

1. Research question

Suppose we have **12 studies** investigating the effect of an intervention on a continuous outcome.

We want to know: **Does the intervention effect change according to participants' mean age, and is the age effect different for randomized vs. non-randomized studies?**

Let

- y_i = observed effect size for study i
- v_i = sampling variance of y_i
- $SE_i = \sqrt{v_i}$
- x_i = mean age of participants
- Z_i = study design indicator

$$Z_i = \begin{cases} 0, & \text{Non-randomized} \\ 1, & \text{Randomized} \end{cases}$$

2. Meta-regression model

A useful model is

$$y_i = \beta_0 + \beta_1 x_i + \beta_2 Z_i + \beta_3 (x_i Z_i) + u_i + \epsilon_i$$

where

$$u_i \sim N(0, \tau^2)$$

and

$$\epsilon_i \sim N(0, v_i).$$

Therefore,

$$y_i \sim N(\beta_0 + \beta_1 x_i + \beta_2 Z_i + \beta_3 x_i Z_i, v_i + \tau^2).$$

This is a **mixed-effects meta-regression**.

3. Mathematical interpretation

The model is

$$y_i = \beta_0 + \beta_1 \text{Age}_i + \beta_2 \text{Design}_i + \beta_3 (\text{Age}_i \times \text{Design}_i) + u_i + \epsilon_i.$$

For non-randomized studies

Since $Z_i = 0$,

$$E(y_i | Z_i = 0) = \beta_0 + \beta_1 \text{Age}_i.$$

So the age slope is

$$\beta_1$$

For randomized studies

Since $Z_i = 1$,

$$E(y_i | Z_i = 1) = \beta_0 + \beta_2 + (\beta_1 + \beta_3)Age_i.$$

Therefore, the age slope for randomized studies is

$$\beta_1 + \beta_3.$$

This is the important part.

The interaction coefficient

$$\beta_3$$

tells us whether the relationship between age and effect size differs between randomized and non-randomized studies.

4. Hypothetical data

Suppose we obtain the following study-level results.

Study	Effect y_i	SE	Age	Design
S1	0.42	0.12	25	Non-RCT
S2	0.31	0.10	28	RCT
S3	0.35	0.11	30	Non-RCT
S4	0.18	0.09	32	RCT
S5	0.27	0.10	35	Non-RCT
S6	0.14	0.08	38	RCT
S7	0.22	0.09	40	Non-RCT
S8	0.05	0.08	43	RCT
S9	0.15	0.09	45	Non-RCT
S10	-0.02	0.07	48	RCT
S11	0.08	0.08	50	Non-RCT
S12	-0.10	0.07	55	RCT

For each study,

$$v_i = SE_i^2.$$

For example, for S1,

$$v_1 = (0.12)^2 = 0.0144.$$

5. Center the continuous moderator

For interpretation, it is better to center age around its mean.

Let

$$Age_i^* = Age_i - \overline{Age}.$$

This changes the intercept interpretation.

Now the model becomes

$$y_i = \beta_0 + \beta_1 Age_i^* + \beta_2 Design_i + \beta_3 (Age_i^* \times Design_i) + u_i + \epsilon_i.$$

Now:

$$\beta_0$$

represents the estimated effect for the **reference group at average age**.

This is much easier to interpret.

6. R analysis using metafor

```
install.packages("metafor")
```

```
library(metafor)
```

Create the dataset:

```
dat <- data.frame(  
  study = paste0("S", 1:12),  
  yi = c(  
    0.42, 0.31, 0.35, 0.18,  
    0.27, 0.14, 0.22, 0.05,  
    0.15, -0.02, 0.08, -0.10  
  ),
```

```
  sei = c(  
    0.12, 0.10, 0.11, 0.09,  
    0.10, 0.08, 0.09, 0.08,  
    0.09, 0.07, 0.08, 0.07  
  ),
```

```
  age = c(  
    25, 28, 30, 32,
```

35, 38, 40, 43,

45, 48, 50, 55

),

```
design = c(
```

```
"Non-RCT", "RCT", "Non-RCT", "RCT",
```

```
"Non-RCT", "RCT", "Non-RCT", "RCT",
```

```
"Non-RCT", "RCT", "Non-RCT", "RCT"
```

```
)
```

```
)
```

```
dat$vi <- dat$sei^2
```

7. Center age

```
dat$age_c <- dat$age - mean(dat$age)
```

```
dat$design <- factor(
```

```
  dat$design,
```

```
  levels = c("Non-RCT", "RCT")
```

```
)
```

Check:

```
mean(dat$age)
```

The mean age is approximately

39.92.

Therefore,

$$Age_c = Age - 39.92.$$

8. Fit the meta-regression

The advanced model is:

```
model <- rma(
```

```
  yi = yi,
```

```
  vi = vi,
```

```
  mods = ~ age_c * design,
```

```
  data = dat,
```

```
  method = "REML"
```

```
)
```

summary(model)

The formula

So mathematically:

$$y_i = \beta_0 + \beta_1 Age_c + \beta_2 RCT + \beta_3 (Age_c \times RCT) + u_i + \epsilon_i.$$

9. What the coefficients mean

Suppose your fitted model produces approximately:

Parameter	Interpretation
β_0	Effect for Non-RCT at mean age
β_1	Age slope among Non-RCT studies
β_2	RCT vs Non-RCT difference at mean age
β_3	Difference in age slopes between RCT and Non-RCT

For example, imagine the results were:

$$\hat{\beta}_0 = 0.22$$

$$\hat{\beta}_1 = -0.012$$

$$\hat{\beta}_2 = -0.06$$

$$\hat{\beta}_3 = -0.008.$$

Then for **Non-RCT**:

$$\hat{y} = 0.22 - 0.012 Age_c.$$

Thus, effect decreases with age among Non-RCTs

For **RCT**:

$$\hat{y} = 0.22 - 0.06 + (-0.012 - 0.008) Age_c$$

so

$$\hat{y} = 0.16 - 0.020 Age_c$$

Thus, the effect decreases more rapidly with age among RCTs.

10. Test of the interaction

The important hypothesis is

$$H_0: \beta_3 = 0$$

against

$$H_A: \beta_3 \neq 0.$$

In R:

```
summary(model)
```

Look at:

```
age_c:designRCT
```

If

$$p < 0.05,$$

there is evidence that the age-effect relationship differs according to study design.

11. Test the overall moderator

You can also test whether the moderators jointly explain heterogeneity.

```
anova(model)
```

The omnibus test evaluates whether the moderator coefficients collectively contribute to explaining differences in effect sizes.

12. Heterogeneity

The model estimates:

$$\tau^2$$

which represents residual between-study heterogeneity after accounting for the moderators.

You can extract it using:

```
model$tau2
```

You can also examine:

```
model$QE
```

```
model$QEp
```

where Q_E is the residual heterogeneity statistic.

13. Calculate R^2

A useful meta-regression quantity is the proportion of between-study heterogeneity explained by the moderators:

$$R^2 = \frac{\tau_{\text{unconditional}}^2 - \tau_{\text{model}}^2}{\tau_{\text{unconditional}}^2} \times 100\%.$$

First fit the intercept-only model:

```
model0 <- rma(  
  yi = yi,  
  vi = vi,  
  data = dat,  
  method = "REML"  
)
```

Then:

```
tau20 <- model0$tau2  
tau2m <- model$tau2
```

```
R2 <- 100 * (tau20 - tau2m) / tau20
```

R2

For example, if

$$\tau_0^2 = 0.020$$

and

$$\tau_m^2 = 0.008,$$

then

$$R^2 = \frac{0.020 - 0.008}{0.020} \times 100 = 60\%.$$

So, the moderators explain approximately **60% of the between-study heterogeneity**.

14. Prediction equation

For a new study with moderator vector

$$x_0 = (1, A, g, e_c \text{DesignAge}_c \times \text{Design}),$$

the predicted effect is

$$\hat{y}_0 = x_0^T \hat{\beta}.$$

The approximate prediction interval is based on

$$\hat{y}_0 \pm z_{0.975} \sqrt{x_0^T \text{Var}(\hat{\beta}) x_0 + \tau^2}.$$

This is important because a **prediction interval** incorporates residual between-study heterogeneity, whereas the confidence interval for the mean regression line does not.

15. Meta-regression plot

For an advanced plot, we can show:

- individual study effect sizes
- study precision
- separate regression lines for RCT and Non-RCT
- 95% confidence bands.

First create predictions.

```
newdat <- expand.grid(  
  age_c = seq(  
    min(dat$age_c),  
    max(dat$age_c),  
    length.out = 100  
  ),  
  design = levels(dat$design)  
)
```

```
pred <- predict(  
  model,  
  newmods = model.matrix(  
    ~ age_c * design,  
    newdat  
  )[, -1]  
)
```

```
newdat$pred <- pred$pred  
newdat$ci.lb <- pred$ci.lb  
newdat$ci.ub <- pred$ci.ub
```

Now plot using ggplot2:

```
install.packages("ggplot2")  
library(ggplot2)
```

```
ggplot(dat,  
  aes(x = age_c,  
    y = yi,  
    size = 1/sei,  
    shape = design)) +
```

```
geom_point(alpha = 0.8) +
```

```
geom_ribbon(  
  data = newdat,  
  aes(  
    x = age_c,  
    ymin = ci.lb,  
    ymax = ci.ub,  
    fill = design  
  ),  
  alpha = 0.15,  
  inherit.aes = FALSE  
) +
```

```
geom_line(  
  data = newdat,  
  aes(  
    x = age_c,  
    y = pred,  
    linetype = design  
  ),  
  linewidth = 1,  
  inherit.aes = FALSE  
) +
```

```
labs(  
  x = "Centered mean age",  
  y = "Effect size",  
  size = "Precision",  
  shape = "Study design",
```

```
linetype = "Study design",  
fill = "Study design",  
title = "Advanced Meta-Regression"  
) +
```

```
theme_minimal()
```

What the plot tells you

Illustrative meta-regression: effect size vs centered age

Illustrative study-level relationship; the fitted interaction model allows different age slopes by study design.

-0.200.20.40.6-16-80816

The important feature is that the two study-design groups can have **different slopes**.

*****Thanks to all*****