An Alternative Method for Meta-Analysis

JOACHIM HARTUNG
Department of Statistics, SFB475
University of Dortmund, Germany

Summary

In many fields of applications, test statistics are obtained by combining estimates from several experiments, studies or centres of a multi-centre trial. The commonly used test procedure to judge the evidence of a common overall effect can result in a considerable overestimation of the significance level, leading to a high rate of too liberal decisions. An alternative test statistic is presented and a better approximating test distribution is derived. Explicitely discussed are the methods in the unbalanced heteroscedastic 1-way random ANOVA model and for the probability difference method, including interaction treatment by centres. Numerical results are presented by simulation studies.

Key words: meta-analysis, combining experiments, multi-centre study, interaction treatment by centres, heteroscedastic ANOVA model, random effects, probability difference method

Zusammenfassung

In vielen Anwendungsgebieten werden Teststatistiken mittels Kombination von Schätzungen aus verschiedenen Experimenten, Studien oder Zentren einer multizentrischen Studie erhalten. Die allgemein verwandte Testmethode zur Beurteilung der Evidenz eines Gesamteffektes kann zu einer beträchtlichen Überschätzung des Signifikanz-Niveaus führen, mit der Folge eines hohen Anteils zu liberaler Entscheidungen. Es wird eine alternative Teststatistik vorgestellt und eine besser approximierende Testverteilung wird hergeleitet. Die Verfahren werden in der unbalanzierten heteroskedastischen Einfachklassifikation der Varianzanalyse mit zufälligen Effekten und für die Wahrscheinlichkeits-Differenzen-Methode ausführlich diskutiert, einschließlich einer Wechselwirkung zwischen Behandlung und Zentren. Numerische Ergebnisse werden aufgezeigt mittels Simulationsstudien.

1 Introduction

The problem of judging an overall effect from several studies or experiments arises in a variety of application fields. However, not only in meta-analysis of composed individual results but also in analysing multi-centre trials, for instance one has separated samples with heterogenous error variances and possibly an interaction of treatments with centres to asses the overall effect. Taking the effects of this interaction as random, one gets the so-called random effects model of meta-analysis, cf. sec. 2.

In the commonly used method of meta-analysis, tracing back to Cochran (1937, 1954), one gets the variance of the overall effect by estimating its components seperately, and for the corresponding test statistic the standard normal distribution is taken as test distribution, cf. sec. 3. Now, this procedure is observed not to hold the prescribed significance level, which can lead to a high rate of too liberal decisions. That phenomenon mainly is not a problem of the type of variance estimator involved. Computational experience with other estimators than the usual unbiased one, for instance also with a kind of nonnegative minimum biased estimator, as discussed by Hartung (1981), yield qualitatively no essential improvements in the significance levels obtained.

Therefore in the following, cf. sec. 4, an estimation function is introduced that estimates the variance of the weighted mean directly, based in its realisation on weights which on a first stage are estimated upon some other estimation principle, for instance here is cosen the classical one.

Further, its distribution is approximated by equating the first two moments to that one of a χ^2-distribution, such that the test of significance for the overall effect becomes an approximate t-test that is more able to hold the actual significance level near the prescribed one.

The performance of the test procedures is discussed, including simulation studies, in the unbalanced heteroscedastic random 1-way ANOVA model, cf. sec. 5, and, finally, in order to demonstrate the application to data that don't follow an ANOVA model, to the probability difference method, cf. sec. 6, comparing two proportions that are observed several times.

Sometimes there is the opinion that one should avoid the random effects model because it would be too conservative, and one should better work with the so-called fixed effects model, neglecting the interaction effect and assuming homogenity with respect to a common mean. Of course, this would lead to a higher actual siginificance level. However, here it is to say that, even if the fixed effects model is the correct one for the data situation given, also the method commonly used there can yield a high rate of too liberal decisions, e. g. Li/ Shi/ Roth (1994), Böckenhoff/ Hartung (1998).

2 The Model

Let $\hat{\mu}_i$ for $i = 1, \ldots, k, k \geq 2$, be stochastically independent normally distributed unbiased estimators for the common mean μ of k independent experiments, studies or centres of a multi-centre trial, which also let provide stochastically independent unbiased estimators $\widehat{\sigma_i^2} > 0$, (a. e.), for the partial variance $\sigma_i^2 > 0$ of $\widehat{\mu_i}, i = 1, \ldots, k$. Due, for instance, to an 'interaction of response with centres', there may be a common part $\sigma_a^2 \geq 0$ of the variance of the $\widehat{\mu_i}$ that can not be estimated in the i-th study, $i = 1, \ldots, k$; that is, we have the so-called random effects model of parametric meta-analysis, respectively of combining of experiments:

$$\widehat{\mu_i} \sim N(\mu, \sigma_a^2 + \sigma_i^2), \widehat{\sigma_i^2} \text{ given }, i = 1, \ldots, k, \tag{1}$$

e. g. Cochran (1937, 1954), Yates/ Cochran (1938), Hedges/ Olkin (1985), DerSimonian/ Laird (1986), Whitehead/ Whitehead (1991), Draper et al. (1992).
Of main interest here is to test a hypothesis like

$$\mathrm{H}_0 : \mu \leq 0 \text{ against } \mathrm{H}_1 : \mu > 0, \tag{2}$$

respectively to derive a confidence interval for the common mean μ. Denote

$$\tau_i = \frac{1}{\sigma_a^2 + \sigma_i^2}, i = 1, \ldots, k, \tau = \sum_{i=1}^{k} \tau_i, \tag{3}$$

the best unbiased estimator of μ would be

$$\tilde{\mu} = \sum_{i=1}^{k} \frac{\tau_i}{\tau} \widehat{\mu_i} \tag{4}$$

with the variance $\mathrm{var}(\tilde{\mu}) = 1/\tau$, leading, under $\mu = 0$, to the test statistic

$$\frac{\tilde{\mu}}{\sqrt{1/\tau}} \sim N(0, 1). \tag{5}$$

Now, for a realisation, the involved parameters have to be estimated.
Note that in applications often $\widehat{\mu_i}$ is a function of further parameter estimates, for instance a mean difference or an effect size of two treatments, or e. g. the difference, cf. sec. 6, the (log) odds ratio or relative risk of two observed proportions, cf. the references cited above, and $\widehat{\sigma_i^2}$ frequently is only an approximation, e.g. via the delta-method. The general assumptions for $\widehat{\mu_i}$ and $\widehat{\sigma_i^2}$ then can be fulfilled only in approximation, of course, implying the same for resulting properties.

3 The commonly used method

Let denote $\omega_i = 1/\sigma_i^2, i = 1, \ldots, k$, and $\omega = \sum_{i=1}^k \omega_i$, then, e. g. Chochran (1954), DerSimonian/ Laird (1986), Whitehead/ Whitehead (1991), an unbiased estimator of σ_a^2 is given by

$$\bar{\sigma_a^2} := \frac{\omega}{\omega^2 - \sum_{i=1}^k \omega_i^2} \left\{ \sum_{i=1}^k \omega_i \left(\widehat{\mu}_i - \sum_{j=1}^k \frac{\omega_j}{\omega} \widehat{\mu_j} \right)^2 - k + 1 \right\}, \tag{6}$$

with the realisation $\widetilde{\sigma_a^2}$, replacing σ_i^2 by $\widehat{\sigma_i^2}, i = 1, \ldots, k$, in $\bar{\sigma_a^2}$. This estimator can become negative with positive probability, and so one truncates it at zero,

$$\widehat{\sigma_a^2} := \max\{0, \widetilde{\sigma_a^2}\},$$

such that with

$$\widehat{\tau}_i = \frac{1}{\widehat{\sigma_a^2} + \widehat{\sigma_i^2}}, i = 1, \ldots, k, \text{ and } \hat{\tau} = \sum_{i=1}^k \widehat{\tau}_i \tag{7}$$

the common mean is estimated by

$$\hat{\mu} = \sum_{i=1}^k \frac{\widehat{\tau}_i}{\hat{\tau}} \widehat{\mu_i}, \tag{8}$$

and the test statistic under $\mu = 0$ is taken as

$$T_1 := \frac{\hat{\mu}}{\sqrt{1/\hat{\tau}}} \overset{appr.}{\sim} N(0, 1). \tag{9}$$

Caused by distributional deficiencies, cf. also Li et al.(1994), Böckenhoff/ Hartung (1998), the resulting test procedure is not satisfactory, because of the observation that the actual levels attained by the test can arise much above the prescribed level, leading to a high rate of too liberal decisions, cf. the simulation results in sec. 5 and 6.

4 A refined method

Let $\beta_i = \tau_i/\tau, i = 1, \ldots, k, \beta = (\beta_1, \ldots, \beta_k)'$, and $x_i = \widehat{\mu_i}, i = 1, \ldots, k, x = (x_1, \ldots, x_k)'$ (where c' denotes the transpose of a vector c), then we consider the following quadratic form in x:

$$S(\beta) := \sum_{i=1}^k \beta_i (x_i - \beta' x)^2. \tag{10}$$

Theorem 4.1

(i). $\tau \cdot S(\beta)$ has a (central) χ^2-distribution with (k-1) degrees of freedom,

(ii). $\tilde{\mu} = \beta' x$ and $S(\beta)$ are stochastically independent.

Proof: Denote $D = diag(\sigma_a^2 + \sigma_i^2, i = 1, \ldots, k)$ the diagonal covariance matrix of x and $\mathbf{1} := (1, \ldots, 1)' \in I\!R^k$, i. e.

$$x \sim N(\mu \cdot \mathbf{1}, D),$$

and let $\mathbf{1_i} := (0, \ldots, 0, 1, 0, \ldots, 0)' \in I\!R^k$, with the 1 at the i-th place, $m_i := \mathbf{1_i} - \beta$ and the matrix $M := \tau \cdot \sum_{i=1}^k \beta_i m_i m_i'$. So we can write

$$\begin{aligned} \tau S(\beta) &= \tau \cdot \sum_{i=1}^k \beta_i (m_i' x)^2 \\ &= \tau \cdot \sum_{i=1}^k \beta_i x' m_i m_i' x \\ &= x' M x, \end{aligned}$$

and if $MD = (MD)^2$, then $x'Mx$ is χ^2-distributed with $trace(MD)$ degrees of freedom, and if $MD\beta = 0$, then $x'Mx$ and $\beta' x$ are stochastically independent, e. g. Mathai/Provost (1992, p. 197, 227); note that $m_i' \mathbf{1} = 0$ and thus $ME(x) = 0$.

To (i): We have

$$\begin{aligned} MD &= \sum_{i=1}^k \tau \beta_i m_i m_i' D \\ &= \sum_{i=1}^k \tau \beta_i (\mathbf{1_i} - \beta)(\mathbf{1_i} - \beta)' diag\left(\frac{1}{\tau_i}, i = 1, \ldots, k\right) \\ &= \sum_{i=1}^k \tau (\beta_i \mathbf{1_i} - \beta_i \beta) \left(\frac{1}{\tau_i} \mathbf{1_i} - \frac{1}{\tau} \mathbf{1}\right)' \\ &= \sum_{i=1}^k \tau \left(\frac{1}{\tau} \mathbf{1_i} \mathbf{1_i}' - \frac{1}{\tau} \beta \mathbf{1_i}' - \frac{1}{\tau} \beta_i \mathbf{1_i} \mathbf{1}' + \frac{1}{\tau} \beta_i \beta \mathbf{1}'\right) \\ &= \sum_{i=1}^k \mathbf{1_i} \mathbf{1_i}' - \beta \sum_{i=1}^k \mathbf{1_i}' - \left(\sum_{i=1}^k \beta_i \mathbf{1_i}\right) \mathbf{1}' + \left(\sum_{i=1}^k \beta_i\right) \beta \mathbf{1}' \\ &= I - \beta \mathbf{1}' - \sum_{i=1}^k \beta_i \mathbf{1_i} \mathbf{1}' + \beta \mathbf{1}' \\ &= I - \sum_{i=1}^k \beta_i \mathbf{1_i} \mathbf{1}', \end{aligned}$$

where I denotes the $(k \times k)$ identity matrix and

$$\begin{aligned}
(MD)^2 &= (I - \sum_{i=1}^{k} \beta_i \mathbf{1_i 1'})(I - \sum_{i=1}^{k} \beta_i \mathbf{1_i 1'}) \\
&= I - 2\sum_{i=1}^{k} \beta_i \mathbf{1_i 1'} + \sum_{i=1}^{k}\sum_{j=1}^{k} \beta_i \beta_j \mathbf{1_i 1' 1_j 1'} \\
&= I - 2\sum_{i=1}^{k} \beta_i \mathbf{1_i 1'} + \left(\sum_{j=1}^{k} \beta_j\right) \sum_{i=1}^{k} \beta_i \mathbf{1_i 1'} \\
&= I - \sum_{i=1}^{k} \beta_i \mathbf{1_i 1'} \\
&= MD,
\end{aligned}$$

noting that $\sum_{j=1}^{k} \beta_j = 1$ and $\mathbf{1'1_j} = 1$. Further, we get $trace(MD) = k - \sum_{i=1}^{k} \beta_i = k - 1$, which completes the proof of (i).
To (ii): There is with (i)

$$\begin{aligned}
MD\beta &= \left(I - \sum_{i=1}^{k} \beta_i \mathbf{1_i 1'}\right) \beta \\
&= \beta - \sum_{i=1}^{k} \beta_i \mathbf{1_i 1'} \beta \\
&= \beta - \sum_{i=1}^{k} \beta_i \mathbf{1_i} \\
&= 0,
\end{aligned}$$

which yields (ii).

Note that a result like (i) (for $\sigma_a^2 = 0$) is stated already by Cochran (1937, p. 111); however, a direct proof that refers to a χ^2- criterion for quadratic forms we could not find in the literature.
Now $S(\beta)/(k-1)$ is an unbiased estimator of $\text{var}(\beta' x) = 1/\tau$, but for a realisation, one has to replace β by an estimate, and then in numerical experiences it proves to be much more sensitive with respect to alterations in the β-estimates than the following estimation function.
Defining now

$$\lambda(\beta) = \frac{\beta'\beta}{1 - \beta'\beta}, \text{ and } \psi_i(\beta) = \beta_i - \frac{\beta_i - \beta_i^2}{1 - \beta'\beta}, i = 1, \ldots, k, \tag{11}$$

we consider the affine quadratic (with respect to the random variables) form

$$Q(\beta) := \lambda(\beta) S(\beta) + \sum_{i=1}^{k} \psi_i(\beta) \widehat{\sigma_i^2}, \tag{12}$$

and get the following

Theorem 4.2

(i). $Q(\beta)$ is an unbiased estimator of $1/\tau = Var(\tilde{\mu})$.

(ii). If the estimators $\widehat{\sigma_i^2}$ of σ_i^2 are also stochastically independent of the estimators $\widehat{\mu_j}$ of $\mu, i, j = 1, \ldots, k$, then

(a) $Q(\beta)$ and $\tilde{\mu}$ are stochastically independent, and

(b) an approximate (central) $\chi^2(\nu)$-distribution of $\nu \cdot Q(\beta)/(1/\tau)$ has the degrees of freedom

$$\nu = \nu_{Q(\beta)} = 2 \cdot \frac{(1/\tau)^2}{2(k-1)(1/\tau)^2\lambda(\beta)^2 + \sum_{i=1}^k \psi_i(\beta)^2\, Var(\hat{\sigma}_i^2)}.$$

Proof: To (i): For the expectation we get with (i) of theorem 4.1, noting that

$$\sum_{i=1}^k \psi_i = 1 - \frac{1 - \sum_{i=1}^k \beta_i^2}{(1 - \beta'\beta} = 0,$$

$$\begin{aligned}
\mathrm{E}(Q(\beta)) &= \lambda(\beta)\frac{1}{\tau}(k-1) + \sum_{i=1}^k \psi_i(\beta)\sigma_i^2 + \left(\sum_{i=1}^k \psi_i(\beta)\right)\sigma_a^2 \\
&= (k-1)\frac{1}{\tau}\lambda(\beta) + \sum_{i=1}^k \left(\frac{\tau_i}{\tau} - \frac{(\tau_i/\tau) - (\tau_i^2/\tau^2)}{1-\beta'\beta}\right)\frac{1}{\tau_i} \\
&= (k-1)\frac{1}{\tau}\lambda(\beta) + \sum_{i=1}^k \left(\frac{1}{\tau} - \frac{(1/\tau) - (\tau_i/\tau^2)}{1-\beta'\beta}\right) \\
&= \frac{1}{\tau}(k-1)\frac{\beta'\beta}{1-\beta'\beta} + \frac{1}{\tau}k - \frac{(k/\tau) - (1/\tau)}{1-\beta'\beta} \\
&= \frac{1}{\tau}\left(k + \frac{(k-1)\beta'\beta - k + 1}{1-\beta'\beta}\right) \\
&= \frac{1}{\tau}\left(\frac{k - k\beta'\beta + k\beta'\beta - \beta'\beta - k + 1}{1-\beta'\beta}\right) \\
&= \frac{1}{\tau}.
\end{aligned}$$

To (ii): Part (a) follows from theorem 4.1 (ii), together with the additional assumptions above. Now we come to part (b).

If the random variable $Q^* = \nu \cdot Q/\mathrm{E}(Q)$ follows $\chi^2(\nu)$- distribution, then for the variance we have

$$\mathrm{Var}(Q^*) = \frac{\nu^2 \cdot \mathrm{Var}(Q)}{(\mathrm{E}(Q))^2} = 2 \cdot \nu,$$

that is

$$\nu = 2 \cdot \frac{(\mathrm{E}(Q))^2}{\mathrm{Var}(Q)}, \tag{13}$$

which conversely can be used as an estimate of the degrees of freedom of an approximate χ^2-distribution, i. e. the first two moments of the distributions are equated, cf. Patnaik (1949). So again, with (i) of theorem 4.1, we get for the variance of $Q(\beta)$

$$V_{Q(\beta)} := \lambda(\beta)^2 \left(\frac{1}{\tau}\right)^2 2(k-1) + \sum_{i=1}^{k} \psi_i(\beta)^2 \mathrm{Var}(\widehat{\sigma_i^2}), \tag{14}$$

and by (i): $\mathrm{E}(Q(\beta)) = 1/\tau$, yielding now the desired estimate for the degrees of freedom, which completes the proof.

So, if $Q > 0$, then under $\mu = 0$, the test statistic $\tilde{\mu}/\sqrt{Q} \overset{appr.}{\sim} t(\nu)$, where $t(\nu)$ denotes the (central) t-distribution with ν degrees of freedom, and H_0 can be tested, replacing the parameters by their estimates.
If the $\beta_1, \ldots, \beta_k$ take on different values, then $Q(\beta)$ can become negative with a positive probability, too, which is implied by the following

Corollary 4.1
Either at least one of the $\psi_i(\beta), i = 1, \ldots, k$, is negative or all $\psi_i(\beta)$ are zero and all $\beta_i = 1/k, i = 1, \ldots, k$.

Proof: Assume that for all $i = 1, \ldots, k : \psi_i(\beta) \geq 0$. Now, cf. (11), $\sum_{i=1}^{k} \psi_i(\beta) = 0$, implying, by our assumption, that for all $i = 1, \ldots, k$, there holds $\psi_i(\beta) = 0$, and therefore for all $i = 1, \ldots, k$,

$$\beta_i(1 - \beta'\beta) = \beta_i - \beta_i^2,$$

respectively dividing by $\beta_i (\beta_i > 0)$ yields

$$1 - \beta'\beta = 1 - \beta_i,$$

i. e. all β_i are identical and by $\sum_{i=1}^{k} \beta_i = 1$ there has to be $\beta_i = 1/k$ for all $i = 1, \ldots, k$, completing the proof.

Thus we have also to truncate $Q(\beta)$ in a suitable way, considering simultanously the idea of preserving a pointwise order for two estimators as induced by their expectations and the concept of combining estimators. Now there is

$$\mathrm{var}(\tilde{\mu}) \quad = \quad \mathrm{var}\left(\sum_{i=1}^{k} \beta_i \widehat{\mu_i}\right)$$

$$
\begin{aligned}
&= \sum_{i=1}^{k} \beta_i^2(\sigma_a^2 + \sigma_i^2) \\
&\geq \sum_{i=1}^{k} \beta_i^2 \sigma_i^2,
\end{aligned}
$$

so defining

$$R(\beta) := \sum_{i=1}^{k} \beta_i^2 \widehat{\sigma_i^2}, (> 0 \text{ a. e.}), \tag{15}$$

we have by $R(\beta)$ a lower estimator in the following sense:

$$
\begin{aligned}
\mathrm{E}(Q(\beta)) &= \mathrm{var}(\tilde{\mu}) \\
&\geq \mathrm{E}(R(\beta)).
\end{aligned}
$$

Hence, a truncated estimator of $\mathrm{var}(\tilde{\mu})$ can be given by the order (pointwise) preserving estimator

$$q_m(\beta) = \max\{Q(\beta), R(\beta)\}, \tag{16}$$

or, more generally, by a linear interpolation near the switching point. That means, for some real values A and B with

$$0 < A \leq 1 \leq B \tag{17}$$

let denote

$$L(\beta) := \min\left\{1, \max\left\{0, \frac{(Q(\beta)/R(\beta)) - A}{B - A}\right\}\right\}, \tag{18}$$

where $x/0 := +\infty$ for $x \geq 0$ and $x/0 := -\infty$ for $x < 0$, then $0 \leq L(\beta) \leq 1$, and we define the convex combination of Q and R:

$$q_{L(\beta)}(\beta) := L(\beta)Q(\beta) + \{1 - L(\beta)\}R(\beta), \tag{19}$$

and, regarding β, τ, and $\mathrm{var}(\hat{\sigma_i^2}), i = 1, \ldots, k$, as known, the variance of q_L can approximately be estimated ($L(\beta)$ random) by

$$
\begin{aligned}
V_q &:= L(\beta)^2 V_{Q(\beta)} + \{1 - L(\beta)\}^2 \sum_{i=1}^{k} \beta_i^4 \mathrm{var}(\widehat{\sigma_i^2}) \\
&\quad + L(\beta)\{1 - L(\beta)\} \sum_{i=1}^{k} \psi_i(\beta)\beta_i^2 \mathrm{var}(\widehat{\sigma_i^2}) \\
&=: v\left(\beta, \tau, \mathrm{var}(\widehat{\sigma_1^2}), \ldots, \mathrm{var}(\widehat{\sigma_k^2})\right),
\end{aligned} \tag{20}
$$

where the last term in (20) corresponds to the approximate covariance of $L \cdot Q$ and $(1 - L) \cdot R$, and $V_{Q(\beta)}$ is given in (14). Note that $L = 0$ for $Q \leq A \cdot R$ and $L = 1$ for

$Q \geq B \cdot R$, if $A < B$; for $A = B = 1 : q_L = q_m$. Further, $q_L > 0$, a. e.
Applying again the Patnaik-approximation, cf. (13), an approximate $\chi^2(\mathrm{E}\tilde{\nu})$-distribution of $(\mathrm{E}\tilde{\nu}) \cdot q_{L(\beta)}(\beta)/\mathrm{E}q_{L(\beta)}(\beta)$ is given for

$$\tilde{\nu} = \tilde{\nu}_{q_{L(\beta)}(\beta)} = 2 \cdot \frac{\{\mathrm{E}q_{L(\beta)}(\beta)\}^2}{V_q}. \tag{21}$$

If the $\widehat{\sigma_i^2}$ are stochastically independent of the $\widehat{\mu_j}, i, j = 1, \ldots, k$, then q_L and $\tilde{\mu}$ are also stochastically independent, cf. theorem 4.2 (ii), and V_q is a better estimate of $\mathrm{var}(q_L)$ with respect to bias.
For a realisation, now β is replaced in q_L and in $\tilde{\nu}_{q_L}$ by its estimator, cf. (7),

$$\hat{\beta} = \frac{1}{\hat{\tau}}(\widehat{\tau_1}, \ldots, \widehat{\tau_k})'$$

and with

$$\hat{q} := q_{L(\hat{\beta})}(\hat{\beta}), \text{ cf. (19)}, \tag{22}$$

$$\hat{v}(\hat{q}) := v(\hat{\beta}, \hat{\tau}, \widehat{\mathrm{var}}(\widehat{\sigma_1^2}), \ldots, \widehat{\mathrm{var}}(\widehat{\sigma_k^2})), \text{ cf. (20)}, \tag{23}$$

where $\widehat{\mathrm{var}}(\widehat{\sigma_i^2})$ denotes an estimator of $\mathrm{var}(\widehat{\sigma_1^2}), i = 1, \ldots, k$, the results above can be summarized to state the *test procedure* in a compromized form as follows:

Theorem 4.3
Under $\mu = 0$, there holds for the test statistic, cf. (8), (22), (23),

$$T_2 := \frac{\hat{\mu}}{\sqrt{\hat{q}}} \overset{appr.}{\sim} t(\hat{\nu}), \tag{24}$$

with

$$\hat{\nu} = 2 \cdot \frac{\hat{q}^2}{\hat{v}(\hat{q})}. \tag{25}$$

Remark 4.1
The approximation in (24) is better for the case that the $\widehat{\sigma_i^2}$ are stochastically independent of $\widehat{\mu_j}$, implying also that $\widehat{\beta_i}$ depends on $\widehat{\mu_j}$ only via $\widehat{\sigma_a^2}, i, j = 1, \ldots, k$.

Remark 4.2

(i). If an estimate of var$(\widehat{\sigma_i^2})$ is not reported in the i-th study or can not be computed upon the knowledge of $\widehat{\mu_i}$ and $\widehat{\sigma_i^2}$, for instance by the delta-method, cf. sec. 6, then in the formulas above we put $\widehat{var}(\widehat{\sigma_i^2}) \equiv 0, i = 1, \ldots, k$, defining then in $\hat{\nu} : x/0 := \infty$ for $x > 0$ and the $t(\infty)$-distribution as the standard normal distribution, cf. also sec. 5 and 6.

(ii). Defining, cf. (15),

$$\hat{\nu}_R = \frac{2R(\hat{\beta})^2}{\sum_{i=1}^{k} \beta_i^4 \widehat{var}(\widehat{\sigma_i^2})}, \tag{26}$$

then $\chi^2(E\hat{\nu}_R)$ is an approximate distribution of $(E\hat{\nu}_R) \cdot R(\beta)/E(R(\beta))$, cf. (13), so that for $0 < 2\kappa < 1$, an approximate $(1-2\kappa)$-confidence interval for $E(R(\beta))$ is given by

$$\frac{\hat{\nu}_R}{\chi^2(\hat{\nu}_R)_{1-\kappa}} R(\hat{\beta}) \leq E(R(\beta)) \leq R(\hat{\beta}) \frac{\hat{\nu}_R}{\chi^2(\hat{\nu}_R)_\kappa},$$

where $\chi^2(\nu)_\kappa$ denotes the κ-quantile of the $\chi^2(\nu)$-distribution.
Now $E(R(\beta)) = var(\tilde{\mu})$ under $\sigma_a^2 \equiv 0$, and $E(R(\beta)) \leq E(Q(\beta))$, so one can use the bounds of that interval to define the 'changing points' A and B,

$$A = \frac{\hat{\nu}_R}{\chi^2(\hat{\nu}_R)_{1-\kappa}}, \text{ and } B = \frac{\hat{\nu}_R}{\chi^2(\hat{\nu}_R)_\kappa}, \tag{27}$$

where in the following sections we have chosen $\kappa = 0.25$. There also we have ignored partially the knowledge about the estimates of $var(\widehat{\sigma_i^2})$ and worked with fixed values for A and B: $A = 0.8, B = 1.2$, and $A = 0.95, B = 1.05$.
The different choices of A and B 'in the neighbourhood of 1' do not seem to have much influence on the results.

5 The unbalanced heteroscedastic 1-way random ANOVA model

Here let us consider the model

$$y_{ij} = \mu + a_i + e_{ij}, \quad i = 1, \ldots, k; j = 1, \ldots, n_i \geq 2, \tag{28}$$

where $a_1, \ldots, a_k, e_{11}, \ldots, e_{1n_1}, \ldots, e_{kn_k}$ are stochastically independent normally distributed random variables with $\mathrm{E}(a_i) = \mathrm{E}(e_{ij}) = 0, \mathrm{var}(a_i) = \sigma_a^2 \geq 0, \mathrm{var}(e_{ij}) = \xi_i^2 > 0$, and $\mu = \mathrm{E}(y_{ij}), i = 1, \ldots, k, j = 1, \ldots, n_i$.
For the i-th estimate $\hat{\mu}_i, i = 1, \ldots, k$, of μ we get

$$\widehat{\mu_i} = \frac{1}{n_i} \sum_{j=1}^{n_i} y_{ij} \sim N(\mu, \sigma_a^2 + \sigma_i^2) \quad \text{with } \sigma_i^2 = \frac{1}{n_i} \xi_i^2,$$

and σ_i^2 is estimated by

$$\widehat{\sigma_i^2} = \frac{1}{n_i} \cdot \frac{1}{n_i - 1} \sum_{j=1}^{n_i} (y_{ij} - \widehat{\mu_i})^2, i = 1, \ldots, k,$$

Table 5.1. Unbalanced heteroscedastic random 1-way ANOVA model: sample designs $D(d,k)$, for $d = 1,2,3,4, k = 3$, and $k = 6$, with $\sigma_a^2 = 0.1, 1.0$, and 10 for the simulation results in table 5.2.

$D(d,k)$		$k=3$			$k=6$					
d	i	1	2	3	1	2	3	4	5	6
1	n_i	5	10	15	5	10	15	5	10	15
	ξ_i^2	1	3	5	1	3	5	1	3	5
2	n_i	10	20	30	10	20	30	10	20	30
	ξ_i^2	1	3	5	1	3	5	1	3	5
3	n_i	5	10	15	5	10	15	5	10	15
	ξ_i^2	5	3	1	5	3	1	5	3	1
4	n_i	10	20	30	10	20	30	10	20	30
	ξ_i^2	5	3	1	5	3	1	5	3	1

which is stochastically independent of $\hat{\mu}_i, i = 1,\ldots,k$, and further $\mathrm{var}(\widehat{\sigma_i^2}) = 2\frac{1}{n_i - 1}(\sigma_i^2)^2$, of which an unbiased estimator, e. g. Hartung/ Voet (1986), is given by

$$\widehat{\mathrm{var}}(\widehat{\sigma_i^2}) = 2 \cdot \frac{1}{n_i + 1}(\widehat{\sigma_i^2})^2, i = 1,\ldots,k. \tag{29}$$

In $k = 3$ illustrative samples of sizes n_i, respectively $2n_i$, and in $k = 6$ samples by independent replications of the first samples, cf. table 5.1, with different constellations of the residual variances ξ_i^2, for $\sigma_a^2 = 0.1, 1.0$, and 10, a simulation study (10 000 runs each) is performed in order to get estimates $\hat{\alpha}$ of the actual levels attained by the various test statistics, at the prescribed nominal level $\alpha = 0.05$, for the one-sided hypothesis $\mathrm{H}_{01} : \mu \leq 0$ against $\mathrm{H}_{11} : \mu > 0$ and for the two-sided hypothesis $\mathrm{H}_{02} : \mu = 0$ against $\mathrm{H}_{12} : \mu \neq 0$. This is done for the commonly used statistic T_1, cf. sec. 3, and for some variants of T_2, cf. (24), where $T_{2,1} = T_2$ with $A = 0.8, B = 1.2$, $T_{2,2} = T_2$ with $A = 0.95, B = 1.05$, and in the correspondent test procedures, the knowledge of an estimate for $\mathrm{var}(\widehat{\sigma_i^2})$ is ignored for both, i. e. we put there $\widehat{\mathrm{var}}(\widehat{\sigma_i^2}) \equiv 0, i = 1,\ldots,k$, cf. remark 4.2 (i). Finally, $T_{2,3} = T_2$ with A, B chosen in accordance with remark 4.2 (ii), where κ in (27) is taken as $\kappa = 0.25$; here the information $\widehat{\mathrm{var}}(\widehat{\sigma_1^2})$, given by (29), is used.

The simulation results are shown in table 5.2, where in each package the first number gives $\hat{\alpha}$ for H_{01} and the second number (*cursive*) $\hat{\alpha}$ for H_{02}.

Table 5.2. Unbalanced heteroscedastic random 1-way ANOVA model: realized significance levels $\hat{\alpha}\%$ at the nominal level $\alpha = 5\%$ for the one sided H_{01} [1st number] and for the two-sided H_{02} [*second number, cursive*] with the test statistics $T_1, T_{2,1}, T_{2,2}$ and $T_{2,3}$ in the sample designs $D(d, k)$ from table 5.1.

$\alpha = 5\%$			$\hat{\alpha}\%$							
σ_a^2	d	H_{0l}	$k = 3$				$k = 6$			
		l	T_1	$T_{2,1}$	$T_{2,2}$	$T_{2,3}$	T_1	$T_{2,1}$	$T_{2,2}$	$T_{2,3}$
0.1	1	1	8.0	6.1	5.8	5.2	7.8	4.7	5.2	4.5
		2	*10.4*	*7.4*	*7.6*	*5.9*	*9.8*	*6.0*	*6.1*	*5.2*
	2	1	7.7	5.3	5.5	5.2	7.1	4.9	4.6	4.7
		2	*9.6*	*7.0*	*7.4*	*6.8*	*8.6*	*5.3*	*4.7*	*5.6*
	3	1	9.2	6.6	7.2	6.2	8.4	4.6	4.9	5.3
		2	*12.4*	*9.5*	*9.3*	*8.3*	*11.0*	*6.0*	*6.6*	*6.9*
	4	1	11.0	7.8	6.7	7.5	8.7	4.8	4.8	4.7
		2	*15.7*	*11.7*	*9.8*	*11.5*	*12.3*	*6.2*	*6.6*	*6.4*
1.0	1	1	11.4	5.5	5.2	5.1	8.0	4.9	4.5	4.9
		2	*16.7*	*7.7*	*7.6*	*6.9*	*10.9*	*5.1*	*4.6*	*4.7*
	2	1	12.2	5.1	5.3	5.1	8.3	4.8	5.1	5.2
		2	*18.4*	*6.3*	*6.9*	*6.8*	*11.3*	*4.9*	*4.9*	*4.9*
	3	1	12.8	6.2	5.9	6.5	9.5	4.4	4.2	4.4
		2	*20.2*	*10.3*	*10.5*	*10.4*	*13.1*	*4.3*	*4.2*	*4.5*
	4	1	13.2	5.0	4.9	5.2	9.8	4.8	4.5	4.5
		2	*20.6*	*8.1*	*8.2*	*8.4*	*13.6*	*4.2*	*4.3*	*4.4*
10	1	1	12.2	4.9	5.0	5.0	8.7	5.2	4.9	5.2
		2	*19.1*	*5.6*	*5.6*	*5.5*	*12.2*	*5.3*	*5.1*	*5.2*
	2	1	12.4	5.4	4.8	4.9	8.3	4.7	5.0	4.9
		2	*19.3*	*5.5*	*5.2*	*5.1*	*11.2*	*5.0*	*5.0*	*5.2*
	3	1	13.4	4.4	5.0	4.4	9.4	4.9	4.9	4.5
		2	*21.4*	*5.6*	*6.1*	*5.4*	*13.6*	*5.1*	*5.6*	*5.2*
	4	1	13.7	5.0	4.7	4.8	9.8	5.0	5.1	4.9
		2	*21.6*	*5.9*	*5.5*	*5.2*	*14.1*	*5.5*	*5.3*	*5.2*

For T_1, the realized levels $\hat{\alpha}$ are lying partially much above the 5%-level, whereas the results of the other test procedures don't differ very much and are satisfactory, at the whole, cf. table 5.2.

6 The probability difference method

Just for demonstrating the application of the procedures discussed in sec. 3 and 4 to data not following an ANOVA model, we here consider the problem of testing the difference of two proportions. To this let for $i = 1, \ldots, k$ and $j = 1, 2$ the stochastically independent random variables z_{ij} be binomially distributed with parameters $n_{ij} \geq 2$ and $p_{ij}, 0 < p_{ij} < 1$, where it is assumed that in all studies i the probability difference is identical: $p_{i1} - p_{i2} =: \mu$. An unbiased estimator of p_{ij} is $\widehat{p_{ij}} = z_{ij}/n_{ij}$ with $\mathrm{var}(\widehat{p_{ij}}) = p_{ij}(1 - p_{ij})/n_{ij} =: \sigma^2_{ij}$, and $\widehat{p_{ij}} \overset{appr.}{\sim} N(p, \sigma^2_{ij})$. For the difference of the empirical rates there is allowed that one really observes

$$\widehat{\mu_i} = a_i + (\widehat{p_{i1}} - \widehat{p_{i2}}), \quad i = 1, \ldots, k, \tag{30}$$

where the stochastically independent $a_i \overset{appr.}{\sim} N(0, \sigma^2_a), \sigma^2_a \geq 0$, corresponding to 'random deviations' of the assumption of identical probability differences, represent the random interaction effects, also assumed as stochastically independent of the $z_{ij}, i = 1, \ldots, k, j = 1, 2$. Thus we have

$$\widehat{\mu_i} \overset{appr.}{\sim} N(\mu, \sigma^2_a + \sigma^2_i), \sigma^2_i := \sigma^2_{i1} + \sigma^2_{i2}, i = 1, \ldots, k. \tag{31}$$

An unbiased estimator of σ^2_{ij} is

$$\widehat{\sigma^2_{ij}} = \frac{1}{n_{ij} - 1} \left(\widehat{p_{ij}} - \widehat{p^2_{ij}} \right), \tag{32}$$

and for its variance it is sufficient here to take the approximation given by the delta-method,

$$\mathrm{var}(\widehat{\sigma^2_{ij}}) \overset{appr.}{=} \left(\left. \frac{\partial \widehat{\sigma^2_{ij}}}{\partial \widehat{p_{ij}}} \right|_{p_{ij}} \right)^2 \sigma^2_{ij} = \left(\frac{1 - 2p_{ij}}{n_{ij} - 1} \right)^2 \frac{1}{n_{ij}} \cdot p_{ij} \cdot (1 - p_{ij}), \tag{33}$$

which is estimated by replacing p_{ij} with $\widehat{p_{ij}}$.
Now $\widehat{\sigma^2_i} = \widehat{\sigma^2_{i1}} + \widehat{\sigma^2_{i2}}$ and $\mathrm{var}(\widehat{\sigma^2_i}) = \mathrm{var}(\widehat{\sigma^2_{i1}}) + \mathrm{var}(\widehat{\sigma^2_{i2}})$, so that (33) yields an estimate $\widehat{\mathrm{var}}(\widehat{\sigma^2_i})$.

Table 6.1. Probability difference method: realized significance levels $\hat{\alpha}\%$ at the nominal level $\alpha = 5\%$ for the one sided H_{01} [1st number] and for the two-sided H_{02} [*2nd number, cursive*] with the test statistics $T_1, T_{2,1}, T_{2,2}$ and $T_{2,3}$ in the sample designs $SD(k)$ for $k = 3$ and $k = 6$.

Sample designs $SD(k)$ for $\sigma_a^2 = 0.01, 0.1, 0.5$, and $p_{ij} = 0.2$ $k = 3$: (15,25),(20,15),(30,20) $k = 6$: (15,25),(20,15),(30,20),(15,25),(20,15)(30,20)									
$\alpha = 5\%$		$\hat{\alpha}\%$							
σ_a^2	H_{0l}	$k = 3$				$k = 6$			
	l	T_1	$T_{2,1}$	$T_{2,2}$	$T_{2,3}$	T_1	$T_{2,1}$	$T_{2,2}$	$T_{2,3}$
0.01	1	7.3	5.0	5.0	4.8	6.7	4.9	4.8	4.8
	2	*8.8*	*6.0*	*6.1*	*5.8*	*7.7*	*5.1*	*5.0*	*4.9*
0.1	1	11.8	5.5	5.6	5.3	8.1	5.1	5.5	5.0
	2	*17.4*	*7.1*	*6.8*	*6.8*	*11.1*	*5.1*	*5.0*	*5.0*
0.5	1	12.1	4.5	4.3	4.3	8.4	5.1	5.1	5.1
	2	*18.7*	*3.9*	*3.7*	*4.0*	*11.7*	*5.2*	*5.4*	*5.2*

Hence, all our test procedures can approximately be applied. This is illustrated in a simulation study (10 000 runs each) for $k = 3$ groups of paired samples (n_{i1}, n_{i2}), with sizes (15,25), (20,15), (30,20), and for $k = 6$ groups by an independent replication of the first samples to get estimates $\hat{\alpha}$ of the actual levels attained by the various test statistics - with the prescribed nominal level $\alpha = 0.05$ - for the one-sided hypothesis $\mathrm{H}_{01} : \mu \leq 0$ versus $\mathrm{H}_{11} : \mu > 0$ and for the two-sided hypothesis $\mathrm{H}_{02} : \mu = 0$ versus $\mathrm{H}_{12} : \mu \neq 0$, where under $0 = \mu = p_{i1} - p_{i2}$, the probabilities are taken as $p_{ij} = 0.2$, whereas σ_a^2 here is chosen as 0.01, 0.1, and 0.5. The test procedures, corresponding to $T_1, T_{2,1}, T_{2,2}$ and $T_{2,3}$, are chosen identically as in sec. 5, where for $T_{2,3}$ here the approximate estimate $\widehat{\mathrm{var}}(\widehat{\sigma_i^2})$ derived from (33) is used. The results are put together in table 6.1.

Now again we observe the realized levels $\hat{\alpha}$ for T_1 partially to increase much over the 5%-level, and that the T_2-variants produce, at the whole, satisfactory results.

7 Final remark

In this paper we have shown the consequences of the commonly used method for testing hypotheses about the common effect in combining estimates from several independent studies, experiments or centres of a multi-centre trial, where the occurence of a random interaction of response with centres or studies is included in the considerations.
We recommend the use of the proposed alternative test procedure with the better approximating test distribution.

Acknowledgement: Thanks are due to Kepher Makambi for assistance with the simulations.

References

BÖCKENHOFF, A. and HARTUNG, J., 1998: Some corrections of the significance level in meta-analysis. *Biometrical Journal* **40**, vol. 8, 1-11.

COCHRAN, W. G., 1937: Problems arising in the analysis of a series of similar experiments. *Journal of the Royal Statistical Society B* **4**, 102-118.

COCHRAN, W. G., 1954: The combination of estimates from different experiments. *Biometrics* **10**, 101-129.

DERSIMONIAN, R. and LAIRD, N., 1986: Meta-analysis in clinical trials. *Controlled Clinical Trials* **7**, 177-188.

DRAPER, D., GAVER, D. R. jr., GOEL P. K., GREENHOUSE, J. B., HEDGES, L. V., MORRIS, C. N., TUCKER, J. R. and WATERNAUX, C. M., 1992: *Combining information.* National Academic Press, Washington, D. C..

HARTUNG, J., 1981: Non-negative minimum biased invariant estimation in variance component models. *The Annals of Statistics* **9**, 278-292.

HARTUNG, J. and VOET, B., 1986: Best invariant unbiased estimators for the mean squared error of variance component estimators. *Journal of the American Statistical Association* **81**, 689-691.

HEDGES, L. V. and OLKIN, I., 1985: *Statistical methods for meta-analysis.* Academic Press, Orlando.

Li, Y., Shi, L., and 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.

Mathai, A. M. and Provost, S. B., 1992: *Quadratic forms in random variables.* Marcel Decker, New York.

Patnaik, P. B., 1949: The non-central χ^2- and F-distributions and their applications. *Biometrika* **36**, 202-232.

Whitehead, A. and Whitehead, J., 1991: A general parametric approach to meta-analysis of controlled clinical trials. *Statistics in Medicine* **10**, 1665-1677.

Yates, F. and Cochran, W. G., 1938: The analysis of groups of experiments. *Journal of Agricultural Sciences* **28**, 556-580.

Joachim Hartung
Department of Statistics, SFB 475
University of Dortmund
Vogelpothsweg 87
D-44221 Dortmund
Germany