Meta-Regression

Guido Knapp
Department of Statistics, TU Dortmund University

Gerta Rücker and Guido Schwarzer
Center for Medical Biometry and Medical Informatics
University of Freiburg

DAGStat 2016
Göttingen, March 14, 2016

Meta-Regression

- In case of substantial heterogeneity between the studies, possible causes of the heterogeneity should be explored.
- In the context of meta-analysis this can be done by either covariates on the study level that could explain the differences between the studies or by covariates on the subject level.
- However, the latter approach is only possible when individual data are available.
- Since often only information on the study level is available, explaining and investigating heterogeneity by covariates on the study level has drawn much attention in applied sciences.

Meta-Regression

- Since the number of studies in a meta-analysis is usually quite small, there is a great danger of overfitting.
- So, there is only room for a few explanatory variables in a meta-regression, whereas a lot of characteristics of the studies may be identified as potential causes of heterogeneity.
- Investigations of differences between the studies and their results are observational associations and are subject to biases (such as aggregation bias) and confounding (resulting from correlation between study characteristics).
- Consequently, there is a clear danger of misleading conclusions if P-values from multiple meta-regression analyses are interpreted naïvely.

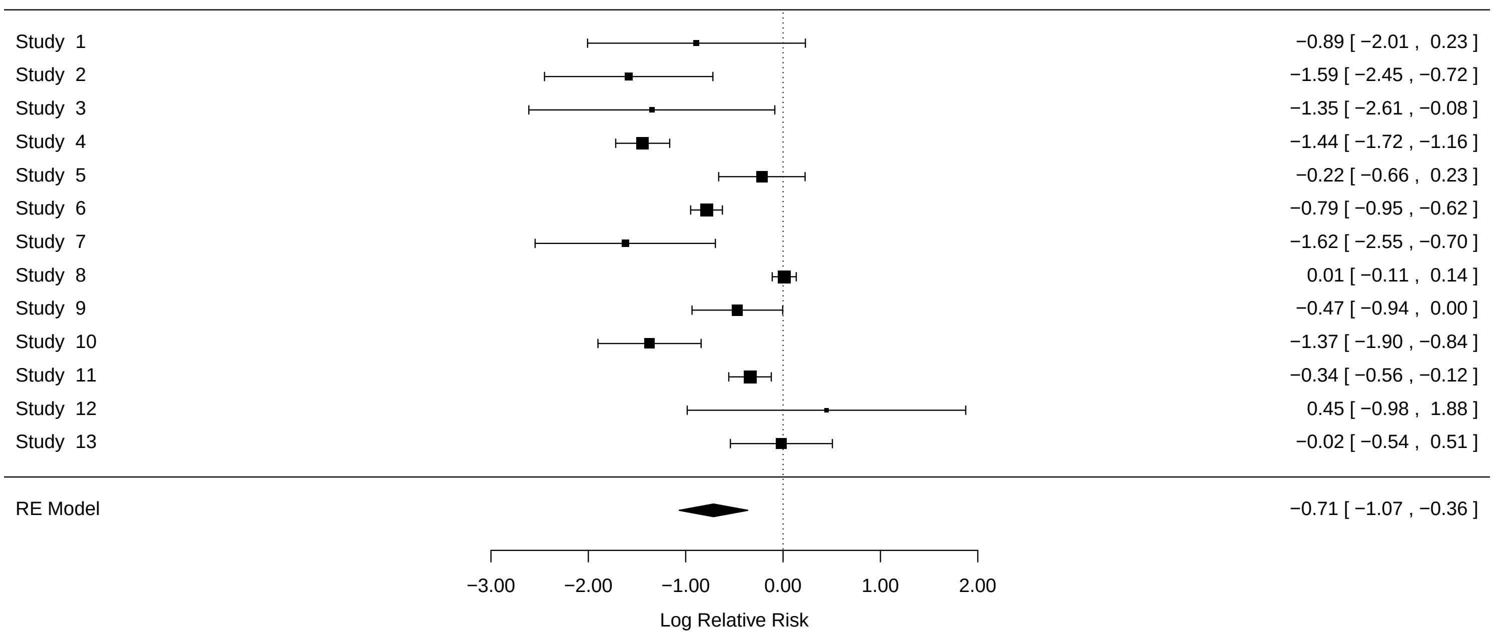
Example

Data on 13 trials on the prevention of tuberculosis using BCG vaccination

Trial	Vaccinated		Not vaccinated		Latitude
	Disease	No disease	Disease	No Disease	
1	4	119	11	128	44
2	6	300	29	274	55
3	3	228	11	209	42
4	62	13536	248	12619	52
5	33	5036	47	5761	13
6	180	1361	372	1079	44
7	8	2537	10	619	19
8	505	87886	499	87892	13
9	29	7470	45	7232	27*
10	17	1699	65	1600	42
11	186	50448	414	27197	18
12	5	2493	3	2338	33
13	27	16886	29	17825	33

Further covariates available: Year of publication, type of allocation (alternate, random, systematic)

Example



Study 1
Study 2
Study 3
Study 4
Study 5
Study 6
Study 7
Study 8
Study 9
Study 10
Study 11
Study 12
Study 13
RE Model
−0.89 [−2.01 , 0.23]
−1.59 [−2.45 , −0.72]
−1.35 [−2.61 , −0.08]
−1.44 [−1.72 , −1.16]
−0.22 [−0.66 , 0.23]
−0.79 [−0.95 , −0.62]
−1.62 [−2.55 , −0.70]
0.01 [−0.11 , 0.14]
−0.47 [−0.94 , 0.00]
−1.37 [−1.90 , −0.84]
−0.34 [−0.56 , −0.12]
0.45 [−0.98 , 1.88]
−0.02 [−0.54 , 0.51]
−0.71 [−1.07 , −0.36]
−3.00
−2.00
−1.00
0.00
1.00
2.00
Log Relative Risk

Meta-Regression (one covariate)

Let us consider k independent trials (experiments) and each trial provides an estimate, say $\hat{\theta}_i$, $i = 1, \ldots, k$, of a parameter of interest, say θ, and an estimate of the variance of $\hat{\theta}_i$, say $\hat{\sigma}_i^2$, $i = 1, \ldots, k$. Moreover, one covariate on study-level, say x_i, is known for each trial.

Normal-normal hierarchical model:

$$\hat{\theta}_i \sim N\left(\theta_i, \hat{\sigma}_i^2\right)$$

and

$$\theta_i \sim N\left(\theta_{MA}, \tau_{MA}^2\right)$$

OR

$$\theta_i \sim N\left(\theta_{MR} + \beta x_i, \tau_{MR}^2\right)$$

Meta-Regression (one covariate)

Random effects meta-analysis model

$$\hat{\theta}_i \sim N\left(\theta_{MA}, \tau^2_{MA} + \hat{\sigma}^2_i\right).$$

- θ_{MA} – mean effect size
- τ^2_{MA} – between-study variability (heterogeneity parameter)

Random effects meta-regression

$$\hat{\theta}_i \sim N\left(\theta_{MR} + \beta x_i \ , \ \tau^2_{MR} + \hat{\sigma}^2_i\right).$$

- θ_{MR} – effect size given that the covariate is zero
- τ^2_{MR} – residual heterogeneity

Meta-Regression (one covariate)

Random effects meta-regression

$$\hat{\theta}_i \sim N\left(\theta_{MR} + \beta x_i \;,\; \tau^2_{MR} + \hat{\sigma}^2_i\right).$$

Objectives:

- Fixed effects or random effects meta-regression?
  Test of $H_0 : \tau^2_{MR} = 0$!
- Estimate and confidence interval for τ^2_{MR}
- Estimates and confidence intervals for θ_{MR} and β

Extend analysis methods from meta-analysis to meta-regression:

- (conditional) restricted maximum likelihood estimation
- method of moments estimation

R Package `metafor`

- "Classical" meta-analysis and meta-regression (weighted least squares method)

- Knapp-Hartung (2003) approach for meta-analysis and meta-regression

- Q-profiling confidence interval for τ^2

- A lot of estimators for the heterogeneity parameter τ^2

- ...

Note: The $metareg$ function in R package `meta` can also be used (wrapper function that calls $rma.uni$ function from R package `metafor`).

Inference on the Fixed Effects

- Let $\tilde{\theta}$ and $\tilde{\beta}$ be the weighted least-squares estimators with known variances.

- Knapp and Hartung (2003) considered the quadratic form

$$Q_2 = \frac{1}{k-2} \sum_{i=1}^{k} w_i \, (Y_i - \tilde{\theta} - \tilde{\beta}\, x_i)^2 \, , \; k > 2 \, .$$

  that is, a mean sum of the weighted least-squares residuals.

- Under normality of Y_i, the quadratic form Q_2 is stochastically independent of the $\tilde{\theta}$ and $\tilde{\beta}$, and $(k-2)\ Q_2$ is χ^2-distributed with $k-2$ degrees of freedom.

Inference on the Fixed Effects

- Hence, unbiased and non-negative estimators of the variances of $\tilde{\theta}$ and $\tilde{\beta}$ are given by

$$Q_2(\tilde{\theta}) = \frac{1}{k-2} \sum_{i=1}^{k} g_i \, (Y_i - \tilde{\theta} - \tilde{\beta}\, x_i)^2$$

with $\quad g_i = w_i \,/\, [\sum w_j - (\sum w_j \, x_j)^2 / \sum w_j \, x_j^2], \; i = 1, \ldots, k$, and

$$Q_2(\tilde{\beta}) = \frac{1}{k-2} \sum_{i=1}^{k} h_i \, (Y_i - \tilde{\theta} - \tilde{\beta}\, x_i)^2$$

with $\quad h_i = w_i \,/\, [\sum w_j \, x_j^2 - (\sum w_j \, x_j)^2 / \sum w_j], \; i = 1, \ldots, k$.

Inference on the Fixed Effects

- Replacing the unknown variance components in $Q_2(\tilde{\theta})$ and $Q_2(\tilde{\beta})$ by appropriate estimates, Knapp and Hartung (2003) proposed the following approximate $(1-\alpha)$-confidence intervals on θ and β:

$$\hat{\theta} \pm \sqrt{\hat{Q}_2(\hat{\theta})}\, t_{k-2;\alpha/2}$$

and

$$\hat{\beta} \pm \sqrt{\hat{Q}_2(\hat{\beta})}\, t_{k-2;\alpha/2}\ .$$

Example

Results for slope with covariate latitude

Method $\hat{\tau}^2$	Estimate	95% CI (classical)	95% CI (KH)
Hunter-Schmidt	-0.0296	[-0.0398, -0.0193]	[-0.0447, -0.0144]
Hedges	-0.0282	[-0.0489, -0.0075]	[-0.0493, -0.0071]
DerSimonian-Laird	-0.0292	[-0.0424, -0.0160]	[-0.0467, -0.0118]
Sidik-Jonkman	-0.0281	[-0.0497, -0.0065]	[-0.0495, -0.0067]
ML	-0.0295	[-0.0403, -0.0188]	[-0.0452, -0.0139]
REML	-0.0291	[-0.0432, -0.0150]	[-0.0472, -0.0111]
Paule-Mandel	-0.0286	[-0.0463, -0.0108]	[-0.0485, -0.0086]

Explaining Heterogeneity

Method $\hat{\tau}^2$	MA	MR	Reduction (in %)
Hunter-Schmidt	0.2284	0.0291	87.26
Hedges	0.3285	0.2090	36.38
DerSimonian-Laird	0.3087	0.0633	79.50
Sidik-Jonkman	0.3455	0.2318	32.90
ML	0.2800	0.0344	87.73
REML	0.3132	0.0764	75.62
Paule-Mandel	0.3180	0.1421	55.31

Categorical Covariate

```
### load package
load(metafor)
### load BCG vaccine data
data(dat.bcg)
### calculate log relative risks and corresponding sampling variances
dat <- escalc(measure="RR", ai=tpos, bi=tneg, ci=cpos,
                            di=cneg, data=dat.bcg)
### using a model formula to specify the same model
rma(yi, vi, mods = ~ factor(alloc), data=dat, method="REML", btt=c(2,3))
```

Categorical Covariate

```
Mixed-Effects Model (k = 13; tau^2 estimator: REML)

tau^2 (estimated amount of residual heterogeneity):     0.3615
tau (square root of estimated tau^2 value):             0.6013
I^2 (residual heterogeneity / unaccounted variability): 88.77%
H^2 (unaccounted variability / sampling variability):   8.91
R^2 (amount of heterogeneity accounted for):            0.00%

Test for Residual Heterogeneity:
QE(df = 10) = 132.3676, p-val < .0001

Test of Moderators (coefficient(s) 2,3):
QM(df = 2) = 1.7675, p-val = 0.4132
```

Categorical Covariate

```
Model Results:

            estimate      se     zval    pval    ci.lb   ci.ub
intrcpt      -0.5180  0.4412  -1.1740  0.2404  -1.3827  0.3468
random       -0.4478  0.5158  -0.8682  0.3853  -1.4588  0.5632
systematic    0.0890  0.5600   0.1590  0.8737  -1.0086  1.1867

---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1
Results for type of allocation
```

Further Applications

- Methods can be easily extended to more than one covariate.
- Meta-regression is primarily used for explaining heterogeneity between study results.
- Meta-regression technique can be also used for other applications of combining results; e.g. combining results from controlled and uncontrolled studies in meta-regression model where the covariate indicates whether the result comes from a controlled or from an uncontrolled study