The Common Mean Problem and Inference in Random-Effects Meta-Analysis Model with Normal Outcomes

Guido Knapp

Department of Statistics, TU Dortmund University, Dortmund, Germany

Symposium
"Recent Advances in Meta-Analysis Methods and Software"
Göttingen, Germany, 23 – 24 August 2023

Contents

Common Mean Problem

- Let us consider k independent normal populations where the ith population follows a normal distribution with mean $\mu \in \mathbb{R}$ and variance $\sigma_i^2 > 0$, $i = 1, \ldots, k$.
- Let $\bar{Y}_i$ denote the sample mean in the ith population, S_i^2 the sample variance, and n_i the sample size, $i = 1, \ldots, k$.
- Then, we have

$$\bar{Y}_i \sim N\left(\mu \, , \, \frac{\sigma_i^2}{n_i}\right) \quad \text{and} \quad \frac{(n_i - 1)\, S_i^2}{\sigma_i^2} \sim \chi^2_{n_i - 1}, \; i = 1, \ldots, k,$$

and the statistics are all mutually independent.
Note that $(\bar{Y}_i, S_i^2, i = 1, \ldots, k)$ is minimal sufficient for $(\mu, \sigma_1^2, \ldots, \sigma_k^2)$ even though it is not complete.

Estimates of μ

- If the population variances $\sigma_1^2, \ldots, \sigma_k^2$ are completely known, the maximum likelihood estimator of μ is given by

$$\hat{\mu} = \frac{\sum_{i=1}^{k} \frac{n_i}{\sigma_i^2} \; \bar{Y}_i}{\sum_{j=1}^{k} \frac{n_j}{\sigma_j^2}}.$$

- The above estimator is also the minimum variance unbiased estimator under normality as well as the best linear unbiased estimator without normality for estimating μ.
- The variance of $\hat{\mu}$ is given by $\quad \mathrm{Var}\,(\hat{\mu}) = \dfrac{1}{\sum_{i=1}^{k} \frac{n_i}{\sigma_i^2}}.$

Estimates of μ

- Graybill-Deal (1959) estimator of μ is given as

$$\hat{\mu}_{GD} = \frac{\sum_{i=1}^{k} \frac{n_i}{S_i^2} \ \bar{Y}_i}{\sum_{j=1}^{k} \frac{n_j}{S_j^2}}.$$

  Clearly, $\hat{\mu}_{GD}$ is an unbiased estimator of the common mean μ.

- For calculating the variance of $\hat{\mu}_{GD}$, it holds

$$\begin{aligned} \mathrm{Var}\left(\hat{\mu}_{GD}\right) &= \mathrm{E}\left[\mathrm{Var}\left(\hat{\mu}_{GD}|S_1,\ldots,S_k\right)\right] + \mathrm{Var}\left[\mathrm{E}\left(\hat{\mu}_{GD}|S_1,\ldots,S_k\right)\right] \\ &= \mathrm{E}\left[\left(\sum_{i=1}^{k} \frac{n_i \ \sigma_i^2}{S_i^4}\right) \Big/ \left(\sum_{i=1}^{k} \frac{n_i}{S_i^2}\right)^2\right]. \end{aligned}$$

Estimates of μ

Meier (1953) derived a first order approximation of the variance of $\hat{\mu}_{GD}$ as

$$\mathrm{Var}\left(\hat{\mu}_{GD}\right) = \frac{1}{\sum_{i=1}^{k} \frac{n_i}{\sigma_i^2}} \left[1 + 2 \sum_{i=1}^{k} \frac{1}{n_i - 1} \; c_i \; (1 - c_i) + O\left(\sum_{i=1}^{k} \frac{1}{(n_i - 1)^2} \right) \right]$$

with

$$c_i = \frac{n_i \;/\; \sigma_i^2}{\sum_{j=1}^{k} n_j \;/\; \sigma_j^2}, \;\; i = 1, \ldots, k.$$

Variance Estimates

Sinha (1985) derived an unbiased estimator of the variance of $\hat{\mu}_{GD}$ that is a convergent series. A first order approximation of this estimator is

$$\widehat{\mathrm{Var}}_{(1)}\left(\hat{\mu}_{GD}\right) = \frac{1}{\sum_{i=1}^{k} \frac{n_i}{S_i^2}} \left[1 + \sum_{i=1}^{k} \frac{4}{n_i + 1} \left(\frac{n_i \; / \; S_i^2}{\sum_{j=1}^{k} n_j \; / \; S_j^2} - \frac{n_i^2 \; / \; S_i^4}{\left(\sum_{j=1}^{k} n_j \; / \; S_j^2\right)^2} \right) \right].$$

This estimator is comparable to Meier's (1953) approximate estimator:

$$\widehat{\mathrm{Var}}_{(2)}\left(\hat{\mu}_{GD}\right) = \frac{1}{\sum_{i=1}^{k} \frac{n_i}{S_i^2}} \left[1 + \sum_{i=1}^{k} \frac{4}{n_i - 1} \left(\frac{n_i \; / \; S_i^2}{\sum_{j=1}^{k} n_j \; / \; S_j^2} - \frac{n_i^2 \; / \; S_i^4}{\left(\sum_{j=1}^{k} n_j \; / \; S_j^2\right)^2} \right) \right].$$

Variance Estimates

Two further estimates

- The "classical" meta-analysis variance estimator

$$\widehat{\mathrm{Var}}_{(3)}\left(\hat{\mu}_{GD}\right) = \frac{1}{\sum_{i=1}^{k} \frac{n_i}{S_i^2}}.$$

- Hartung (1999): approximate variance estimator

$$\widehat{\mathrm{Var}}_{(4)}\left(\hat{\mu}_{GD}\right) = \frac{1}{\sum_{i=1}^{k} \frac{n_i}{S_i^2}} \left[\frac{1}{k-1} \sum_{i=1}^{k} \frac{n_i}{S_i^2} \left(\bar{Y}_i - \hat{\mu}_{GD}\right)^2 \right].$$

Random-Effects Model

- Let us consider k independent normal populations where the ith population follows a normal distribution with mean $\mu_i \in \mathbb{R}$ and variance $\sigma_i^2 > 0$, $i = 1, \ldots, k$.
- For the expected values, we assume

$$\mu_i \sim N(\mu, \tau^2), i = 1, \ldots, k.$$

  Here, μ is the grand mean and τ^2 the between-population variance (heterogeneity parameter).
- Let $\bar{Y}_i$ denote the sample mean in the ith population, S_i^2 the sample variance, and n_i the sample size, $i = 1, \ldots, k$.
- Then, we obtain the random-effects model

$$\bar{Y}_i \sim N\left(\mu\ ,\ \tau^2 + \frac{\sigma_i^2}{n_i}\right),\ i = 1, \ldots, k.$$

Random-Effects Model

- Practically in meta-analysis, we work with

$$\bar{Y}_i \sim N\left(\mu\ ,\ \tau^2 + \frac{S_i^2}{n_i}\right),\ i = 1, \ldots, k.$$

- Parameter space of (μ, τ^2):

$$\Theta = \mathbb{R} \times [0, \infty) \quad \text{or} \quad \tilde{\Theta} = \mathbb{R} \times (0, \infty)$$

  Do we allow $\tau^2 = 0$ or not?

- Note: The variances S_i^2/n_i determine the order of the weights of the populations in the meta-analysis, in other words, the order of the importance of the populations.

Random-Effects Model

- In simulation studies, however, for studying properties of statistical methods, the parameter space of the data-generating model is for $(\mu, \tau^2, \sigma_1^2, \ldots, \sigma_k^2)$

$$\Theta^* = \mathbb{R} \times [0, \infty) \times (0, \infty)^k$$

using $\dfrac{(n_i - 1)\ S_i^2}{\sigma_i^2} \sim \chi^2_{n_i - 1}$, for generating S_i^2, $i = 1, \ldots, k$.

- Consequently, the order of the importance of the populations may change from simulation run to simulation run.

Heterogeneity estimates

- How important is the estimation of the heterogeneity parameter?
- What is the criterion for a *good* estimator?
  To describe well the underlying heterogeneity
  OR
  to produce a *good* statistical analysis about the grand mean?
- Confidence intervals for the heterogeneity parameter exist. If we accept the values in the intervals as feasible values for the heterogeneity, what will be their impact on the statistical analysis about the grand mean?

Weights in RE Model

Example: Results of eight randomized controlled trials comparing the effectiveness of amlodipine and a placebo on work capacity
(here only the results of the control group, Li et al. (1994))

Protocol	Placebo (C)		
	n_{Ci}	$\bar{y}_{Ci}$	s^2_{Ci}
154	48	-0.0027	0.0007
156	26	0.0270	0.1139
157	72	0.0443	0.4972
162A	12	0.2277	0.0488
163	34	0.0056	0.0955
166	31	0.0943	0.1734
303A	27	-0.0057	0.9891
306	47	-0.0057	0.1291

What would be a good guess for a common mean?

Weights in RE Model

- Weights depend on the empirical variances and the sample sizes.
- Not only large sample size may lead to a larger precision of the study-specific estimate but also a small empirical variance.
- Should a result from a rather homogeneous study population (small variance) really be the most important result in meta-analysis?
- Hartung et al. (2003): Methods forcombining results of experiments with general weights including inverse variances in the random-effects model.
- Simulation study: Choosing the random-effects model as data-generating model, which Type I error rate is then acceptable for using general weights?

Pre-Test Estimation

- If you accept $\tau^2 \in [0, \infty)$, you will allow that the random-effects model may reduce to the common effect model for $\tau^2 = 0$.
- If the heterogeneity estimate is zero, the meta-analysis is done in the common effect model.
- Should we use approximate confidence in the common mean problem? No!
- In the common mean problem, exact confidence intervals for μ exist.

Pre-Test Estimation

- Since

$$t_i = \frac{\sqrt{n_i}\ \left(\bar{Y}_i - \mu\right)}{S_i} \sim t_{n_i - 1}$$

or, equivalently,

$$F_i = \frac{n_i\ \left(\bar{Y}_i - \mu\right)^2}{S_i^2} \sim F_{1, n_i - 1}$$

are test statistics for testing hypotheses about μ based on the ith sample, suitable linear combinations of these test statistics or other functions thereof can be used as a pivotal quantity to construct exact confidence intervals for μ.

Pre-Test Estimation

Fairweather (1972): weighted linear combination of the t_i's, namely

$$W_t = \sum_{i=1}^{k} u_i \; t_i, \quad u_i = \frac{[\mathrm{Var}(t_i)]^{-1}}{\sum_{j=1}^{k} [\mathrm{Var}(t_i)]^{-1}}, i = 1, \ldots, k.$$

Let $b_{\alpha/2}$ denote the upper critical value of the distribution of W_t satisfying the equation $1 - \alpha = \mathrm{P}\left(|W_t| \leq b_{\alpha/2}\right)$, then the exact $100(1-\alpha)\%$ confidence interval for μ is given by

$$CI_{(7)}(\mu) : \frac{\sum_{i=1}^{k} \sqrt{n_i} \; u_i \; \bar{Y}_i \; / \; S_i}{\sum_{i=1}^{k} \sqrt{n_i} \; u_i \; / \; S_i} \mp \frac{b_{\alpha/2}}{\sum_{i=1}^{k} \sqrt{n_i} \; u_i \; / \; S_i}.$$

Pre-Test Estimation

- At least seven exact intervals exist. Which one should we use?
- All the intervals except Fairweather's and Hartung and Knapp (2005) interval do not necessarily produce a genuine interval.
- For some intervals, sufficient conditions exist to produce a genuine interval.
- What is a good strategy to reduce the random-effects model to the common mean problem? Use of an estimator of τ^2 or a hypothesis test for $H_0 : \tau^2 = 0$?

Example

Example: Results of eight randomized controlled trials comparing the effectiveness of amlodipine and a placebo on work capacity
(here only the results of the control group, Li et al. (1994))

	Placebo (C)		
Protocol	n_{Ci}	$\bar{y}_{Ci}$	s^2_{Ci}
154	48	-0.0027	0.0007
156	26	0.0270	0.1139
157	72	0.0443	0.4972
162A	12	0.2277	0.0488
163	34	0.0056	0.0955
166	31	0.0943	0.1734
303A	27	-0.0057	0.9891
306	47	-0.0057	0.1291

What would be a good guess for a common mean?

Analysis in the example (R package meta)

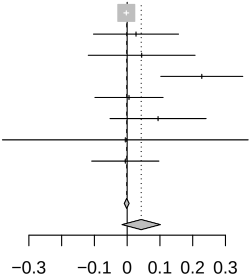
		95%-CI	%W(common)	%W(random)
1	-0.0027	[-0.0102; 0.0048]	97.8	26.5
2	0.0270	[-0.1027; 0.1567]	0.3	11.5
3	0.0443	[-0.1186; 0.2072]	0.2	8.7
4	0.2277	[0.1027; 0.3527]	0.4	12.0
5	0.0056	[-0.0983; 0.1095]	0.5	14.5
6	0.0943	[-0.0523; 0.2409]	0.3	10.0
7	-0.0057	[-0.3808; 0.3694]	0.0	2.2
8	-0.0057	[-0.1084; 0.0970]	0.5	14.6

		95%-CI	z	p-value
Common effect model	-0.0014	[-0.0088; 0.0060]	-0.38	0.7058
Random effects model	0.0428	[-0.0156; 0.1013]	1.44	0.1510

Quantifying heterogeneity:
tau2 = 0.0033 [0.0000; 0.0220]; tau = 0.0578 [0.0000; 0.1482]

Analysis in the example (R package meta)

Study	TE	SE(TE)	95%−CI	Weight (common)	Weight (random)
1	−0.0027	0.0038	−0.00 [−0.01; 0.00]	97.8%	26.5%
2	0.0270	0.0662	0.03 [−0.10; 0.16]	0.3%	11.5%
3	0.0443	0.0831	0.04 [−0.12; 0.21]	0.2%	8.7%
4	0.2277	0.0638	0.23 [0.10; 0.35]	0.4%	12.0%
5	0.0056	0.0530	0.01 [−0.10; 0.11]	0.5%	14.5%
6	0.0943	0.0748	0.09 [−0.05; 0.24]	0.3%	10.0%
7	−0.0057	0.1914	−0.01 [−0.38; 0.37]	0.0%	2.2%
8	−0.0057	0.0524	−0.01 [−0.11; 0.10]	0.5%	14.6%
Common effect model			**−0.00 [−0.01; 0.01]**	**100.0%**	**--**
Random effects model			**0.04 [−0.02; 0.10]**	**--**	**100.0%**



Heterogeneity: $I^2 = 54\%$, $\tau^2 = 0.0033$, $p = 0.03$

Analysis in the example (R package meta)

Subgroup analysis (delete study 4)

		95%-CI	%W(common)	%W(random)
1	-0.0027	[-0.0102; 0.0048]	98.1	98.1
2	0.0270	[-0.1027; 0.1567]	0.3	0.3
3	0.0443	[-0.1186; 0.2072]	0.2	0.2
4	0.0056	[-0.0983; 0.1095]	0.5	0.5
5	0.0943	[-0.0523; 0.2409]	0.3	0.3
6	-0.0057	[-0.3808; 0.3694]	0.0	0.0
7	-0.0057	[-0.1084; 0.0970]	0.5	0.5

		95%-CI	z	p-value
Common effect model	-0.0022	[-0.0096; 0.0052]	-0.59	0.5552
Random effects model	-0.0022	[-0.0096; 0.0052]	-0.59	0.5552

Is the model $\mu_i \sim N(\mu, \tau^2), i = 1, \ldots, k.$ justified in the first analysis?

Analysis in the example (R package meta)

We accept the random-effects model for the eight studies.
Which estimate of τ^2 should we use?

Method	$\hat{\tau}$	$\hat{\mu}$ and 95%-CI on μ	p-value
REML	0.0578	0.0428 [-0.0156; 0.1013]	0.1510
ML	0	-0.0014 [-0.0088; 0.0060]	0.7058
EB	0.0512	0.0408 [-0.0137; 0.0953]	0.1420
DL	0.0518	0.0410 [-0.0138; 0.0958]	0.1426
HE	0	-0.0014 [-0.0088; 0.0060]	0.7058
SJ	0.0611	0.0437 [-0.0168; 0.1041]	0.1569
HS	0.0101	0.0060 [-0.0136; 0.0255]	0.5485

REML=Restricted Maximum Likelihood, ML = Maximum Likelihood, EB = Empirical Bayes,

DL =DerSimonian-Laird, HE= Hedges, SJ= Sidik-Jonkman, HS = Hunter-Schmid

Specific Remarks

- Generalized confidence intervals are a viable alternative to frequentist and Bayesian approaches for the meta-analysis of normal means or difference of normal means; distributions of the statistics with respect to the nuisance parameters are included and no prior distributions for the parameters are needed.
- One approach was implemented in the R package metagen (Not on CRAN in the moment).
- Exact confidence intervals in the common mean problem can be easily extended the meta-analysis for a common difference of normal means.
- Even for a common standardized difference of means, an exact confidence interval can be determined (Knapp, 2017).

General Remarks

- Meta-analysis is retrospective data analysis.
- Each new meta-analysis is a new challenge in data analysis.
- A *best* statistical method for meta-analysis does not exist.
- Performance of the statistical methods should be only discussed with respect to a effect size of interest, not generally.
- Use several available statistical methods for meta-analysis to make decisions.

References

Fairweather, W.R. (1972). A method of obtaining an exact confidence interval for the common mean of several normal populations. *Applied Statistics*, 21, 229–233.

Graybill, F.A. and Deal, R.B. (1959). Combining unbiased estimators. *Biometrics*, 15, 543–550.

Hartung, J. (1999). An alternative method for meta-analysis. *Biometrical Journal*, 41, 901–916.

Hartung, J., Knapp, G. (2005). Models for combining results of different experiments: retrospective and prospective. *American Journal of Mathematical and Management Sciences, 25, 149–188.*

Knapp, G. (2017). An exact confidence interval for a common effect size. Journal of Statistical Theory and Practice, DOI: 10.1080/15598608.2016.1278060.

Li, Y., Shi, L., Roth, H.D. (1994). The bias of the commonly used estimate of variance in meta-analysis. Communications in Statistics – Theory and Methods, 23, 1063–1085.

Meier, P. (1953). Variance of a weighted mean. Biometrics, 9, 59–73.

Sinha, B.K. (1985). Unbiased estimation of the variance of the Graybill-Deal estimator of the common mean of several normal distributions. Canadian Journal of Statistics, 13, 243–247.

Weerahandi, S. (1993). Generalized confidence intervals. Journal of the American Statistical Association, 88, 899–905.