Lab 12: Meta-Analysis

Contents

Meta-analysis

Warm-up questions: Why do a meta-analysis?

1. Instead of conducting a meta-analysis, many review papers contain statements like: "Eight of the ten studies provided significant evidence for XXX effect." What do you think about this?
2. Instead of doing a meta-analysis, why not just average the studies' point estimates?

Basics

As discussed in lecture, there are two types of meta-analyis: **fixed effects** and **random effects**. Today I'll focus only on random effects meta-analysis since it is by far the more common approach, and it subsumes fixed effects meta-analysis (the latter is just random effects meta-analysis with zero heterogeneity). Let θ_i be the true effect in study i (an unknown statistical parameter) and SE_i^2 is the within-study variance (squared standard error), assumed fixed and known. Then random effects meta-analysis framework assumes, at minimum:

$$E[\theta_i] = \theta \tag{1}$$

$$\text{Var}[\theta_i] = T^2 \tag{2}$$

There are multiple estimation approaches. The classical Dersimonian-Laird approach came first historically. It is nonparametric approach that does *not* assume normality on the true effects. **Despite being the default in most non-R softwares, the DL method has anti-conservative inference and poor heterogeneity estimation, so it is no longer the tool of choice** (Veroniki et al. 2016).

Happily, there are now parametric alternatives to the Dersimonian-Laird approach that are unbiased for the grand mean (θ) and variance (T^2) and that have correct inference. These variations add normality assumptions on the true effects:

$$\begin{aligned} y_i &= \underbrace{\theta + \gamma_i}_{\theta_i} + \epsilon_i \\ \gamma_i &\sim N(0, T^2) \\ \epsilon_i &\sim N(0, SE_i^2) \end{aligned} \tag{3}$$

Thus, marginally, we have $y_i \sim N(\theta, T^2 + SE_i^2)$.

The usual targets of statistical inference are θ and T^2. Since this is a parametric model, the usual estimation approach is maximum likelihood (ML) or restricted maximum likelihood (REML). (REML is the quite sensible default method in the R package we'll use, `metafor`.) Either way, the pooled point estimate of θ is:

$$\widehat{\theta} = \frac{\sum_i w_i y_i}{\sum_i w_i} \tag{4}$$

$$w_i = \frac{1}{T^2 + SE_i^2} \tag{5}$$

The ML/REML solutions for $\widehat{T}^2$ are complicated, and I'm not going to cover them (see Veroniki et al. (2016)).

For inference, one option for an approximate confidence interval is to use the asymptotic normal distribution of the MLE:

$$\widehat{\theta} \approx N\left(\theta, \frac{1}{\sum_{i=1}^{k} \frac{1}{\widehat{T}^2 + SE_i^2}}\right)$$

where $\widehat{T}^2$ estimates T^2 and the variance is usual inverse Fisher information. In practice, **these CIs should be tweaked via the Knapp-Hartung adjustment** to avoid anticonservative inference, **but note that this is not the default in most software, including the package we will use**[1].

Effect size conversions and normality

There are two reasons for which we often need to convert effect sizes to different scales when doing meta-analysis. First, typically studies will report effect sizes on different scales (e.g., risk ratios vs. standardized mean differences), and the effect sizes need to be put onto a common scale in order to synthesize them. Second, not all effect scales will fulfill the normality assumptions; you should choose a scale on which the normality assumption is plausible (but you should still check the assumption visually). For example, you'd want to meta-analyze RRs on the log scale. For Pearson's correlations, there is a conversion (which you will see in the pre-written homework code) to put them on an unbounded, approximately normal scale called Fisher's z scale:

$$z = \frac{1}{2} ln\left(\frac{1+r}{1-r}\right) = \text{arctanh}(r)$$

[1] Optional details: In the standard MLE confidence interval, we just plug in the estimated $\widehat{T}^2$ in place of T^2, which assumes that it is known rather than estimated. Pretending that we know T^2 exactly is bad because in fact it is quite hard to estimate precisely. For this reason, you should always use Knapp-Hartung SEs to account for uncertainty in T^2, which are often much wider than naïve SEs (IntHout, Ioannidis, and Borm 2014).

$$r = \frac{\exp(2z) - 1}{\exp(2z) + 1} = \tanh(z)$$

I've posted a textbook chapter ("Converting among effect sizes" from Borenstein et al. (2009)) that provides some conversions. If you actually use these conversions in practice, please see the Optional Supplement for some important caveats.

Example: School-based writing interventions

The `metafor` package contains an example dataset of 48 experiments teaching student writing skills:

"In each of the studies included in this meta-analysis, an experimental group (i.e., a group of students that received instruction with increased emphasis on writing tasks) was compared against a control group (i.e., a group of students that received conventional instruction) with respect to some content-related measure of academic achievement (e.g., final grade, an exam/quiz/test score). The effect size measure for this meta-analysis was the standardized mean difference (with positive scores indicating a higher mean level of academic achievement in the intervention group)."

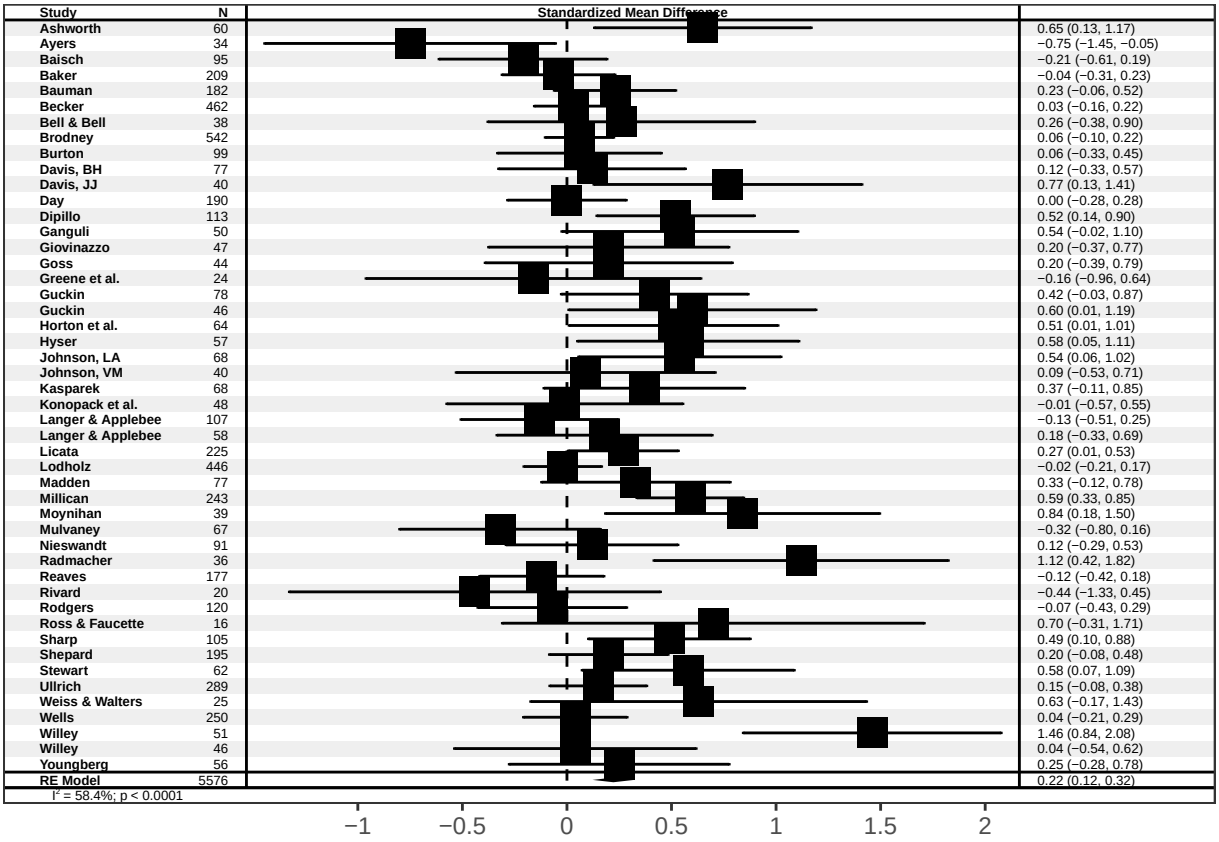
These effect sizes are standardized mean differences, and we are going to estimate standard errors using the Knapp-Hartung adjustment (`knha=TRUE`). Let's fit a random-effects meta-analysis model using the `metafor` package and use it to create a distateful-looking forest plot[2].

```
library(metafor)
d = dat.bangertdrowns2004[ !is.na( dat.bangertdrowns2004$yi ), ]

# data dictionary
?dat.bangertdrowns2004

( m = rma( yi = d$yi,
           vi = d$vi,
           ni = d$ni,  # sample sizes only needed for later forest plot
           knha = TRUE ) )
```

```
##
## Random-Effects Model (k = 48; tau^2 estimator: REML)
##
## tau^2 (estimated amount of total heterogeneity): 0.0499 (SE = 0.0197)
## tau (square root of estimated tau^2 value):      0.2235
## I^2 (total heterogeneity / total variability):   58.37%
## H^2 (total variability / sampling variability):  2.40
##
## Test for Heterogeneity:
## Q(df = 47) = 107.1061, p-val < .0001
##
## Model Results:
##
## estimate      se    tval    pval   ci.lb   ci.ub
##   0.2219  0.0495  4.4810  <.0001  0.1223  0.3216  ***
##
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

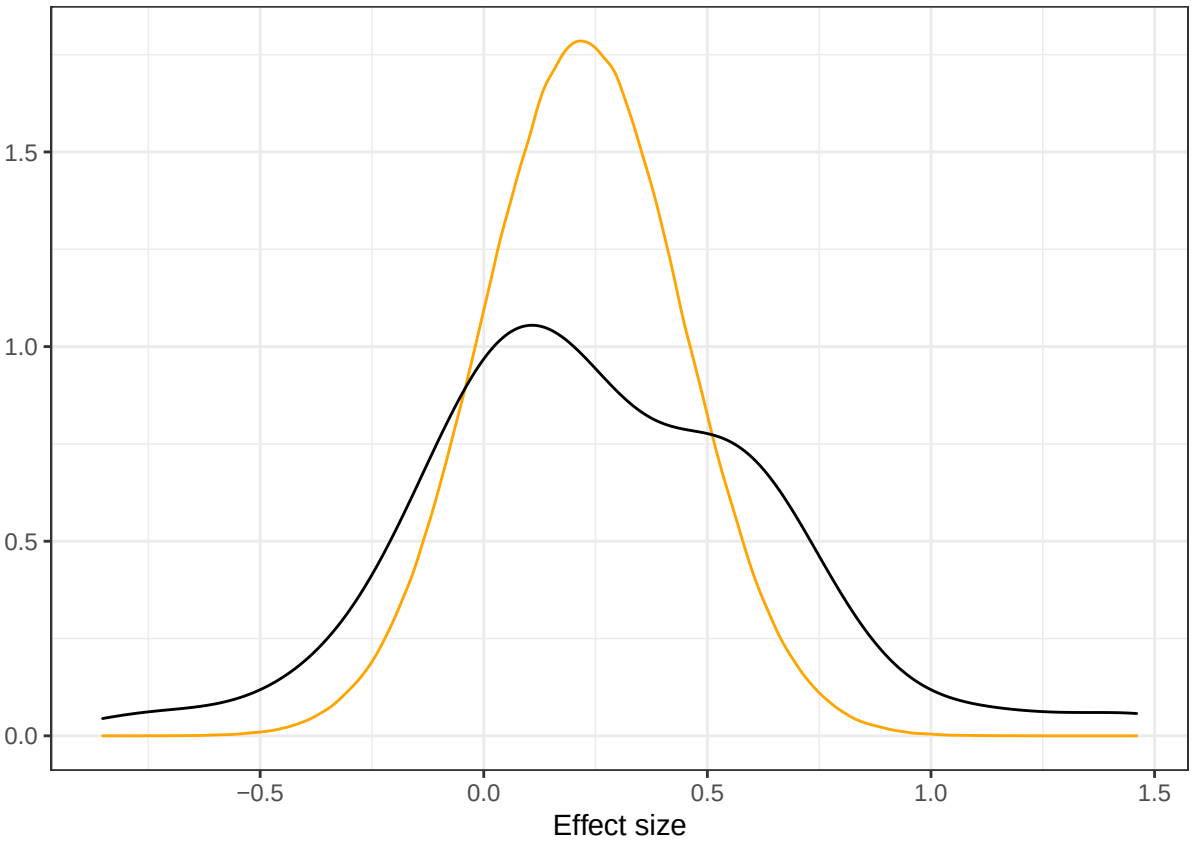
```
library(forestmodel)
forest_rma(m, study_labels = d$author)
```

[2] I don't think there is a way to get tasteful forest plots other than to custom-write them in `ggplot2`, but that's beyond the scope of this lab.



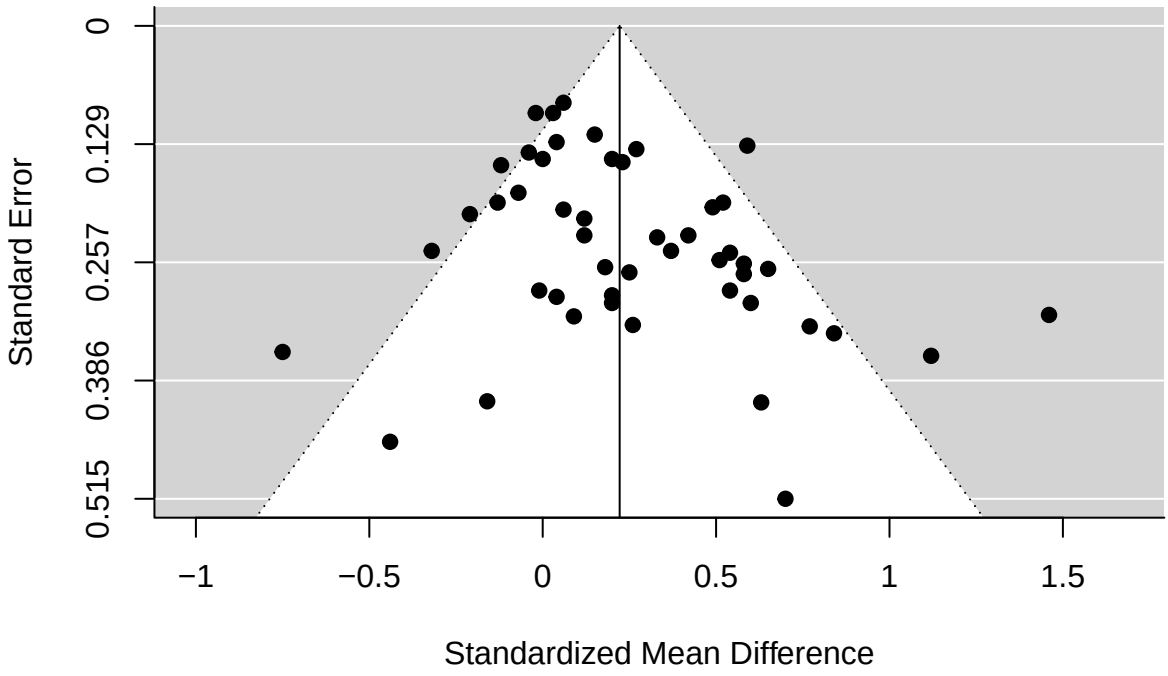
Let's visualize the estimated true effect distribution as well as the observed point estimates.

```
# generate "true effects" from estimated density
muhat = m$b
vhat = m$tau2
trues = rnorm( n=1000000, m = muhat, sd = sqrt(vhat) )

# compare observed density of point estimates to estimated density of true effects
library(ggplot2)
ggplot() +
  geom_density( aes(x=trues), color = "orange" ) +
  geom_density( aes(x=d$yi) ) +
  theme_bw() +
  xlab("Effect size") +
  ylab("")
```



Now let's assess for possible publication bias. Let's start with a traditional approach: the funnel plot and its corresponding hypothesis test and correction.

```
# funnel plot
funnel(m)
```



```r
# Egger's regression
regtest(m, model="lm", predictor="sei")
```

```
##
## Regression Test for Funnel Plot Asymmetry
##
## model:     weighted regression with multiplicative dispersion
## predictor: standard error
##
## test for funnel plot asymmetry: t = 2.7077, df = 46, p = 0.0095
```

The funnel plot is informative under a model of publication bias in which studies are selectively published when they have large effect sizes in a particular direction. If that's the mechanism, then asymmetry in the funnel plot would suggest publication bias. You can formally test this via Egger's regression-based test. (If you plan to conduct a meta-analysis in practice, note that these methods are no longer the best practice; see the Optional Supplement for details.)

Optional Supplement to Lab 12

Optional: Caveats about effect size conversions

If you actually conduct a meta-analysis, **be careful about the following, usually unstated, caveats**:

- The conversion from log-OR to Cohen's d only works well for a common binary outcome (Chinn 2000).
- The standard conversion from r to d assumes that the exposure variable, X, is binary. However, usually correlations are computed using continuous variables, not binary variables. For a continuous X, there is a different conversion (Mathur and VanderWeele, n.d.).

There are many more types of effect size conversions besides those in Borenstein et al. (2009). For a review of conversions for the t-family of statistics (e.g., ANOVA, regressions), see Lakens (2013). For conversions that are useful when meta-analyzing 2-group designs along with multiple regression results, see Aloe and Becker (2012).

Optional: Improved methods to assess publication bias

The funnel plot and Egger's test have important limitations, and you should interpret them with a grain of salt. Given the above assumption, these approaches do not perform well in literatures in which publication bias operates on the p-values rather than the effect sizes (which is probably the case more often than not).

What happens under a more realistic selection model? There are many different models along these lines, so I am showing you one that I believe uses reasonable assumptions (unlike the funnel plot and Egger's test) and has accessible software tools (Vevea and Hedges 1995). The model assumes that the true effects are normally distributed, that selection operates on the one-tailed p-values (i.e, that the literature prefers "significant" results in a particular direction[3]), and that the relative probability that a result is published is a step function of the p-value, such as:

$$P(\text{publish}) \propto \left\{ \begin{array}{ll} 1, & \text{for } p_{\text{1-tail}} < 0.025 \\ w_1, & \text{for } 0.025 \leq p_{\text{1-tail}} \end{array} \right\}$$

Technical point: I chose this specification because I'm guessing the literature selects for studies with *positive* effects (hence the one-sided p-values), but that the magical cutoff happens when effects become "significant" based on 2-sided tests reported in the papers (hence the 0.025 cutoff).

First we'll use R code:

```
library(weightr)
( wf = weightfunct( d$yi, d$vi, steps = c(0.025, 1) ) )
```

```
##
## Unadjusted Model (k = 48):
##
## tau^2 (estimated amount of total heterogeneity): 0.0471 (SE = 0.0220)
## tau (square root of estimated tau^2 value):  0.2169
##
## Test for Heterogeneity:
## Q(df = 47) = 107.1061, p-val = 2.117905e-06
##
## Model Results:
##
##           estimate std.error z-stat      p-val ci.lb  ci.ub
```

[3]If selection is actually two-tailed, this turns out not to matter too much especially if the true effect is fairly large. Other selection models have explicitly considered two-tailed selection criteria.

```
## Intercept   0.2207   0.04628  4.769 1.8518e-06  0.13 0.3114
##
## Adjusted Model (k = 48):
##
## tau^2 (estimated amount of total heterogeneity): 0.0276 (SE = 0.0238)
## tau (square root of estimated tau^2 value):  0.1660
##
## Test for Heterogeneity:
## Q(df = 47) = 107.1061, p-val = 2.117905e-06
##
## Model Results:
##
##               estimate std.error z-stat    p-val     ci.lb  ci.ub
## Intercept       0.1477   0.07311  2.020 0.043428  0.004358 0.2909
## 0.025 < p < 1   0.4664   0.34924  1.336 0.181690 -0.218065 1.1509
##
## Likelihood Ratio Test:
## X^2(df = 1) = 1.163544, p-val = 0.28073
```

```
# sanity check: reproduce fn's p-values
#p1 = 1 - pnorm( d$yi / sqrt(d$vi) ); wf$p
# how many positive and 1-directionally significant p-vals do we have?
#table( p1 < 0.025 & d$yi > 0 )
```

In this literature, compared to studies with "significant" one-directional results, "nonsignificant" results $p_{\text{1-tail}} > 0.05$ are about 47% as likely to be published. However, a likelihood-ratio test comparing the model allowing for publication bias to the model not allowing publication bias does not suggest strong publication bias ($p = 0.28$). (There is a lot of uncertainty, though: the SEs are huge here.)

(These analyses can also be conducted with Vevea & Hedges' very sleek Shiny app: https://vevealab.shinyapps.io/WeightFunctionModel/.)

Optional: Choose your own R adventure

The effect sizes in the writing-interventions example looked a bit bimodal. Let's examine sensitivity of parametric meta-analysis mean estimation to two violations of the normality assumption.

Level 3: Black belt

1. Simulate true effect sizes that are bimodal (scenario 1) or right-skewed (scenario 2) with a known mean and variance of your choosing.
2. Simulate point estimates from the true effect sizes with SEs of your choosing.
3. Meta-analyze the simulated point estimates and save the pooled point estimate.
4. Repeat for a number of simulation reps and report the average bias in the pooled point estimate.

Level 2: Purple belt

Now I'll give you more instruction about how to accomplish the above. I am also rearranging the instructions to be more like "pseudo-code", i.e., you can roughly translate the prose into code line-by-line.

1. Define variables setting a true mean and variance.
2. Set up a data structure to store your results.
3. For each of a large-ish number of simulation reps:

(a) For the bimodal scenario, generate true effect sizes from a mixture distribution (e.g., there is a 50% probability that the true ES comes from normal distribution #1 and a 50% probability that it comes from normal distribution #2, where the overall mean of the 2 distributions is what you set above).

(b) Simulate SEs from, for example, a uniform distribution.

(c) Simulate point estimates using the true effect sizes and the SEs. (Hint: point estimates are generally normal around their true effect sizes.)

(d) Meta-analyze the point estimates and record the pooled point estimate.

(e) Do something similar for the skewed scenario (e.g., generating true effect sizes from an exponential distribution).

4. For each scenario, calculate the average bias in the pooled point estimate.

Level 1: Green belt

Here I've also given you some code from the solution, but with gaps that you'll need to fill in.

```
# known mean and variance
M = 1
V = .5^2

# number of simulation reps
sim.reps =

# initialize vectors that will hold the results for each of 2 scenarios
res1 =
res2 =

# number of studies in meta-analysis
# interesting to vary this
k = 10

# for each simulation rep, simulate for both scenarios
for ( i in 1:sim.reps ) {

  ######## Scenario 1: Bimodal ########
  # simulate true effect sizes bimodally
  # by using a mixture of normals
  # first draw indicator for which distribution each study comes from
  ind =

  # draw true effects depending on the above indicator
  true = rep( NA, k )
  true[ ind == 1 ] =
  true[ ind == 0 ] =

  # simulate SEs from uniform with basically arbitrary endpoints
  SE =

  # simulate point estimates
  # these will be normal around their own true parameters
  ests =
```

```
  # meta-analyze
  m =

  # put the pooled estimate in the vector of results
  res1[i] =


  ######## Scenario 2: Exponential ########
  # simulate true effect sizes skewishly
  # use same SEs as before (reduces simulation error)

  # try to write this part yourself
  # it will look very similar to the above
}

# assess average bias
# write this part yourself
```

One possible solution

This isn't the most generalizable or the most computationally efficient solution, but I want to show you how to write the simulation as simply as possible.

```
# known mean and variance
M = 1
V = .5^2
# note that when we simulate from bimodal normal, this will be variance
#  WITHIN each group; doesn't matter because we're only looking at mean estimation

# number of simulation reps
sim.reps = 1000

# initialize vectors that will hold the results for each of 2 scenarios
res1 = rep( NA, sim.reps )
res2 = rep( NA, sim.reps )

# number of studies in meta-analysis
# could try varying this
k = 10

# for each simulation rep, simulate for both scenarios
for ( i in 1:sim.reps ) {

  ######## Scenario 1: Bimodal ########
  # simulate true effect sizes bimodally
  # by using a mixture of normals
  # first draw indicator for which distribution each study comes from
  ind = rbinom( n=k, size=1, prob=0.4 )
  true = rep( NA, k )

  if ( any(ind == 1) ) {
      true[ ind == 1 ] = rnorm( n = length( true[ ind == 1 ] ),
                                mean = M - 0.5,
                                sd = sqrt(V) )
```

```
  }

  if ( any(ind == 0) ) {
      true[ ind == 0 ] = rnorm( n = length( true[ ind == 0 ] ),
                                mean = M + 0.5,
                                sd = sqrt(V) )
  }

  # simulate SEs from uniform with basically arbitrary endpoints
  SE = runif( n=k, sqrt(V)/2, sqrt(V)*1.1 )

  # simulate point estimates
  # these will be normal around their own true parameters
  ests = rnorm( n = k, mean = true, sd = SE)

  # meta-analyze
  m = rma.uni( yi = ests,
               vi = SE^2,
               measure = "SMD",
               knha = TRUE )

  # put the pooled estimate in the vector of results
  res1[i] = m$b


  ######## Scenario 1: Exponential ########
  # simulate true effect sizes skewishly
  # use same SEs as before (reduces simulation error)
  # generate true effect sizes
  true = rexp( n = k )

  # simulate point estimates as before
  ests = rnorm( n = k, mean = true, sd = SE)

  # meta-analyze
  m = rma.uni( yi = ests,
               vi = SE^2,
               measure = "SMD",
               knha = TRUE )

  # put the pooled estimate in the vector of results
  res2[i] = m$b
}

# assess average bias (absolute)
mean(res1) - M
mean(res2) - M

# and relative bias (percent)
100 * abs( mean(res1) - M ) / M
100 * abs( mean(res2) - M ) / M
```

References

Aloe, Ariel M, and Betsy Jane Becker. 2012. "An Effect Size for Regression Predictors in Meta-Analysis." *Journal of Educational and Behavioral Statistics* 37 (2): 278–97.

Borenstein, M, LV Hedges, J Higgins, and HR Rothstein. 2009. *Introduction to Meta-Analysis*. Wiley Online Library.

Chinn, Susan. 2000. "A Simple Method for Converting an Odds Ratio to Effect Size for Use in Meta-Analysis." *Statistics in Medicine* 19 (22): 3127–31.

IntHout, Joanna, John PA Ioannidis, and George F Borm. 2014. "The Hartung-Knapp-Sidik-Jonkman Method for Random Effects Meta-Analysis, Is Straightforward, and Considerably Outperforms the Standard Dersimonian-Laird Method." *BMC Medical Research Methodology* 14 (1): 25.

Lakens, Daniël. 2013. "Calculating and Reporting Effect Sizes to Facilitate Cumulative Science: A Practical Primer for T-Tests and Anovas." *Frontiers in Psychology* 4: 863.

Mathur, MB, and TJ VanderWeele. n.d. "A Simple, Interpretable Conversion from Pearson's Correlation to Cohen's d for Continuous Exposures." *Epidemiology*.

Veroniki, AA, D Jackson, W Viechtbauer, R Bender, J Bowden, G Knapp, O Kuss, J Higgins, D Langan, and G Salanti. 2016. "Methods to Estimate the Between-Study Variance and Its Uncertainty in Meta-Analysis." *Research Synthesis Methods* 7 (1): 55–79.

Vevea, Jack L, and Larry V Hedges. 1995. "A General Linear Model for Estimating Effect Size in the Presence of Publication Bias." *Psychometrika* 60 (3): 419–35.