Some Aspects of Combining Mean Differences of a few Studies

Guido Knapp

Department of Statistics, TU Dortmund University, Dortmund, Germany

Symposium
"Recent Advances in Meta-Analysis"
Dortmund, Germany, 30 June – 1 July 2025

Contents

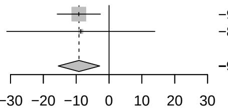
Motivation

- Walter and Balakrishnan (2022). A method was developed for correcting the bias in the usual study weights in meta-analyses. Journal of Clinical Epidemiology.
- Effect size: Common difference of (normal) means
- Examples from Cochrane Database with a few studies and small sample sizes.

Example 1

Ker et al. (2010). Caffeine for the prevention of injuries and errors in shift workers.

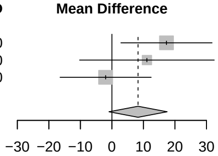
Study	Experimental Total	Experimental Mean	Experimental SD	Control Total	Control Mean	Control SD	Mean Difference	MD	95%−CI	Weight
1	4	−51.00	6.0000	4	−41.80	3.0000		−9.20	[−15.77; −2.63]	92.2%
2	6	−79.90	18.9000	6	−71.30	20.9000		−8.60	[−31.15; 13.95]	7.8%
Common effect model	**10**			**10**				**−9.15**	**[−15.46; −2.84]**	**100.0%**



Example 2

Rietberg et al. (2017). Respiratory muscle training for multiple sclerosis.

Study	Experimental Total	Experimental Mean	Experimental SD	Control Total	Control Mean	Control SD	MD	95%−CI	Weight
1	9	40.30	21.0000	9	23.00	7.0000	17.30	[2.84; 31.76]	40.6%
2	10	73.30	19.8000	5	62.20	20.1000	11.10	[−10.26; 32.46]	18.6%
3	23	98.00	28.0000	25	100.00	23.0000	−2.00	[−16.45; 12.45]	40.7%
Common effect model	**42**			**39**			**8.29**	**[−0.93; 17.51]**	**100.0%**



Model

- Let us consider k comparative trials, treatment (T) vs. control (C).
- Let be $\bar{Y}_{Ti}$, S_{Ti}^2 and $\bar{Y}_{Ci}$, S_{Ci}^2 sample mean and sample variance in the corresponding groups and n_{Ti} and n_{Ci} the sample sizes.
- $$\bar{Y}_{Ti} \sim \mathcal{N}\left(\mu_{Ti}, \frac{\sigma_i^2}{n_{Ti}}\right) \tag{1}$$

  and

  $$\bar{Y}_{Ci} \sim \mathcal{N}\left(\mu_{Ci}, \frac{\sigma_i^2}{n_{Ci}}\right). \tag{2}$$

  Note the population variances are equal within the trial.

Model

- Let S_i^2 be estimated by the pooled sample variance, namely

$$S_i^2 = \frac{(n_{Ti} - 1)\, S_{Ti}^2 + (n_{Ci} - 1)\, S_{Ci}^2}{n_{Ti} + n_{Ci} - 2}. \tag{3}$$

- Let be the parameter of interest in each trial

$$\mu_{Di} = \mu_{Ti} - \mu_{Ci}$$

and the unbiased estimator is

$$D_i = \bar{Y}_{Ti} - \bar{Y}_{Ci}.$$

- Under the homogeneity assumption, that is,

$$H_0 : \mu_{D1} = \mu_{D2} = \cdots = \mu_{Dk} =: \mu_D, \tag{4}$$

we have the scenario of the well-known common mean problem.

Model

- Common effect model

$$D_i \sim \mathcal{N}\left(\mu_D, \frac{n_{Ti} + n_{Ci}}{n_{Ti} n_{Ci}}\ \sigma_i^2\right) \tag{5}$$

with estimated variance

$$\widehat{\text{var}}(D_i) = \frac{n_{Ti} + n_{Ci}}{n_{Ti} n_{Ci}}\ S_i^2 =: \hat{w}_i^{-1} \tag{6}$$

- $\{D_i, S_i^2\}_{i=1,\dots,k}$ are minimal sufficient, but not complete.
- Consequently, UMVU estimator of μ_D does not exist.

Model

- The inverse-variance weighted estimator of μ_D is

$$\hat{\mu}_D = \frac{\sum_{i=1}^k \hat{w}_i D_i}{\sum_{i=1}^k \hat{w}_i}. \tag{7}$$

- The estimator is unbiased, because (law of iterated expectation)

$$\mathrm{E}(\mu_D) = \mathrm{E}\left[\mathrm{E}(\hat{\mu}_D | S_1^2, \ldots, S_k^2)\right] = \mu_D \; \mathrm{E}\left[\frac{\sum_{i=1}^k \hat{w}_i}{\sum_{i=1}^k \hat{w}_i}\right] = \mu_D$$

- Unbiased estimator of the variance (first-order approximation, HKS (2008, p. 100)))

$$\widehat{var}(\hat{\mu}_D) = \frac{1}{\sum_i \hat{w}_i}\left[1 + \sum_i \frac{4}{n_{Ti} + n_{Ci}}\left(\frac{\hat{w}_i}{\sum_i \hat{w}_i} - \frac{\hat{w}_i^2}{(\sum_i \hat{w}_i)^2}\right)\right]$$

Model

- The exact confidence intervals are based on test statistics for hypothesis testing about the common effect size μ_D in the i-th study.
- Note that, under H_0 from (4), it holds

$$t_i = \sqrt{\frac{n_{Ti}\; n_{Ci}}{n_{Ti} + n_{Ci}}}\, \frac{D_i - \mu_D}{S_i} \sim t_{n_{Ti}+n_{Ci}-2} \tag{8}$$

or, equivalently,

-
$$F_i = \frac{n_{Ti}\; n_{Ci}}{n_{Ti} + n_{Ci}}\, \frac{(D_i - \mu_D)^2}{S_i^2} \sim F_{1,n_{Ti}+n_{Ci}-2}. \tag{9}$$

Methods for building intervals

Two procedures:

- Intervals based on linear combinations
- Intervals based on probability integral transformations

Interval based on linear combinations

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

$$W_t = \sum_{i=1}^{k} v_i \; t_i , \quad v_i = \frac{[\mathrm{Var}(t_i)]^{-1}}{\sum_{j=1}^{k} [\mathrm{Var}(t_j)]^{-1}} , \quad i = 1, \ldots, k. \tag{10}$$

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

$$\mathrm{CI}_F : \frac{\sum\limits_{i=1}^{k} \sqrt{\tilde{n}_i} \; v_i \; D_i \; / \; S_i}{\sum\limits_{j=1}^{k} \sqrt{\tilde{n}_j} \; v_j \; / \; S_j} \mp \frac{b_{1-\alpha/2}}{\sum\limits_{j=1}^{k} \sqrt{\tilde{n}_j} \; v_j \; / \; S_j} \quad , \tilde{n}_i := \frac{n_{Ti} \; n_{Ci}}{n_{Ti} + n_{Ci}} , i = 1, \ldots, k. \tag{11}$$

Interval based on linear combinations

- Fairweather provided an approximation of $b_{1-\alpha/2}$.
- Fairweather's interval is always a genuine interval for $0 < \alpha < 0.5$.
- Center of the interval

$$\hat{\mu}_{D,F} = \frac{\sum_{i=1}^{k} \sqrt{\tilde{n}_i}\ v_i\ D_i\ /\ S_i}{\sum_{j=1}^{k} \sqrt{\tilde{n}_j}\ v_j\ /\ S_j}$$

is another unbiased estimator of μ_D.

Intervals based on probability integral transformations

- HK (2005) use t test statistics t_i.
- Let $F_{t_{n_{Ti}+n_{Ci}-2}}$ be the cdf of a $t_{n_{Ti}+n_{Ci}-2}$-variate.
- It holds

$$F_{t_{n_{Ti}+n_{Ci}-2}}(t_i) =: u_i \sim U(0,1) \quad \text{and} \quad \Phi^{-1}(u_i) \sim N(0,1),$$

  where $U(0,1)$ stands for the uniform distribution on the unit interval.
- Let us consider the inverse normal combination statistic

$$Z(\mu_D) = \sum_{i=1}^{k} \Phi^{-1}\left(F_{t_{n_{Ti}+n_{Ci}-2}}(t_i)\right) / \sqrt{k}. \qquad (12)$$

  Clearly, $Z(\mu_D)$ is a standard normal random variable.

Intervals based on probability integral transformations

- Since the functions $F_{t_{n_{Ti}+n_{Ci}-2}}(\cdot)$ and $\Phi^{-1}(\cdot)$ are monotone increasing functions in their arguments $(\cdot)$, $Z(\mu_D)$ is a monotone decreasing function in μ_D.
- Consequently, an exact $100(1-\alpha)\%$ confidence interval for μ_D is given by

$$\mathrm{CI}_{IN}: \quad [\ \mu_{D,L,Z}\ ;\ \mu_{D,U,Z}\ ], \tag{13}$$

where the bounds $\mu_{D,L,Z}$ and $\mu_{D,U,Z}$ are the unique solutions for μ_D of the equations

$$Z(\mu_D) = \Phi^{-1}(1-\alpha/2) \quad \text{and} \quad Z(\mu_D) = \Phi^{-1}(\alpha/2).$$

Intervals based on probability integral transformations

- Solving $Z(\mu_D) = 0$ leads to a point estimate of μ_D.
- Given fixed weights, say f_i, for each trial, a weighted combination function can also be used, namely

$$\tilde{Z}(\mu_D) = \sum_{i=1}^{k} \sqrt{f_i}\ \Phi^{-1}\left(F_{t_{n_{Ti}+n_{Ci}-2}}(t_i)\right) / \sum_{i=1}^{k} \sqrt{f_i}.$$

Example 1

Caffeine for the prevention of injuries and errors in shift workers
Effect size: Mean differences (MD)

Method	Estimate MD	95%-CI for MD
Standard MA	-9.1530	[-15.4641 – -2.8419]
Fairweather approx. critical value	-9.0445	[-18.0394 – -0.0540]
Fairweather simulated crit. value	-9.0445	[-17.9976 – -0.1152]
Inverse Normal	-9.0630	[-17.7072 – -0.2981]

Simulation using setting in example

Put $\mu_D = 0$ and $\sigma_i^2 = 1$, $i = 1, 2$, m= 100.000 simulation runs
Estimated actual levels (in %; nominal level: 95%) and average lengths (in parentheses) of the confidence intervals

Sample sizes	Standard MA	Fairweather	Inverse Normal
(4, 4, 6, 6)	89.08	95.01	94.94
	(1.6179)	(1.9408)	(1.9014)
(4, 4, 12, 12)	90.96	95.08	94.91
	(1.3150)	(1.5051)	(1.5064)
(4, 4, 24, 24)	92.28	95.11	94.96
	(1.0142)	(1.1507)	(1.1723)

Simulation using setting in example

Robustness in random-effects model?
Put $\mu_D = 0$ and $\sigma_i^2 = 1$, Ker sample sizes, $\sum_i \mathrm{var}(D_i)/2 = 5/12$, m= 100.000 simulation runs

τ^2	$\frac{\tau^2}{\tau^2+5/12}$	Standard MA	Fairweather	Inverse Normal
0	0	89.08	95.01	94.94
0.01	0.0234	88.70	94.78	94.70
0.02	0.0458	88.34	94.54	94.46
0.05	0.1071	87.22	93.81	93.78
0.10	0.1935	85.41	92.54	92.58
0.50	0.5455	73.21	82.65	83.94

Example 2

Respiratory muscle training for multiple sclerosis
Effect size: Mean differences (MD)

Method	Estimate MD	95%-CI for MD
Standard MA	8.2855	[-0.9340 – 17.5051]
Fairweather approx. critical value	8.1411	[-1.7392 – 18.0215]
Fairweather simulated crit. value	8.1411	[-1.7377 – 18.0200]
Inverse Normal	8.4473	[-1.6742 – 18.2508]

+

Summary

- A set of exact confidence interval for a common difference of means exist in meta-analysis, see HKS (2008).
- The two approaches presented here always yield genuine intervals and are easy to calculate.
- A slight amount of heterogeneity seems to be acceptable in this context.

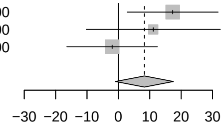
Problems induced by meta-analyses

Chalmers (1991). "Heterogeneity of the control event rates and the treatment differences need to be dealt with statistically".

Here: Homogeneity/heterogeneity of treatment differences and means of control groups. Recall Example 2.

Study	Experimental Total	Experimental Mean	Experimental SD	Control Total	Control Mean	Control SD	MD	95%−CI	Weight
1	9	40.30	21.0000	9	23.00	7.0000	17.30	[2.84; 31.76]	40.6%
2	10	73.30	19.8000	5	62.20	20.1000	11.10	[−10.26; 32.46]	18.6%
3	23	98.00	28.0000	25	100.00	23.0000	−2.00	[−16.45; 12.45]	40.7%
Common effect model	**42**			**39**			**8.29**	**[−0.93; 17.51]**	**100.0%**

Mean Difference



References

Chalmers, T.C. (1991). Problems induced by meta-analyses. *Statistics in Medicine*, 10, 971–980.

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.

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.*

Hartung, J., Knapp, G., Sinha, B.K. (2008). Statistical Meta-analysis with Applications. Wiley, Hoboken NJ.

Ker, K., Edwards, P.J., Felix, L.M., Blackhall, K., Roberts, I. (2010). Caffeine for the prevention of injuries and errors in shift workers (analysis 1.3.4). Cochrane Database of Systematic Reviews 2010, Issue 5. Art. No.: CD008508. DOI: 10.1002/14651858.CD008508

Rietberg, M.B., Veerbeek, J.M., Gosselink, R., Kwakkel, G., van Wegen, E.E.H. (2017). Respiratory muscle training for multiple sclerosis (analysis 2.1). Cochrane Database Systematic Reviews 2017. Art. No. CD009424.

Walter, S.D., Balakrishnan ,N . (2022). A method was developed for correcting the bias in the usual study weights in meta-analyses. Journal of Clinical Epidemiology, 152, 23-29.