

# Identify interaction effects between multiple moderators in R by metacart package

*Xinru Li and Elise Dusseldorp*

*2017-10-16*

This is an introduction to show how to use the r-package {metacart} for interaction detection in meta-analysis. Meta-CART analyses will be performed on two example datasets, including one simulated data set and one real-world data sets.

## The simulated data example

First, we start with a simulated data set generated from a true model with a three-way interaction between three moderators. The description of the simulated data can be found in the help file by typing “?SimData” or “help(SimData)”.

```
library(metacart)

## Loading required package: rpart
## Loading required package: rpart.plot
## Loading required package: ggplot2
## Loading required package: gridExtra

?SimData
summary(SimData)

##      trail          efk          vark          m1          m2
## Min.   : 1.00   Min.   :-0.3554   Min.   :0.006855   A:61   A:56
## 1st Qu.: 30.75  1st Qu.:-0.0682   1st Qu.:0.010165   B:59   B:64
## Median : 60.50  Median : 0.0280   Median :0.011994
## Mean   : 60.50  Mean   : 0.0876   Mean   :0.016211
## 3rd Qu.: 90.25  3rd Qu.: 0.1145   3rd Qu.:0.016162
## Max.   :120.00  Max.   : 1.1346   Max.   :0.226828
## m3      m4      m5
## A:39   A:39   A:34
## B:41   B:43   B:43
## C:40   C:38   C:43
##
##
##

set.seed(1)
```

To perform a meta-CART analysis, there are two model assumptions to choose from: the fixed-effect model and the random-effects model. To perform a RE meta-CART analysis, we need to use the function REMrt() and specify the effect size variable, the moderators, the sampling variance variable, the data set, and the pruning parameter. The argument “formula =” can be used to specify the effect size and the moderators. In this case, the effect size vector is called “efk” in the data set, and the interested moderators are “m1”, “m2”, “m3”, “m4” and “m5”. Therefore, we can use “formula = efk ~ m1 + m2 + m3 + m4 + m5” to specify the outcome variable (i.e., effect size) and the moderators. The argument “vi =” can be used to specify the sampling variance parameter, and we use “vi = vark” here because our sampling variance is called

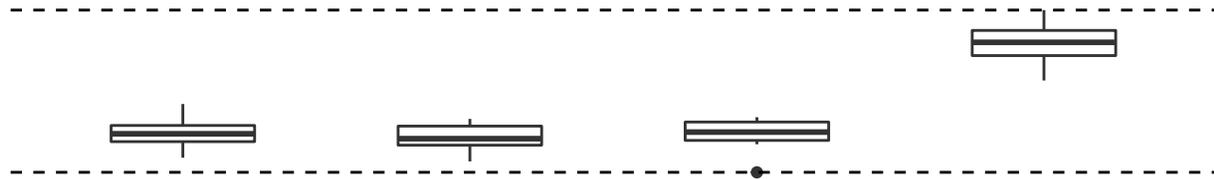
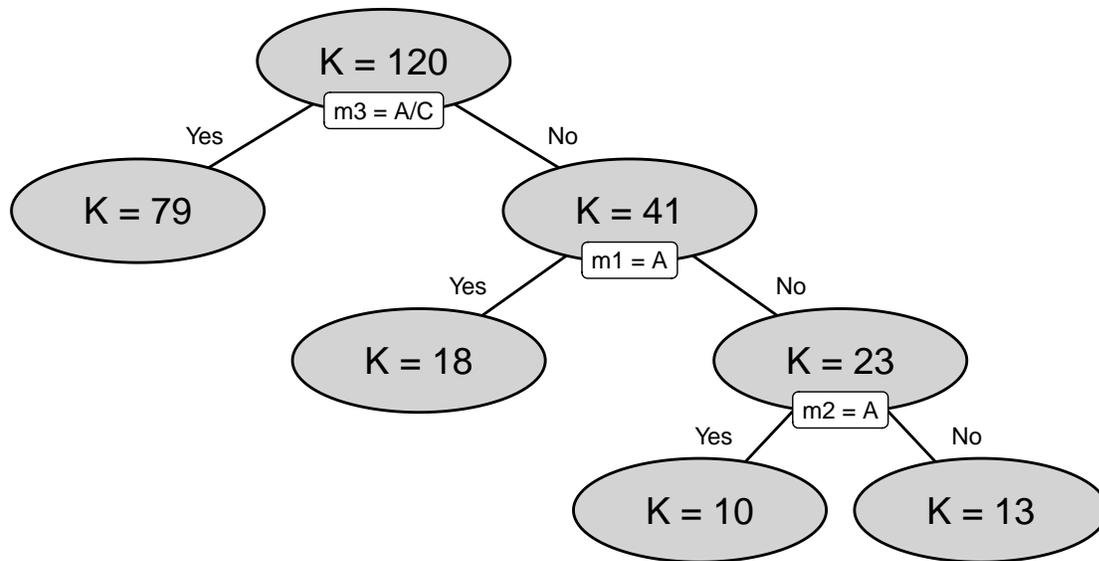
“var $k$ ”. The argument “data = SimData” is used to specify that we are going to analysis the data set called “SimData”. The pruning parameter “c =” is a tuning parameter that ranges from 0 to 1. It influences the pruning process of meta-CART. The larger the pruning parameter is, the more pruning meta-CART will have, and the more conservative the analysis will be (i.e. more control of Type I error and sacrifice some power). In Li et al. (2017), it is recommend to use  $c = 1$  for RE meta-CART if the number of studies is smaller than 120 (to control Type I error below 0.05). Thus, we use pruning rule:  $c = 0.5$  here.

```
res.simRE <- REmrt(formula = efk ~ m1 + m2 + m3 + m4 +m5, data = SimData, vi = vark, c = 0.5)
res.simRE
```

```
##
## Random Effects Meta-tree (K = 120 studies);
## REmrt(formula = efk ~ m1 + m2 + m3 + m4 + m5, data = SimData,
##       vi = vark, c = 0.5)
##
## A tree with 4 terminal nodes was detected
## The moderators are m3, m1, m2
## use summary() and plot() to see the moderator analysis results and the tree structure
##      Qb      tau2    split  mod pleaf
## 1  0.00000 0.051465622    <NA> <NA>   NA
## 2 21.13725 0.044753498 m3 = A/C  m3     1
## 3 93.60737 0.024651334 m1 = A   m1     3
## 4 415.25681 0.001997694 m2 = A   m2     5
```

The analysis results indicate that a tree with three moderators was detected. And we can use the function `plot()` to see the interaction effects between the moderators by the parsimonious tree model. The tree model indicates that the effect size is highest when the moderator “m3” is not A or C, and moderators “m1” and “m2” are not A. This is the same as our true model. By using the function `summary()`, a standard subgroup analysis is performed and the between-subgroups  $Q$ -statistic indicates that the moderator effects are significant ( $Q_b = 415.257$ ,  $p$ -value  $< 0.0001$ ). The estimates of effect size in each subgroup and the confidence intervals are also give.

```
plot(res.simRE)
```



```
summary(res.simRE)
```

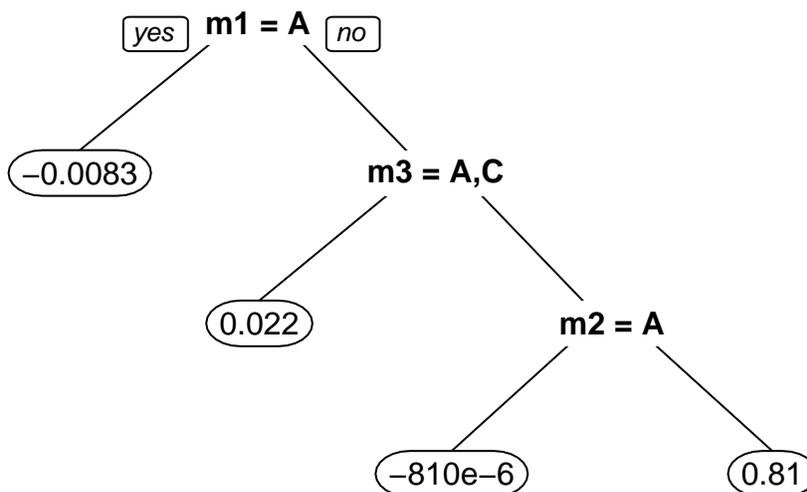
```
##
## Random Effects meta-tree (K = 120 studies);
## REmrt(formula = efk ~ m1 + m2 + m3 + m4 + m5, data = SimData,
##       vi = vark, c = 0.5)
##
## A tree with 4 terminal nodes was detected
## Moderators were detected as: m3, m1, m2
##
## Test for Between-Subgroups Heterogeneity under RE assumption:
## Qb = 415.257 (df = 3), p-value < 1e-04;
## The estimate for the residual heterogeneity tau2 = 0.002;
##
## Subgroup Meta-analysis Results:
##   K    g    se  zval  pval  ci.lb  ci.ub
## 2 79  0.011 0.013  0.797 0.426 -0.016 0.037
## 4 18 -0.032 0.027 -1.195 0.232 -0.084 0.020
## 6 10 -0.002 0.045 -0.034 0.973 -0.090 0.087
## 7 13  0.807 0.038 21.305 0.000  0.733 0.882 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

To perform a FE meta-CART analysis, we need to use the function `FEmrt()`. The arguments of `FEmrt()` is similar to `REmrt()`, with “formula =” specifying the effect size variable and the moderators, “vi =” specifying the sampling variance variable, “data =” specifying the data set to be analyzed, and “c =” specifying the pruning parameter. In Li et al. (2017), it is recommend to use  $c = 1$  for FE meta-CART if the number of studies is smaller than 80, and  $c = 0.5$  if there number of studies is equal or larger than 80. So we use  $c = 0.5$  in this case.

```
res.simFE <- FEmrt(formula = efk ~ m1 + m2 + m3 + m4 + m5, data = SimData, vi = vark, c = 0.5)
res.simFE
```

```
##
## Fixed Effects Meta-tree (K = 120 studies);
## FEmrt(formula = efk ~ m1 + m2 + m3 + m4 + m5, data = SimData,
##       vi = vark, c = 0.5)
##
## A tree with 4 terminal nodes was detected
## The moderators are m1, m3, m2
## Use summary() and plot() to inspect the moderator analysis results and the tree structure.
## n= 120
##
## node), split, n, deviance, yval
##       * denotes terminal node
##
## 1) root 120 607.865200 0.0688800100
## 2) m1=A 61 73.080390 -0.0083067430 *
## 3) m1=B 59 463.723100 0.1649009000
## 6) m3=A,C 36 32.841200 0.0224452200 *
## 7) m3=B 23 241.785000 0.4755726000
## 14) m2=A 10 8.223932 -0.0008098938 *
## 15) m2=B 13 21.080780 0.8076203000 *
```

```
plot(res.simFE)
```



```
summary(res.simFE)
```

```
##
## Fixed Effects meta-tree (K = 120 studies);
## FEmrt(formula = efk ~ m1 + m2 + m3 + m4 + m5, data = SimData,
##       vi = vark, c = 0.5)
##
## A tree with 4 terminal nodes was detected
## Moderators were detected as: m1, m3, m2
##
## Test for Between-Subgroups Heterogeneity under FE assumption:
## Qb = 472.639 (df = 3), p-value < 1e-04;
##
```

```
## Subgroup Meta-analysis Results:
##      K      Qw      g      se      zval      pval      ci.lb      ci.ub
## 2   61 73.080 -0.008 0.014 -0.606 0.545 -0.035 0.019
## 6   36 32.841  0.022 0.018  1.215 0.224 -0.014 0.059
## 14  10  8.224 -0.001 0.043 -0.019 0.985 -0.084 0.083
## 15  13 21.081  0.808 0.036 22.722 0.000  0.738 0.877 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

The same as the RE meta-CART analysis, the true model is recovered with a three-way interaction between  $m_1$ ,  $m_2$ , and  $m_3$ . Note that the between-subgroup  $Q_b$  is larger and the CIs are more narrow under the FE assumption comparing to the RE assumption. This is because that FE assumption ignores the uncertainty introduced by the residual heterogeneity.

## The real-world data example

In this example, we will analyze the health psychology data by Michie et al. (2009). The detailed of this data set can be find by type in r: `?dat.BCT2009` or `help(dat.BCT2009)`. The data set consists of 106 interventions and five moderators (motivation-enhancing BCTs). We will investigate the interaction effects between these moderators, and identify the effective combination(s) of BCTs.

```
data("dat.BCT2009")
summary(dat.BCT2009)
```

```
##              study              g              vi              T1
## ALDANA 2005 (PA) : 1   Min.      :-0.1696   Min.      :0.00129   0:69
## ANDERSON 2006   : 1   1st Qu.: 0.1291   1st Qu.:0.01533   1:37
## ARAO 2007 (PA)  : 1   Median : 0.3033   Median :0.02796
## BABAZONO 2007 (PA): 1   Mean    : 0.3308   Mean    :0.04985
## BAKER 2008      : 1   3rd Qu.: 0.4932   3rd Qu.:0.06188
## BENNETT 2008   : 1   Max.    : 1.2831   Max.    :0.45788
## (Other)        :100
## T2      T3      T4      T25
## 0:42    0:106   0:33    0:89
## 1:64          1:73    1:17
##
##
##
##
##
```

By inspecting the data, we noticed that BCT3 “Provide information about other’s approval” was not used in any interventions. Thus, this moderator will not be included in the meta-CART analysis.

To perform a RE meta-CART analysis, we use “`formula = g ~ T1 + T2 + T4 + T5`” to specify the outcome variable (i.e., effect size) and the moderators. The argument “`vi = vi`” was used to specify the sampling variance parameter. The argument “`data = dat.BCT2009`” was used to specify the data set. We start with the conservative pruning rule  $c = 1$ .

```
set.seed(2017)
REres1 <- REmrt(formula = g ~ T1 + T2 + T4 + T25, vi = vi, data = dat.BCT2009, c = 1)
```

```
## Warning in REmrt(formula = g ~ T1 + T2 + T4 + T25, vi = vi, data =
## dat.BCT2009, : no moderator effect was detected
```

We received a warning message that “no moderator effect was detected”. That means that RE meta-CART did not detect any moderator in this case. We can still get the RE meta-analysis results by using the function

summary(). The  $Q = 253.0357$  ( $df = 105$ ,  $p$ -value  $< 1e-04$ ) indicates that there is significant heterogeneity between the interventions. And the estimate for the overall effect size and the confidence interval (CI) under the RE assumption are also given.

```
summary(REres1)
```

```
##
## Random Effects meta-tree (K = 106 studies);
## REEmrt(formula = g ~ T1 + T2 + T4 + T25, data = dat.BCT2009, vi = vi,
##       c = 1)
##
## No moderator effect was detected
##
## Test for Heterogeneity
## Q = 253.0357 (df = 105), p-value < 1e-04;
## The estimate for the residual heterogeneity tau2 = 0.023;
##
## Random Effects Meta-analysis Results:
##      K      g      se  zval  pval  ci.lb  ci.ub
##    106 0.270 0.022 12.344 0.000 0.227 0.313   ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

If we want to perform a more liberal meta-CART analysis and risk higher Type I error, we can perform the analysis using the pruning parameter  $c = 0$ . This time RE meta-CART detected two moderators T1 and T4, and a tree with three terminal nodes was identified.

```
REres0 <- REEmrt(formula = g ~ T1 + T2 + T4 + T25, vi = vi, data = dat.BCT2009, c = 0)
REres0
```

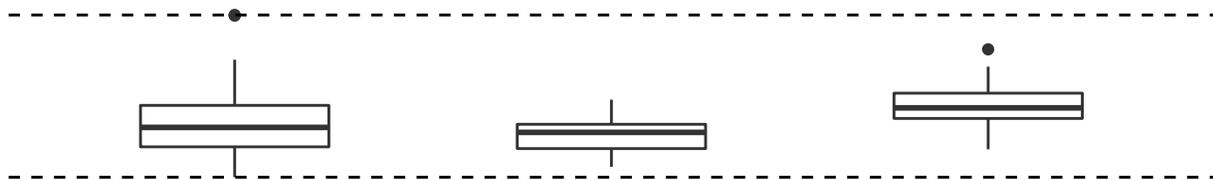
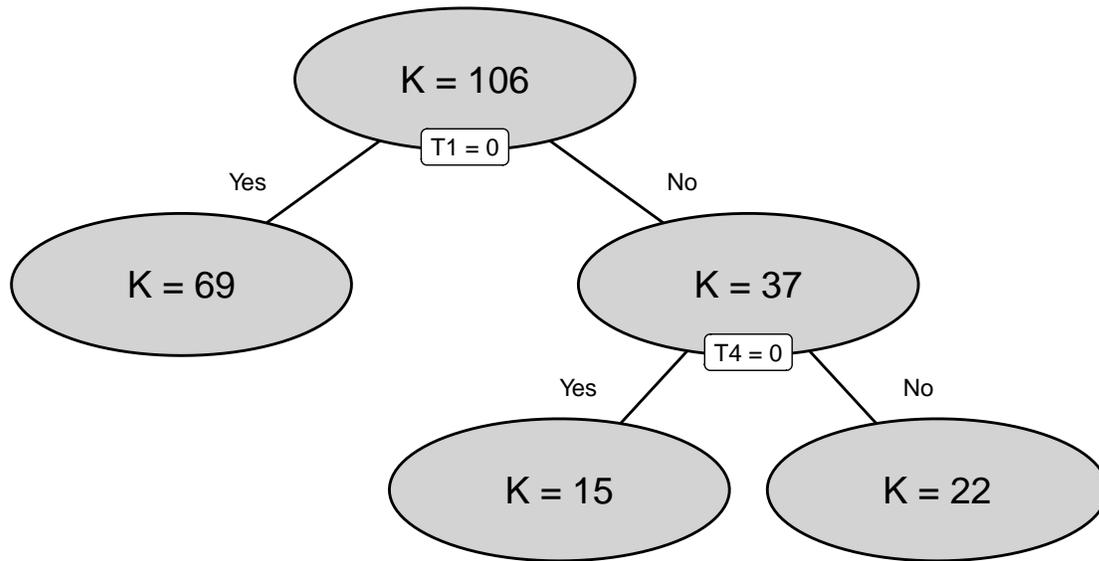
```
##
## Random Effects Meta-tree (K = 106 studies);
## REEmrt(formula = g ~ T1 + T2 + T4 + T25, data = dat.BCT2009, vi = vi,
##       c = 0)
##
## A tree with 3 terminal nodes was detected
## The moderators are T1, T4
## use summary() and plot() to see the moderator analysis results and the tree structure
##      Qb      tau2  split  mod  pleaf
## 1  0.000000 0.02278241  <NA> <NA>   NA
## 2  2.741234 0.02066011  T1 = 0  T1    1
## 3 13.207668 0.01734858  T4 = 0  T4    3
```

```
summary(REres0)
```

```
##
## Random Effects meta-tree (K = 106 studies);
## REEmrt(formula = g ~ T1 + T2 + T4 + T25, data = dat.BCT2009, vi = vi,
##       c = 0)
##
## A tree with 3 terminal nodes was detected
## Moderators were detected as: T1, T4
##
## Test for Between-Subgroups Heterogeneity under RE assumption:
## Qb = 13.208 (df = 2), p-value 0.0013552;
## The estimate for the residual heterogeneity tau2 = 0.017;
##
```

```
## Subgroup Meta-analysis Results:
##   K   g   se  zval  pval  ci.lb  ci.ub
## 2 69 0.241 0.025 9.791 0.000 0.193 0.289 ***
## 4 15 0.199 0.053 3.795 0.000 0.096 0.302 ***
## 5 22 0.429 0.050 8.553 0.000 0.331 0.528 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
plot(REres0)
```



We can obtain the subgroup analysis results by using the function `summary`. The between-subgroups  $Q_b = 13.208$  ( $df = 2$ ,  $p$ -value  $0.0013552$ ) indicated that there is significant difference in the effect sizes between the three identified subgroups, and the moderator effects are significant. The estimates for the summary effect sizes and CIs for each subgroup are given as well.

```
summary(REres1)
```

```
##
## Random Effects meta-tree (K = 106 studies);
## REEmrt(formula = g ~ T1 + T2 + T4 + T25, data = dat.BCT2009, vi = vi,
##       c = 1)
##
## No moderator effect was detected
##
## Test for Heterogeneity
## Q = 253.0357 (df = 105), p-value < 1e-04;
## The estimate for the residual heterogeneity tau2 = 0.023;
##
## Random Effects Meta-analysis Results:
##   K   g   se  zval  pval  ci.lb  ci.ub
```

```
##      106  0.270  0.022 12.344  0.000  0.227  0.313    ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Then we use `FEmrt()` to perform a FE meta-CART analysis. The number of studies is larger than 80, so we use  $c = 0.5$ . The FE meta-CART also detected an interaction effect between BCT1 and BCT4. When the two BCTs are both included, the effect size will be higher. By using the `summary()` function, we can obtain the heterogeneity test, and the estimates for effect sizes and CIs for each subgroup under the FE model assumption.

```
FERes <- FEmrt(formula = g ~ T1 + T2 + T4 + T25, vi = vi, data = dat.BCT2009, c = 0.5)
FERes
```

```
##
## Fixed Effects Meta-tree (K = 106 studies);
## FEmrt(formula = g ~ T1 + T2 + T4 + T25, data = dat.BCT2009, vi = vi,
##       c = 0.5)
##
## A tree with 3 terminal nodes was detected
## The moderators are T1, T4
## Use summary() and plot() to inspect the moderator analysis results and the tree structure.
## n= 106
##
## node), split, n, deviance, yval
##       * denotes terminal node
##
## 1) root 106 253.03570 0.2294837
##   2) T1=0 69 168.39650 0.2018800 *
##   3) T1=1 37  67.44715 0.3222563
##     6) T4=0 15  20.19503 0.1906789 *
##     7) T4=1 22  23.85478 0.4377661 *
```

```
summary(FERes)
```

```
##
## Fixed Effects meta-tree (K = 106 studies);
## FEmrt(formula = g ~ T1 + T2 + T4 + T25, data = dat.BCT2009, vi = vi,
##       c = 0.5)
##
## A tree with 3 terminal nodes was detected
## Moderators were detected as: T1, T4
##
## Test for Between-Subgroups Heterogeneity under FE assumption:
## Qb = 40.589 (df = 2), p-value < 1e-04;
##
## Subgroup Meta-analysis Results:
##   K      Qw      g      se      zval      pval      ci.lb      ci.ub
## 2 69 168.397 0.202 0.014 14.521 0.000 0.175 0.229 ***
## 6 15  20.195 0.191 0.037  5.115 0.000 0.118 0.264 ***
## 7 22  23.855 0.438 0.035 12.534 0.000 0.369 0.506 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
plot(FEres)
```

