

# Zweistufige Schätzung in der Meta-Analyse

**Annette Böckenhoff, Joachim Hartung**

Universität Dortmund

Fachbereich Statistik

44221 Dortmund

boecken@statistik.uni-dortmund.de

hartung@statistik.uni-dortmund.de

## Zusammenfassung

Verschiedenartige Schätzverfahren zur Kombination von Einzelresultaten können in der Meta-Analyse zu teilweise widersprüchlichen Gesamtergebnissen führen. Es wird ein neues zweistufiges Verfahren vorgestellt, das eine Angleichung der p-Werte aus den verschiedenen Schätzverfahren in Richtung des nominellen Niveaus bewirkt. Entsprechende SAS-Programmstatements der Prozedur MIXED werden diskutiert.

**Keywords:** Meta-Analyse, t-Verteilung, zweistufige Schätzverfahren, stetige Zielgrößen, Heterogenitätsvarianz, Fehler 1. Art.

## 1 Beispiele

Anhand von zwei Beispielen wird zunächst demonstriert, wie unterschiedlich die Ergebnisse bei der Kombination von Studien sein können, wenn verschiedene Methoden zur Schätzung und Testung eines gemeinsamen Effektes zur Anwendung kommen. Alle Berechnungen wurden mit der Prozedur MIXED bzw. dem Teilpaket IML von SAS durchgeführt.

### 1.1 Beispiel 1

Bei Normand [1] wird eine Kombination von Studien zitiert, in denen Schlaganfallpatienten mit zwei unterschiedlichen Pflegemethoden im Krankenhaus

behandelt werden. Diese Methoden umfassen zum einen die Pflege von Spezialisten und zum anderen die konventionelle Pflege. Dazu werden neun randomisierte Studien von der Cochrane Database of Systematic Reviews (1995) herangezogen.

Die Zielgröße für den Vergleich ist die Dauer des Krankenhausaufenthaltes. Dazu wurde in jeder Studie die mittlere Dauer der akuten stationären Behandlung (LOS, length of stay) in den beiden Gruppen bestimmt und deren Differenz gebildet. Es zeigten vier der neun Studien statistisch signifikant kürzere Aufenthalte in der Gruppe mit der Spezialistenpflege.

Zur Kombination dieser neun Studienergebnisse haben wir verschiedene statistische Methoden angewendet, wobei neben der Normalverteilung auch die t-Verteilung zur Konstruktion der Tests verwendet wurde, siehe Kapitel 2. Die resultierenden p-Werte des Tests auf gleichen Pflegeeffekt sind in Tabelle 1 zu finden.

**Tabelle 1:** Ergebnisse der verschiedenen Schätzverfahren für Beispiel 1

Methoden	Heterogenitätsvarianz	Gesamteffekt	p-Wert
FIX, z	0	-3.49	< 0.0001
FIX, t			0.0021
DSL, z	218.72	-14.10	0.0076
DSL, t			0.0284
REML, z	685.20	-15.12	0.0913
REML, t			0.1297
HEDGES, z	771.32	-15.19	0.1088
HEDGES, t			0.1475
POS <sub>new</sub> , t	738.97	-15.17	0.1386

Die Tabelle 1 zeigt, dass die Methoden nur zum Teil ein signifikantes Gesamtergebnis liefern. Die letzte Zeile repräsentiert ein neues analytisches Verfahren, das hier vorgestellt werden soll.

## 1.2 Beispiel 2

Bei Li, Shi und Roth [2] werden Studienergebnisse aus Studien kombiniert, die die Wirksamkeit des Medikaments Amlodipine im Vergleich zu Placebo bei der Behandlung von Angina untersuchen.

Die Zielgröße ist die Veränderung in der körperlichen Leistungsfähigkeit als Quotient der Dauer für eine Übung nach Behandlung zur Dauer vor der Behandlung. Die Logarithmen der beobachteten Veränderungen wurden als normalverteilt angesehen.

Es lagen Daten aus acht randomisierten kontrollierten Studien vor, von denen vier Studien signifikante Ergebnisse zeigten. Die Studien wurden als heterogen angesehen ( $Q = 12.36, p = 0.0893$ ).

Zur Kombination der Studienergebnisse haben wir auch hier unterschiedliche statistische Verfahren angewendet, vgl. Beispiel 1. Die resultierenden p-Werte bzgl. des Tests auf gleichen Behandlungseffekt sind in Tabelle 2 zu finden.

**Tabelle 2:** Ergebnisse der verschiedenen Schätzverfahren für Beispiel 2

Method	Heterogenitätsvarianz	Gesamteffekt	p-Wert
FIX, z	0	0.1624	< 0.0001
FIX, t			0.0015
DSL, z	0.0066	0.1590	0.0004
DSL, t			0.0093
REML, z	0.0003	0.1620	< 0.0001
REML, t			0.0017
HEDGES, z	0.0353	0.1654	0.0306
HEDGES, t			0.0674
POS <sub>new</sub> , t	0.0315	0.1644	0.0334

Nur ein Verfahren (HEDGES mit der t-Verteilung) zeigt zum 5%-Niveau kein signifikantes Gesamtergebnis.

Dabei gab es mit der Prozedur MIXED von SAS in der Version 6.12 Probleme, wenn man die von Normand [1] vorgeschlagenen Statements verwendet. Denn bei kleiner Heterogenitätsvarianz wie in diesem Beispiel erhält man in der Version 6.12 bei Angabe der PREDICTED-Option im MODEL-Statement andere Ergebnisse (0.1423 für den Gesamteffekt (REML)) als ohne PREDICTED-Option (0.1111 für den Gesamteffekt (REML)). Für die SOLUTION-Option im RANDOM-Statement gilt das Gleiche. Nur mit dem Zusatz des PARMS-Statements für den Startwert erhält man mit dem Normand-Code korrekte REML-Schätzer (0.1620 für den Gesamteffekt (REML)).

In der Version SAS 8e ist die PREDICTED-Option im MODEL-Statement abgeschafft worden. Die Probleme mit der SOLUTION-Option bestehen nicht mehr.

## 2 Einleitung

Bei der analytischen Meta-Analyse werden Einzelergebnisse aus verschiedenen Studien zur gleichen Fragestellung kombiniert. Das Hauptziel ist, den Gesamteffekt  $\theta$  zu schätzen, die Hypothese  $\theta = 0$  zu testen und ein Konfidenzintervall für diesen Gesamteffekt anzugeben.

Für die statistische Inferenz in Modellen mit festen bzw. mit zufälligen Effekten in der Meta-Analyse existieren verschiedene Schätzverfahren, z.B. die Verfahren nach DerSimonian/Laird (DSL), Restricted Maximum Likelihood (REML) und Hedges [3]. Sie liefern verschiedene p-Werte mit möglicherweise widersprüchlichen Schlußfolgerungen für den Test bzgl. des Gesamteffektes.

Ein zweites Problem besteht darin, dass die Standardverfahren die Nullverteilung der Teststatistik durch eine Normalverteilung approximieren. Dieses Vorgehen ist asymptotisch in der Anzahl  $k$  der Studien. Liegen nur wenige Studien vor, was häufig im medizinischen Bereich der Fall ist, dann unterliegt man einer wesentlichen Erhöhung des Fehlers 1. Art, vgl. [4] und [5]. Ein Ausweg besteht darin, die  $t_{k-1}$ -Verteilung für die übliche Teststatistik im Modell mit zufälligen Effekten anzuwenden, vgl. [6]. Dies liefert allerdings keine globale Verbesserung, denn sind die Studien ähnlich, d.h. die Heterogenitätsvarianz ist klein, dann erhält man zu breite Konfidenzintervalle und man kommt zu selten zu Entscheidungen, siehe Kapitel 8.

Wir stellen hier verbesserte Verfahren zur Schätzung der Varianz des Effektschätzers und zur Schätzung der Heterogenitätsvarianz vor, so dass die resultierenden Tests mit den Standardverfahren p-Werte liefern, die in Richtung des nominellen Niveaus verschoben sind. Dies wird durch Simulationsergebnisse in Kapitel 8 validiert.

### 3 Das Modell

An dieser Stelle wollen wir Studien mit stetigen Zielgrößen betrachten, wie z.B. die Differenz der Mittelwerte der Behandlungs- (T) und der Kontrollgruppe (C) bei stetigen Variablen.

Wir betrachten also in Studie  $i$  den Schätzer  $\hat{\theta}_i = \bar{y}_{Ti} - \bar{y}_{Ci}$  der Zielgröße  $\theta$ , wobei wir homogene Varianzen in den Medikationsgruppen unterstellen, d.h.  $\text{var}(y_{Tij}) = \text{var}(y_{Cij'})$ . Die geschätzte Interstudienvarianz, also die geschätzte Varianz des Schätzers  $\hat{\theta}_i$  lässt sich darstellen als

$$\hat{\nu}_i^2 = \left( \frac{1}{n_{Ti}} + \frac{1}{n_{Ci}} \right) s_{pi}^2 .$$

Dies ist ein Vielfaches der empirischen gepoolten Varianz  $s_{pi}^2$  in jeder Studie. Wir unterstellen ein Modell mit zufälligen Effekten für  $\hat{\theta}_i$  in jeder Studie mit

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

wobei  $\theta$  den gemeinsamen Behandlungseffekt für alle Studien,  $a_i$  den Studieneffekt mit  $a_i \sim \mathcal{N}(0, \tau^2)$  und  $e_i$  den Intrastudieneffekt mit  $e_i \sim \mathcal{N}(0, \nu_i^2)$  darstellen für  $i = 1, \dots, k$ . Man erhält so die folgende Darstellung

$$\hat{\theta}_i \sim \mathcal{N}(\theta, \tau^2 + \nu_i^2) \quad ,$$

wobei die Normalverteilung zumindest asymptotisch gelten soll. Im Spezialfall  $\tau^2 = 0$  erhalten wir das sogenannte Fixed effects model, das Modell mit festen Effekten.

## 4 Schätzung des gemeinsamen Behandlungseffektes

Der übliche Schätzer des gemeinsamen Behandlungseffektes ist das gewichtete Mittel, das von der Schätzmethode bezüglich  $\tau^2$  abhängt mit

$$\hat{\theta} = \hat{\theta}(\hat{\tau}^2) = \sum_{i=1}^k \frac{w_i(\hat{\tau}^2)}{w_{\Sigma}(\hat{\tau}^2)} \hat{\theta}_i \quad ,$$

wobei  $w_i(\hat{\tau}^2) = (\hat{\tau}^2 + \hat{\nu}_i^2)^{-1}$  und  $w_{\Sigma}(\hat{\tau}^2) = \sum_{i=1}^k w_i(\hat{\tau}^2)$ . Sind die Gewichte bekannt, so erhält man den Maximum Likelihood Schätzer bzw. den Besten Linear Unverzerrten Schätzer für  $\theta$ . Im Modell mit festen Effekten reduziert sich dieser Schätzer zu

$$\hat{\theta}_{FIX} = \hat{\theta}(0) = \sum_{i=1}^k \frac{1/\hat{\nu}_i^2}{\sum_{i=1}^k 1/\hat{\nu}_i^2} \hat{\theta}_i \quad .$$

## 5 Schätzung der Heterogenitätsvarianz $\tau^2$

Es existieren verschiedene analytische Verfahren zur Schätzung der Heterogenitätsvarianz  $\tau^2$ . Der Schätzer nach dem Verfahren von DerSimonian-Laird [7] wird auch Momentenschätzer genannt, beruht auf einer quadratischen Form und wird bei Null abgeschnitten. Dieser Schätzer wird bezeichnet mit

$$\hat{\tau}_{DSL,+}^2 = \max \left\{ 0, \frac{Q - (k - 1)}{\sum_{i=1}^k \hat{\nu}_i^{-2} - \sum_{i=1}^k \hat{\nu}_i^{-4} / \sum_{i=1}^k \hat{\nu}_i^{-2}} \right\} \quad ,$$

wobei  $Q = \sum_{i=1}^k \frac{1}{\hat{\nu}_i^2} \left\{ \hat{\theta}_i - \hat{\theta}(0) \right\}^2$  ist. Dieser Schätzer hat einen kleinen Bias und eine größere Varianz im Vergleich zu den anderen Schätzverfahren.

Das REML-Verfahren (Restricted Maximum Likelihood) (s. [8]), liefert nicht-negative Schätzer mit

$$\hat{\tau}_{REML}^2 = \frac{\sum_{i=1}^k w_i^2(\hat{\tau}_{REML}^2) \left( \left( \hat{\theta}_i - \hat{\theta}_{REML} \right)^2 - \nu_i^2 \right)}{\sum_{i=1}^k w_i^2(\hat{\tau}_{REML}^2)} + \frac{1}{w_{\Sigma}(\hat{\tau}_{REML}^2)} \quad ,$$

dessen Lösung auf iterativem Weg erhalten wird. Dieser Schätzer wird von SAS PROC MIXED berechnet.

Der HEDGES-Schätzer [3] ist ein abgeschnittener ungewichteter Quadratsummenschätzer mit

$$\hat{\tau}_{HEDGES}^2 = \max \left\{ 0, \frac{1}{k-1} \sum_{i=1}^k (\hat{\theta}_i - \bar{\hat{\theta}})^2 - \frac{1}{k} \sum_{i=1}^k \hat{\nu}_i^2 \right\} ,$$

wobei  $\bar{\hat{\theta}}$  das arithmetische Mittel der  $\hat{\theta}_i$ ,  $i = 1, \dots, k$ , darstellt.

An dieser Stelle möchten wir einen neuen nichtnegativen Schätzer für Varianzkomponenten vorstellen, der auf dem (Uniformly) Minimum Bias Prinzip nach Hartung [9] beruht. Er enthält Komponenten des Hedges-Schätzers mit

$$\hat{\tau}_{POSnew}^2 = \frac{A_1^2}{A_1 + 2A_2} ,$$

wobei

$A_1 = \frac{1}{k-1} \sum_{i=1}^k (\hat{\theta}_i - \bar{\hat{\theta}})^2$  und  $A_2 = \frac{1}{k} \sum_{i=1}^k \hat{\nu}_i^2$  mit  $E(A_1) = \tau^2 + \frac{1}{k} \sum_{i=1}^k \nu_i^2$ , vgl. auch [10]. Hiermit lässt sich auch ein Konfidenzintervall für  $\tau^2$  angeben, worauf an dieser Stelle nicht weiter eingegangen werden soll.

## 6 Schätzung der Varianz von $\hat{\theta}$

Die Varianz des geschätzten gemeinsamen Effektes  $\hat{\theta}$  ist

$$\text{var}(\hat{\theta}) = \left( \sum_{i=1}^k \frac{1}{\tau^2 + \nu_i^2} \right)^{-1} .$$

Üblicherweise werden in den Gewichten die geschätzten Parameter eingesetzt und man erhält

$$\widehat{\text{var}}(\hat{\theta}) = \left( \sum_{i=1}^k \frac{1}{\hat{\tau}^2 + \hat{\nu}_i^2} \right)^{-1} = w_{\Sigma}^{-1}(\hat{\tau}^2) .$$

Dies kann allerdings zu einer erheblichen Erhöhung des Fehlers erster Art führen, was später in den Simulationstabellen zu sehen ist, siehe auch [4], [2] und [5].

Wir empfehlen, einen neuen Schätzer aus [11] zu verwenden, der einige wünschenswerte Eigenschaften besitzt (Erwartungstreue, Vielfaches einer  $\chi^2$ -verteilten Variablen und Unabhängigkeit vom Effektschätzer, wenn die wahren

Gewichte bekannt sind). Dieser Schätzer für die Varianz des geschätzten gemeinsamen Effektes  $\hat{\theta}$  lautet

$$s_{new}^2 = \widehat{\text{var}}_{new}(\hat{\theta}) = \frac{1}{k-1} \sum_{i=1}^k \frac{w_i(\hat{\tau}^2)}{w_{\Sigma}(\hat{\tau}^2)} \left\{ \hat{\theta}_i - \hat{\theta}(\hat{\tau}^2) \right\}^2 .$$

Für  $\hat{\tau}^2$  können die Schätzer aus den verschiedenen Verfahren nach Kapitel 5 eingesetzt werden. Das hier vorgeschlagene zweistufige Verfahren schätzt zunächst die Heterogenitätsvarianz mit  $\hat{\tau}_{POSnew}^2$  und setzt diesen Wert dann in  $s_{new}^2$  ein.

## 7 Konfidenzintervalle für $\theta$

Das Standardverfahren zur Berechnung von Konfidenzintervallen für  $\theta$  verwendet die Quantile der Normalverteilung. So erhält man ein  $(1 - \alpha)$ -Konfidenzintervall zu

$$\hat{\theta}(\hat{\tau}^2) \pm w_{\Sigma}^{-1/2}(\hat{\tau}^2) \cdot z_{1-\alpha/2} .$$

Bei Follmann/Proschan [6] wird hier die Verwendung der Quantile der  $t$ -Verteilung vorgeschlagen, was auch bereits in der Prozedur MIXED von SAS implementiert wurde. Bei homoskedastischen Studienvarianzen liegt als Nullverteilung eine exakte  $t$ -Verteilung mit dem nicht abgeschnittenen DerSimonian-Laird Schätzer für  $\tau^2$  vor.

Wir schlagen die Verwendung der  $t$ -Verteilung mit dem neuen Schätzer für die Varianz von  $\hat{\theta}$  vor, d.h.

$$\hat{\theta}(\hat{\tau}^2) \pm \sqrt{s_{new}^2} \cdot t_{k-1, 1-\alpha/2} .$$

Dieses Intervall hält zumindest approximativ das nominelle Niveau im Gegensatz zu den anderen Verfahren, was durch die folgende Simulationsstudie untermauert wird.

## 8 Simulationsstudie

In einer Simulationsstudie haben wir die tatsächlichen Niveaus verschiedener Testverfahren für den gemeinsamen Effekt  $\theta$  als Differenz von Mittelwerten geschätzt. Das nominelle Niveau war 5%. Wir haben die verschiedenen Methoden zur Schätzung der Heterogenitätsvarianz  $\tau^2$ , siehe Kapitel 5, und verschiedene Testmethoden, siehe Kapitel 7, angewendet. Die geschätzten Niveaus haben wir für unterschiedliche Konstellationen der Interstudienvarianzen

$(\nu_1^2, \dots, \nu_k^2)$ , der Studienumfänge  $(N_1, \dots, N_k)$  und der Heterogenitätsvarianz  $\tau^2$  bestimmt. Für jede Konstellation gab es 10000 Durchläufe.

Die SAS-Codes zur Berechnung der Schätzer in diesen Modellen sind bei verschiedenen Autoren, z.B. [12], [13] und [14], zu finden. Dabei kann man die Intrastudienvarianzen entweder über die Option `PARMSDATA` oder mit einem `WEIGHT`-Statement angeben. Die Schätzer für die Heterogenitätsvarianz wurden außer bei dem REML-Schätzer über IML bestimmt und dann mit dem `PARMS`-Statement an die Prozedur `MIXED` übergeben, um die entsprechenden Schätzer für den Gesamteffekt zu erhalten. Die Testergebnisse, die aus den Konfidenzintervallen resultieren, können für die  $t$ -Verteilung wie für die Normalverteilung, siehe [13], mit `MIXED` bestimmt werden. Nur die Tests mit dem neuen Varianzschätzer  $s_{new}^2$  mussten mit IML berechnet werden.

Es wurde von einer Datei `Daten` ausgegangen, die die Variablen `styid` für die Durchnummerierung der Studien, `theta_i` für die einzelnen Studieneffekte und `w_i` für die Inverse der Varianzschätzer ( $1/\hat{\nu}_i^2$ ) enthält. Ein möglicher SAS-Code zur Berechnung des REML-Schätzers im Modell mit festen Effekten ist

```
PROC MIXED DATA = Daten;
  CLASS styid;
  WEIGHT w_i;
  MODEL theta_i = / S;
  PARMS (1) / EQCONS = 1;
RUN;
```

Für das Modell mit zufälligen Effekten kann man die REML-Schätzer mit folgendem Programm bestimmen.

```
PROC MIXED DATA = Daten;
  CLASS styid;
  WEIGHT w_i;
  MODEL theta_i = / S;
  RANDOM styid;
  PARMS (1) (1) / EQCONS = 2;
RUN;
```

Zur Berechnung der anderen Gesamteffektschätzer wurden die Schätzer der Heterogenitätsvarianz ( $\hat{\tau}_*^2$ ) zunächst mit IML berechnet und dann im `PARMS`-Statement mit `PARMS ( $\hat{\tau}_*^2$ ) (1) / EQCONS = 1 TO 2;` an `MIXED` übergeben.

Die Ergebnisse sind beispielhaft in der Tabelle 3 für  $k = 6$  Studien, in der Tabelle 4 für die Konstellation des Beispiels 1 ( $k = 9$ ) und in der Tabelle 5 für die Konstellation des Beispiels 2 ( $k = 8$ ) aufgeführt.

Man erkennt in Tabelle 3, dass bei Vorliegen eines Modells mit festen Effekten ( $\tau^2 = 0$ ) das Standardverfahren (FIX) mit dem Normalverteilungsquantil ein geschätztes tatsächliches Niveau von 9.21% beim 5%-Niveau Test aufweist. Die Verfahren mit dem  $t$ -Quantil liefern zu kleine Niveaus ( $< 3\%$ ). Im dritten



**Tabelle 3:** Geschätzte Niveaus (%) verschiedener Tests bzgl.  $\theta$  zum 5%-Niveau ( $k = 6$  Studien, Intrastudienvarianzen (0.1, 0.1, 0.15, 0.15, 0.2, 0.2), Studienumfänge (30, 30, 20, 20, 10, 10))

Quantil	Methode	$\tau^2 = 0$	$\tau^2 = 0.1$	$\tau^2 = 0.5$	$\tau^2 = 1$	$\tau^2 = 10$
$z$	FIX	9.21	19.17	42.41	54.21	84.42
	DSL	5.40	8.48	10.48	10.75	10.42
	REML	5.38	8.56	10.41	10.37	10.37
	HEDGES	5.90	9.14	10.79	10.67	10.41
$t$	FIX	2.92	8.82	29.55	42.66	79.57
	DSL	1.28	2.96	4.58	5.03	4.86
	REML	1.22	2.89	4.60	4.96	4.67
	HEDGES	1.54	3.53	4.83	5.00	4.67
$t, s_{new}^2$	FIX	5.90	6.11	6.02	6.27	6.03
	DSL	5.84	5.73	5.02	5.13	4.69
	REML	5.77	5.71	5.07	5.09	4.68
	HEDGES	5.82	5.73	4.92	4.99	4.67
	POS <sub>new</sub>	5.48	5.40	4.87	4.95	4.67

unteren Teil der Tabelle sind die Verfahren mit dem  $t$ -Quantil und dem neuen vorgeschlagenen Schätzer für die Varianz des gemeinsamen Effektschätzers dargestellt. Wird hier mit dem Modell mit festen Effekten auch ausgewertet, so erhält man jetzt ein geschätztes Niveau von 5.90%.

Wertet man dagegen das Modell mit festen Effekten im allgemeineren Modell mit zufälligen Effekten aus, so hat man ein Niveau bei Verwendung des neuen Schätzers für die Heterogenitätsvarianz  $\tau^2$  von 5.48% in der hier vorgestellten Parameterkonstellation. Liegt tatsächlich eine Heterogenitätsvarianz ungleich Null vor, so hält das neue Verfahren vergleichsweise gut das nominelle Niveau. Werden die klassischen Methoden zur Schätzung von  $\tau^2$  verwendet, so gleichen sich die Niveaus unter Verwendung der  $t$ -Verteilung und  $s_{new}^2$  relativ gut dem vorgegebenen Niveau an.

Im mittleren Teil der Tabelle wird ersichtlich, dass die Verwendung nur des  $t$ -Quantils bei kleiner Interstudienvarianz zu konservativ ist.

In der Simulationsstudie wurden weitere Konstellationen betrachtet, die ähnliche Ergebnisse zeigten.

In Tabelle 4 ist die Konstellation des ersten Beispiels mit Intrastudienvarianzen (40.586, 2.047, 15.281, 150.222, 20.192, 1.224, 95.376, 8.032, 20.694) und Studienumfängen (312, 64, 246, 36, 22, 110, 68, 292, 112) dargestellt. Als Heterogenitätsvarianz wurde zunächst der geschätzte REML-Wert und zum Vergleich die Null unterstellt. Wird  $\tau^2 = 685$  unterstellt, so erhält man zu schmale Konfidenzintervalle mit dem Normalverteilungsquantil, insbesondere beim

Der Simonian-Laird-Schätzer. Der REML-Schätzer zeigt gute Ergebnisse mit dem  $t$ -Quantil, wie auch die neuen Verfahren.

**Tabelle 4:** Geschätzte Niveaus (%) verschiedener Tests bzgl.  $\theta$  zum 5%-Niveau ( $k = 9$  Studien, Konstellation aus Beispiel 1)

Quantil	Methode	$\tau^2 = 0$	$\tau^2 = 685$
$z$	FIX	5.32	92.39
	DSL	3.46	12.84
	REML	3.79	8.77
	HEDGES	3.37	8.76
$t$	FIX	2.23	91.00
	DSL	1.37	8.55
	REML	1.50	5.16
	HEDGES	1.36	5.35
$t, s_{new}^2$	FIX	5.08	34.26
	DSL	4.86	5.34
	REML	4.97	5.19
	HEDGES	4.56	5.20
	POS <sub>new</sub>	4.20	5.15

Die Ergebnisse für die Konstellation aus Beispiel 2 mit Intrastudienvarianzen (0.005, 0.009, 0.009, 0.016, 0.006, 0.010, 0.063, 0.005) und Studienumfängen (94, 56, 147, 24, 66, 62, 54, 93) sind in Tabelle 5 dargestellt. Hier werden für  $\tau^2$  unterschiedliche Werte unterstellt, da für kleine  $\tau^2$  die Niveaus stark schwanken. Man erhält wieder für kleine  $\tau^2$  zu konservative Ergebnisse mit dem  $t$ -Quantil, während die neuen Verfahren gute Niveaus erzielen.

Aufgrund der vorliegenden Simulationsergebnisse empfehlen wir die Verwendung des  $t$ -Quantils mit dem neuen Schätzer für die Varianz des Gesamteffektschätzers  $\hat{\theta}$  und der Schätzung der Heterogenitätsvarianz  $\tau^2$  mit den Standardverfahren bzw. auch mit dem neuen nichtnegativen Schätzer.

**Tabelle 5:** Geschätzte Niveaus (%) verschiedener Tests bzgl.  $\theta$  zum 5%-Niveau ( $k = 8$  Studien, Konstellation aus Beispiel 2)

$\alpha = 5\%$	$\tau^2 =$	0	0.00026	0.0066	0.0353	1
<b>Quantil</b>	<b>Methode</b>	(FIX)	(REML)	(DSL)	(HEDGES)	
$z$	FIX	5.52	6.03	17.36	44.03	88.05
	DSL	4.00	4.22	7.80	9.63	10.41
	REML	4.12	4.14	7.96	9.67	9.54
	HEDGES	3.96	4.40	9.34	10.85	9.58
$t$	FIX	2.29	2.45	10.28	35.78	85.87
	DSL	1.46	1.58	3.82	5.52	6.03
	REML	1.53	1.57	3.92	5.46	5.37
	HEDGES	1.54	1.78	5.05	6.56	5.40
$t, s_{new}^2$	FIX	5.15	5.04	6.22	7.64	8.41
	DSL	5.08	4.92	5.72	5.63	5.47
	REML	5.12	4.93	5.78	5.65	5.45
	HEDGES	4.92	4.81	5.57	5.34	5.44
	POS <sub>new</sub>	4.51	4.46	5.04	5.20	5.43

## Literatur

- [1] Normand, S.-L. T. (1999). Tutorial in Biostatistics. Meta-analysis: formulating, evaluating, combining and reporting. *Statistics in Medicine*, **18**, 321-359.
- [2] Li, Y., Shi, L. and Roth, H. D. (1994): The bias of the commonly-used estimate of variance in meta-analysis. *Commun. Statist. - Theory Meth.*, **23**, 1063-1085.
- [3] Hedges, L. V. (1983). A random effects model for effect sizes. *Psychological Bulletin*, **93**, 388-395.
- [4] Böckenhoff, A. and Hartung, J. (1998). Some corrections of the significance level in meta-analysis. *Biometrical Journal*, **40**, 937-947.
- [5] Ziegler, S. und Victor, N. (1999). Gefahren der Standardmethoden für Meta-Analysen bei Vorliegen von Heterogenität. *Informatik, Biometrie und Epidemiologie in Medizin und Biologie*, **30**, 131-140.
- [6] Follmann, D. A. and Proschan, M. A. (1999). Valid inference in random effects meta-analysis. *Biometrics*, **55**, 732-737.

- [7] DerSimonian, R. and Laird, N. (1986). Meta-analysis in clinical trials. *Controlled Clinical Trials*, **7**, 177-188.
- [8] Harville, D. A. (1977). Maximum Likelihood approaches to variance component estimation and to related problems. *Journal of the American Statistical Association*, **72**, 320-338.
- [9] Hartung, J. (1981). Nonnegative minimum biased invariant estimation in variance component models. *Ann. Statist.*, **9**, 278-292.
- [10] Hartung, J. (1999a). A short-cut method for computing positive variance component estimates. In: F. Krumm and V. S. Schwarze (Hrsg.), *Quo vadis geodesia ...? - Festschrift for Erik W. Grafarend on the occasion of his 60th birthday (Part 1)*, Technical Reports, Department of Geodesy and Geoinformatics (Report No. 1999.6-1), University of Stuttgart, 151-154.
- [11] Hartung, J. (1999b). An alternative method for meta-analysis. *Biometrical Journal*, **41**, 901-916.
- [12] Stijnen, T. (2000). Letters to the editor: tutorial in biostatistics. Meta-analysis: formulating, evaluating, combining and reporting, by S.-L. Normand. *Statistics in Medicine*, **19**, 753-761.
- [13] Witte, S. und Kuß, O. (2001). Meta-Analysen mit SAS: Analysen und Graphiken. In: Schumacher, E. und Streichfuss, K. (Hrsg.), *Proceedings der 5. Konferenz der SAS-Anwender in Forschung und Entwicklung (KSFE)*, 8.-9. März 2001, Stuttgart, 481-487.
- [14] Houwelingen, H. C., Arends, L. R. and Stijnen, T. (2002). Tutorial in Biostatistics. Advanced methods in meta-analysis: multivariate approach and meta-regression. *Statistics in Medicine*, **21**, 589-624.