Meta-Analysis:
Different Methods - Different Conclusions?

Annette BÖCKENHOFF and Joachim HARTUNG
Department of Statistics, University of Dortmund, D-44221 Dortmund, Germany

Abstract. Different estimation methods for covariance parameters in meta-analyses can result in conflicting p-values concerning the test of treatment effect. We propose a valid method to overcome this problem at least partially by introducing a new estimator for the standard deviation of the common treatment difference.

1. Introduction

Meta-analyses receive increased attention in the recent years. In different application fields we can make use of the methods of combining information from different studies or experiments, e.g. for estimating the mean effect of a new treatment by various studies. Moreover, there exists methods for a variety of outcome measures.

Recently, there has been a discussion about the comparison of the results of analysing summary data with the use of individual patient data, see [1] and [2]. However, the same deficiencies may occur with standard methods for individual patient data as known for combining summarised data, confer [3],[4],[5],[6]. Here, two examples are used to illustrate that different estimation methods, implemented in some software packages, can lead to extreme different results. In this context we propose methods to overcome this dilemma at least partially. Hence, the tests and confidence intervals for the mean effect will be less dependent of the estimation method for the parameter of heterogeneity in the random effects model of meta-analysis .

2. Individual patient data versus summary data

If in meta-analysis one has the opportunity to obtain individual patient data, apart from the costs, time and effort, one can consider the question, whether analysing individual patient data gives more precise results than summary data, assuming the suitable quality of the data.

For the fixed effects model, see below, it can be shown, that the pooled individual patient data modelled as a two-way ANOVA model without interaction is equivalent to combining the best linear unbiased estimators from each study, provided that observations between studies are independent and have a common variance, confer [1]. Recently, it was stated by [2], that this equivalence is more general. Consequently, it holds also in random effects models were there exists heterogeneity between studies.

However, in the random effects model we have to estimate covariance parameters. If the same estimated covariance parameters are used in both data cases, we get the same estimators for the mean effect, see [2]. So, different results in analysing individual patient data or summarised data can be a problem of different estimators in the covariance matrix, which can be indicated in several software packages as for example in BUGS, S-Plus or PROC MIXED in SAS, see e.g.[7].

Nonetheless, different estimation methods for the covariance matrix can result in extreme different decisions concerning the significance of the treatment effect. In the next section we present a method to overcome this problem.

3. The model

In this context we regard in study i, i=1,...,k, the estimators

$$\hat{\theta}_i = \bar{y}_{1i} - \bar{y}_{2i} ,$$

where $\bar{y}_{1i}$ and $\bar{y}_{2i}$ are the mean treatment effect and the mean effect of the control, respectively, assuming homogeneous variance of the observations in study i. So the estimated variance of this estimator in a first step is the pooled variance

$$\hat{\upsilon}_i^2 = \left(\frac{1}{n_{1i}} + \frac{1}{n_{2i}}\right)\left(\frac{1}{n_{1i} + n_{2i} - 2}\right)\left((n_{1i} - 1)s_{1i}^2 + (n_{2i} - 1)s_{2i}^2\right) ,$$

where n_{1i} and n_{2i} are the numbers of observations in the treatment and control group, respectively. In each group we have the (estimated) standard deviations s_{1i} and s_{2i}.

We consider the random effects model for summarised data

$$\hat{\theta}_i = \theta + a_i + e_i \quad , i = 1,\ldots,k ,$$

where θ is the mean treatment difference, a_i is a random effect with mean zero and variance τ^2, and the random error e_i has mean zero and variance υ_i^2. Hence, a correlation between observations within studies is assumed. In general, mean effect estimators depend on the estimation method of the variance component τ^2. They are computed by

$$\hat{\theta} = \hat{\theta}(\hat{\tau}^2) = \sum_{i=1}^{k} \frac{w_i(\hat{\tau}^2)}{w_\Sigma(\hat{\tau}^2)}\hat{\theta}_i , \text{ with } w_i(\hat{\tau}^2) = (\hat{\tau}^2 + \hat{\upsilon}_i^2)^{-1} \text{ and } w_\Sigma(\hat{\tau}^2) = \sum_{i=1}^{k} w_i(\hat{\tau}^2) .$$

Omitting a_i in the model and $\hat{\tau}^2$ in the estimators, respectively, one gets the (homogeneous) fixed effects model.

The common estimator of the standard deviation of $\hat{\theta}$ is given by

$$\hat{\sigma}(\hat{\theta}) = \left(\frac{1}{w_\Sigma(\hat{\tau}^2)}\right)^{1/2} .$$

In the traditional random effects meta-analysis we usually compare the teststatistic $\hat{\theta}\big/\hat{\sigma}(\hat{\theta})$, which involves the method of moments estimator for τ^2, also called the DerSimonian-Laird estimator, see [8], with the (1-α/2)-quantile of the standard normal distribution. Moreover, there exists proposals to compare the teststatistics with the (1-α/2)-quantile of the central t-distribution with (k-1) degrees of freedom, confer e.g.[5], and it is used by SAS PROC MIXED.

Now, we present a new estimator for the standard deviation of $\hat{\theta}$ based on a method proposed by [9], which depends merely on the estimation method of τ^2. The proposed estimator of the standard deviation of $\hat{\theta}$ is given by

$$\hat{\sigma}_{new}(\hat{\theta}) = \left\{ \frac{1}{k-1} \sum_{i=1}^{k} \frac{w_i(\hat{\tau}^2)}{w_\Sigma(\hat{\tau}^2)} \left(\hat{\theta}_i - \hat{\theta}(\hat{\tau}^2) \right)^2 \right\}^{1/2},$$

confer [9]. The corresponding teststatistic should also be compared with the (1-α/2)-quantile of the central t-distribution with (k-1) degrees of freedom. The results of these methods can be seen in the following examples.

4. Example 1

The author of [7] cites an example of analysing treatment-control differences in a random effects model of meta-analysis. This example compares specialist multidisciplinary team care for managing stroke patients with routine management in general medical wards (from Cochrane Database of Systematic Reviews, 1995). The length of stay was measured for each patient from nine studies. The data is given in table 1.

Table 1. Care for stroke patients from nine studies. (SD = standard deviation)

Study	Specialist care			Routine management		
	n	mean	SD	n	mean	SD
1	155	55.0	47.0	156	75.0	64.0
2	31	27.0	7.0	32	29.0	4.0
3	75	64.0	17.0	71	119.0	29.0
4	18	66.0	20.0	18	137.0	48.0
5	8	14.0	8.0	13	18.0	11.0
6	57	19.0	7.0	52	18.0	4.0
7	34	52.0	45.0	33	41.0	34.0
8	110	21.0	16.0	183	31.0	27.0
9	60	30.0	27.0	52	23.0	20.0

For the following table 2, we computed estimators for the heterogeneity variance τ^2 with different methods. We used the traditional random effects meta-analysis with the DerSimonian-Laird estimator (DL), as well as estimation methods feasible e.g. with PROC MIXED as Restricted Maximum Likelihood (REML) and Maximum Likelihood (ML). Moreover, we computed an estimator proposed by Hedges [10] and the fixed effects estimator (FE). The presented p-values are derived from related t-tests, see section 3.

Table 2. Estimation methods for the heterogeneity variance τ^2 and resulting estimators for the treatment difference, their standard deviations and p-values from t-tests.

Method	$\hat{\tau}^2$	$\hat{\theta}(\hat{\tau}^2)$	$\hat{\sigma}(\hat{\theta})$	p-value	$\hat{\sigma}_{new}(\hat{\theta})$	p_{new}-value
DL	218.72	-14.10	5.28	0.0284	8.78	0.1470
REML	685.09	-15.12	8.95	0.1297	9.20	0.1390
ML	596.04	-15.03	8.38	0.1107	9.17	0.1397
HEDGES	771.32	-15.19	9.48	0.1475	9.23	0.1384
FE	0	-3.49	0.78	0.0021	4.29	0.4392

In the traditional random effects meta-analysis we usually compare the teststatistic with the (1-α/2)-quantile of the standard normal distribution. Then, we get the p-value p=0.0038 for estimation methods with the DerSimonian-Laird estimator and p<0.0001 in the fixed effects model.

We see from table 2 that there are differences in the p-values of the standard t-test by different estimation methods for τ^2. If we apply the new estimator for the standard deviation of $\hat{\theta}$ we get similar results for the p-values despite of different estimation methods for τ^2 in the random effects model.

5. Example 2

The second example is taken from [3]. They use a data set for testing the effectiveness of amlodipine in the treatment of angina. Eight randomized controlled trials have compared the change in work capacity for patients who received either the drug or placebo. The change in work capacity is the ratio of the exercise time after invention to before for each patient. The logarithms of the observed changes are assumed to be approximately normally distributed. The data is given in table 3.

Table 3. Change in work capacity in the treatment of angina.

Study	Amlodipine			Placebo		
	n	mean	variance	n	mean	variance
1	46	0.2316	0.2254	48	-0.0027	0.0007
2	30	0.2811	0.1441	26	0.0270	0.1139
3	75	0.1894	0.1981	72	0.0443	0.4972
4	12	0.0930	0.1389	12	0.2277	0.0488
5	32	0.1622	0.0961	34	0.0056	0.0955
6	31	0.1837	0.1246	31	0.0943	0.1734
7	27	0.6612	0.7060	27	-0.0057	0.9891
8	46	0.1366	0.1211	47	-0.0057	0.1291

We computed the different estimators for the interesting measures by the various methods already mentioned in section 4, see table 4.

Table 4. Estimation methods for the heterogeneity variance τ^2 and resulting estimators for the treatment difference, their standard deviations and p-values from t-tests.

Method	$\hat{\tau}^2$	$\hat{\theta}(\hat{\tau}^2)$	$\hat{\sigma}(\hat{\theta})$	p-value	$\hat{\sigma}_{new}(\hat{\theta})$	p_{new}-value
DL	0.0066	0.1590	0.0447	0.0093	0.0507	0.0165
REML (SAS)	0	0.1111	0.0305	0.0083	0.0465	0.0496
ML (SAS)	0	0.1068	0.0224	0.0020	0.0476	0.0599
HEDGES	0.0353	0.1654	0.0765	0.0674	0.0632	0.0346
FE	0	0.1624	0.0321	0.0015	0.0427	0.0067

In this example, the estimated heterogeneity variance is very small. For the REML and the ML estimation method this parameter is estimated by zero. Therefore, one would expect, that the corresponding estimator for the treatment difference is equal to the fixed effects estimator, but SAS computes another solution.

If we compare the teststatistics with the (1-α/2)-quantile of the standard normal distribution, we get the p-value p=0.0004 for estimation methods with the DerSimonian-Laird estimator and p<0.0001 in the fixed effects model.

The new estimation method results in p-values which are nearly in the same order, while the common method results in p-values which differ around 30-fold in the values.

6. Concluding remarks

As we see, with the usual methods we can come to quite different conclusions depending on the chosen method of estimation. Simulation results show, that the proposed new method equalizes the significance results towards the nominal level. With regard to section 2, where we considered individual versus summarized data, we have to state, that also with the individual data case this estimation problem is not yet solved. Moreover, we conclude, that the “optimal” estimator for the heterogeneity variance seems not to be found yet.

Acknowledgement

The support of the Deutsche Forschungsgemeinschaft (SFB 475, “Reduction of complexity in multivariate data structures”) is gratefully acknowledged.

References

[1] I. Olkin and A. Sampson, Comparison of Meta-Analysis Versus Analysis of Variance of Individual Patient Data, *Biometrics* **54** (1998) 317-322.

[2] T. Mathew and K. Nordström, On the Equivalence of Meta-Analysis Using Literature and Using Individual Patient Data, *Biometrics* **55** (1999) 1221-1223.

[3] Y. Li, L. Shi and H.D. Roth, The Bias of the Commonly-Used Estimate of Variance in Meta-Analysis, *Commun. Statist.-Theory Meth.* **23** (1994) 1063-1085.

[4] A. Böckenhoff and J. Hartung, Some Corrections of the Significance Level in Meta-Analysis, *Biometrical Journal* **40** (1998) 937-947.

[5] D.A. Follmann and M.A. Proschan, Valid Inference in Random Effects Meta-Analysis, *Biometrics* **55** (1999) 732-737.

[6] S. Ziegler and N. Victor, Gefahren der Standardmethoden für Meta-Analysen bei Vorliegen von Heterogenität, *Informatik, Biometrie und Epidemiologie in Medizin und Biologie* **30** (1999) 131-140.

[7] S.-L.T. Normand, Meta-Analysis: Formulating, Evaluating, Combining, and Reporting, *Statistics in Medicine* **18** (1999) 321-359.

[8] R. DerSimonian and N. Laird, Meta-Analysis in Clinical Trials, *Controlled Clinical Trials* **7** (1986) 177-188.

[9] J. Hartung, An Alternative Method for Meta-Analysis, *Biometrical Journal* **41** (1999) 901-916.

[10] L.V. Hedges, A Random Effects Model for Effect Sizes, *Psychological Bulletin* **93** (1983) 388-395.