MA shapbias analysis

Samah

2023-03-14

Contents

Preliminaries and data reading

```
# additional libraries
library("knitr")
library("janitor")
library("broom.mixed")
library("lme4")
library("emmeans")
library("tidyverse")
library("kableExtra")
```

```
## Warning in !is.null(rmarkdown::metadata$output) && rmarkdown::metadata$output
## %in% : 'length(x) = 2 > 1' in coercion to 'logical(1)'
```

```
library("modelr")
library("broom")
library("nlme")
library(wesanderson)
library("meta")
library("metafor")
# library("dmetar")
library(jtools) # Load jtools
theme_set(theme_bw())
```

```
# reading the data file
pilot1_data = read_csv("252.csv")
df_shape= filter(pilot1_data, !is.na(d))
# pilot1_data = pilot1_data %>%
#   select(ID, Title, d, d_var, Author)

# df_shape_summary = df_shape %>%
#   group_by(ID, Title, Author) %>%
#   summarize(mean = mean(d),
#             mean_se = mean(d_var))


df_shape_summary <- df_shape %>%
  group_by(language) %>%
  summarize( count = n())

df_shape$englishgrp <- fct_relevel(as.factor(df_shape$language %in%
                                               c("english")),
                                   "TRUE")

df_shape$mean_age_months_centered36 <- df_shape$mean_age_months - 36
df_shape$log_mean_age_months <- log(df_shape$mean_age_months)

df_shape$indoeuropean <- fct_relevel(as.factor(df_shape$language %in%
                                               c("english","spanish", "german")),
                                   "TRUE")

df_shape_indo <- df_shape %>%
  filter(indoeuropean == TRUE)

df_shape_nonendo <- df_shape %>%
  filter(indoeuropean == FALSE)
```

Initial exploration

First, Visualizing the data to have an initial idea of how it looks.

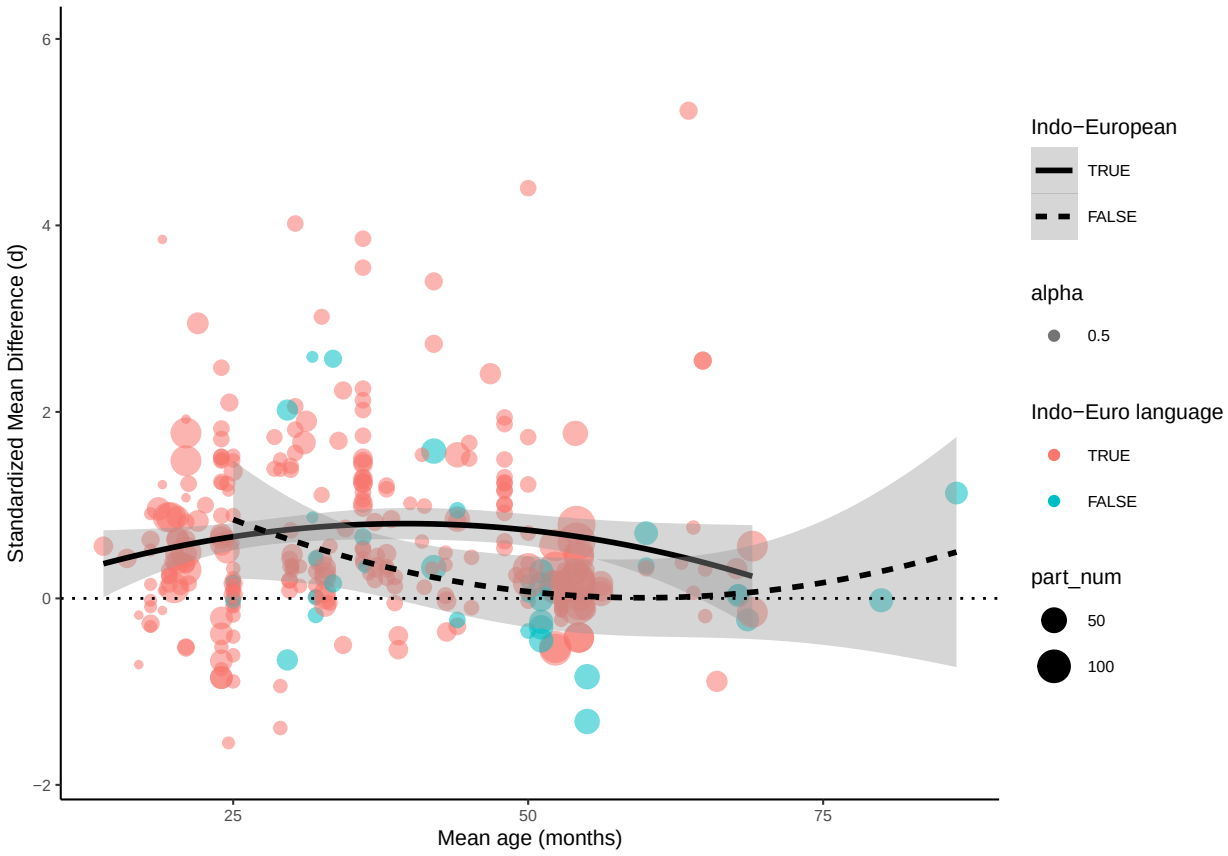
First dividing the language group into two groups: the first one is the indo-european group which includes the English and the Spanish languages. The second group includes the rest of the languages: Japanese, Chinese, Tsimane.

```r
# creating a plot that shows the effects sizes colored per language group as well as the polynomial reg


ggplot(df_shape,
       aes(x = mean_age_months, y = d, color = indoeuropean)) + geom_point(aes(ymin = d - d_var, ymax =
                                                                               alpha = .5, size = part_num)
  geom_smooth(aes(group = indoeuropean,
                  lty  = indoeuropean),
              col = "black",
              method = "lm", se = TRUE,
              formula = y ~ poly(x,2)) +
  geom_hline(yintercept = 0, lty = 3) +
  ylab("Standardized Mean Difference (d)") +
  xlab("Mean age (months)") +
  scale_color_discrete(name = "Indo-Euro language") +
  scale_linetype_discrete(name = "Indo-European") +
  theme(legend.position = "bottom") +
  theme_classic(base_size = 8)
```

```
## Warning in geom_point(aes(ymin = d - d_var, ymax = d + d_var, alpha = 0.5, :
## Ignoring unknown aesthetics: ymin and ymax
```

```
## Warning: Removed 3 rows containing non-finite values (`stat_smooth()`).
```

```
## Warning: Removed 3 rows containing missing values (`geom_point()`).
```

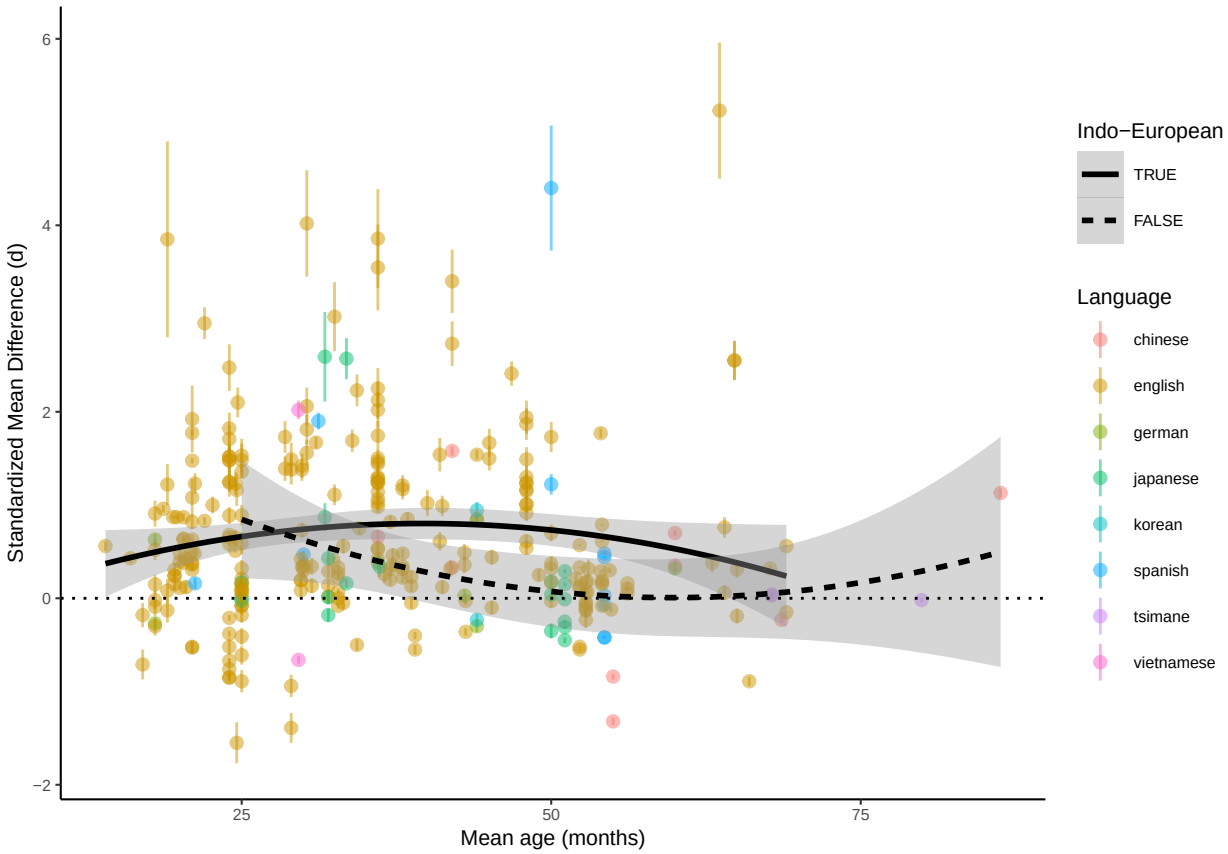


```
ggplot(df_shape,
       aes(x = mean_age_months, y = d, color = language))+
  geom_pointrange(aes(ymin = d - d_var, ymax = d + d_var),
                  alpha = .5, size = 0.3) +
  geom_smooth(aes(group = indoeuropean,
                  lty  = indoeuropean),
              col = "black",
              method = "lm", se = TRUE,
              formula = y ~ poly(x,2)) +
  geom_hline(yintercept = 0, lty = 3) +
  ylab("Standardized Mean Difference (d)") +
  xlab("Mean age (months)") +
  scale_color_discrete(name = "Language") +
  scale_linetype_discrete(name = "Indo-European") +
  theme(legend.position = "bottom") +
  theme_classic(base_size = 8)
```

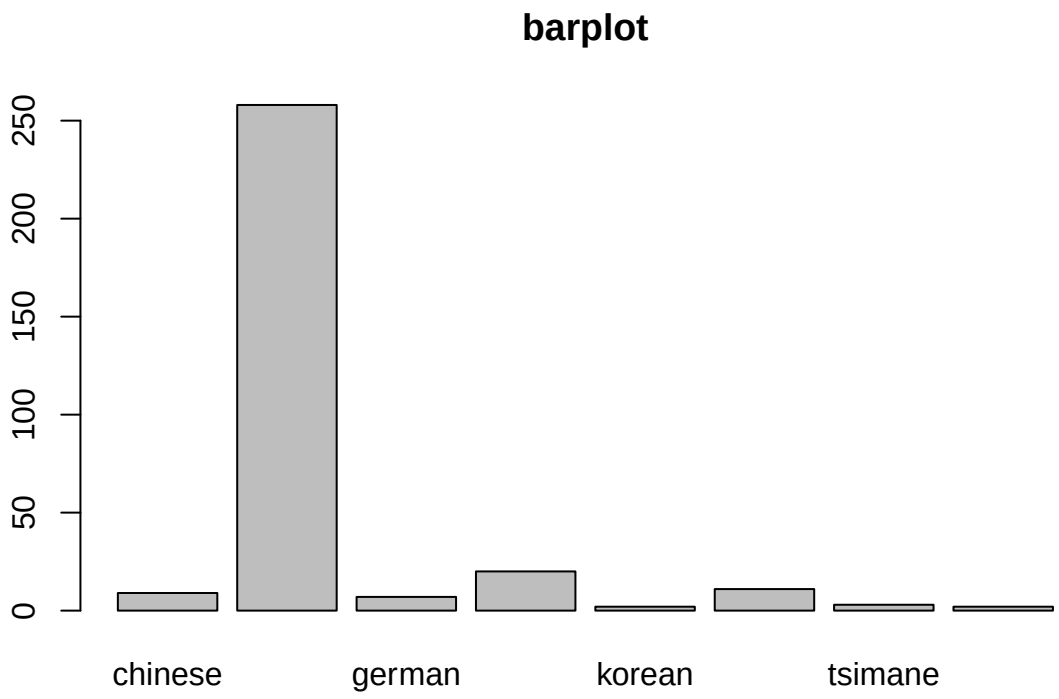
```
## Warning: Removed 3 rows containing non-finite values (`stat_smooth()`).
```

```
## Warning: Removed 3 rows containing missing values (`geom_pointrange()`).
```

```
## Warning: Removed 1 rows containing missing values (`geom_segment()`).
```



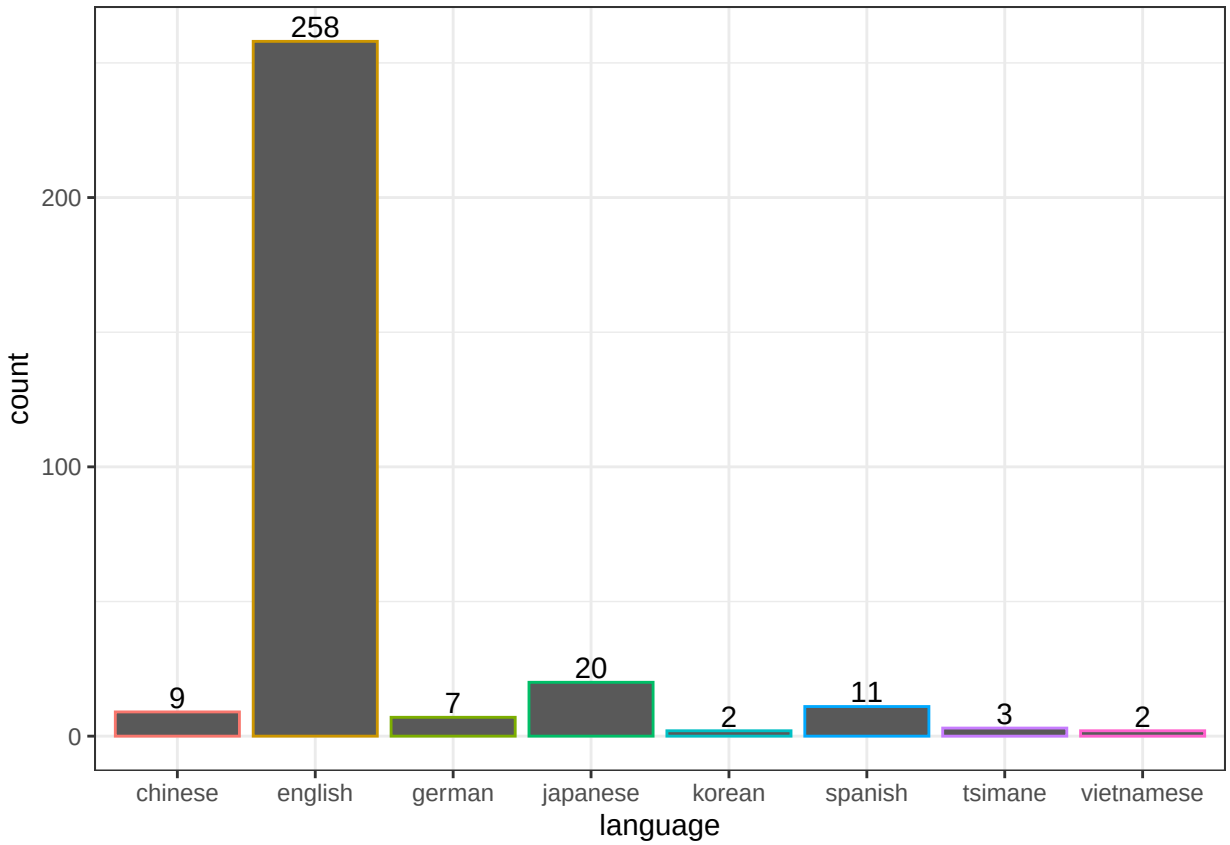
```r
# ggsave("first graph.png", width = 7, height = 4)

barplot(table(df_shape$language), main = "barplot")
```




```r
length(which(df_shape$language=="english"))
```

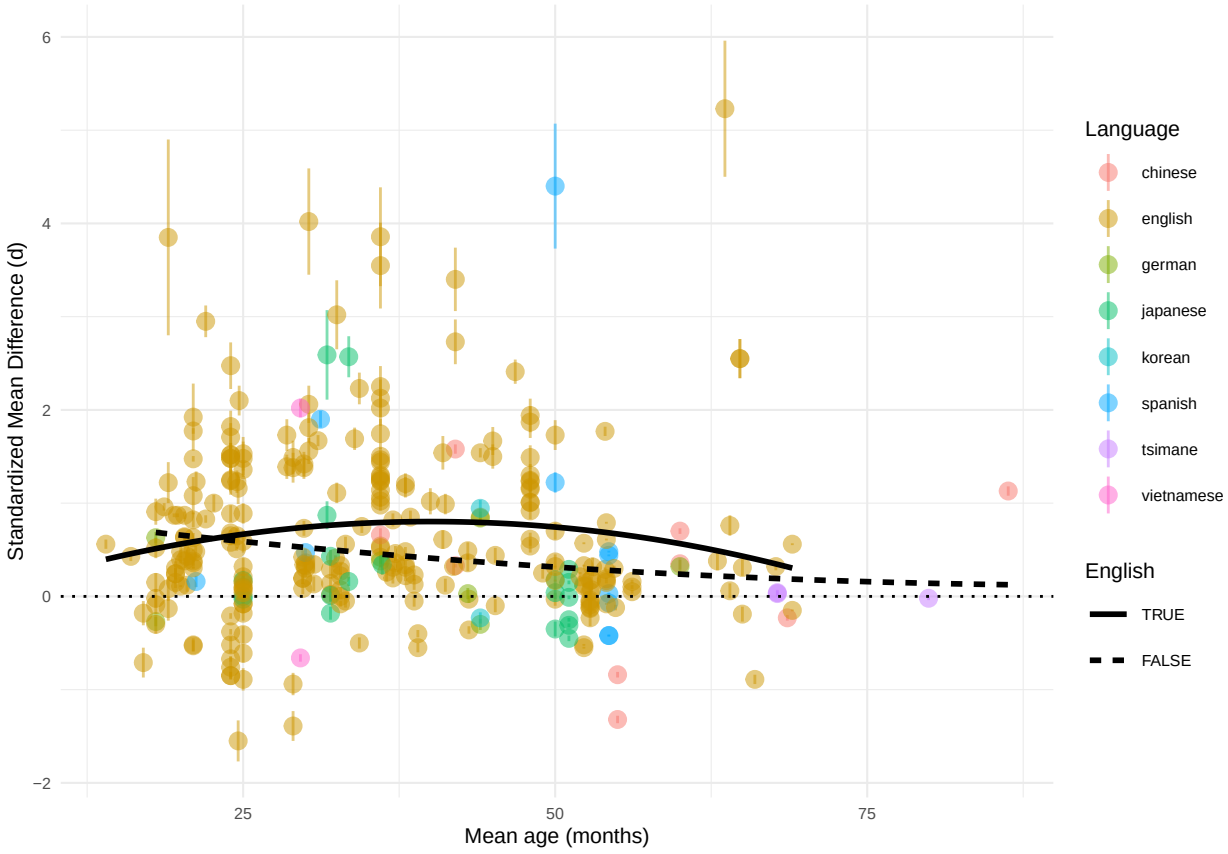
```
## [1] 258
```

```r
ggplot(df_shape_summary, aes(x = language, y = count)) +
  geom_col(aes(color = language), ) +
  theme(legend.position = "none")  +
  geom_text(aes(label = count), vjust = -0.2)
```




```
# creating a plot that shows the effects sizes colored per language group as well as the polynomial reg

ggplot(df_shape,
       aes(x = mean_age_months, y = d, color = language))+
  geom_pointrange(aes(ymin = d - d_var, ymax = d + d_var),
                  alpha = .5) +
  geom_smooth(aes(group = englishgrp,
                  lty  = englishgrp),
              col = "black",
              method = "lm", se = FALSE,
              formula = y ~ poly(x,2)) +
  geom_hline(yintercept = 0, lty = 3) +
  ylab("Standardized Mean Difference (d)") +
  xlab("Mean age (months)") +
  scale_color_discrete(name = "Language") +
  scale_linetype_discrete(name = "English") +
  theme(legend.position = "bottom") +
  theme_minimal(base_size = 8)
```

```
## Warning: Removed 3 rows containing non-finite values (`stat_smooth()`).

## Warning: Removed 3 rows containing missing values (`geom_pointrange()`).

## Warning: Removed 1 rows containing missing values (`geom_segment()`).
```



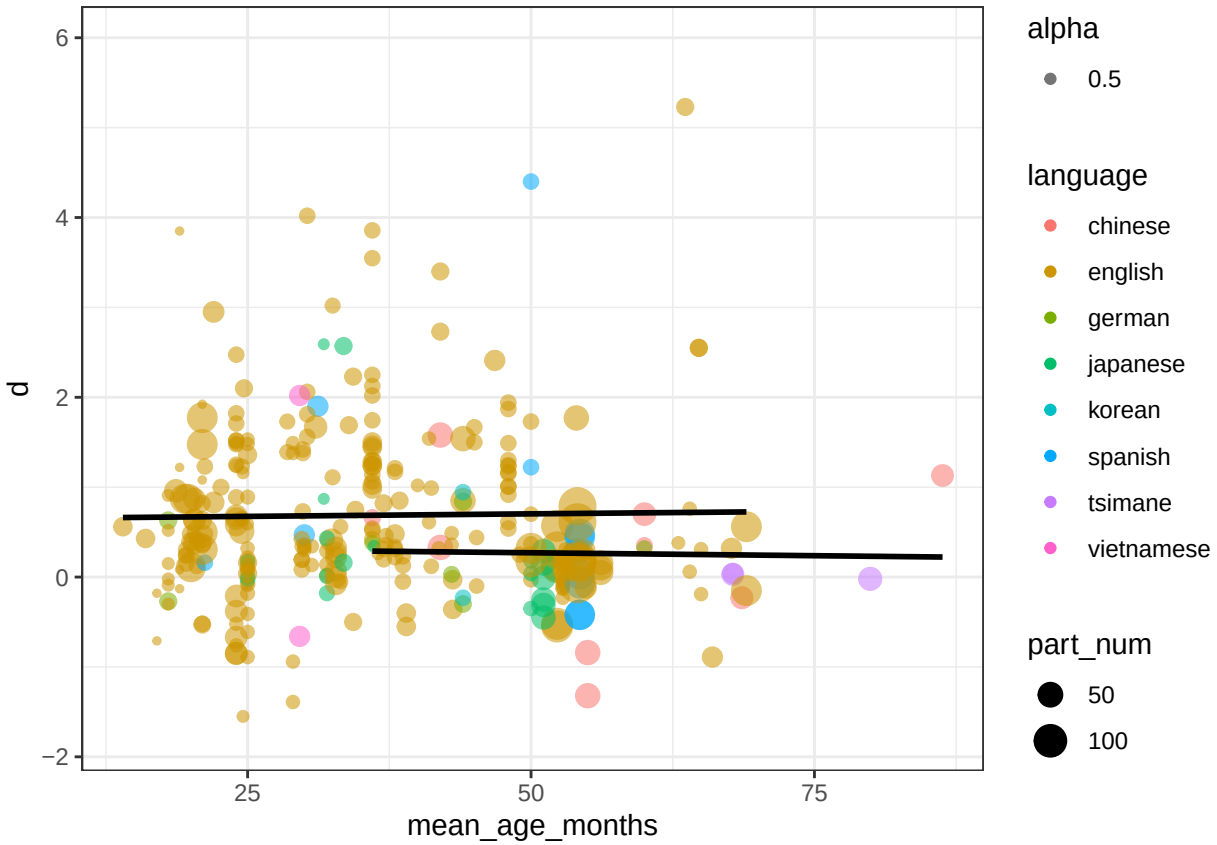

```
ggplot(df_shape,
       aes(x = mean_age_months, y = d, color = language)) + geom_point(aes(ymin = d - d_var, ymax = d +
                  alpha = .5, size = part_num))  +
  geom_smooth(data = filter(df_shape, language == "chinese"),
              col = "black",
              method = "lm", se = FALSE) +
  geom_smooth(data = filter(df_shape, language == "english"),
              col = "black",
              method = "lm", se = FALSE)
```

```
## Warning in geom_point(aes(ymin = d - d_var, ymax = d + d_var, alpha = 0.5, :
## Ignoring unknown aesthetics: ymin and ymax
```

```
## `geom_smooth()` using formula = 'y ~ x'
## `geom_smooth()` using formula = 'y ~ x'
```

```
## Warning: Removed 3 rows containing non-finite values (`stat_smooth()`).
```

```
## Warning: Removed 3 rows containing missing values (`geom_point()`).
```



Basic meta-analysis

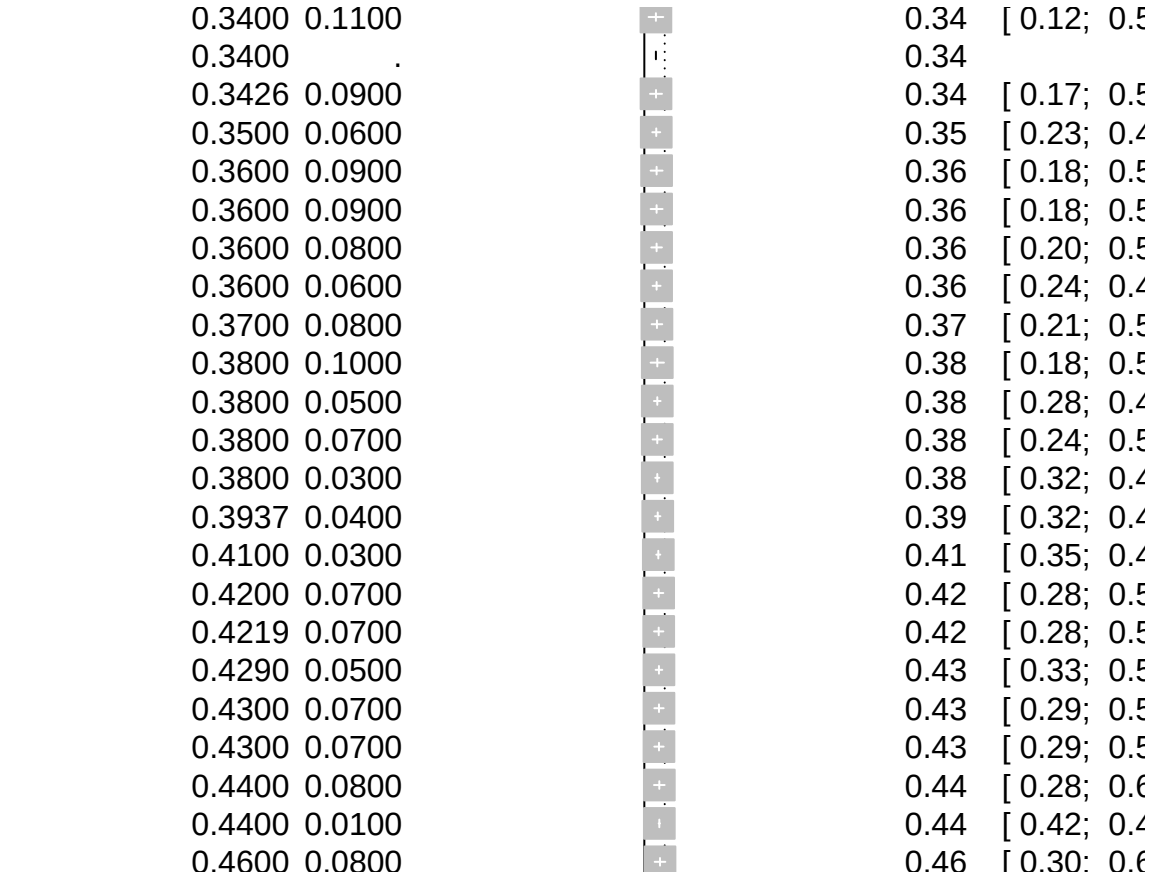
`metagen` approach

using the meta-analytic function meta-gen which calculates the weights for each effects and confidence interval, pooled effect size, the heterogeneity.

```
m.gen <- metagen( TE= d,
                  seTE = d_var,
                  studlab = ID,
                  data = df_shape,
                  sm = "SMD",
                  fixed = FALSE,
                  random = TRUE,
                  method.tau = "REML",
                  hakn = TRUE,
                  title = "pilot shape bias meta-analysis"

)

# summary(m.gen)['TE']
# m.gen["TE.fixed"]
# m.gen["TE.random"]
# m.gen["w.random"]
```

forest plot using the m-gen function object

```
forextobj <- forest.meta(m.gen,
                         sortvar = TE,
                         prediction = TRUE,
                         print.tau2 = FALSE,
                         leftlabs = c("Author", "g", "SE"))
```



forest plot from the rma model

metafor approach

using rma.mv instead of m.gen

```
mod <- rma.mv(yi = d,
              V = d_var,
              random = ~ 1 | ID,
              slab = short_cite,
              data = df_shape)
```

```
## Warning: 1 row with NAs omitted from model fitting.
```

```
summary(mod)
```

```
##
## Multivariate Meta-Analysis Model (k = 311; method: REML)
##
##    logLik   Deviance        AIC        BIC       AICc
## -797.2159  1594.4318  1598.4318  1605.9049  1598.4709
##
## Variance Components:
##
##             estim    sqrt  nlvls  fixed  factor
## sigma^2    0.1656  0.4069     48     no      ID
##
## Test for Heterogeneity:
## Q(df = 310) = 2573.9189, p-val < .0001
##
## Model Results:
##
## estimate      se    zval    pval   ci.lb   ci.ub
##   0.4365  0.0617  7.0794  <.0001  0.3156  0.5573  ***
##
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Nested model.

```
mod_nested <- rma.mv(yi = d,
                     V = d_var,
                     random = ~ 1 | ID/exp_num,
                     slab = short_cite,
                     data = filter(df_shape, !is.na(exp_num)))
```

```
## Warning: 1 row with NAs omitted from model fitting.
```

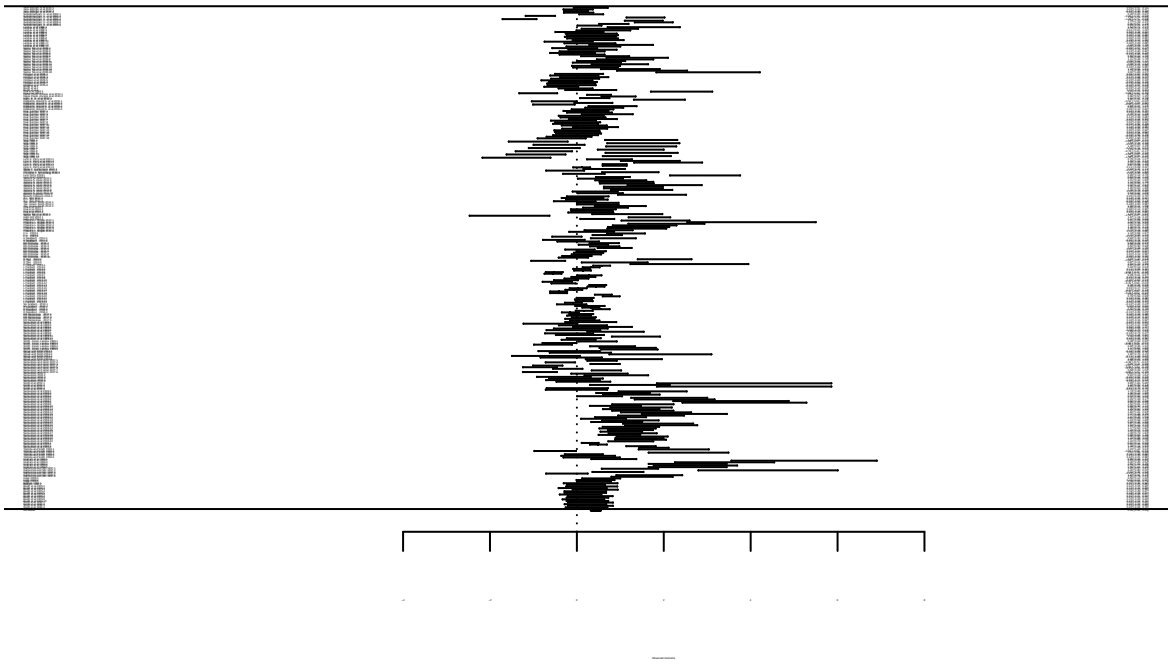
```
summary(mod_nested)
```

```
##
## Multivariate Meta-Analysis Model (k = 304; method: REML)
##
##    logLik   Deviance        AIC        BIC       AICc
## -743.4370  1486.8741  1492.8741  1504.0153  1492.9543
##
## Variance Components:
##
##             estim    sqrt  nlvls  fixed      factor
## sigma^2.1  0.1245  0.3528     48     no          ID
## sigma^2.2  0.0956  0.3093     75     no  ID/exp_num
##
## Test for Heterogeneity:
## Q(df = 303) = 2501.5515, p-val < .0001
##
## Model Results:
##
## estimate      se    zval    pval   ci.lb   ci.ub
```

```
##   0.4530  0.0668  6.7854  <.0001  0.3221  0.5838  ***
##
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Forest plot from `metafor`.

```
forest(mod) +
  theme_minimal(base_size = 8)
```



```
## NULL
```

Try this using ggplot

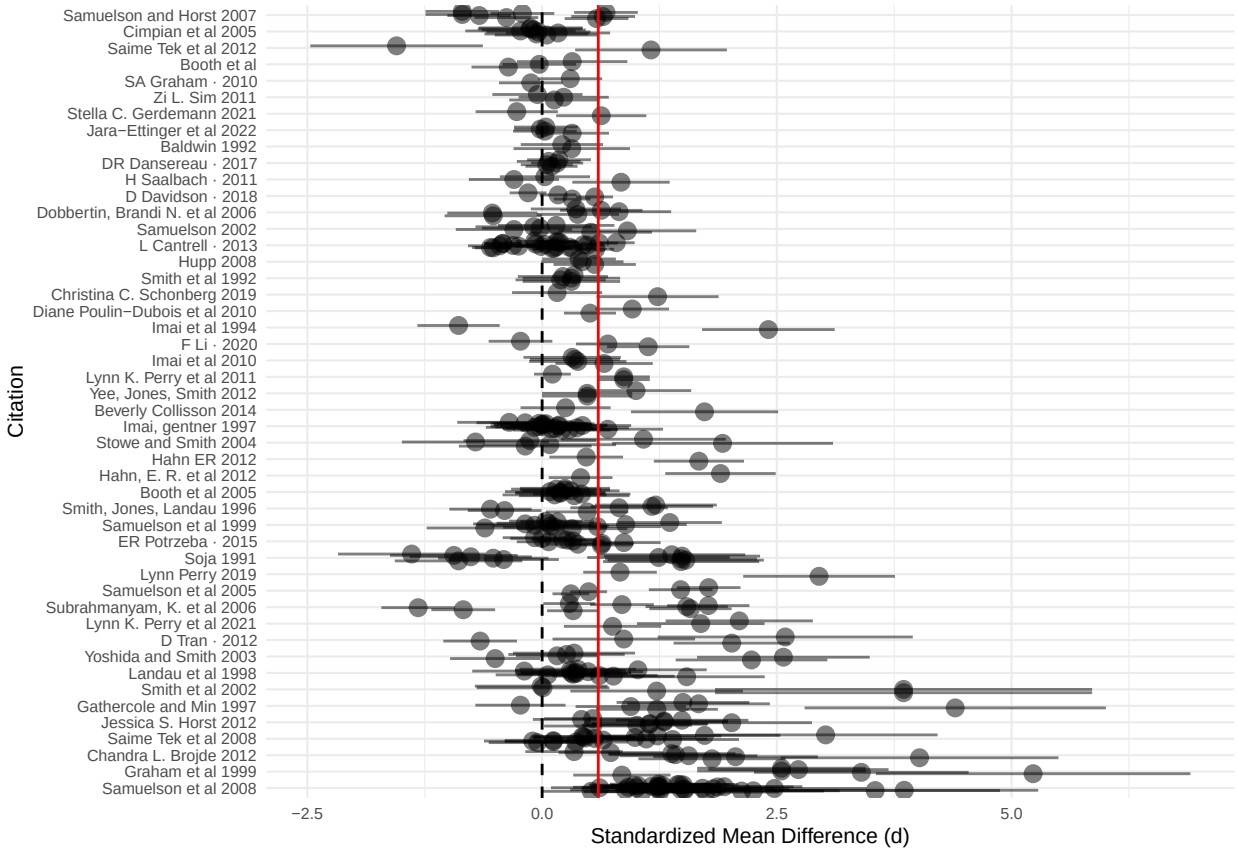
Revised forest plot

```
ggplot(df_shape, aes(x = short_cite, y = d,
                     ymin=d-sqrt(d_var)*1.96,
                     ymax=d+sqrt(d_var)*1.96)) +
  geom_pointrange( alpha = .5, position=position_dodge2(width=.5)) +
  coord_flip() +
  geom_hline(yintercept = 0, lty = 2) +
  geom_hline(data = m.gen ,yintercept = m.gen$TE.random, color = "red") +
```

```
  aes(x=reorder(short_cite,-d, sum)) +
  ylab("Standardized Mean Difference (d)") +
  xlab("Citation") +
  theme_minimal(base_size = 8)
```

```
## Warning: `geom_hline()`: Ignoring `data` because `yintercept` was provided.
```

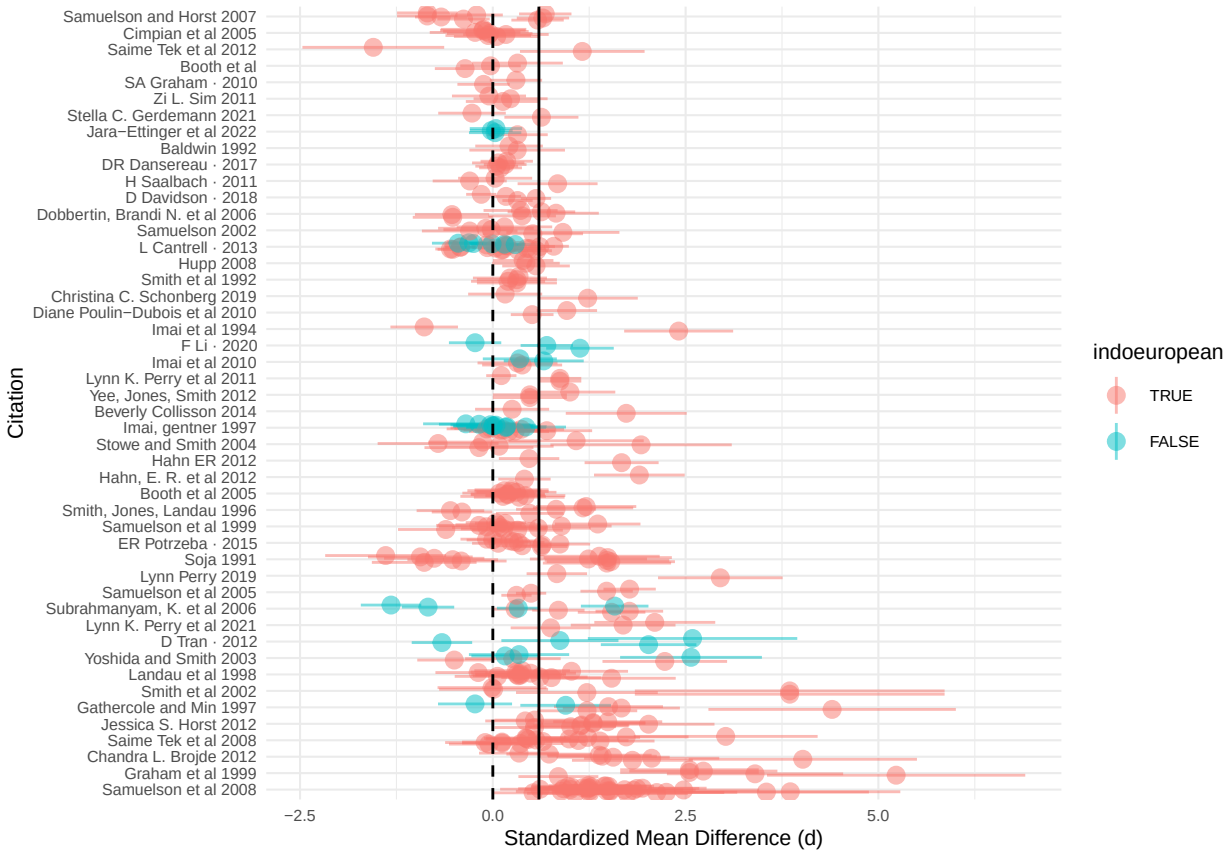
```
## Warning: Removed 1 rows containing missing values (`geom_segment()`).
```




```
 #+geom_text(aes(1.5,m.gen$TE.random,label = round(m.gen$TE.random,2), color = "darkred", size = 0.01 )
```

```
ggplot(df_shape, aes(x = short_cite, y = d,
                     ymin=d-sqrt(d_var)*1.96,
                     ymax=d+sqrt(d_var)*1.96)) +
  geom_pointrange(aes(color=indoeuropean), alpha = .5, position=position_dodge2(width=.5)) +
  coord_flip() +
  geom_hline(yintercept = 0, lty = 2) +
  geom_hline(data = m.gen ,yintercept = m.gen$TE.random) +
  aes(x=reorder(short_cite,-d, sum)) +
  ylab("Standardized Mean Difference (d)") +
  xlab("Citation") +
  theme_minimal(base_size = 8)
```

```
## Warning: `geom_hline()`: Ignoring `data` because `yintercept` was provided.
## Removed 1 rows containing missing values (`geom_segment()`).
```



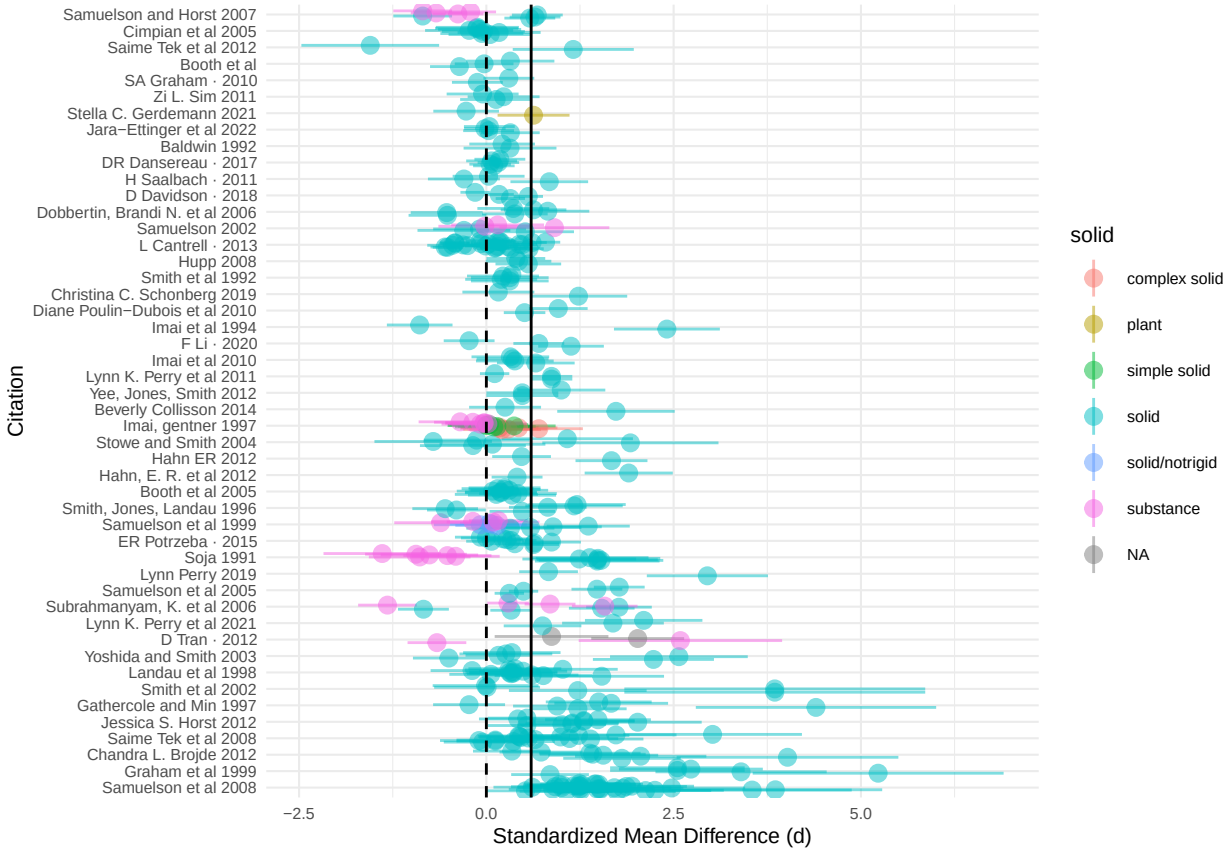

```
#png("secondgraph.png")
```

Forest plot with solidity:

```
ggplot(df_shape, aes(x = short_cite, y = d,
                     ymin=d-sqrt(d_var)*1.96,
                     ymax=d+sqrt(d_var)*1.96)) +
  geom_pointrange(aes(color=solid), alpha = .5, position=position_dodge2(width=.5)) +
  coord_flip() +
  geom_hline(yintercept = 0, lty = 2) +
  geom_hline(data = m.gen ,yintercept = m.gen$TE.random) +
  aes(x=reorder(short_cite,-d, sum)) +
  ylab("Standardized Mean Difference (d)") +
  xlab("Citation") +
  theme_minimal(base_size = 8)
```

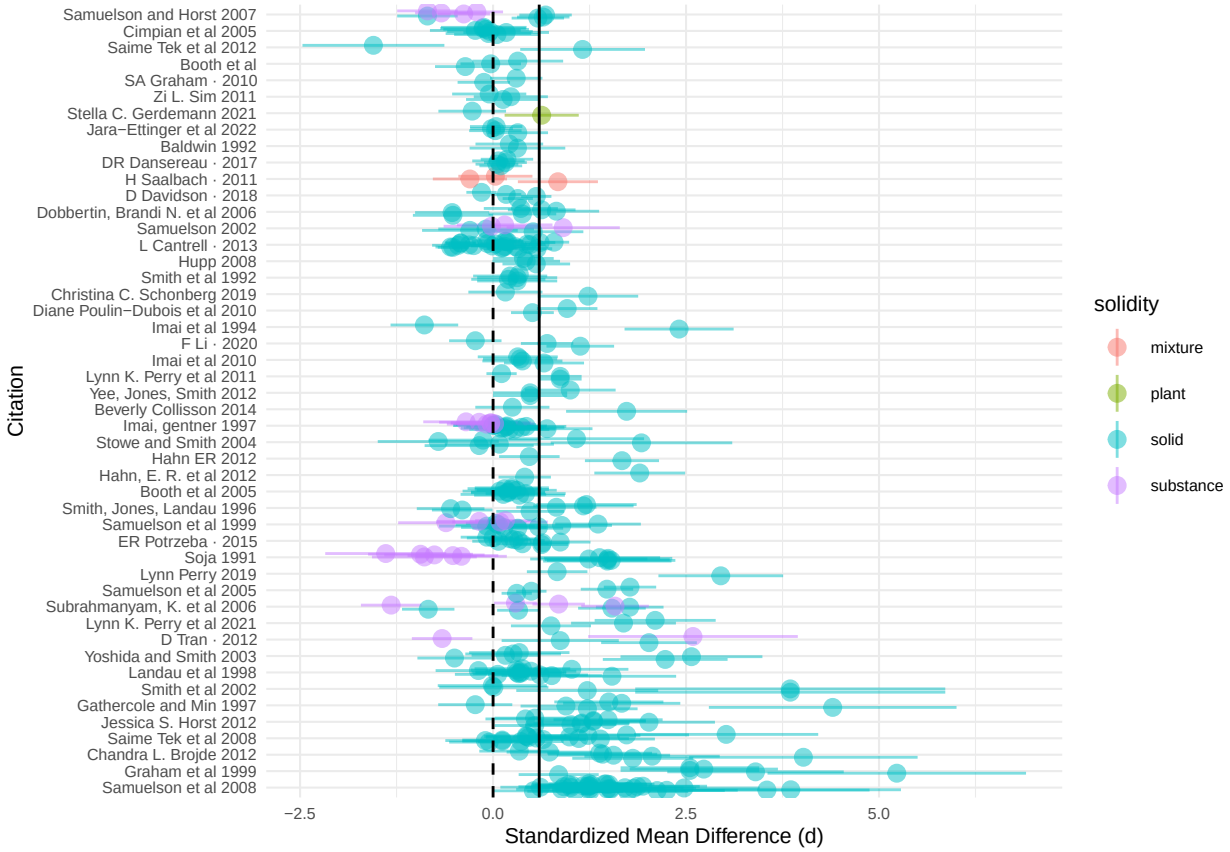
```
## Warning: `geom_hline()`: Ignoring `data` because `yintercept` was provided.
```

```
## Warning: Removed 1 rows containing missing values (`geom_segment()`).
```




```
ggplot(df_shape, aes(x = short_cite, y = d,
                     ymin=d-sqrt(d_var)*1.96,
                     ymax=d+sqrt(d_var)*1.96)) +
  geom_pointrange(aes(color=solidity), alpha = .5, position=position_dodge2(width=.5)) +
  coord_flip() +
  geom_hline(yintercept = 0, lty = 2) +
  geom_hline(data = m.gen ,yintercept = m.gen$TE.random) +
  aes(x=reorder(short_cite,-d, sum)) +
  ylab("Standardized Mean Difference (d)") +
  xlab("Citation") +
  theme_minimal(base_size = 8)
```

```
## Warning: ‘geom_hline()‘: Ignoring ‘data‘ because ‘yintercept‘ was provided.
## Removed 1 rows containing missing values (‘geom_segment()‘).
```



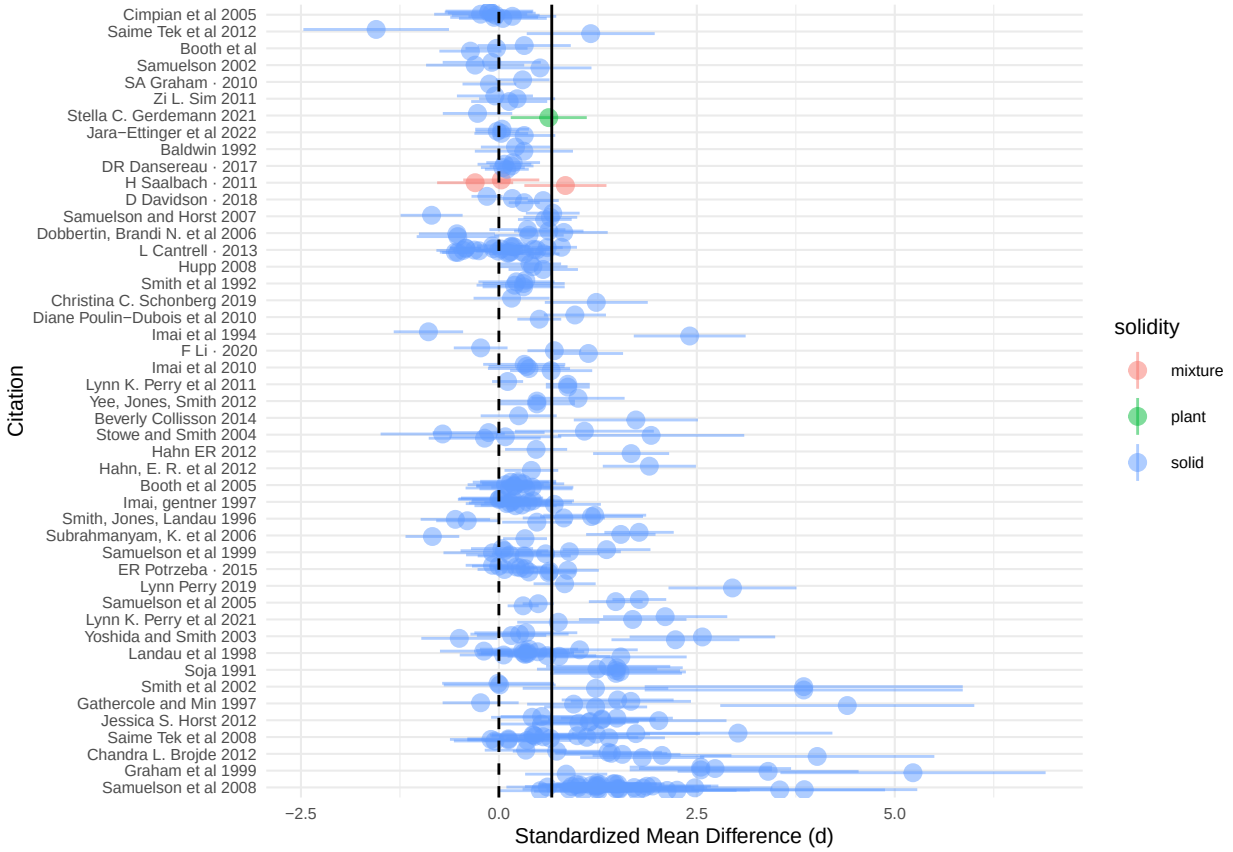

```
#png("secondgraph.png")
```

```
## data with only solidity:
df_shape_solid <- df_shape %>% filter(solid != "substance")

m.gen_solid <- metagen( TE= d,
                  seTE = d_var,
                  studlab = ID,
                  data = df_shape_solid,
                  sm = "SMD",
                  fixed = FALSE,
                  random = TRUE,
                  method.tau = "REML",
                  hakn = TRUE,
                  title = "pilot shape bias meta-analysis")

ggplot(df_shape_solid, aes(x = short_cite, y = d,
                     ymin=d-sqrt(d_var)*1.96,
                     ymax=d+sqrt(d_var)*1.96)) +
  geom_pointrange(aes(color=solidity), alpha = .5, position=position_dodge2(width=.5)) +
  coord_flip() +
  geom_hline(yintercept = 0, lty = 2) +
  geom_hline(data = m.gen ,yintercept = m.gen_solid$TE.random) +
  aes(x=reorder(short_cite,-d, sum)) +
  ylab("Standardized Mean Difference (d)") +
  xlab("Citation") +
```

```
  theme_minimal(base_size = 8)
```

```
## Warning: `geom_hline()`: Ignoring `data` because `yintercept` was provided.
```

```
## Warning: Removed 1 rows containing missing values (`geom_segment()`).
```




```
m.gen_solid
```

```
## Review:     pilot shape bias meta-analysis
##
## Number of studies: k = 280
##
##                              SMD           95%-CI     t  p-value
## Random effects model (HK) 0.6676 [0.5659; 0.7694] 12.92 < 0.0001
##
## Quantifying heterogeneity:
##  tau^2 = 0.6554 [0.6173; 0.8995]; tau = 0.8096 [0.7857; 0.9484]
##  I^2 = 99.5% [99.5%; 99.6%]; H = 14.75 [14.51; 14.99]
##
## Test of heterogeneity:
##         Q d.f. p-value
##  60688.36  279       0
##
## Details on meta-analytical method:
```

```
## - Inverse variance method
## - Restricted maximum-likelihood estimator for tau^2
## - Q-Profile method for confidence interval of tau^2 and tau
## - Hartung-Knapp adjustment for random effects model (df = 279)
```

```
m.gen
```

```
## Review:     pilot shape bias meta-analysis
##
## Number of studies: k = 311
##
##                              SMD           95%-CI     t  p-value
## Random effects model (HK) 0.5978 [0.4980; 0.6975] 11.79 < 0.0001
##
## Quantifying heterogeneity:
##  tau^2 = 0.7112 [0.6668; 0.9485]; tau = 0.8433 [0.8166; 0.9739]
##  I^2 = 99.5% [99.5%; 99.6%]; H = 14.70 [14.47; 14.93]
##
## Test of heterogeneity:
##         Q d.f. p-value
##  66972.27  310       0
##
## Details on meta-analytical method:
## - Inverse variance method
## - Restricted maximum-likelihood estimator for tau^2
## - Q-Profile method for confidence interval of tau^2 and tau
## - Hartung-Knapp adjustment for random effects model (df = 310)
```

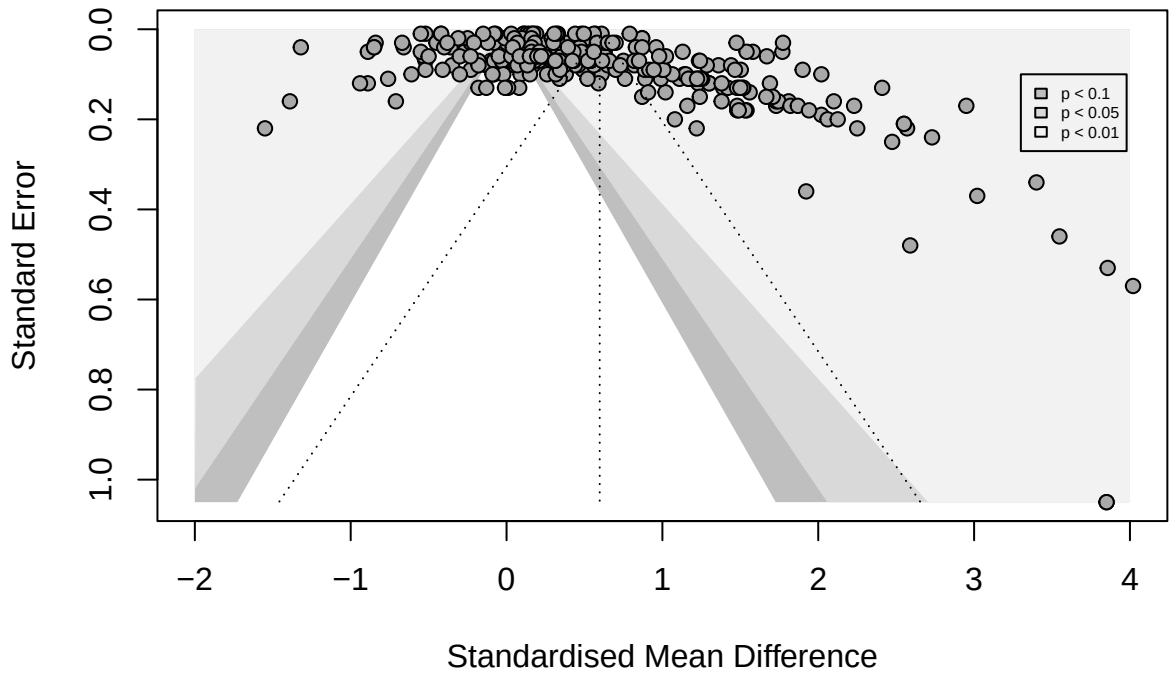
Publication bias

```
col.contour = c("gray75", "gray85", "gray95")

funnel(m.gen,
       comb.random = TRUE,
       xlim = c(-2, 4),
       contour = c(0.9, 0.95, 0.99),
       col.contour = col.contour)

regtest(x = d, vi = d_var,
        data = df_shape)
```

```
## Warning: 1 study with NAs omitted from test.
```

```
##
## Regression Test for Funnel Plot Asymmetry
##
## Model:     mixed-effects meta-regression model
## Predictor: standard error
##
## Test for Funnel Plot Asymmetry: z =  14.1226, p < .0001
## Limit Estimate (as sei -> 0):   b = -0.7016 (CI: -0.8874, -0.5159)
```

```
# Add a legend
legend(x = 3.3, y = 0.1, cex = 0.5,
       legend = c("p < 0.1", "p < 0.05", "p < 0.01"),
       fill = col.contour)
```




```
#png("funnel.png")
```

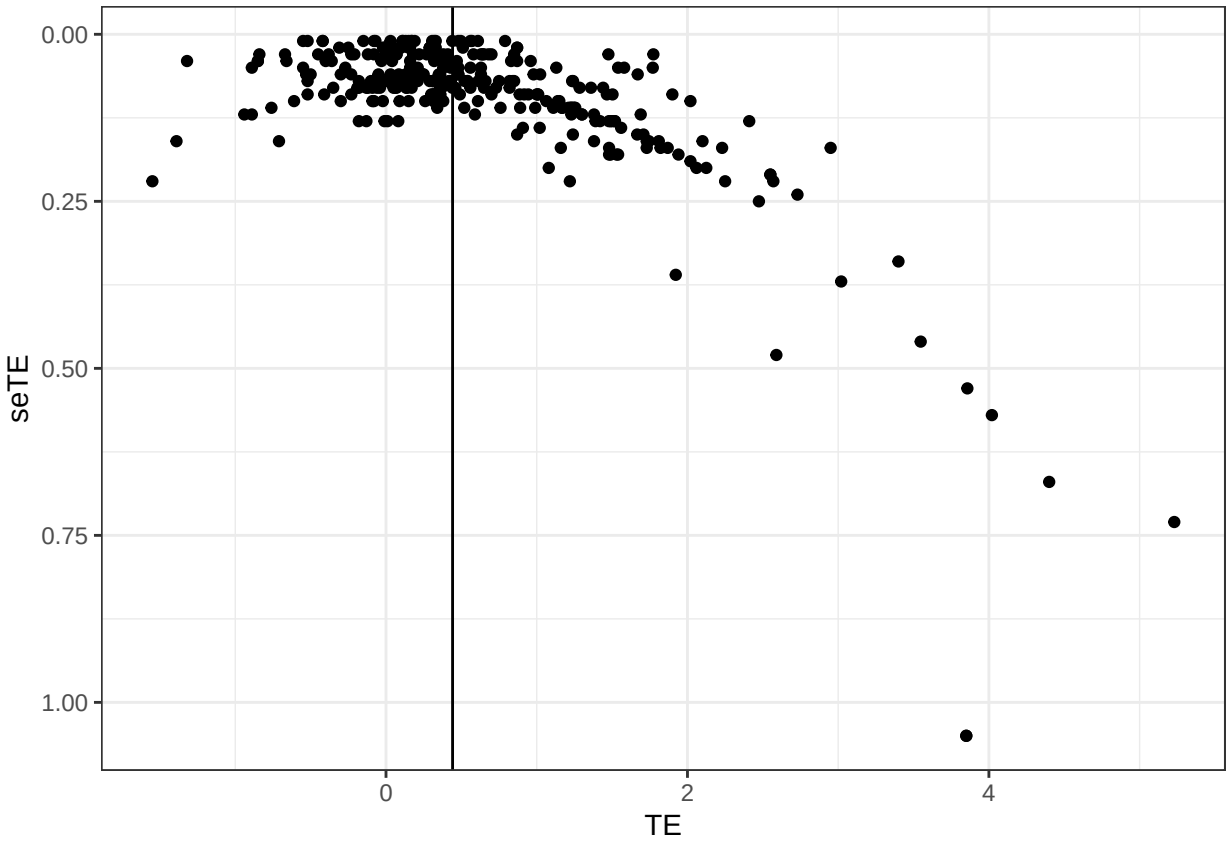
funnel plots using ggplot to account for moderators:

```
x = summary(m.gen)['TE']
y = summary(m.gen)['seTE']
m.gen["TE.fixed"]
```

```
## $TE.fixed
## [1] 0.189402
```

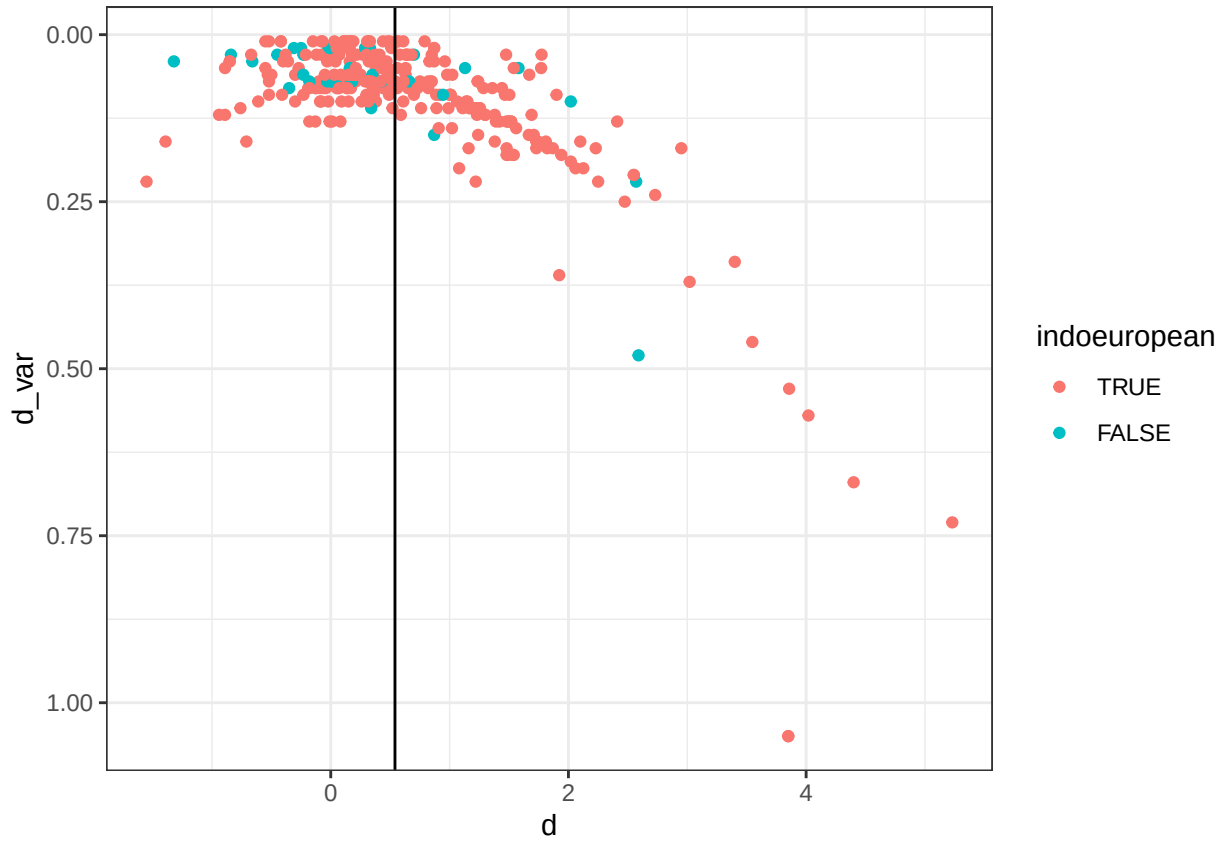
```
ter = m.gen["TE.random"]


data.gen = data.frame(x,y,ter)

ggplot(data = data.gen, mapping = aes(x=TE, y = seTE, color= )) +
  geom_point() +
  geom_vline(xintercept = 0.4418062) +
  scale_y_reverse()
```

```
## Warning: Removed 1 rows containing missing values (`geom_point()`).
```



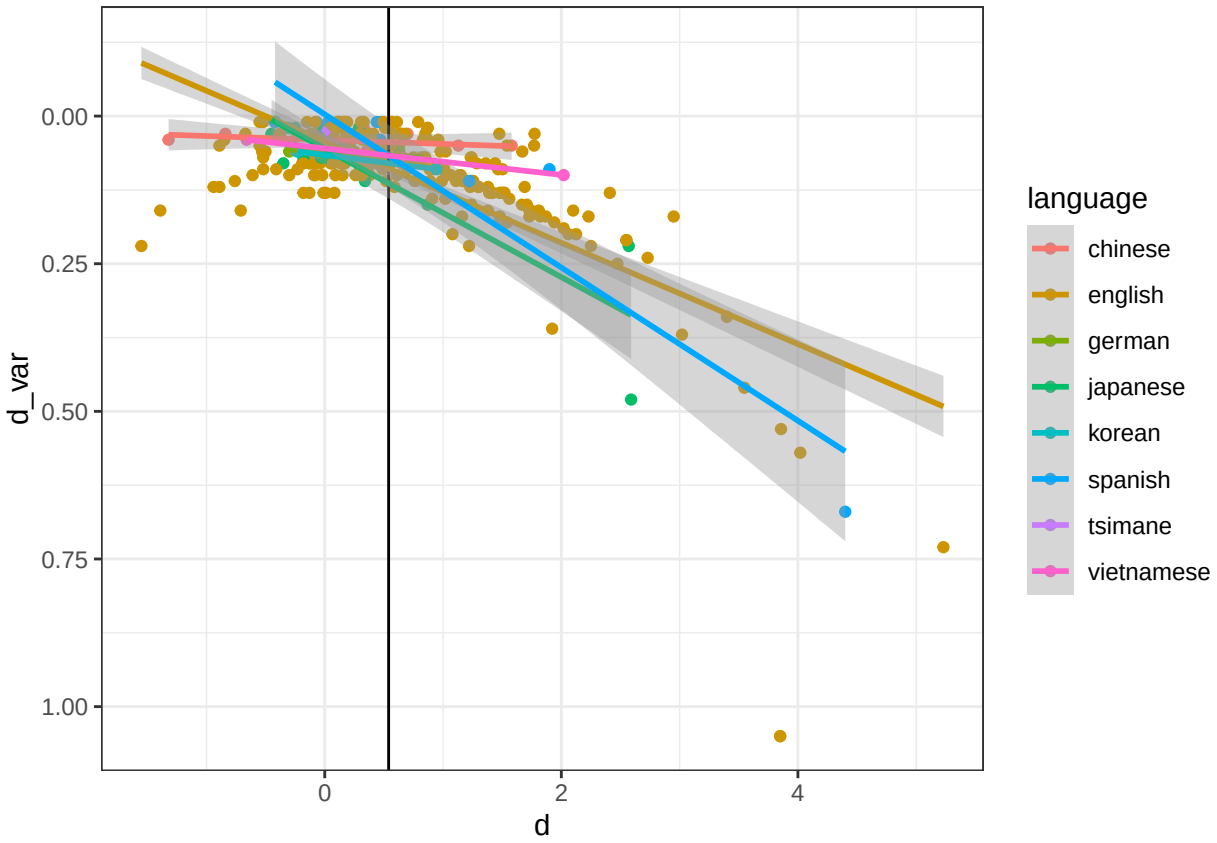

```
ggplot(data = df_shape, mapping = aes(x=d, y = d_var, color= indoeuropean)) +
  geom_point() +
  geom_vline(xintercept = 0.5401759) +
  scale_y_reverse()
```

```
## Warning: Removed 1 rows containing missing values (`geom_point()`).
```




```
ggplot(data = df_shape, mapping = aes(x=d, y = d_var, color= language)) +
  geom_point() +
  geom_vline(xintercept = 0.5401759) +
  scale_y_reverse() +
  geom_smooth(method = "lm")
```

```
## `geom_smooth()` using formula = 'y ~ x'

## Warning: Removed 1 rows containing non-finite values (`stat_smooth()`).

## Warning in qt((1 - level)/2, df): NaNs produced

## Warning in qt((1 - level)/2, df): NaNs produced

## Warning: Removed 1 rows containing missing values (`geom_point()`).

## Warning in max(ids, na.rm = TRUE): no non-missing arguments to max; returning
## -Inf

## Warning in max(ids, na.rm = TRUE): no non-missing arguments to max; returning
## -Inf
```



Eggers regresstion test

```
m.gen$data %>%
  mutate(y = m.gen$TE/m.gen$seTE, x = 1/m.gen$seTE) %>%
  lm(y ~ x, data= .) %>%
  summary()
```

```
##
## Call:
## lm(formula = y ~ x, data = .)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -69.191  -5.659   0.225   5.321  64.809
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  4.89851    1.08913   4.498 9.74e-06 ***
## x            0.09293    0.03203   2.902  0.00398 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 14.26 on 309 degrees of freedom
```

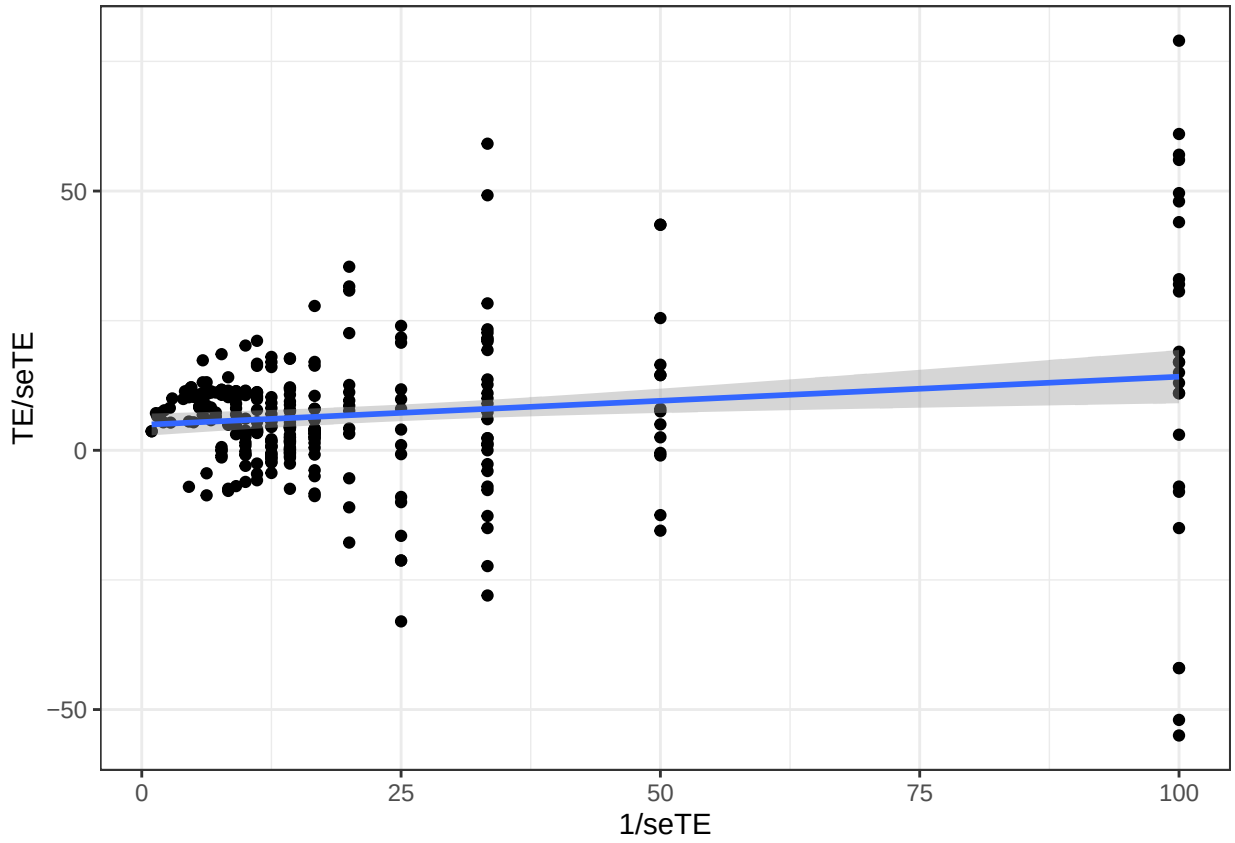
```
##   (1 observation deleted due to missingness)
## Multiple R-squared:  0.02652,    Adjusted R-squared:  0.02337
## F-statistic: 8.419 on 1 and 309 DF,  p-value: 0.00398
```

```
#eggers regression
ggplot(data = data.gen, mapping = aes(x=1/seTE, y = TE/seTE, color= )) +
  geom_point() +
  geom_smooth(method = "lm")
```

```
## `geom_smooth()` using formula = 'y ~ x'
```

```
## Warning: Removed 1 rows containing non-finite values (`stat_smooth()`).
```

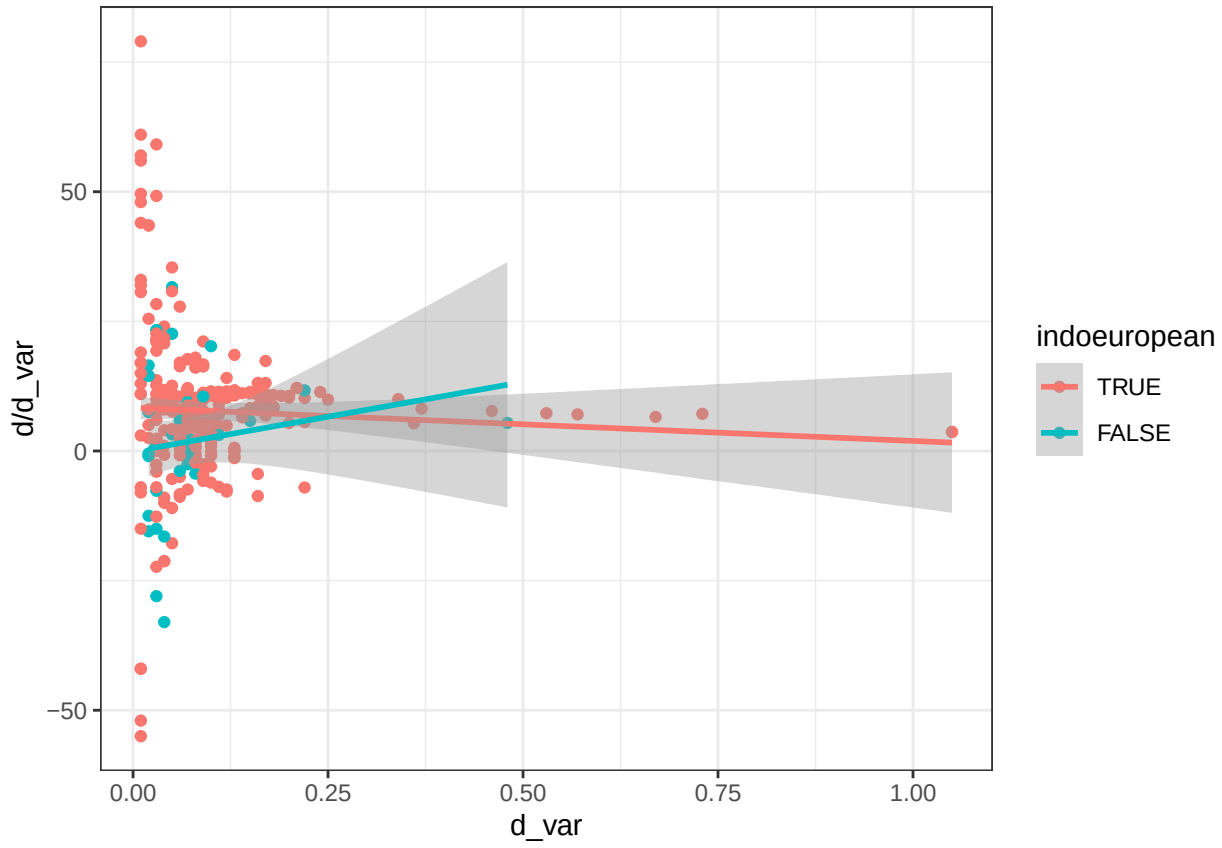
```
## Warning: Removed 1 rows containing missing values (`geom_point()`).
```




```
ggplot(data = df_shape, mapping = aes(x=d_var, y = d/d_var, color= indoeuropean)) +
  geom_point() +
  geom_smooth(method = "lm")
```

```
## `geom_smooth()` using formula = 'y ~ x'
```

```
## Warning: Removed 1 rows containing non-finite values (`stat_smooth()`).
## Removed 1 rows containing missing values (`geom_point()`).
```




```
# eggers.test(m.gen)
```

using rma.mv instead of m.gen

```
mod <- rma.mv(yi = d,
                  V = d_var,
                  random = ~ 1 | ID,
                  slab = short_cite,
                  data = df_shape)
```

```
## Warning: 1 row with NAs omitted from model fitting.
```

```
mod_nested <- rma.mv(yi = d,
                  V = d_var,
                  random = ~ 1 | ID/exp_num,
                  slab = short_cite,
                  data = filter(df_shape, !is.na(exp_num)))
```

```
## Warning: 1 row with NAs omitted from model fitting.
```

```
summary(mod)
```

```
##
```

```
## Multivariate Meta-Analysis Model (k = 311; method: REML)
##
##    logLik   Deviance        AIC        BIC       AICc
## -797.2159  1594.4318  1598.4318  1605.9049  1598.4709
##
## Variance Components:
##
##             estim    sqrt  nlvls  fixed  factor
## sigma^2    0.1656  0.4069     48     no      ID
##
## Test for Heterogeneity:
## Q(df = 310) = 2573.9189, p-val < .0001
##
## Model Results:
##
## estimate      se    zval    pval   ci.lb   ci.ub
##   0.4365  0.0617  7.0794  <.0001  0.3156  0.5573  ***
##
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
summary(mod_nested)
```

```
##
## Multivariate Meta-Analysis Model (k = 304; method: REML)
##
##    logLik   Deviance        AIC        BIC       AICc
## -743.4370  1486.8741  1492.8741  1504.0153  1492.9543
##
## Variance Components:
##
##             estim    sqrt  nlvls  fixed      factor
## sigma^2.1  0.1245  0.3528     48     no          ID
## sigma^2.2  0.0956  0.3093     75     no  ID/exp_num
##
## Test for Heterogeneity:
## Q(df = 303) = 2501.5515, p-val < .0001
##
## Model Results:
##
## estimate      se    zval    pval   ci.lb   ci.ub
##   0.4530  0.0668  6.7854  <.0001  0.3221  0.5838  ***
##
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

plotting coefficients from the rmv model: assuming that those coefficients correspond to effect sizes

Confirmatory analysis

For primary analyses, i will exclude effect sizes from clinical populations and multilingual populations.

I will investigate the hypotheses via multi-level meta-regressions using the metafor package.

In all models, I will include random effects that control for non-independence between effect sizes based on grouping by paper and grouping by experiment.

I will first fit: Shape bias ~ 1 Shape bias ~ age shape bias ~ log(age) shape bias ~ poly(age,2)

intercept:

```
# using the meta and metafor packages to analyze meta-analysis effect sizes
mod_intercept <- rma.mv(d ~ 1,
                        V = d_var,
                        random = ~1 | as.factor(Title) /
                          as.factor(exp_num),
                        slab = Title,
                        data = filter(df_shape, !is.na(exp_num)))
```

```
## Warning: 1 row with NAs omitted from model fitting.
```

```
mod_intercept_nonindo <- rma.mv(d ~ 1,
                        V = d_var,
                        random = ~1 | as.factor(Title) /
                          as.factor(exp_num),
                        slab = Title,
                        data = filter(df_shape_nonendo, !is.na(exp_num)))

mod_intercept_indo <- rma.mv(d ~ 1,
                        V = d_var,
                        random = ~1 | as.factor(Title) /
                          as.factor(exp_num),
                        slab = Title,
                        data = filter(df_shape_indo, !is.na(exp_num)))
```

```
## Warning: 1 row with NAs omitted from model fitting.
```

```
summary(mod_intercept_nonindo)
```

```
##
## Multivariate Meta-Analysis Model (k = 36; method: REML)
##
##    logLik   Deviance        AIC        BIC       AICc
## -118.9638   237.9276   243.9276   248.5937   244.7018
##
## Variance Components:
##
##             estim    sqrt  nlvls  fixed                              factor
## sigma^2.1  0.0019  0.0441      9     no                  as.factor(Title)
## sigma^2.2  0.1019  0.3192     14     no  as.factor(Title)/as.factor(exp_num)
##
## Test for Heterogeneity:
## Q(df = 35) = 308.3847, p-val < .0001
```

```
##
## Model Results:
##
## estimate      se    zval    pval   ci.lb   ci.ub
##   0.2240  0.0979  2.2874  0.0222  0.0321  0.4159  *
##
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
summary(mod_intercept_indo)
```

```
##
## Multivariate Meta-Analysis Model (k = 268; method: REML)
##
##    logLik   Deviance        AIC        BIC       AICc
## -586.7531  1173.5062  1179.5062  1190.2679  1179.5975
##
## Variance Components:
##
##             estim    sqrt  nlvls  fixed                                 factor
## sigma^2.1  0.1548  0.3934     47     no                     as.factor(Title)
## sigma^2.2  0.0890  0.2983     68     no  as.factor(Title)/as.factor(exp_num)
##
## Test for Heterogeneity:
## Q(df = 267) = 2141.0974, p-val < .0001
##
## Model Results:
##
## estimate      se    zval    pval   ci.lb   ci.ub
##   0.4879  0.0724  6.7390  <.0001  0.3460  0.6298  ***
##
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

age

```
mod_age_nonindo <- rma.mv(d ~ mean_age_months_centered36,
                  V = d_var,
                  random = ~1 | as.factor(Title) /
                    as.factor(exp_num),
                  slab = Title,
                  data = filter(df_shape_nonendo, !is.na(exp_num)))

mod_age_indo <- rma.mv(d ~ mean_age_months_centered36,
                  V = d_var,
                  random = ~1 | as.factor(Title) /
                    as.factor(exp_num),
                  slab = Title,
                  data = filter(df_shape_indo, !is.na(exp_num)))
```

```
## Warning: 4 rows with NAs omitted from model fitting.
```

```r
summary(mod_age_nonindo)
```

```
##
## Multivariate Meta-Analysis Model (k = 36; method: REML)
##
##    logLik   Deviance        AIC        BIC       AICc
## -111.6430   223.2860   231.2860   237.3914   232.6653
##
## Variance Components:
##
##             estim    sqrt  nlvls  fixed                              factor
## sigma^2.1  0.0911  0.3018      9     no                  as.factor(Title)
## sigma^2.2  0.1991  0.4462     14     no  as.factor(Title)/as.factor(exp_num)
##
## Test for Residual Heterogeneity:
## QE(df = 34) = 306.0195, p-val < .0001
##
## Test of Moderators (coefficient 2):
## QM(df = 1) = 17.8283, p-val < .0001
##
## Model Results:
##
##                            estimate      se     zval    pval    ci.lb    ci.ub
## intrcpt                      0.5806  0.1868   3.1089  0.0019   0.2146   0.9466
## mean_age_months_centered36  -0.0228  0.0054  -4.2224  <.0001  -0.0334  -0.0122
##
## intrcpt                      **
## mean_age_months_centered36  ***
##
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
summary(mod_age_indo)
```

```
##
## Multivariate Meta-Analysis Model (k = 265; method: REML)
##
##    logLik   Deviance        AIC        BIC       AICc
## -570.5792  1141.1584  1149.1584  1163.4471  1149.3135
##
## Variance Components:
##
##             estim    sqrt  nlvls  fixed                              factor
## sigma^2.1  0.1784  0.4224     47     no                  as.factor(Title)
## sigma^2.2  0.0853  0.2921     67     no  as.factor(Title)/as.factor(exp_num)
##
## Test for Residual Heterogeneity:
## QE(df = 263) = 2049.0995, p-val < .0001
##
## Test of Moderators (coefficient 2):
## QM(df = 1) = 13.5952, p-val = 0.0002
##
```

```
## Model Results:
##
##                             estimate      se     zval    pval    ci.lb    ci.ub
## intrcpt                       0.4990  0.0758   6.5854  <.0001   0.3505   0.6475
## mean_age_months_centered36   -0.0089  0.0024  -3.6872  0.0002  -0.0136  -0.0042
##
## intrcpt                      ***
## mean_age_months_centered36  ***
##
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

log age

```
mod_log_age_nonindo <- rma.mv(d ~ log_mean_age_months,
                      V = d_var,
                      random = ~1 | as.factor(Title) /
                        as.factor(exp_num),
                      slab = Title,
                      data = filter(df_shape_nonendo, !is.na(log_mean_age_months) , !is.na(exp_num)))

mod_log_age_indo <- rma.mv(d ~ log_mean_age_months,
                      V = d_var,
                      random = ~1 | as.factor(Title) /
                        as.factor(exp_num),
                      slab = Title,
                      data = filter(df_shape_indo, !is.na(log_mean_age_months) , !is.na(exp_num)))
```

```
## Warning: 1 row with NAs omitted from model fitting.
```

```
summary(mod_log_age_nonindo)
```

```
##
## Multivariate Meta-Analysis Model (k = 36; method: REML)
##
##    logLik   Deviance        AIC        BIC       AICc
## -112.7611   225.5223   233.5223   239.6277   234.9016
##
## Variance Components:
##
##             estim    sqrt  nlvls  fixed                                factor
## sigma^2.1  0.0451  0.2123      9     no                   as.factor(Title)
## sigma^2.2  0.1386  0.3723     14     no  as.factor(Title)/as.factor(exp_num)
##
## Test for Residual Heterogeneity:
## QE(df = 34) = 303.8279, p-val < .0001
##
## Test of Moderators (coefficient 2):
## QM(df = 1) = 13.5408, p-val = 0.0002
##
## Model Results:
```

```
##
##                       estimate      se     zval    pval    ci.lb    ci.ub
## intrcpt                 3.3799  0.8674   3.8966  <.0001   1.6798   5.0799  ***
## log_mean_age_months    -0.8100  0.2201  -3.6798  0.0002  -1.2414  -0.3786  ***
##
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
summary(mod_log_age_indo)
```

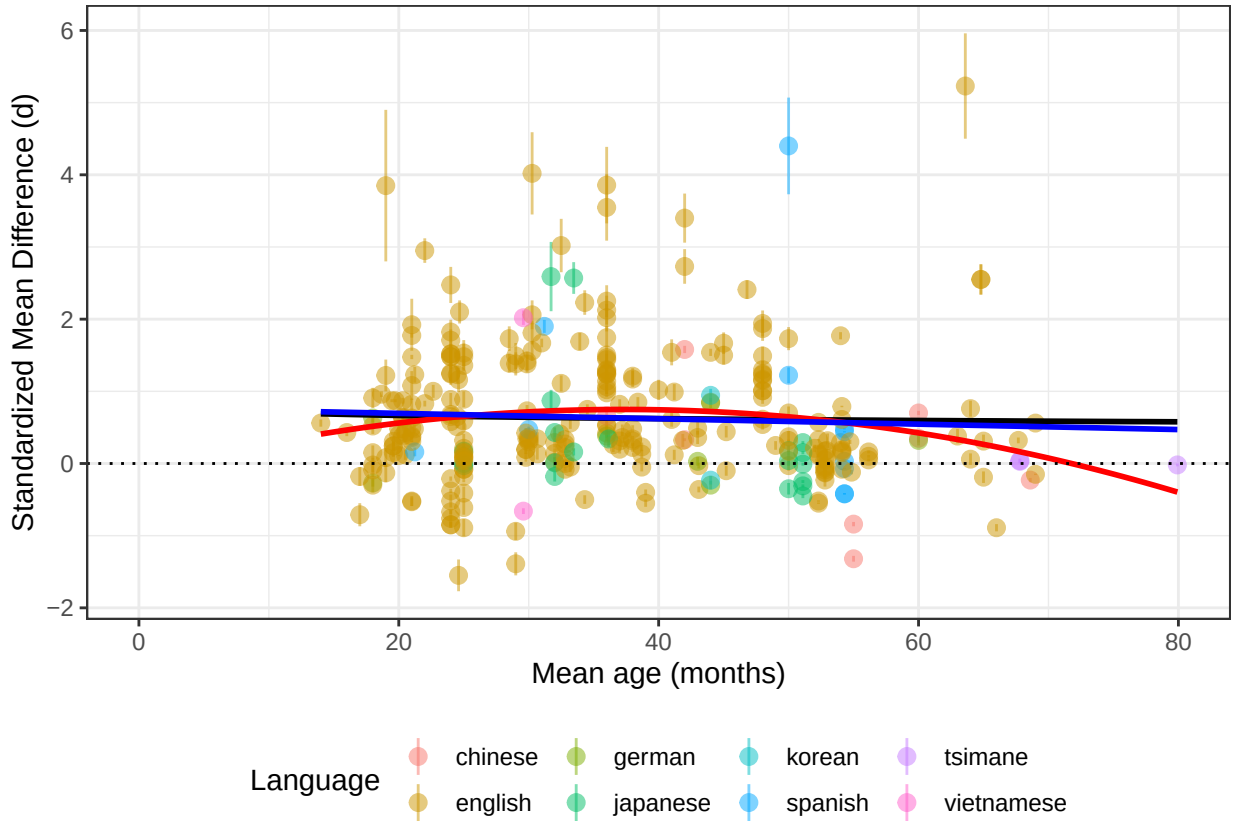
```
##
## Multivariate Meta-Analysis Model (k = 265; method: REML)
##
##    logLik   Deviance        AIC        BIC       AICc
## -570.4428  1140.8855  1148.8855  1163.1742  1149.0406
##
## Variance Components:
##
##             estim    sqrt  nlvls  fixed                                factor
## sigma^2.1  0.1847  0.4297     47     no                   as.factor(Title)
## sigma^2.2  0.0850  0.2916     67     no  as.factor(Title)/as.factor(exp_num)
##
## Test for Residual Heterogeneity:
## QE(df = 263) = 2060.5383, p-val < .0001
##
## Test of Moderators (coefficient 2):
## QM(df = 1) = 13.7542, p-val = 0.0002
##
## Model Results:
##
##                       estimate      se     zval    pval    ci.lb    ci.ub
## intrcpt                 1.7013  0.3330   5.1092  <.0001   1.0487   2.3540  ***
## log_mean_age_months    -0.3434  0.0926  -3.7087  0.0002  -0.5248  -0.1619  ***
##
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Let's look at what this means:

```
ggplot(df_shape,
       aes(x = mean_age_months, y = d, color = language))+
  geom_pointrange(aes(ymin = d - d_var, ymax = d + d_var),
                  alpha = .5) +
  geom_smooth(aes(group = 1),
              col = "black",
              method = "lm", se = FALSE,
              formula = y ~ log(x)) +
    geom_smooth(aes(group = 1),
              col = "red",
              method = "lm", se = FALSE,
              formula = y ~ poly(x,2)) +
  geom_smooth(aes(group = 1),
              col = "blue",
```

```
              method = "lm", se = FALSE,
              formula = y ~ x) +
  geom_hline(yintercept = 0, lty = 3) +
  ylab("Standardized Mean Difference (d)") +
  xlab("Mean age (months)") +
  scale_color_discrete(name = "Language") +
  scale_linetype_discrete(name = "Indo-European") +
  theme(legend.position = "bottom") +
  xlim(0,80)
```

```
## Warning: Removed 4 rows containing non-finite values (`stat_smooth()`).
## Removed 4 rows containing non-finite values (`stat_smooth()`).
## Removed 4 rows containing non-finite values (`stat_smooth()`).

## Warning: Removed 4 rows containing missing values (`geom_pointrange()`).

## Warning: Removed 1 rows containing missing values (`geom_segment()`).
```

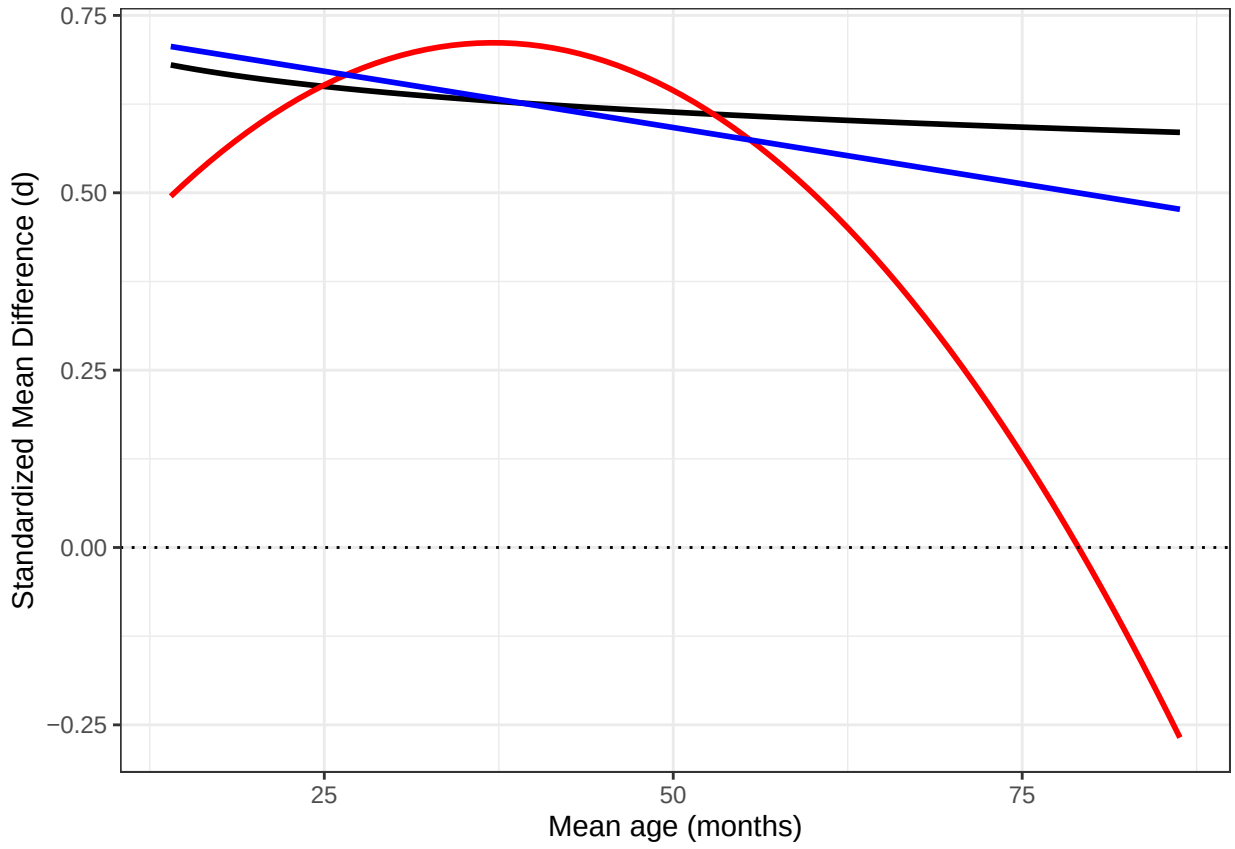



```
ggplot(df_shape,
       aes(x = mean_age_months, y = d))+
  geom_smooth(aes(group = 1),
              col = "black",
              method = "lm", se = FALSE,
              formula = y ~ log(x), show.legend = TRUE) +
```

```
    geom_smooth(aes(group = 1),
                col = "red",
                method = "lm", se = FALSE,
                formula = y ~ poly(x,2)) +
  geom_smooth(aes(group = 1),
                col = "blue",
                method = "lm", se = FALSE,
                formula = y ~ x) +
  geom_hline(yintercept = 0, lty = 3) +
  ylab("Standardized Mean Difference (d)") +
  xlab("Mean age (months)")
```

```
## Warning: Removed 3 rows containing non-finite values (‘stat_smooth()‘).
```

```
## Warning: Removed 3 rows containing non-finite values (‘stat_smooth()‘).
## Removed 3 rows containing non-finite values (‘stat_smooth()‘).
```



polynomial age

```
mod_poly_nonindo <- rma.mv(d ~ mean_age_months_centered36 + I(mean_age_months_centered36^2),
                   V = d_var,
                   random = ~1 | as.factor(ID) /
```

```
                    as.factor(exp_num),
                  slab = Title,
                  data = filter(df_shape_nonendo, !is.na(log_mean_age_months), !is.na(exp_num)))

mod_poly_indo <- rma.mv(d ~ mean_age_months_centered36 + I(mean_age_months_centered36^2),
                  V = d_var,
                  random = ~1 | as.factor(ID) /
                    as.factor(exp_num),
                  slab = Title,
                  data = filter(df_shape_indo, !is.na(log_mean_age_months), !is.na(exp_num)))
```

```
## Warning: 1 row with NAs omitted from model fitting.
```

```
summary(mod_poly_nonindo)
```

```
##
## Multivariate Meta-Analysis Model (k = 36; method: REML)
##
##    logLik   Deviance        AIC        BIC       AICc
## -111.4396   222.8792   232.8792   240.3617   235.1014
##
## Variance Components:
##
##             estim    sqrt  nlvls  fixed                           factor
## sigma^2.1  0.0647  0.2543      9     no                      as.factor(ID)
## sigma^2.2  0.1535  0.3918     14     no  as.factor(ID)/as.factor(exp_num)
##
## Test for Residual Heterogeneity:
## QE(df = 33) = 286.5903, p-val < .0001
##
## Test of Moderators (coefficients 2:3):
## QM(df = 2) = 16.5974, p-val = 0.0002
##
## Model Results:
##
##                                 estimate      se     zval    pval    ci.lb
## intrcpt                           0.5200  0.1723   3.0188  0.0025   0.1824
## mean_age_months_centered36       -0.0249  0.0080  -3.1284  0.0018  -0.0405
## I(mean_age_months_centered36^2)   0.0002  0.0003   0.6881  0.4914  -0.0003
##                                   ci.ub
## intrcpt                          0.8576  **
## mean_age_months_centered36      -0.0093  **
## I(mean_age_months_centered36^2)  0.0007
##
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
summary(mod_poly_indo)
```

```
##
## Multivariate Meta-Analysis Model (k = 265; method: REML)
```

```
##
##    logLik   Deviance        AIC        BIC       AICc
## -570.6134  1141.2268  1151.2268  1169.0686  1151.4612
##
## Variance Components:
##
##             estim    sqrt  nlvls  fixed                          factor
## sigma^2.1  0.1852  0.4303     46     no                     as.factor(ID)
## sigma^2.2  0.0852  0.2918     67     no  as.factor(ID)/as.factor(exp_num)
##
## Test for Residual Heterogeneity:
## QE(df = 262) = 2039.8480, p-val < .0001
##
## Test of Moderators (coefficients 2:3):
## QM(df = 2) = 13.9443, p-val = 0.0009
##
## Model Results:
##
##                                 estimate      se     zval    pval    ci.lb
## intrcpt                           0.4848  0.0806   6.0178  <.0001   0.3269
## mean_age_months_centered36       -0.0099  0.0030  -3.3254  0.0009  -0.0158
## I(mean_age_months_centered36^2)   0.0001  0.0001   0.5625  0.5738  -0.0002
##                                   ci.ub
## intrcpt                          0.6427  ***
## mean_age_months_centered36      -0.0041  ***
## I(mean_age_months_centered36^2)  0.0003
##
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

model comparison and plotting AICc , what is the criteria ? cutoff

Polynomial age models

Let's start with an interaction with `indoeuropean` with standard quadratic terms. This model is very interpretable.

```
rma.mv(d ~ mean_age_months_centered36 * indoeuropean +
         I(mean_age_months_centered36^2) * indoeuropean,
       V = d_var,
       random = ~ 1 | ID/exp_num,
       slab = short_cite,
       data = filter(df_shape, !is.na(exp_num), !is.na(language)))
```

```
## Warning: 4 rows with NAs omitted from model fitting.
```

```
##
## Multivariate Meta-Analysis Model (k = 301; method: REML)
##
## Variance Components:
##
##             estim    sqrt  nlvls  fixed      factor
```

```
## sigma^2.1  0.1582  0.3978     48     no          ID
## sigma^2.2  0.0933  0.3054     74     no  ID/exp_num
##
## Test for Residual Heterogeneity:
## QE(df = 295) = 2326.4383, p-val < .0001
##
## Test of Moderators (coefficients 2:6):
## QM(df = 5) = 56.0555, p-val < .0001
##
## Model Results:
##
##                                                estimate      se     zval
## intrcpt                                          0.4989  0.0757   6.5873
## mean_age_months_centered36                      -0.0099  0.0029  -3.4579
## indoeuropeanFALSE                               -0.1472  0.0737  -1.9967
## I(mean_age_months_centered36^2)                  0.0000  0.0001   0.1988
## mean_age_months_centered36:indoeuropeanFALSE    -0.0143  0.0065  -2.2109
## indoeuropeanFALSE:I(mean_age_months_centered36^2)  0.0003  0.0002   1.1043
##                                                    pval    ci.lb    ci.ub
## intrcpt                                          <.0001   0.3505   0.6474
## mean_age_months_centered36                       0.0005  -0.0155  -0.0043
## indoeuropeanFALSE                                0.0459  -0.2917  -0.0027
## I(mean_age_months_centered36^2)                  0.8424  -0.0002   0.0002
## mean_age_months_centered36:indoeuropeanFALSE     0.0270  -0.0270  -0.0016
## indoeuropeanFALSE:I(mean_age_months_centered36^2)  0.2694  -0.0002   0.0007
##
## intrcpt                                          ***
## mean_age_months_centered36                       ***
## indoeuropeanFALSE                                  *
## I(mean_age_months_centered36^2)
## mean_age_months_centered36:indoeuropeanFALSE       *
## indoeuropeanFALSE:I(mean_age_months_centered36^2)
##
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
rma.mv(d ~ mean_age_months_centered36 * indoeuropean ,
       V = d_var,
       random = ~ 1 | ID/exp_num,
       slab = short_cite,
       data = filter(df_shape, !is.na(exp_num), !is.na(language)))
```

```
## Warning: 4 rows with NAs omitted from model fitting.
```

```
##
## Multivariate Meta-Analysis Model (k = 301; method: REML)
##
## Variance Components:
##
##             estim    sqrt  nlvls  fixed      factor
## sigma^2.1  0.1601  0.4001     48     no          ID
## sigma^2.2  0.1051  0.3242     74     no  ID/exp_num
##
```

```
## Test for Residual Heterogeneity:
## QE(df = 297) = 2355.1191, p-val < .0001
##
## Test of Moderators (coefficients 2:4):
## QM(df = 3) = 54.7650, p-val < .0001
##
## Model Results:
##
##                                          estimate      se     zval    pval
## intrcpt                                    0.5139  0.0738   6.9608  <.0001
## mean_age_months_centered36                -0.0093  0.0023  -4.0273  <.0001
## indoeuropeanFALSE                         -0.1497  0.0737  -2.0327  0.0421
## mean_age_months_centered36:indoeuropeanFALSE  -0.0093  0.0047  -1.9826  0.0474
##                                           ci.lb    ci.ub
## intrcpt                                  0.3692   0.6586  ***
## mean_age_months_centered36               -0.0139  -0.0048  ***
## indoeuropeanFALSE                        -0.2941  -0.0054    *
## mean_age_months_centered36:indoeuropeanFALSE  -0.0184  -0.0001    *
##
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Next try breaking down by language. Here we can see Spanish is sparse and has a huge interaction term for some reason. Probably just overfit.

With the orthogonal polynomials, it blows up completely.

here is the same model but changing the contrasts:

Discussion

References