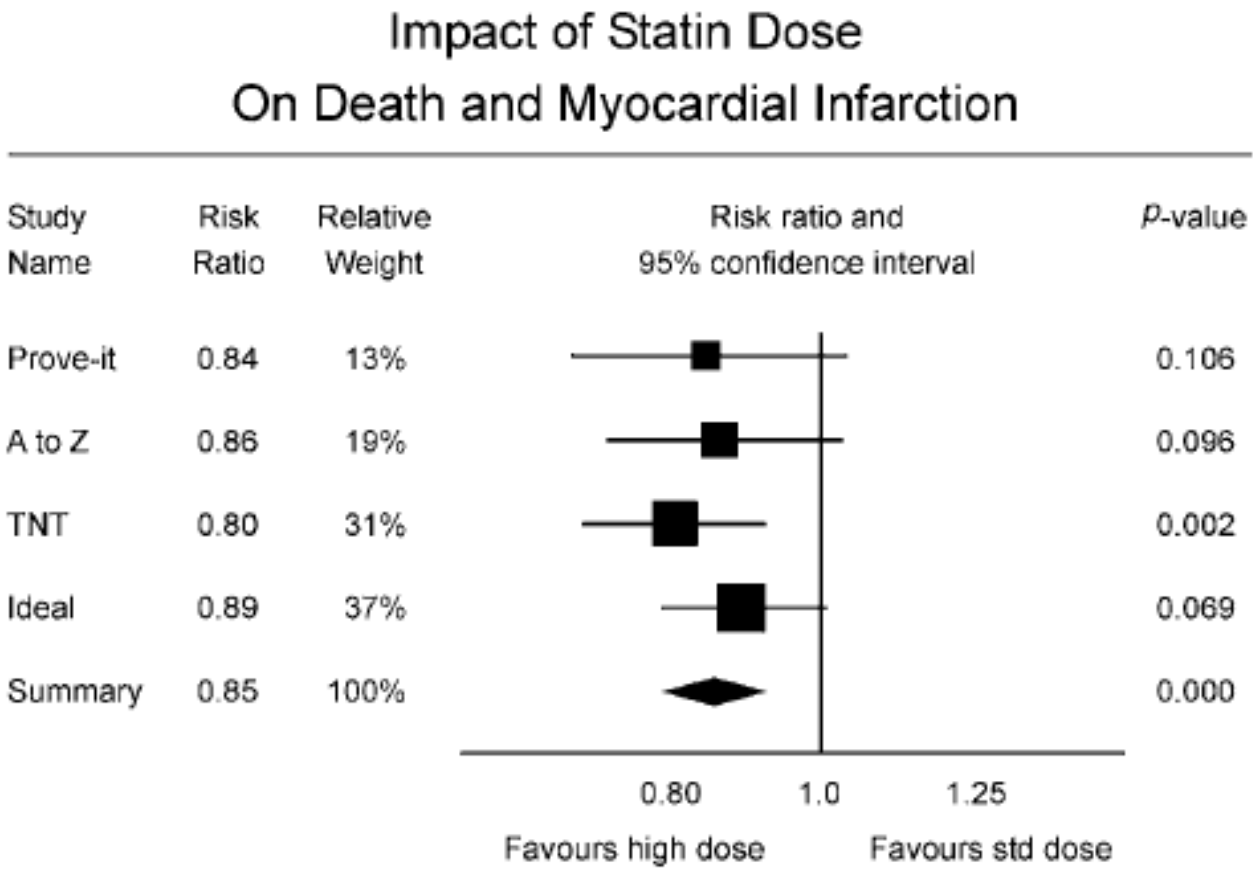
1. Studies are too different
   * Do not meta-analyze

2. Studies are similar, but not 'identical'
   * Random-effects meta-analysis

3. There are known differences between studies
   * Code differences as moderating variables
   * Control for moderators using meta-regression (Higgins et al., 2009)

# Types of meta-analysis

- **Fixed-Effect meta-analysis:**
  - One "true" effect size
  - Observed effect sizes differ due to sampling error
  - Weighted "mean" of effect sizes
  - Big N → more influence



# Types of meta-analysis

* **Random-Effects meta-analysis:**
  * Distribution of true effect sizes
  * Observed effect sizes differ due to:
    * Sampling error (as before)
    * The variance of this distribution of effect sizes
  * Weights based on precision **and** heterogeneity
    * Study weights become more equal, the more between-studies heterogeneity there is

# Meta-regression

* True effect size is a function of moderators
* Weighted regression
  * Fixed-effects or random-effects weights

# Problem with heterogeity

* Differences in terms of samples, operationalizations, and methods might all introduce heterogeneity Liu, Liu, & Xie, 2015
* When the number of studies is small, meta-regression lacks power to test more than a few moderators
* We often lack theory to whittle down the list of moderators to a manageable number Thompson & Higgins, 2002
* If we include too many moderators, we might overfit the data

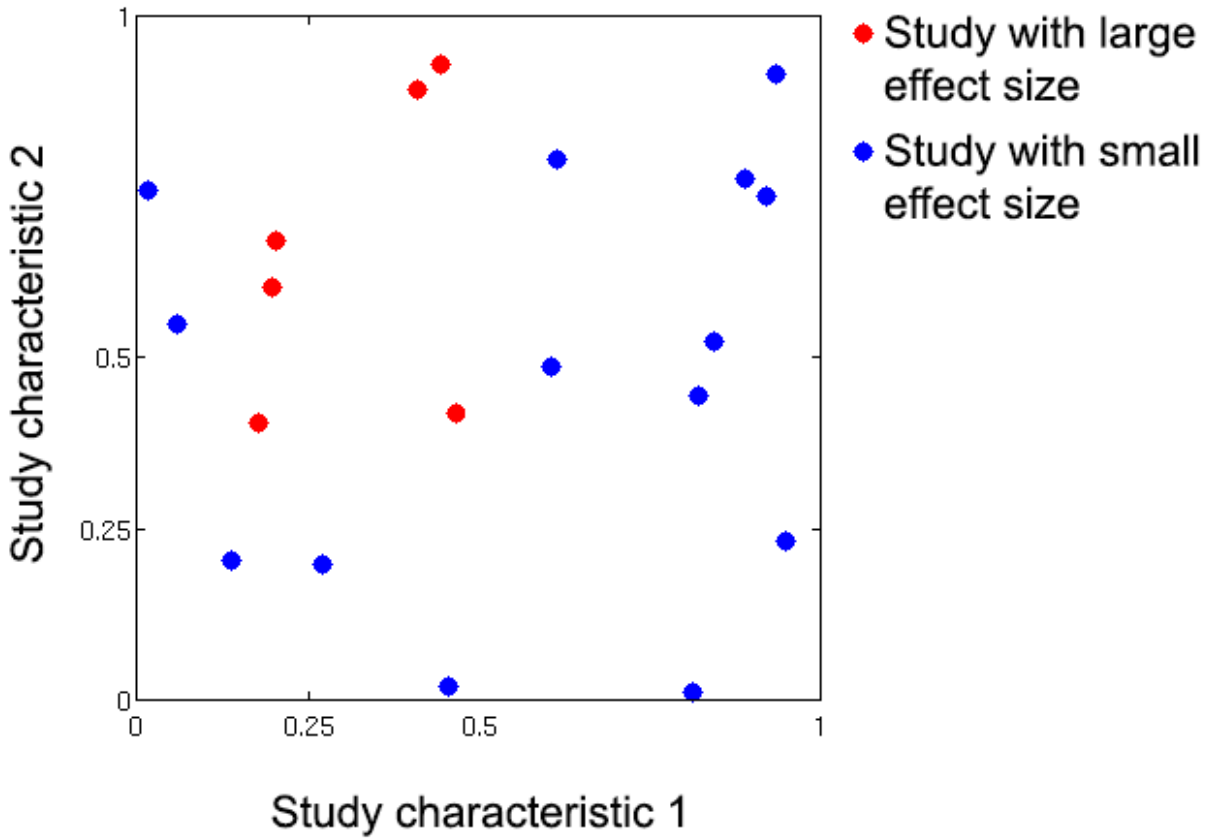
How can we weed out which study characteristics influence effect size?

# A solution has been proposed…

* Dusseldorp and colleagues (2014) used "Classification Trees" to explore which combinations of study characteristics jointly predict effect size
* The Dependent Variable is **Effect Size**
* The Independent Variables are **Study Characteristics** (moderators)

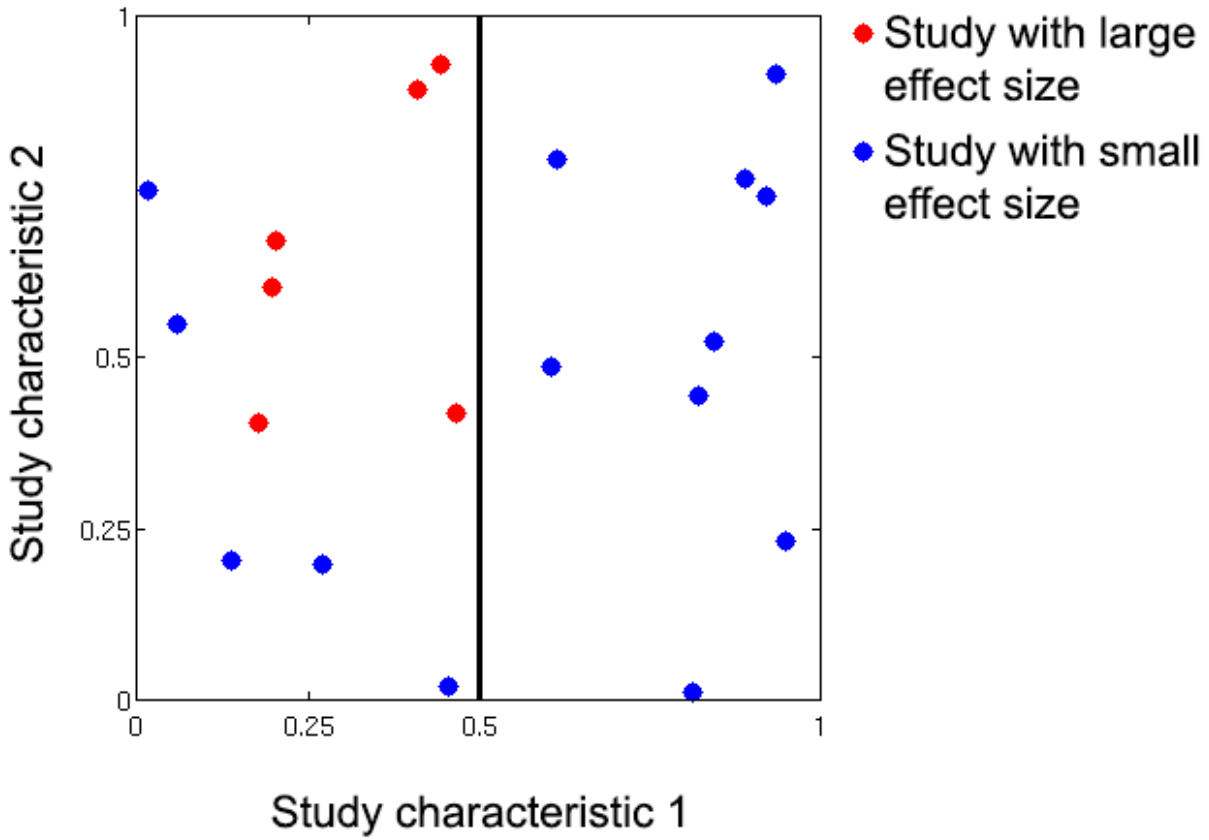
# How do tree-based models work?

* They predict the DV by splitting the data into groups, based on the IV's



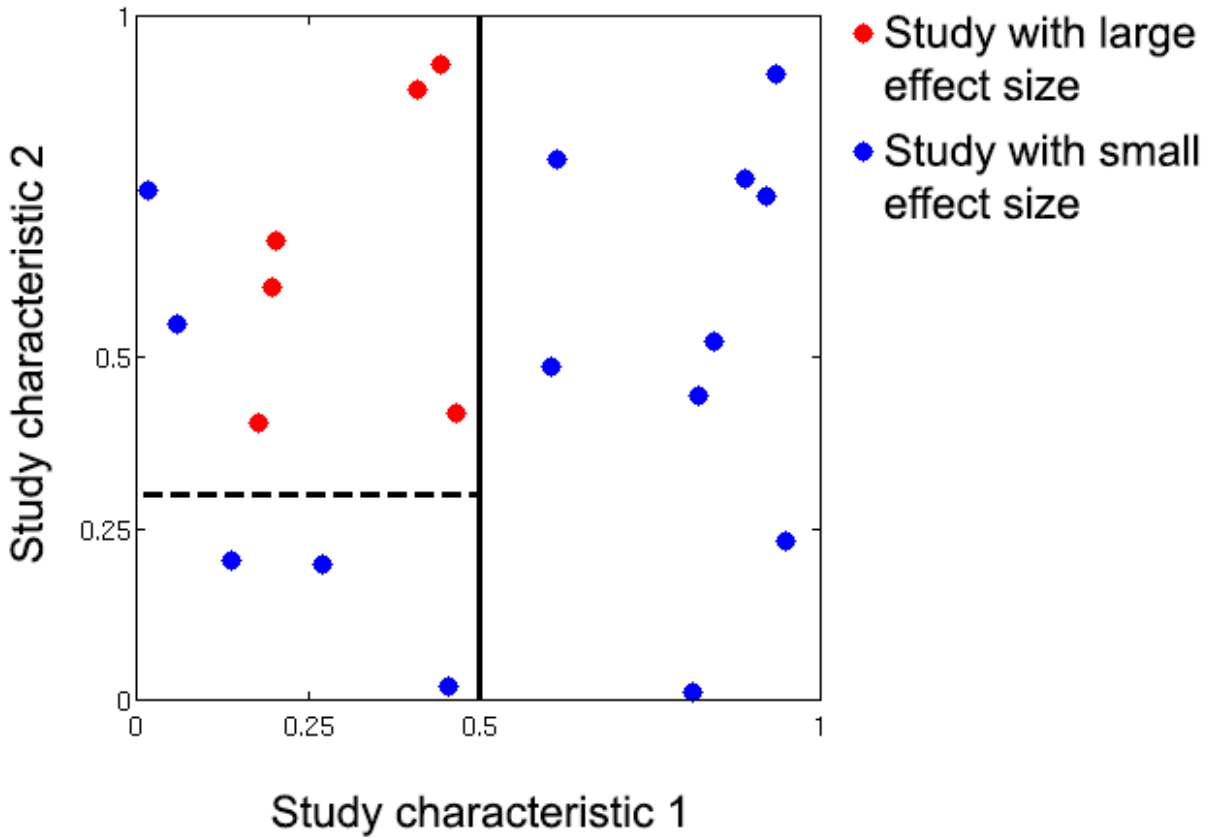
# How do tree-based models work?

* They predict the DV by splitting the data into groups, based on the IV's



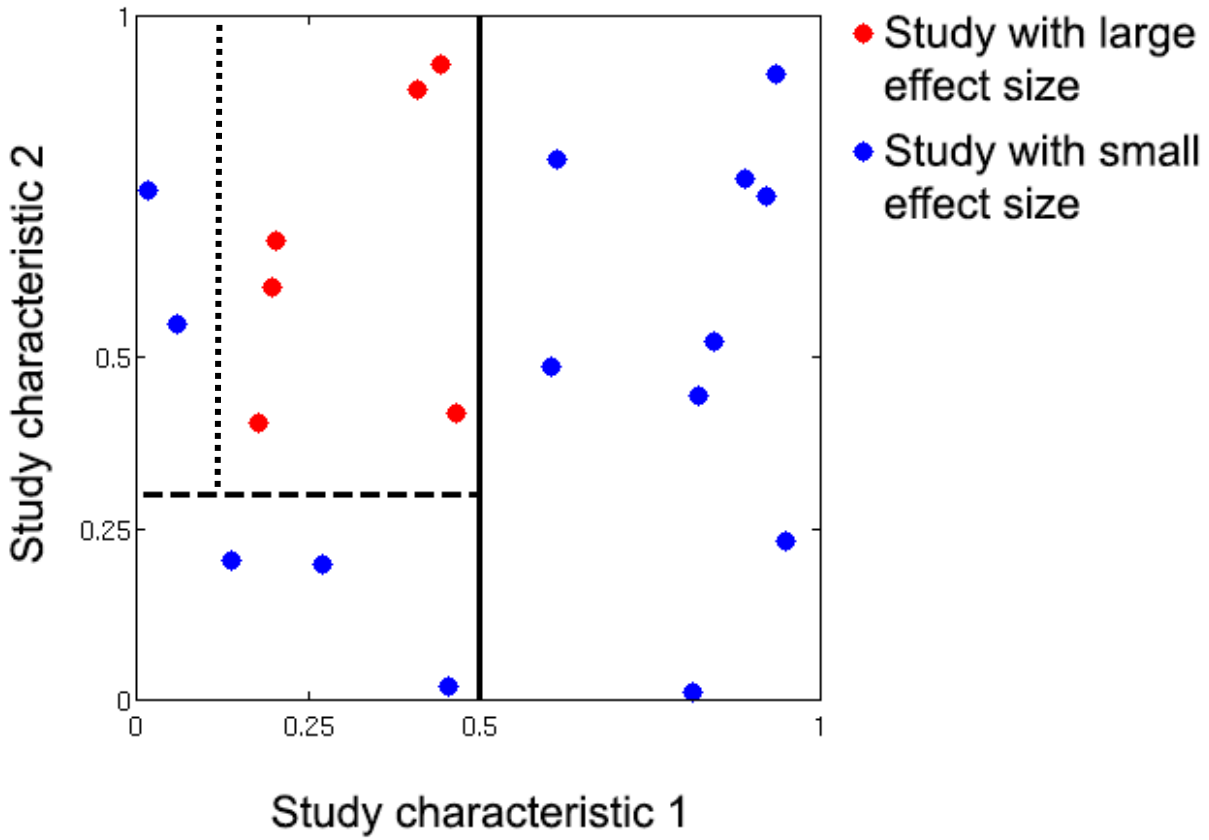
# How do tree-based models work?

* They predict the DV by splitting the data into groups, based on the IV's



# How do tree-based models work?

* They predict the DV by splitting the data into groups, based on the IV's



# Advantages of trees over regression

* Trees easily handle situations where there are many predictors relative to observations
* Trees capture interactions and non-linear effects of moderators
* Both these conditions are likely to be the case when performing meta-analysis in a heterogeneous body of literature

# Limitations of single trees

* Single trees are very prone to overfitting

# Introducing "MetaForest" Van Lissa et al., in preparation

**Random Forests**

1. Draw many (+/-1000) bootstrap samples
2. Grow a trees on each bootstrap sample
3. To make sure each tree learns something unique, they are only allowed to choose the best moderator from a small random selection of moderators at each split
4. Average the predictions of all these trees

# Benefits of random forests

* Random forests are **robust to overfitting**
  * Each tree captures some "true" effects and some idiosyncratic noise
  * Noise averages out across bootstrap samples

* Random forests make **better predictions** than single trees
  * Single trees predict a constant value for each "node"
  * Forests average predictions of many trees, leading to smooth prediction curves

# How does MetaForest work?

* Apply random-effects weights to random forests
* Just like in classic meta-analysis, more precise studies are more influential in building the model

# What do I report in my paper?

* An "$R^2_{oob}$": An estimate of how well this model predicts **new** data

* Variable importance metrics, indicating which moderators most strongly predict effect size

* Partial dependence plots:
  Marginal relationship between moderators and effect size

# Is it any good?

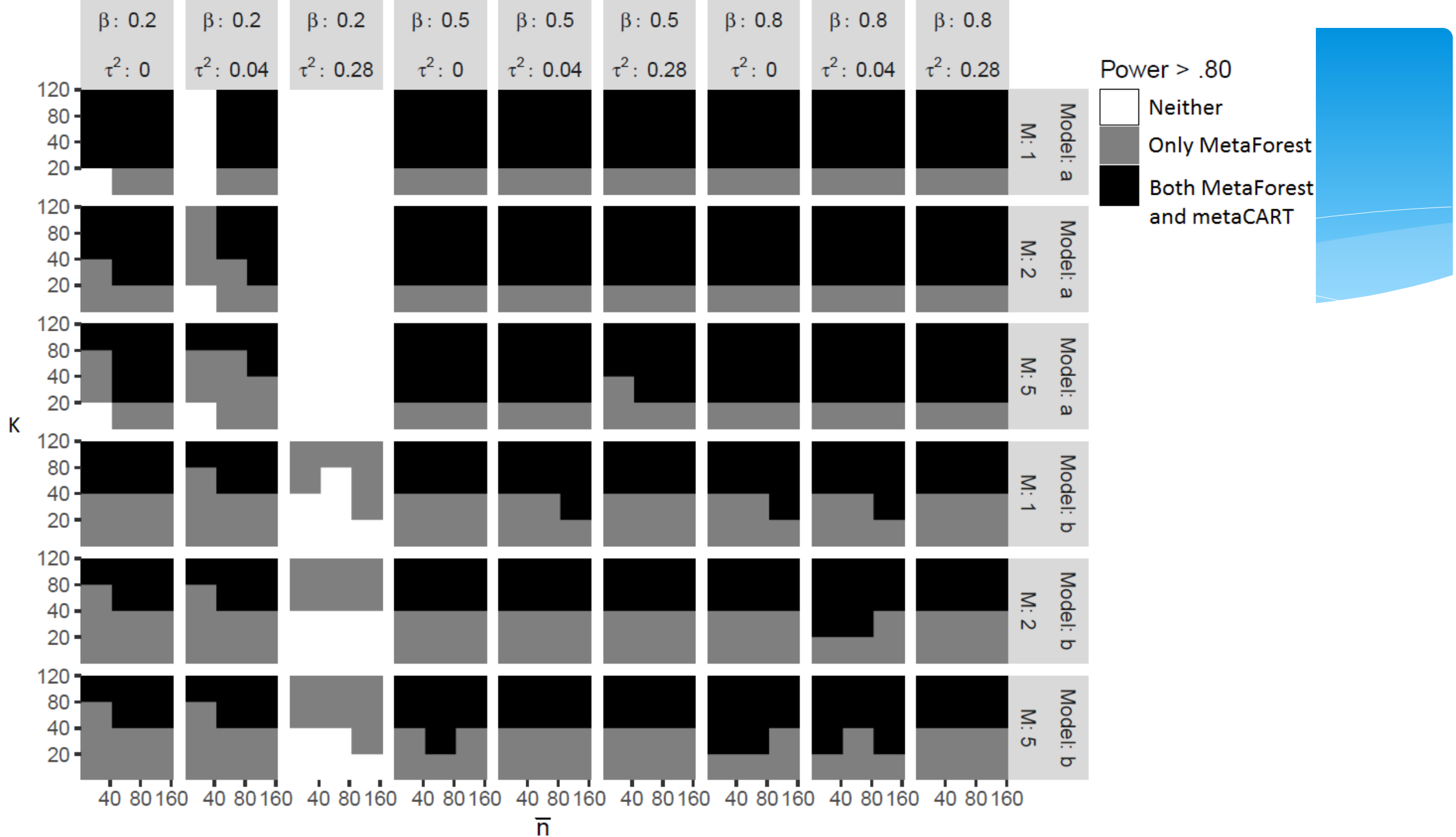
* Several simulation studies examining:
  * Predictive performance
  * Power
  * Ability to identify relevant / irrelevant moderators
* Van Lissa, 2017: https://osf.io/khjgb/

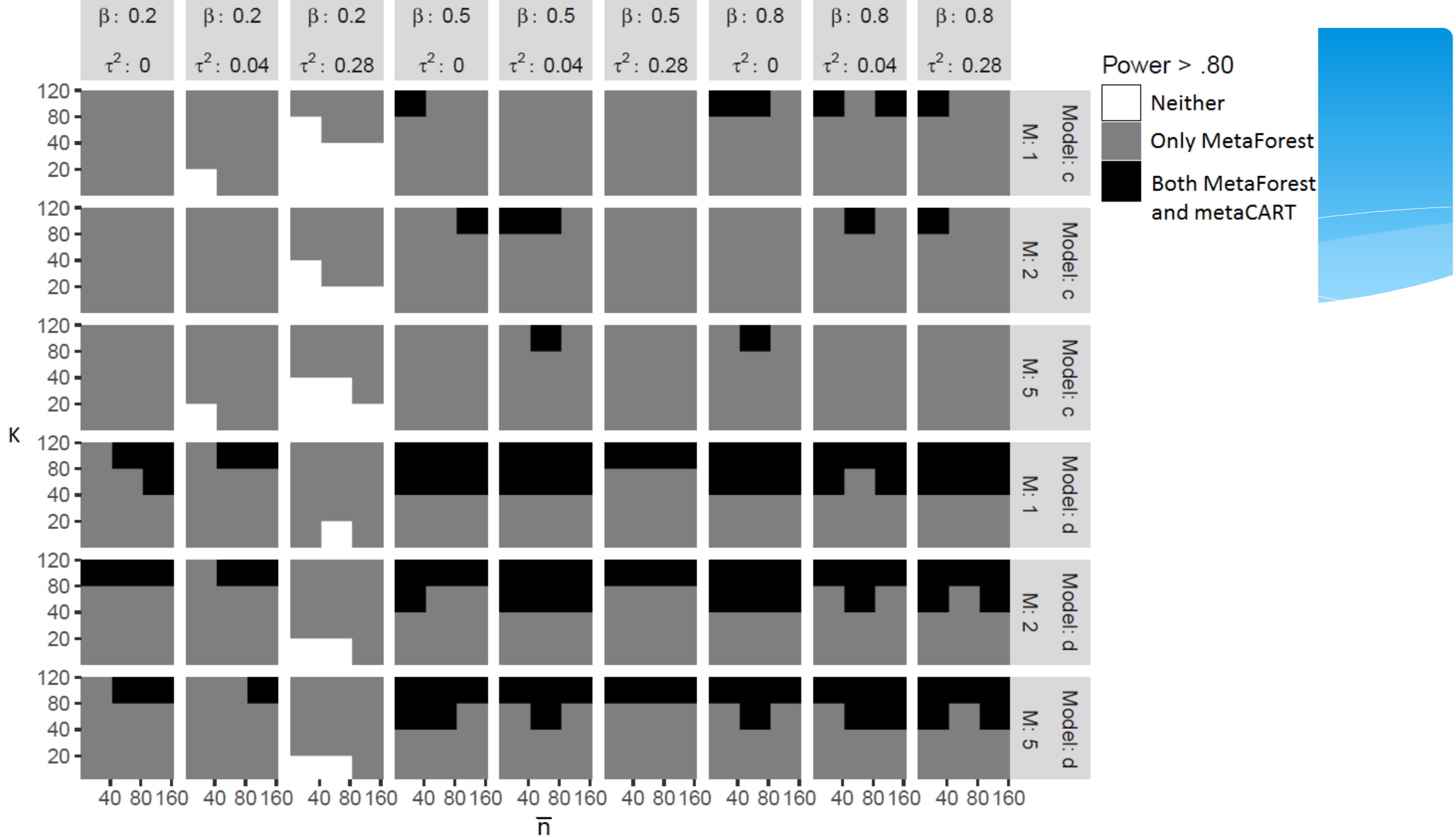
# Focusing on one simulation study

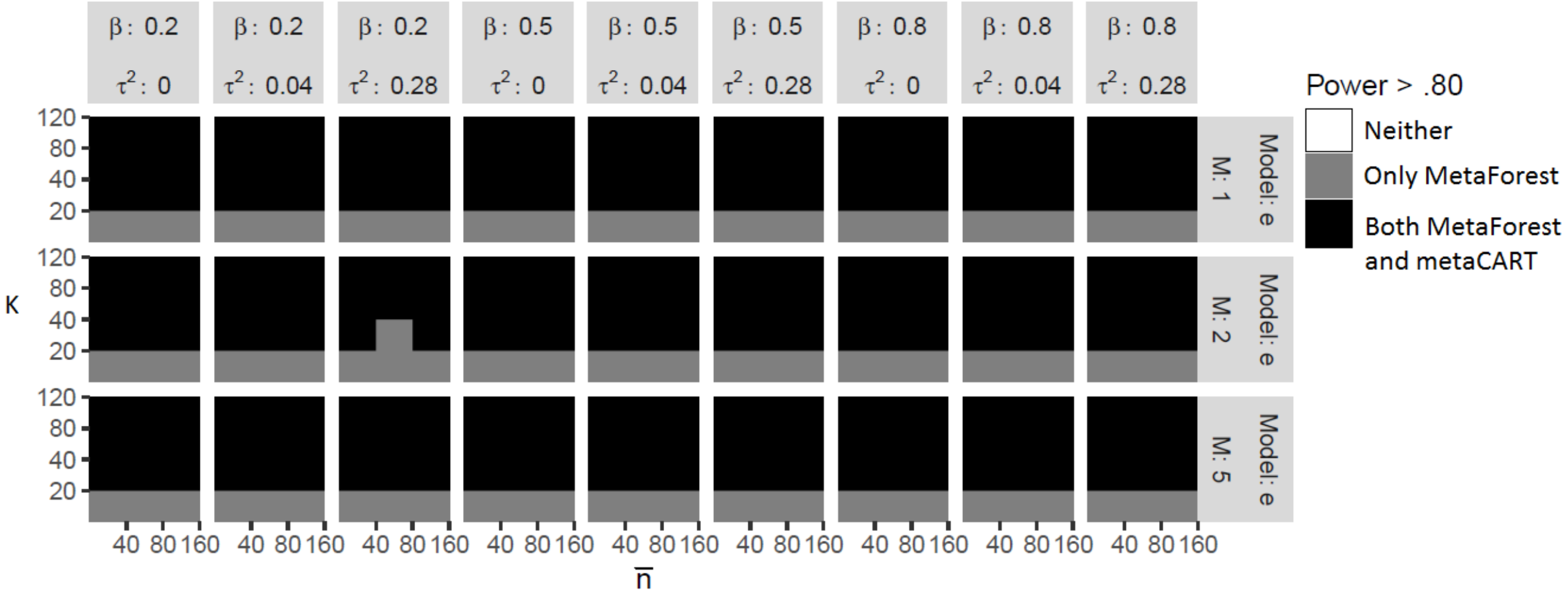
* Design factors:
  * *k*: Number of studies in meta-analysis (20, 40, 80, and 120)
  * *N*: Average within-study sample size (40, 80, and 160)
  * *M*: Number of irrelevant/noise moderators (1, 2, and 5)
  * β: Population effect size (.2, .5, and .8)
  * $\tau^2$: Residual heterogeneity (0, .04, and .28) Van Erp et al., 2017 (0, 50 and 80$^{th}$ percentile)
  * Model:
    * (a) main effect of one moderator
    * (b) two-way interaction
    * (c) three-way interaction
    * (d) two two-way interactions
    * (e) non-linear, cubic relationship

# Power analyses

* To determine practical guidelines, we examined under what conditions MetaForest achieved a positive $R^2$ in new data at least 80% of the time


β: 0.2 τ²: 0
β: 0.2 τ²: 0.04
β: 0.2 τ²: 0.28
β: 0.5 τ²: 0
β: 0.5 τ²: 0.04
β: 0.5 τ²: 0.28
β: 0.8 τ²: 0
β: 0.8 τ²: 0.04
β: 0.8 τ²: 0.28
Model: a M: 1
Model: a M: 2
Model: a M: 5
Model: b M: 1
Model: b M: 2
Model: b M: 5
K
120
80
40
20
40 80 160
n̄
Power > .80
Neither
Only MetaForest
Both MetaForest and metaCART


β: 0.2 τ²: 0
β: 0.2 τ²: 0.04
β: 0.2 τ²: 0.28
β: 0.5 τ²: 0
β: 0.5 τ²: 0.04
β: 0.5 τ²: 0.28
β: 0.8 τ²: 0
β: 0.8 τ²: 0.04
β: 0.8 τ²: 0.28
Model: c M: 1
Model: c M: 2
Model: c M: 5
Model: d M: 1
Model: d M: 2
Model: d M: 5
K
120
80
40
20
40 80 160
n̄
Power > .80
Neither
Only MetaForest
Both MetaForest and metaCART


β: 0.2 τ²: 0
β: 0.2 τ²: 0.04
β: 0.2 τ²: 0.28
β: 0.5 τ²: 0
β: 0.5 τ²: 0.04
β: 0.5 τ²: 0.28
β: 0.8 τ²: 0
β: 0.8 τ²: 0.04
β: 0.8 τ²: 0.28
Model: e M: 1
Model: e M: 2
Model: e M: 5
K
120
80
40
20
40 80 160
n̄
Power > .80
Neither
Only MetaForest
Both MetaForest and metaCART

# Results

* MetaForest had sufficient power in most conditions, even for as little as 20 studies,
  * Except when the effect size was small ($\beta = 0.2$), and residual heterogeneity was high ($\tau^2 = 0.28$)

* Power was most affected by true effect size and residual heterogeneity, followed by the true underlying model

# Integrate in your workflow

* MetaForest is a comprehensive approach to Meta-Analysis.
* You could just report:
  * Variable importance
  * Partial prediction plots
  * Residual heterogeneity

* Alternatively, add it to your existing Meta-Analysis workflow
  * Use it to check for relevant moderators
  * Follow up with classic meta-analysis

# Can you get it published?

**Methodological journal:**

* Received positive Reviews
* Editor: "the field of psychology is simply not ready for this technique"

---

**Applied journal:** (Journal of Experimental Social Psychology, 2018)

* Included MetaForest as a check for moderators
* Accepted WITHOUT QUESTIONS about this new technique
* Editor: "I see the final manuscript as having great potential to inform the field."
* Manuscript, data, and syntax at https://osf.io/sey6x/

# How to do it

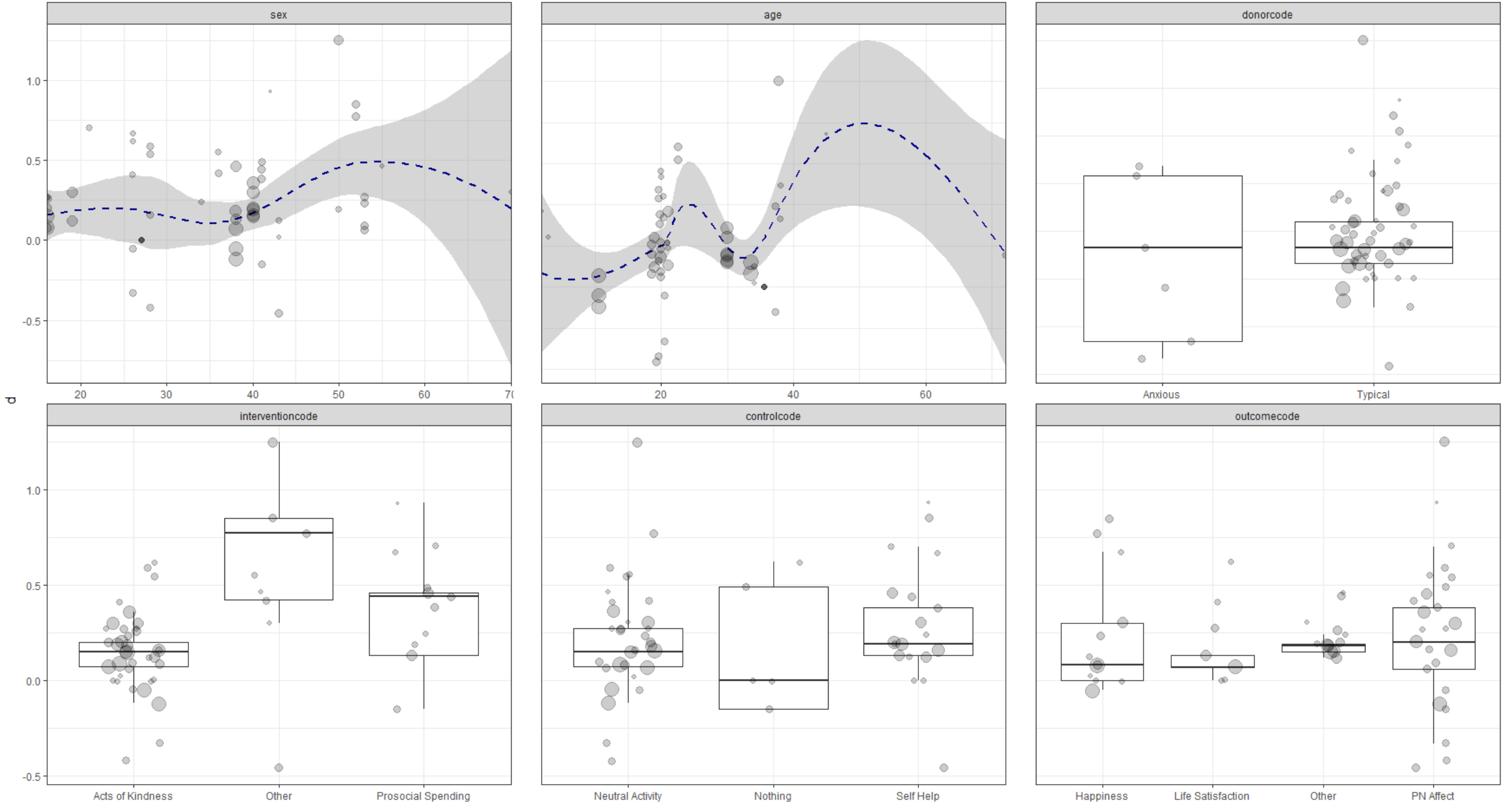
Fukkink, R. G., & Lont, A. (2007). Does training matter? A meta-analysis and review of caregiver training studies. *Early Childhood Research Quarterly*, 22(3), 294-311.

**Small sample: 17 studies** (79 effect sizes)

**Dependent variable:** Intervention effect (Cohen's D)

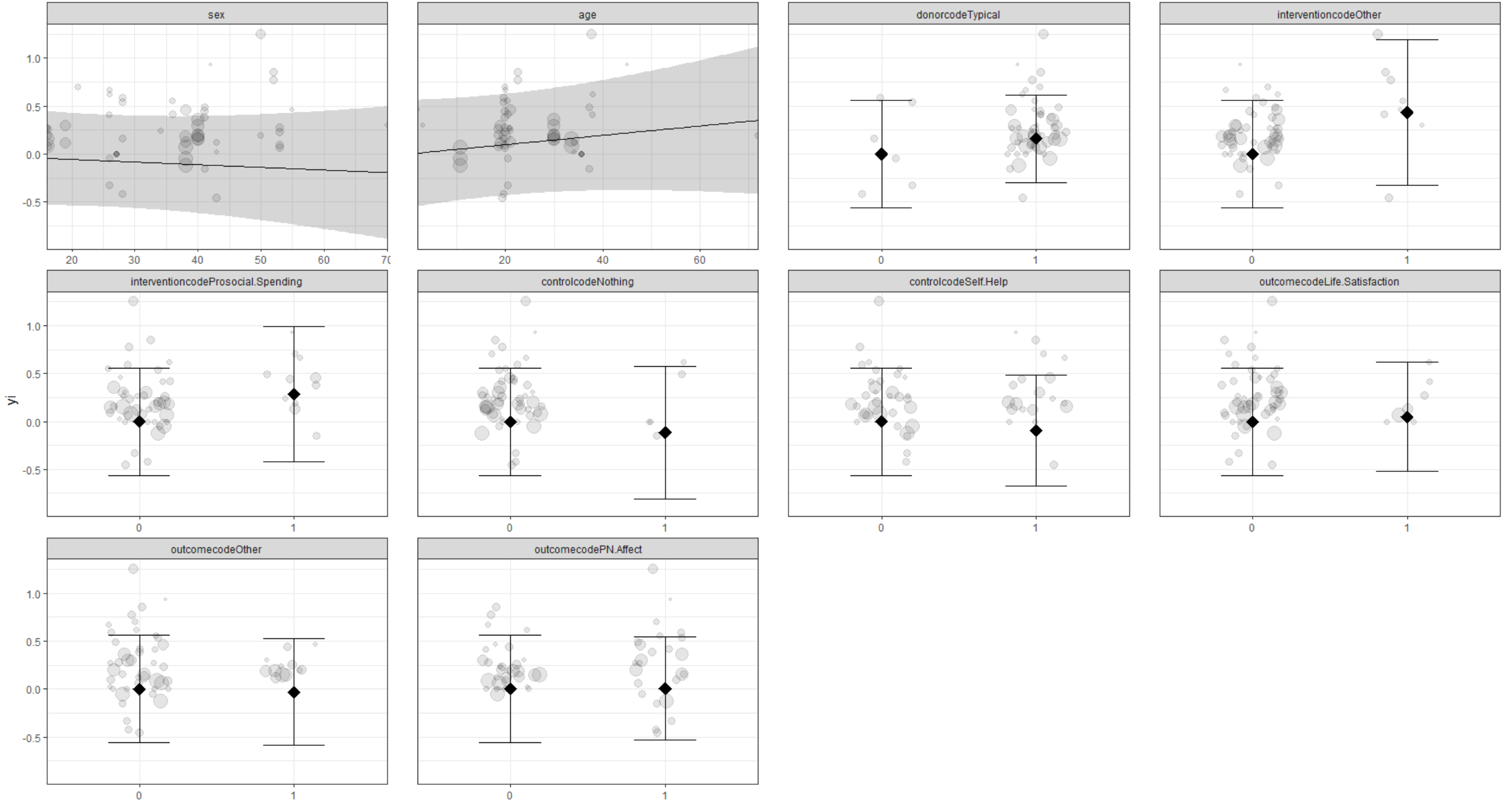
**Moderators:**

* DV_Aligned: Outcome variable aligned with training content?
* Location: Conducted in childcare center or elsewhere?
* Curriculum: Fixed curriculum?
* Train_Knowledge: Focus on teaching knowledge?
* Pre_Post: Is it a pre-post design?
* Blind: Were researchers blind to condition?
* Journal: Is this study published in a peer-reviewed journal?




```
WeightedScatter(data, yi="di")
```

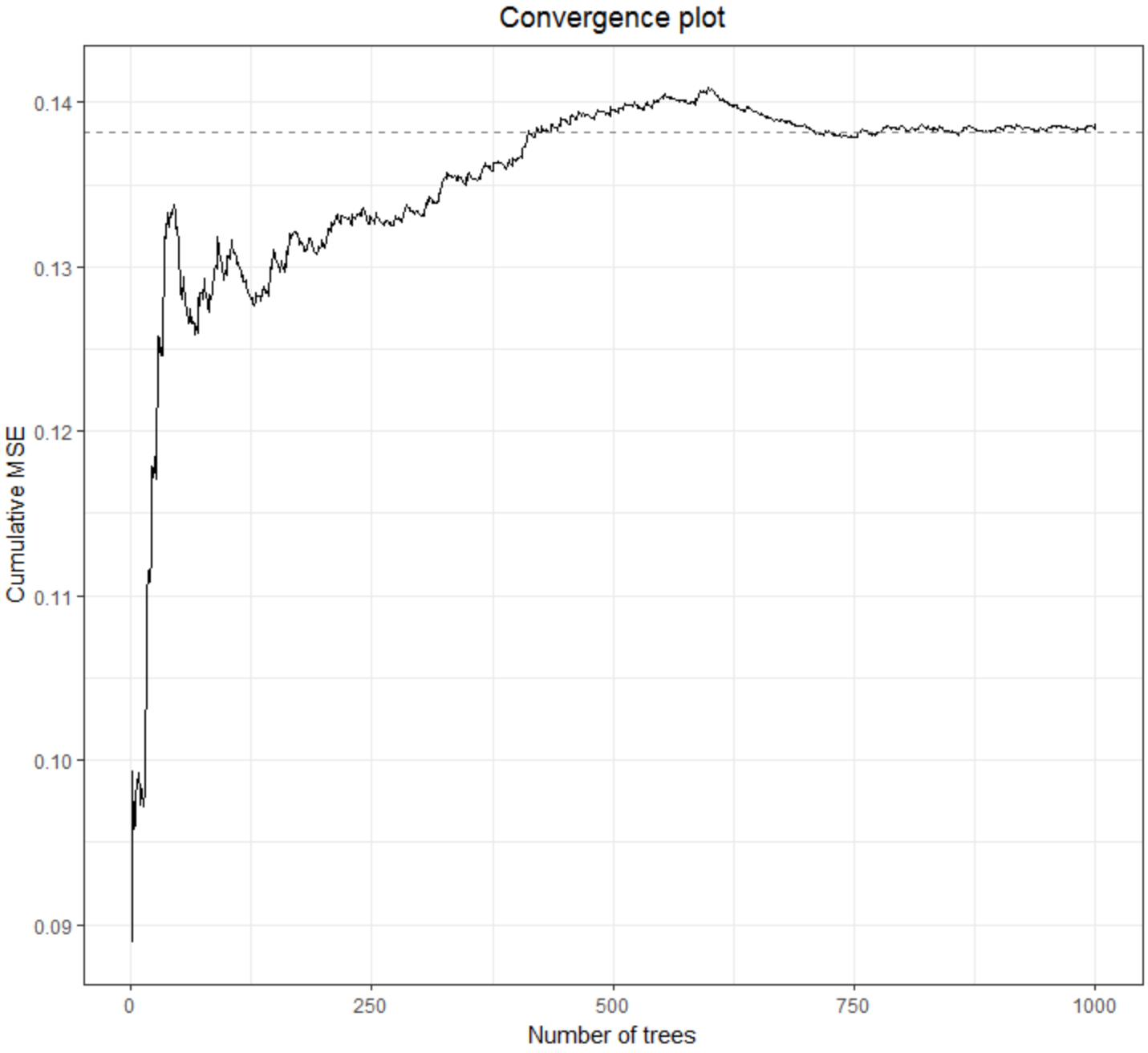
```
res <- rma.mv(d, vi, random = ~ 1 | study_id, mods = moderators, data=data)

                                   estimate      se     zval    pval    ci.lb   ci.ub
intrcpt                             -0.0002  0.2860  -0.0006  0.9995  -0.5607  0.5604
sex                                 -0.0028  0.0058  -0.4842  0.6282  -0.0141  0.0085
age                                  0.0049  0.0053   0.9242  0.3554  -0.0055  0.0152
donorcodeTypical                     0.1581  0.2315   0.6831  0.4945  -0.2956  0.6118
interventioncodeOther                0.4330  0.1973   2.1952  0.0281   0.0464  0.8196  *
interventioncodeProsocial Spending   0.2869  0.1655   1.7328  0.0831  -0.0376  0.6113  .
controlcodeNothing                  -0.1136  0.1896  -0.5989  0.5492  -0.4852  0.2581
controlcodeSelf Help                -0.0917  0.0778  -1.1799  0.2380  -0.2442  0.0607
outcomecodeLife Satisfaction         0.0497  0.0968   0.5134  0.6077  -0.1401  0.2395
outcomecodeOther                    -0.0300  0.0753  -0.3981  0.6906  -0.1777  0.1177
outcomecodePN Affect                 0.0063  0.0794   0.0795  0.9367  -0.1493  0.1619
```



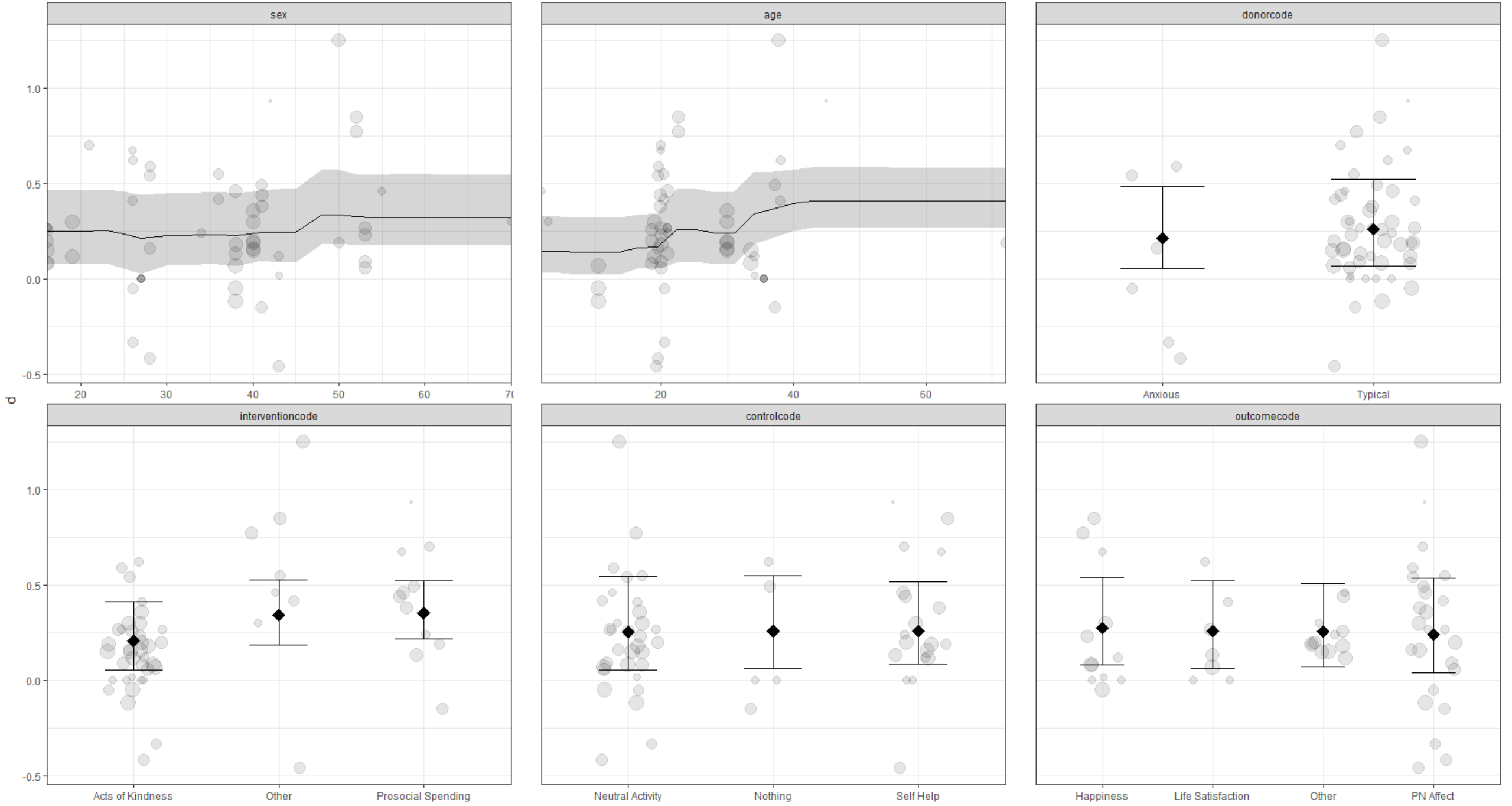

```
PartialDependence(res, rawdata = TRUE, pi = .95)
```

```
mf <- ClusterMF(d ~ ., study = "study_id", data)
```

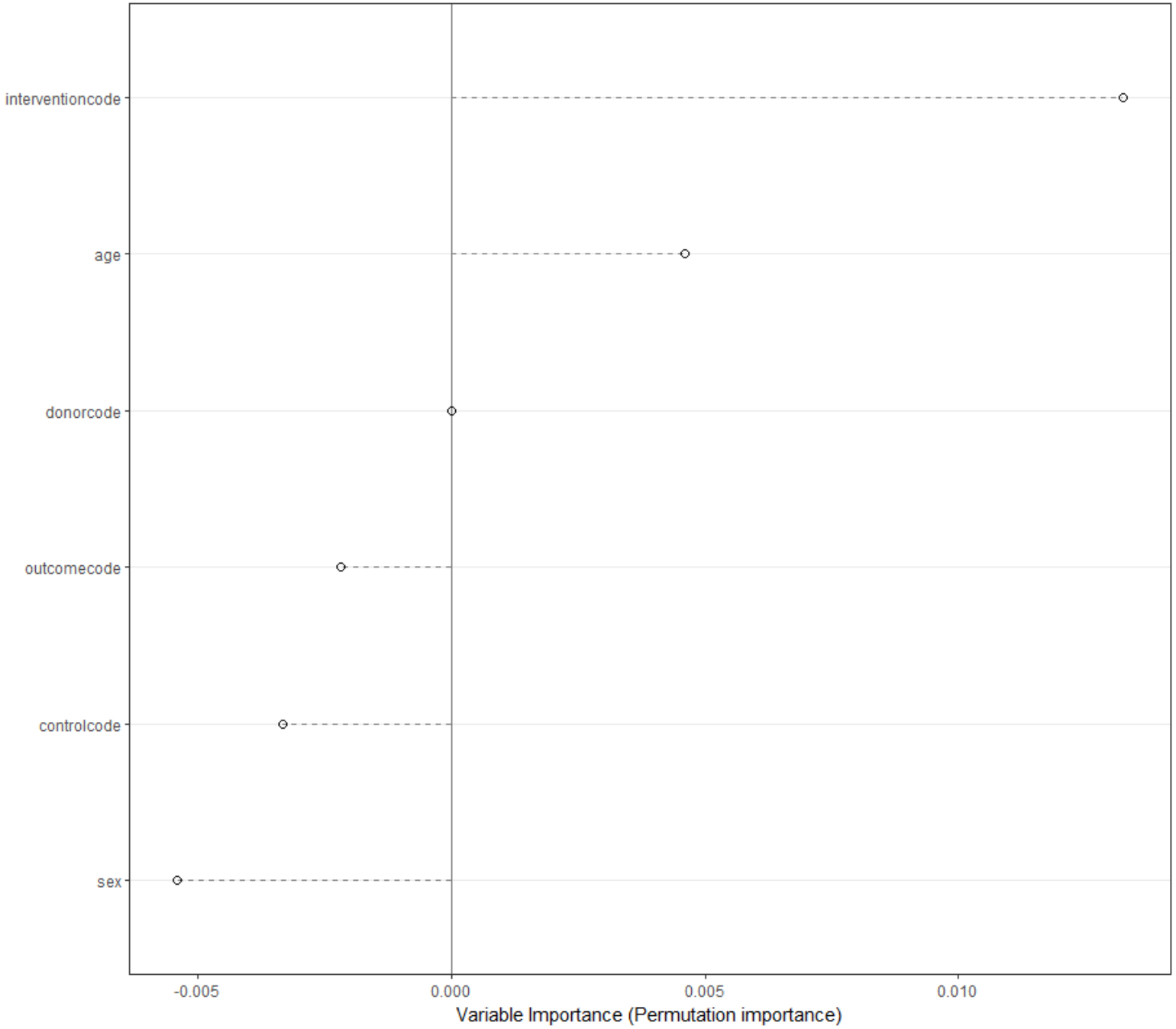
```
Call:
ClusterMF(formula = d ~ ., data = data, study = "study_id")

R squared (OOB): -0.0489
Residual heterogeneity (tau2): 0.0549
```


Convergence plot
Cumulative MSE
0.14
0.13
0.12
0.11
0.10
0.09
0
250
500
750
1000
Number of trees

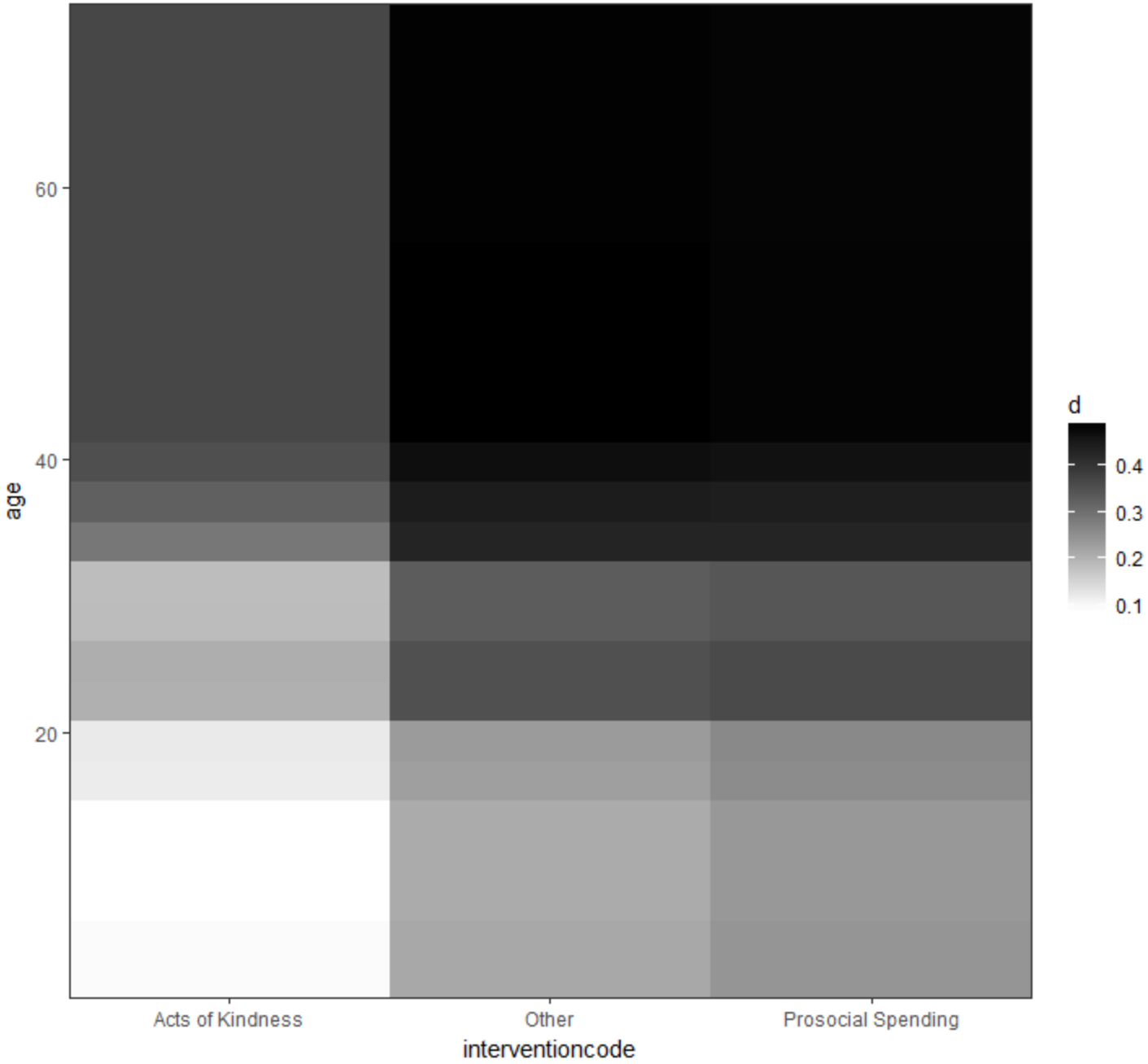
```
plot(mf)
```




```
PartialDependence(mf, rawdata = TRUE, pi = .95)
```




```
PartialDependence(mf, rawdata = TRUE, pi = .95)
```




```
PartialDependence(mf, vars = c("interventioncode", "age"), interaction = TRUE)
```

# Meta-analysis using random forests

Data | Analysis | Explore

**Effect size variable:**

di

**Variance of the effect size:**

vi

**Select moderators:**

DV_Alligned
Location
Curriculum
Train_Knowledge
Pre_Post
Blind
Journal

**Variables to consider at each split:**

2

**How many trees to grow:**

1000

**Which weights to use:**

Random-effects

☑ Data are clustered (multilevel)

**Clustering variable**

exp_id

Results | Exploratory

**MetaForest results:**

| | |
|---|---|
| Type of analysis: | ClusterMF |
| Number of studies: | Forest 1: 38, Forest 2: 40 |
| Number of moderators: | 7 |
| Number of trees in forest: | Two forests of length 1000 |
| Candidate variables per split: | 2 |
| Minimum terminal node size: | 5 |
| OOB prediction error (MSE): | 0.25 |
| R squared (OOB): | 0.25 |

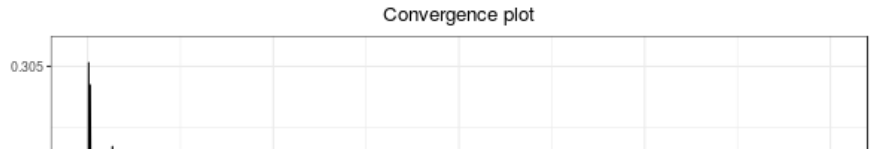
**Tests for Heterogeneity:**

| | tau2 | tau2_SE | I^2 | H^2 | Q-test | df | Q_p |
|---|---|---|---|---|---|---|---|
| Raw effect sizes: | 0.22 | 0.05 | 75.64 | 4.10 | 284.46 | 77 | 0.00 |
| Residuals (after MetaForest): | 0.13 | 0.04 | 64.89 | 2.85 | 198.13 | 77 | 0.00 |

**Random intercept meta-analyses:**

| | Intercept | se | ci.lb | ci.ub | p |
|---|---|---|---|---|---|
| Raw effect sizes: | 0.41 | 0.06 | 0.28 | 0.54 | 0.00 |
| Residuals (after MetaForest): | -0.01 | 0.05 | -0.12 | 0.09 | 0.80 |

**Convergence plot:**



# Get MetaForest

* `install.packages("metaforest")`
  `??MetaForest`
* www.developmentaldatascience.org/metaforest

* Other cool features:
  * Functions for model tuning using the `caret` package