Meta-Regression

Elena Kulinskaya

School of Computing Sciences,
University of East Anglia

November 2020

Meta-regression

In case of substantial heterogeneity between the studies, possible causes of heterogeneity should be explored. This can be done by investigating covariates, or so-called moderators of the effect of interest.

The moderators can be either at the study level, or at the subject level. The latter is possible only when the raw data are available. Meta-regression is the analysis of moderators at the study level.

Since the number of studies in meta-analysis is comparatively small, and there is a large degree of confounding (resulting from association between study characteristics), there is a great danger of overfitting. Only a few moderators can be usefully included in meta-regression.

Outline
One moderator
Contrasts in R
Several moderators
One continuous or binary moderator

Heterogeneity ... a moderator?

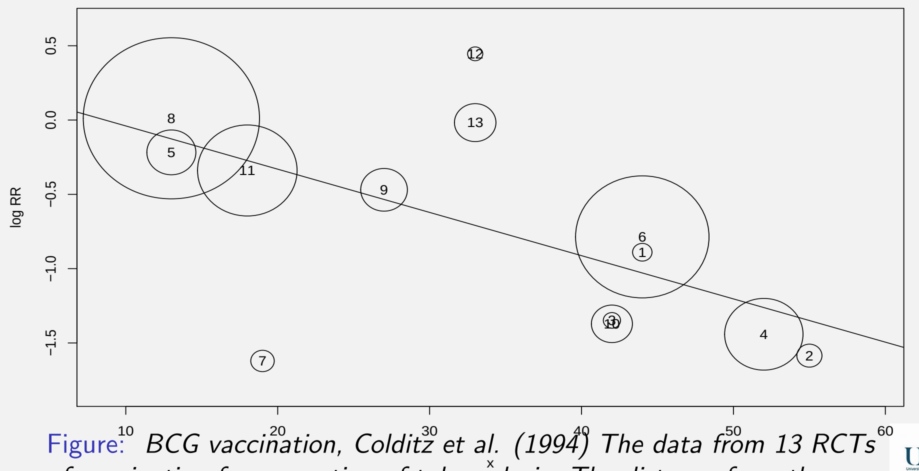

Figure: *BCG vaccination, Colditz et al. (1994) The data from 13 RCTs of vaccination for prevention of tuberculosis. The distance from the equator affects the efficacy of the vaccine! (Meta regression)*

Elena Kulinskaya Meta-regression

One continuous moderator

In the meta-regression, the study means $\theta_i = \theta + \beta x_i$.

Here x_i is a continuous covariate or an indicator for a factor with only two levels.

For continuous covariate, β is the change in effect per unit change in the covariate.

For a binary factor, let us, say, $x_i = 0$ is the 1st (baseline) level, and $x_i = 1$ is the 2nd level.

Then θ is the baseline effect, and β is the difference of effects between the two levels.

Fixed and mixed effects models

In meta-regression, in the fixed effects model, we observe, for the i-th study, normally distributed study level statistics

$$Y_i \sim N(\theta + \beta x_i, v_i).$$

In the mixed effects model we assume

$$Y_i \sim N(\theta_i + \beta x_i, v_i), \text{ and, additionally } \theta_i \sim N(\theta, \tau^2).$$

Thus $Y_i \sim N(\theta + \beta x_i, v_i + \tau^2)$.

Here β is fixed slope and θ is random intercept, hence mixed effects model.

Meta-regression in metafor

We shall use the same procedure rma as previously. But now we need to also specify the moderators of the effect of interest, i.e. possible predictors or confounding variables.

For a single moderator x just add mods=x subcommand.

Or use an R formula: mods $=\sim x$
For two moderators the first option is mods=cbind(x,y) and the formula approach is much more flexible:

mods $=\sim x$ (regression with an intercept and slope)
mods $=\sim x - 1$ (no intercept)
mods $=\sim x + y$ (additive model)
mods $=\sim x * y$ (interaction of x and y)

Example: Meta-regression using BCG vaccine data

```
data(dat.bcg)  ### load BCG vaccine data
##### plot the studies with the cirles diameter
##### inversely proportionate to SD.

symbols(dat$ablat,dat$yi,circles=dat$vi^{-1/2},
xlab="x",ylab="log RR")

#### put study numbers on the plot
text(dat$ablat,dat$yi,labels=trial)
#### meta-regression line
abline(res$b)
```

Distance to equator as a moderator

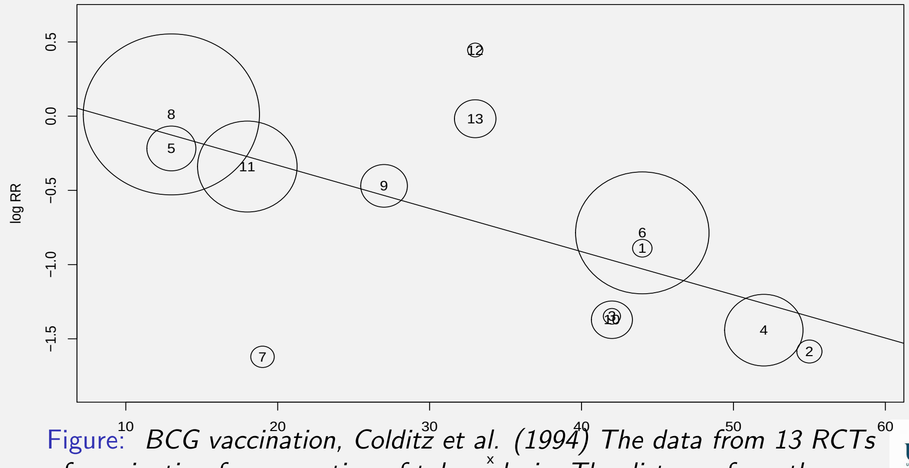

Figure: *BCG vaccination, Colditz et al. (1994) The data from 13 RCTs of vaccination for prevention of tuberculosis. The distance from the equator affects the efficacy of the vaccine! (Meta regression)*

Meta-regression for BCG data (continuous moderator)

```
### calculate log relative risks and corresponding sampling variances
dat <- escalc(measure="RR", ai=tpos, bi=tneg, ci=cpos, di=cneg,
              data=dat.bcg, append=TRUE)
res <- rma(yi, vi, mods=ablat, data=dat, method="REML") ### meta-regression
res
Mixed-Effects Model (k = 13; tau^2 estimator: REML)
tau^2 (estimated amount of residual heterogeneity): 0.0764 (SE = 0.0591)
tau (square root of estimated tau^2 value):             0.2763
I^2 (residual heterogeneity / unaccounted variability): 68.39%
H^2 (unaccounted variability / sampling variability):   3.16
R^2 (amount of heterogeneity accounted for):            75.62%

Test for Residual Heterogeneity:
QE(df = 11) = 30.7331, p-val = 0.0012
Test of Moderators (coefficient(s) 2):
QM(df = 1) = 16.3571, p-val < .0001
```

Meta-regression for BCG data (continued)

```
Model Results:

         estimate      se     zval    pval    ci.lb    ci.ub
intrcpt    0.2515  0.2491   1.0095  0.3127  -0.2368   0.7397
ablat     -0.0291  0.0072  -4.0444  <.0001  -0.0432  -0.0150  ***
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1

> predict(res, newmods=cbind(seq(from=10, to=60, by=10)), transf=exp)
    pred  ci.lb  ci.ub  cr.lb  cr.ub
1 0.9612 0.6668 1.3857 0.5000 1.8478
2 0.7185 0.5526 0.9343 0.3936 1.3117
3 0.5371 0.4355 0.6623 0.3005 0.9600
4 0.4015 0.3151 0.5115 0.2218 0.7266
5 0.3001 0.2144 0.4201 0.1586 0.5678
6 0.2243 0.1423 0.3538 0.1105 0.4552
```


Example: Biodiversity effects of agri-environmental management (categorical moderators)

Data-set 2: Batary et al. (2011) studied landscape-moderated biodiversity effects of agri-environmental management.

Experiment: Agri-environmental management (AEM)

Control: no AEM

Effect size metric: Hedges' d

Possible moderators:

Landscape (Complex (1) vs Simple (2))

Habitat (Cropland (1) vs Grassland (2))

One categorical moderator

Categorical moderators are called factors in the ANOVA, and in R.

It is possible to create dummy variables, indicators of each level, but no need: a variable with character levels is by default a factor.
Character levels of a factor are always ordered alphabetically.

```
> levels(batary$Habitat)
[1] "Cropland"  "Grassland"
```

For numeric variables a factor x needs to be specified as.factor(x).

```
> levels(batary$habitat_n)
NULL
> levels(as.factor(batary$habitat_n))
[1] "1" "2"
```


Meta-regression with Habitat as a moderator

```
> B_R1<-rma(yi=d,vi=Var_d, dat=batary, mods=~Habitat,method="REML")
> summary(B_R1)
Mixed-Effects Model (k = 109; tau^2 estimator: REML)
   logLik   Deviance        AIC        BIC
-145.7299   291.4598   297.4598   305.4783

tau^2 (estimate of residual amount of heterogeneity): 0.5722 (SE = 0.1048)
tau (sqrt of the estimate of residual heterogeneity): 0.7564
Test for Residual Heterogeneity:
QE(df = 107) = 514.7203, p-val < .0001
Test of Moderators (coefficient(s) 2):
QM(df = 1) = 0.3796, p-val = 0.5378

Model Results:
          estimate      se     zval    pval    ci.lb   ci.ub
intrcpt     0.7927  0.1222   6.4879  <.0001   0.5532  1.0321  ***
Habitat2   -0.1043  0.1693  -0.6161  0.5378  -0.4362  0.2275
```

What does this mean?

Our regression is $\theta_i = \alpha + \beta x_i$.

```
Model Results:
          estimate      se     zval    pval    ci.lb   ci.ub
intrcpt     0.7927  0.1222   6.4879  <.0001   0.5532  1.0321  ***
Habitat2   -0.1043  0.1693  -0.6161  0.5378  -0.4362  0.2275
```

The levels of a binary factor are always coded 0 and 1.
For level 1, Cropland, $\theta_1 = 0.7937$, coincides with intercept
For level 2, the difference level 2 - level 1 = -0.1043.
For Grassland, $\theta_2 = 0.7927 - 0.1043 = 0.6883$, .
There are no differences in effect due to Habitat.

Contrasts

Contrasts are linear functions used to represent categorical variables in the regression setting. Categorical variables cannot be represented by just one coefficient (slope), unless the categorical variable has just two levels, such as gender or Habitat.

When we are interested in finding how effect varies with Habitat, we have a model

$$Effect = \theta_0 + \beta_1 f(Habitat) + e.$$

Note that the variable Habitat has just one coefficient β_1 because Habitat has only 2 levels; this generalizes to $K - 1$ coefficients for a categorical variable with K levels.

Contrasts specify how the linear function $f(\cdot)$ looks like.

Treatment contrasts

The common types of contrasts are Treatment, sum, Helmert and Polynomial (orthogonal).

Remember that a categorical variable with K levels will have $K-1$ coefficients and K-1 contrast functions $f_i(\cdot),\ i=1,\cdots,K-1$.

Treatment contrasts are the default contrasts in R. They correspond to corner point constraint.

Parameters are estimated as
intercept$=\theta_1,\ \ \beta_1=\theta_2-\theta_1,\ \cdots,\ \alpha_i=\theta_i-\theta_1,\cdots$

We shall see $K-1$ tests for $\beta_i=0$ vs $\beta_i\neq 0$,$i=2,3,\cdots,K$ in the standard output. These tests compare each level to the first level.

An example with 4 levels: Dataset 6, Gilbert-Norton(2010)

Data-set 6: effectiveness of habitat corridors on animal and plant movement (35 studies). Outcomes are measures of movement of species in the habitats. Positive values of Hedges' d indicate positive effects of corridors on animal and plant movement.

```
taxa:
1  bird
2  invertebrate
3  invertebrate and nonavian vertebrate
            (just 1 study here, we shall remove it)
4  nonavian vertebrate
5  plant
```

For this data first calculate the variance of d (ex.1.6.2).

For 4 levels we will be considering an equation of the form

$$y=\theta+\beta_1 I(group=2)+\beta_2 I(group=3)+\beta_3 I(group=4)$$

Effectiveness of habitat corridors with taxa as a moderator

```
> GN1<-GN[GN$taxa_num!=3,]
> GN_R1<-rma(yi=d,vi=var, dat=GN1,mods=~as.factor(taxa_num),
>method="REML")
> summary(GN_R1)
Mixed-Effects Model (k = 76; tau^2 estimator: REML)
   logLik   Deviance        AIC        BIC
-154.3600   308.7199   318.7199   330.1032
tau^2 (estimate of residual amount of heterogeneity): 0.0000 (SE = 0.7385)
tau (sqrt of the estimate of residual heterogeneity): 0.0015
Test for Residual Heterogeneity:
QE(df = 72) = 50.9449, p-val = 0.9716
Test of Moderators (coefficient(s) 2,3,4):
QM(df = 3) = 2.8216, p-val = 0.4200
Model Results:           estimate    se     zval    pval    ci.lb   ci.ub
intrcpt                    0.4112  0.7287   0.5642  0.5726  -1.0171  1.8394
as.factor(taxa_num)2      -0.1160  0.8227  -0.1410  0.8879  -1.7285  1.4966
as.factor(taxa_num)4       0.8236  0.8634   0.9539  0.3401  -0.8686  2.5158
as.factor(taxa_num)5       0.5416  0.9158   0.5913  0.5543  -1.2534  2.3366
```


Changing the reference level

It is possible to change the reference level from the default value of 1:

```
contrasts(x)<-contr.treatment(4,base=4)
contrasts(x)
  1 2 3
1 1 0 0
2 0 1 0
3 0 0 1
4 0 0 0
```

Now all levels are compared to the last level.

Contrast sum

Another useful contrast is the contrast sum.

In this case $\beta_0 = \theta$ and $\beta_i = \theta - \theta_i$ for $i = 1, \cdots, K-1$.

To specify contrast sum for a factor, write

```
contrasts(x)<-contr.sum
```

This is a standard parametrization for ANOVA.

Analysis of biodiversity with Habitat as a moderator,4

```
 contrasts(batary$Habitat)<-contr.sum
> B_R1<-rma(yi=d,vi=Var_d, dat=batary, mods=~Habitat,method="REML")
> summary(B_R1)
Mixed-Effects Model (k = 109; tau^2 estimator: REML)
   logLik   Deviance        AIC        BIC
-146.4231   292.8461   298.8461   306.8646
tau^2 (estimate of residual amount of heterogeneity): 0.5722 (SE = 0.1048)
tau (sqrt of the estimate of residual heterogeneity): 0.7564
Test for Residual Heterogeneity:
QE(df = 107) = 514.7203, p-val < .0001

Test of Moderators (coefficient(s) 2):
QM(df = 1) = 0.3796, p-val = 0.5378
Model Results:
          estimate      se    zval    pval    ci.lb   ci.ub
intrcpt     0.7405  0.0847  8.7471  <.0001   0.5746  0.9064  ***
Habitat1    0.0522  0.0847  0.6161  0.5378  -0.1138  0.2181
```

Several moderators in meta-regression

Suppose we have a set of moderators $x_1, x_2, \ldots, x_p$ for an effect y. We have a set of effect measurements y_k for $k = 1, \ldots, K$ studies, and for each effect measure we have a set of values for the moderators $x_{1k}, x_{2k}, \ldots, x_{pk}$. Here x_{jk} is the kth observation on variable x_j. We hope to model the relationship as a linear form

$$y_k = \beta_0 + \beta_1 x_{1k} + \beta_2 x_{2k} + \ldots + \beta_p x_{pk} + \epsilon_k \tag{1}$$

If we write in vector form

$$\boldsymbol{x}_k^T = (1, x_{1k}, x_{2k}, \ldots, x_{pk})$$

and

$$\boldsymbol{\beta}^T = (\beta_0, \beta_1, \ldots, \beta_p)$$

we have

$$y_k = \boldsymbol{x}_k^T \boldsymbol{\beta} + \epsilon_k, \text{ where } \epsilon_k \sim N(0, v_k + \tau^2) \tag{2}$$

Two categorical factors

Model: $X_{kij} = \mu + \alpha_i + \beta_j + \alpha\beta_{ij} + e_k$

$e_k \sim N(0, v_i + \tau^2)$ errors

$\alpha_i,\ i = 1, \cdots, I$ effects for factor A

$\beta_j,\ j = 1, \cdots, J$ effects for factor B

$\alpha\beta_{ij},\quad ij = (1,1) \cdots (I, J)$ interactions

$k = 1, \cdots, K$ – studies,

We need to have some constraints, for example ,

$$\text{constraints}: \quad \begin{array}{l} \sum_{i=1}^{I} \alpha_i = 0,\ \sum_{j=1}^{J} \beta_j = 0 \\ \sum_i \alpha\beta_{ij} = 0,\ \sum_j \alpha\beta_{ij} = 0. \end{array}$$

are achieved by specifying contrast sum in R.

In any case we need to understand which contrasts we use to be able to interpret the results.

2-way analysis of Batary(2010) data, 1

```
> B_R3<-rma(yi=d,vi=Var_d, mods=~Landscape*Habitat,
>dat=batary,method="REML")
> summary(B_R3)
Mixed-Effects Model (k = 109; tau^2 estimator: REML)
   logLik   Deviance        AIC        BIC
-143.7095   287.4190   297.4190   310.6888
tau^2 (estimate of residual amount of heterogeneity): 0.5683 (SE = 0.1053)
tau (sqrt of the estimate of residual heterogeneity): 0.7538
Test for Residual Heterogeneity:
QE(df = 105) = 510.6594, p-val < .0001
Test of Moderators (coefficient(s) 2,3,4):
QM(df = 3) = 6.0314, p-val = 0.1101
Model Results:
                          estimate      se     zval    pval    ci.lb    ci.ub
intrcpt                     0.2658  0.2650   1.0030  0.3158  -0.2536   0.7852
LandscapeSimple             0.6682  0.2984   2.2389  0.0252   0.0832   1.2531  *
Habitat2                    0.4981  0.3048   1.6342  0.1022  -0.0993   1.0956
LandscapeSimple:Habitat2   -0.8586  0.3823  -2.2463  0.0247  -1.6078  -0.1094  *
```


Interpretation for 2-way analysis of Batary(2011) data

```
> contrasts(batary$Habitat)
          2
Cropland  0
Grassland 1
> contrasts(batary$Landscape)
        Simple
Complex      0
Simple       1
Model Results:              estimate    se     zval    pval    ci.lb    ci.ub
intrcpt                     0.2658  0.2650   1.0030  0.3158  -0.2536   0.7852
LandscapeSimple             0.6682  0.2984   2.2389  0.0252   0.0832   1.2531  *
Habitat2                    0.4981  0.3048   1.6342  0.1022  -0.0993   1.0956
LandscapeSimple:Habitat2   -0.8586  0.3823  -2.2463  0.0247  -1.6078  -0.1094  *
                                Estimated effect
Landscapecomplex, Habitat 1     0.2658
Landscapecomplex, Habitat 2     0.2658+0.4981=0.7639
Landscapesimple, Habitat 1      0.2658+0.6682=0.9340
Landscapesimple, Habitat 2      0.2658+0.4981+0.6682-0.8586=0.5735
```

2-way analysis of Batary(2011) data, 2

```
> contrasts(batary$Landscape)<-contr.sum
> contrasts(batary$Habitat)<-contr.sum
> B_R4<-rma(yi=d,vi=Var_d, mods=~Landscape*Habitat,dat=batary)
> summary(B_R4)
Mixed-Effects Model (k = 109; tau^2 estimator: REML)
   logLik   Deviance        AIC        BIC
-146.4821   292.9642   302.9642   316.2340
tau^2 (estimate of residual amount of heterogeneity): 0.5683 (SE = 0.1053)
tau (sqrt of the estimate of residual heterogeneity): 0.7538
Test for Residual Heterogeneity:
QE(df = 105) = 510.6594, p-val < .0001
Test of Moderators (coefficient(s) 2,3,4):
QM(df = 3) = 6.0314, p-val = 0.1101
Model Results:         estimate    se     zval    pval    ci.lb    ci.ub
intrcpt                 0.6343  0.0956   6.6376  <.0001   0.4470   0.8216  ***
Landscape1             -0.1194  0.0956  -1.2497  0.2114  -0.3067   0.0679
Habitat1               -0.0344  0.0956  -0.3601  0.7188  -0.2217   0.1529
Landscape1:Habitat1    -0.2147  0.0956  -2.2463  0.0247  -0.4020  -0.0274    *
```


2-way analysis of Batary(2011) data, 3

```
> contrasts(batary$Landscape)<-contr.sum
> contrasts(batary$Habitat)<-contr.sum
Model Results:         estimate    se     zval    pval    ci.lb    ci.ub
intrcpt                 0.6343  0.0956   6.6376  <.0001   0.4470   0.8216  ***
Landscape1             -0.1194  0.0956  -1.2497  0.2114  -0.3067   0.0679
Habitat1               -0.0344  0.0956  -0.3601  0.7188  -0.2217   0.1529
Landscape1:Habitat1    -0.2147  0.0956  -2.2463  0.0247  -0.4020  -0.0274    *

                                Estimated Effect
Mean combined effect            0.6343
Landscapecomplex, Habitat 1     0.6343-0.1194-0.0344-0.2147 =  0.2658
Landscapecomplex, Habitat 2     0.6343-0.1194+0.0344+0.2147 = 0.7639
Landscapesimple, Habitat 1      0.6343+0.1194-0.0344+0.2147 = 0.9340
Landscapesimple, Habitat 2      0.6343+0.1194+0.0344-0.2147 = 0.5735
```