

# Meta-Analysen mit SAS: Analysen und Graphiken

Steffen Witte

Universität Heidelberg, Abteilung Medizinische Biometrie  
INF 305, 69120 Heidelberg  
E-Mail: witte@imbi.uni-heidelberg.de

Oliver Kuß

Universität Halle-Wittenberg, Institut für Medizinische Epidemiologie, Biometrie und  
Informatik  
Magdeburger Str. 27, 06097 Halle/Saale  
E-Mail: oliver.kuss@medizin.uni-halle.de

## Abstract

Es werden drei SAS-Makros für Meta-Analysen vorgestellt. Zum einen (%metacont) handelt es sich dabei um ein Makro für die numerische Analyse, wenn stetige Zielgrößen in den Einzelstudien herangezogen werden, zum anderen um zwei Makros zur graphischen Unterstützung: erstens zur Darstellung des Funnel-Plots zur Identifizierung von Publikationsbias (%metafunn) und zweitens zur Darstellung des Forest-Plots (%metaki) zur Unterstützung der Ergebnisbeurteilung. Auf die Verwendung von PROC MIXED wird kurz eingegangen.

## 1. Einführung

Meta-Analysen (MA) dienen der - insbesondere numerischen - Zusammenfassung von einzelnen Studienergebnissen. Diese Fragestellung ist insbesondere dann interessant, wenn sich die Einzelergebnisse widersprechen. In den letzten 30 Jahren des vergangenen Jahrhunderts wurden viele Methoden dazu entwickelt und Software-Pakete bereitgestellt (EpiMeta [2], Meta-Analysis [7], RevMan [9]) und zu einigen Verfahren SAS-Makros veröffentlicht [6]. Die Software-Pakete zeichnen sich durch zum Teil einfache Handhabung aus, sind aber i.d.R. in ihrer Anwendung auf festgelegte Methoden beschränkt. Die numerische Auswertung der MA (siehe Kapitel 3) und die graphische Darstellung der Ergebnisse (siehe Kapitel 4) mit SAS soll in dieser Arbeit für stetige Zielgrößen vorgestellt werden. Dabei ist eine einfache Verwendung ebenso wichtig, wie die informative Ausgabe und die flexible Anpassungsfähigkeit der Makros. Dabei wird im wesentlichen die Methode von Whitehead & Whitehead [11] verwendet und ein klinisches Beispiel zur Illustration herangezogen, bei dem die Ergebnisse aus randomisierten klinischen Studien (RCTs) zusammengefasst werden sollen.

Beispiel: Arthrosebehandlung

- Indikation: Arthrose (Gelenkverschleiß)
- Behandlung: i.d.R. symptomatisch, d.h. schmerzlindernd und entzündungshemmend
- Fragestellung: Wirkung von S-Adenosylmethionin (SAM)?
- Identifikation von 8 RCTs (versus NSAID, d.h. aktive Kontrollgruppe)
- Zielgröße: standardisierte Differenz der Messwerte am Ende der Behandlungsphase
- Skalierung: stetig (sonst bei Meta-Analysen oft binomial)

Es soll hier nicht auf die Problem von MA im allgemeinen und dieser im speziellen eingegangen werden, nur soviel, dass ein Wirksamkeitsnachweis durch die Studien mit aktiver Kontrolle nur erfolgen könnte, wenn eine Überlegenheit gegenüber den Kontrollmedikationen gezeigt werden könnte.

Folgende Daten können aus den 8 identifizierten Studien gewonnen werden, dabei bezeichnet \*\_e die Daten der Experimentalgruppe (SAM) und \*\_c die der Kontrollgruppe.

Tabelle 1: Datenstruktur der Einzelstudien für die Meta-Analyse

refid	no_e	no_c	mean_e	mean_c	std_e	std_c
CAP1985	50	53	4.26	5.25	3.28	3.77
CAR1980	29	25	1.69	2.24	0.54	0.66
CAR1987N	197	195	0.70	0.68	0.54	0.66
CEC1980	48	39	3.06	3.90	2.79	2.73
CUC1980	114	102	0.86	1.09	0.77	0.87
GLO1985	64	59	5.89	6.13	3.90	3.61
MAC1987	22	23	9.50	7.00	2.63	2.16
PEL1980	24	22	1.29	1.97	0.85	0.69

## 2. Analysen

Für die Berechnung von MA-Schätzern für stetige Zielgrößen wurden bereits mehrere Verfahren vorgeschlagen, unter anderem von [4, 8, 11]. In dieser Arbeit wird die Vorgehensweise von Whitehead & Whitehead [11] verwendet.

Die Ergebnisse werden sowohl unter Zugrundelegung des "fixed effect model" (FEM) als auch unter Zugrundelegung des "random effects model" (REM) berechnet. Das FEM basiert auf der Annahme, dass der "wahre" Behandlungseffekt in den Studien ( $\theta_i$ ) für alle Studien gleich ist ( $\theta = \theta_1 = \dots = \theta_k$ ). Im REM wird angenommen, dass der Behandlungseffekt in den einzelnen Studien ( $\theta_i$ ) eine Zufallsvariable ist (normalverteilt um einen globalen Effekt  $\theta$ ); die Studien in der MA werden als Zufallsstichprobe aus der Grundgesamtheit aller möglichen Studien aufgefaßt. Die REM-Analyse kann als Sensitivitätsanalyse für die Ergebnisse der FEM-Analyse benutzt. Falls beide Ergebnisse stark differieren, kann dem FEM-Ergebnis nicht vertraut werden. Für bestimmte Formen der Heterogenität ist das REM das adäquate Modell. Nicht in allen Situationen stellt jedoch das REM-Resultat den Sachverhalt richtig dar. Bezeichne  $k$  die Anzahl der Studien,  $n_{ei}$  bzw.  $n_{ci}$  die Stichprobenumfänge,  $\bar{x}_{ei}$  bzw.  $\bar{x}_{ci}$  die empirischen Mittelwerte,  $s_{ei}$  und  $s_{ci}$  die Summen und  $q_{ei}$  bzw.  $q_{ci}$  die Quadratsummen der Messwerte, jeweils aus der  $i$ -ten Studie. Die Indizes  $e$  und  $c$  kennzeichnen die Experimental- bzw. Kontrollgruppe. Damit sind

$$D_i^2 = \frac{(q_{ei} - s_{ei}^2/n_{ei}) + (q_{ci} - s_{ci}^2/n_{ci})}{(n_{ei} + n_{ci} - 2)} \quad \text{und} \quad (D_i^*)^2 = \frac{(q_{ei} + q_{ci}) - (s_{ei} + s_{ci})^2 / (n_{ei} + n_{ci})}{(n_{ei} + n_{ci})}$$

die gepoolte empirische Varianz und die ungepoolte empirische Varianz in der  $i$ -ten Studie. Mit den Bezeichnungen

$$Z_i = \frac{n_{ei}n_{ci}}{(n_{ei} + n_{ci})D_i^*} (\bar{x}_{ei} - \bar{x}_{ci}) \quad \text{und} \quad V_i = \frac{n_{ei}n_{ci}}{n_{ei} + n_{ci}} - \frac{Z_i^2}{2(n_{ei} + n_{ci})}$$

ergeben sich als Schätzer für die Effekte in den einzelnen Studien  $\hat{\theta}_i = Z_i / V_i$  und als Schätzer für deren Varianzen  $w_i^{-1} = V_i^{-1}$ . Asymptotisch sind diese Schätzer mit der gewöhnlichen standardisierten Differenz bzw. deren Varianz gleich:

$$\hat{\theta}_i = Z_i / V_i \cong (\bar{x}_{ei} - \bar{x}_{ci}) / D_i \quad \text{bzw.} \quad w_i^{-1} = V_i^{-1} \cong \frac{n_{ei} + n_{ci}}{n_{ei}n_{ci}}.$$

Die zweiseitigen 95% Konfidenzintervalle für  $\theta_i$  berechnen sich unter der Normalverteilungsannahme nach  $\hat{\theta}_i \pm 1.96\sqrt{1/w_i}$ .

Im FEM wird für die Effektschätzer angenommen:  $\hat{\theta}_i \sim N(\theta, w_i^{-1})$ . Der gemeinsame Effekt  $\theta$  wird geschätzt durch  $\hat{\theta} = \sum w_i \hat{\theta}_i / \sum w_i$ . Das dazugehörige zweiseitige 95% Konfidenzintervall berechnet sich aus der Normalapproximation durch  $\hat{\theta} \pm 1.96\sqrt{1/\sum w_i}$ . Die Null-Hypothese  $H_0 : \theta = 0$ , d.h. kein globaler Effekt, wird mit der Teststatistik

$$T = \sum w_i \hat{\theta}_i / \sqrt{\sum w_i} \stackrel{H_0}{\sim} N(0,1)$$

überprüft.

Die Null-Hypothese der Homogenität  $H_0 : \theta_1 = \Lambda = \theta_k$  wird mit der Heterogenitätsstatistik

$$Q = \sum w_i (\hat{\theta}_i - \hat{\theta})^2 \stackrel{H_0}{\sim} \chi_{k-1}^2$$

überprüft. Falls Heterogenität vorliegt, liefern die Methoden des FEM unzulässige Ergebnisse.

Für bestimmte Formen der Heterogenität (Normalverteilungsannahme) ist das REM das passende Modell. In diesem Modell ist der wahre Effekt der einzelnen Studien  $\theta_i$  - im Gegensatz zum FEM - eine normalverteilte Zufallsvariable mit Erwartungswert  $\theta$  und Varianz  $\tau^2$ . Die Schätzung für  $\tau^2$  (Varianz zwischen den Studien) wird nach DerSimonian und Laird [1] vorgenommen:

$$\hat{\tau}^2 = \max \left\{ 0, (Q - (k-1)) / (\sum w_i - \sum w_i^2 / \sum w_i) \right\}.$$

Die Schätzer für den globalen Effekt  $\theta$  und die Teststatistik zur Prüfung der Hypothese  $H_0 : \theta = 0$  werden wie im FEM gebildet; lediglich die Gewichte  $w_i$  sind durch

$w_i^* = (w_i^{-1} + \hat{\tau}^2)^{-1}$  zu ersetzen:

$$\hat{\theta}^* = \sum w_i^* \hat{\theta}_i / \sum w_i^* \quad \text{und} \quad T^* = \sum w_i^* \hat{\theta}_i / \sqrt{\sum w_i^*} \stackrel{H_0}{\sim} N(0,1).$$

Das SAS-Makro %metacont

- berechnet alle Parameter, die für die MA von Wichtigkeit sind,
- berechnet einen oder mehrere Typen von MA-Schätzern (types),
- berechnet immer den FE-Test und den RE-Test,
- verwendet lediglich "basic"-Programmierung,
- leitet die Ergebnisse in SAS-Datensätze zur Weiterverarbeitung (z.B. Graphiken) und
- liefert einen normalen output.

Die Verwendung des Makros in obigem Beispiel lautet %metacont(inset = select1, method = 2, nt = no\_e, nc = no\_c, mt = mean\_e, mc = mean\_c, stdt = std\_e, stdc = std\_c, types = pet rough white coc glass); der Parameter method spezifiziert dabei, ob Quadratsummen (method=1) oder Standardabweichungen (method=2) übergeben werden.

Die Ausgabe erfolgt in zwei Teilen, zum einen werden die Ergebnisse auf Studienbasis mit dem Effekt-Schätzer ( $\hat{\theta}_i$ , theta\_i), dem 95%-Konfidenzintervall auf der Normalverteilung basierend, den Gewichten für die Berechnung des FE-Tests ( $w_i$ , femw\_i) und des RE-Tests ( $w_i^*$ , weights) sowie der Beiträge der einzelnen Studien zur Heterogenitätsstatistik ( $w_i(\hat{\theta}_i - \hat{\theta})^2$ , q\_het\_i) ausgegeben. Die verschiedenen ausgewählten Typen von Berechnungsmethoden werden dabei untereinander ausgegeben.

Tabelle 2: Ausgabe von %metacont auf Studienebene

*** METACONT (2): Effects and contributions to heterogeneity ***							
type	refid	lower_i	theta_i	upper_i	femw_i	weights	q_het_i
Whitehead	Capretto 1985	0.67209	0.28377	0.10456	25.4745	4.93408	0.4736
...							
Whitehead	Pellegrini1980	1.58842	-0.97946	-0.37051	10.3592	3.84690	7.1718

Der zweite Teil der Ausgabe beinhaltet die Ergebnisse der Meta-Analyse, wobei jeder Typ in zwei Zeilen dargestellt wird, eine für das FEM und eine für das REM. Darin enthalten ist die Heterogenitätsstatistik (Q, t\_het), die Freiheitsgrade (q\_het\_df) und der dazugehörige p-Wert

( $p_{het}$ ), der Schätzer für die Varianz zwischen den Studien ( $\hat{\tau}^2$ , tau), der MA-Effekt-Schätzer (FEM:  $\hat{\theta}$ , REM:  $\hat{\theta}^*$ , theta) mit dem 95%-Konfidenzintervall unter der Annahme der Normalverteilung (lower, upper), die Teststatistik (FEM: T, REM:  $T^*$ ,  $t_{theta}$ ) und der dazugehörige p-Wert für den FE- bzw. RE-Test ( $p_{theta}$ ).

Tabelle 3: Ausgabe der Meta-Analyse-Ergebnisse von %metacont

*** METACONT (2): tests and estimates for overall effects ***										
type	model	t_het	p_het	p_theta	q_het_df	Lower	upper	tau	theta	t_theta
White-head	FEM	40.3165	<.0001	0.0173	7	-0.2688	-0.0260	0.0000	-0.1474	5.6663
White-head	REM	40.3165	<.0001	0.1948	7	-0.5305	0.1081	0.1634	-0.2112	1.6812

Die Verwendung von PROC MIXED für die Auswertung von Meta-Analysen wird vor allem bei [5] beschrieben. Die Meta-Analyse wird dabei als gemischtes lineares Modell mit spezieller (fester) Kovarianzstruktur (REPEATED-Statement) aufgefasst. Die geschätzten Studienvarianzen  $V_i^{-1}$  werden im PARMS-Statement übergeben und mit dem EQCONS=Statement als feste, nicht zu schätzende Werte gesetzt. PROC MIXED berechnet standardmäßig den Test auf Effekt mit einer t-Verteilung [3], kann aber mit einem Trick dazu gebracht werden, die Normalverteilung als Prüfverteilung zu benutzen (model theta\_i=int / s cl noint ddf=10000; dabei ist „int“ eine Variable, die für jede Einzelstudie den Wert 1 hat).

```
proc mixed method=ml data=all1;
  class study;
  * model theta_i= / s cl;
  model theta_i=int / s cl noint ddf=10000;
  repeated / group=id;
  parms (0.03925) (0.08341) (0.01021) (0.04702)
        (0.01876) (0.03259) (0.10323) (0.09653) / eqcons=1 to 8;
run;
```

Für das REM muss zusätzlich noch ein RANDOM-Statement eingefügt werden, um die zufällige Verteilung der Studieneffekte um den gemeinsamen Mittelwert zu modellieren. Der Schätzer für die Varianz zwischen den einzelnen Studien,  $\tau^2$ , wird als erster Parameter im PARMS-Statement, ebenfalls als fester Wert, gesetzt.

```
proc mixed method=ml data=all1;
  class study;
  * model theta_i= / s cl;
  model theta_i=int / s cl noint ddf=10000;
  random int / subject=id s;
  repeated / group=id;
  parms (0.16342)
        (0.03925) (0.08341) (0.01021) (0.04702)
        (0.01876) (0.03259) (0.10323) (0.09653) / eqcons=1 to 9;
run;
```

Die Verwendung von PROC MIXED erlaubt des weiteren noch einige Spezialanwendungen: So können z.B. verbesserte Schätzer für  $\tau^2$  berechnet werden, indem man mit dem PARMS-Statement ein Grid von Werten absuchen lässt und dann denjenigen Wert von  $\tau^2$  als neuen Schätzer verwendet, der die Likelihood des Modells maximiert oder indem man METHOD=REML setzt, wodurch man einen iterierten DerSimonian-Laird-Schätzer erhält. Durch Verwendung der s-Option im RANDOM-Statement erhält man eine neue Klasse von Schätzern für die Effekte in den Einzelstudien, die so genannten Empirical-Bayes-Schätzer. In deren Berechnung fließen jeweils Information aus den anderen Studien mit ein, wodurch

diese zum Gesamtschätzer hin verschoben („shrunked“) sind. Kleine Studien werden dabei stärker verschoben als große.

### 3. Graphiken

Die Darstellung von Studienergebnissen in Graphiken ist vor allem bei MA eine wichtige Methode. Dabei spielen Funnel-, Forest- und Radial-Plot eine besondere Rolle. Hier sollen lediglich die beiden ersten Graphiken verwendet werden.

Der Funnelplot ("Trichter-Graphik") wird vor allem als Methode zum Entdecken von Publikationsbias (Studien werden z.T. selektiv publiziert, unter Umständen vom Ergebnis abhängig) verwendet. Bei Nicht-Vorliegen von Publikationsbias erwartet man, dass die Punkte der dargestellten Studien einen Trichter bilden. Dabei wird auf der x-Achse ein Maß für Einfluss der Studie in die MA gewählt (Stichprobenumfang  $n$ , Gewicht im FEM  $w_i$ ) und auf der y-Achse der geschätzte Effekt der Studie, so dass sich ein Scatterplot ergibt. Als Referenzlinie wird in der Regel der gemeinsame Effekt (MA-Effekt-Schätzer,  $\hat{\theta}$ ) verwendet, um diesen sollten die Punkte gleichmäßig streuen.

Bei dem Macro %metafunn handelt es sich im wesentlichen ein Aufruf von PROC GPLOT (proc gplot; plot effect \* n;), bei Bedarf wird das autocall Makro %label verwendet. Der Aufruf in dem Beispiel lautet folgendermaßen, wobei die Makrovariable estfem den MA-Effekt-Schätzer  $\hat{\theta}$  enthält.

```
%metafunn(inset =all1(where=(type="Whitehead")), effect= theta_i, efflow= -2, effupp=2, n = femw_i, ntxt = Weights (FEM), nlow = 0, nupp = 120, nby = 20, vref = %quote(&estfem.);
```

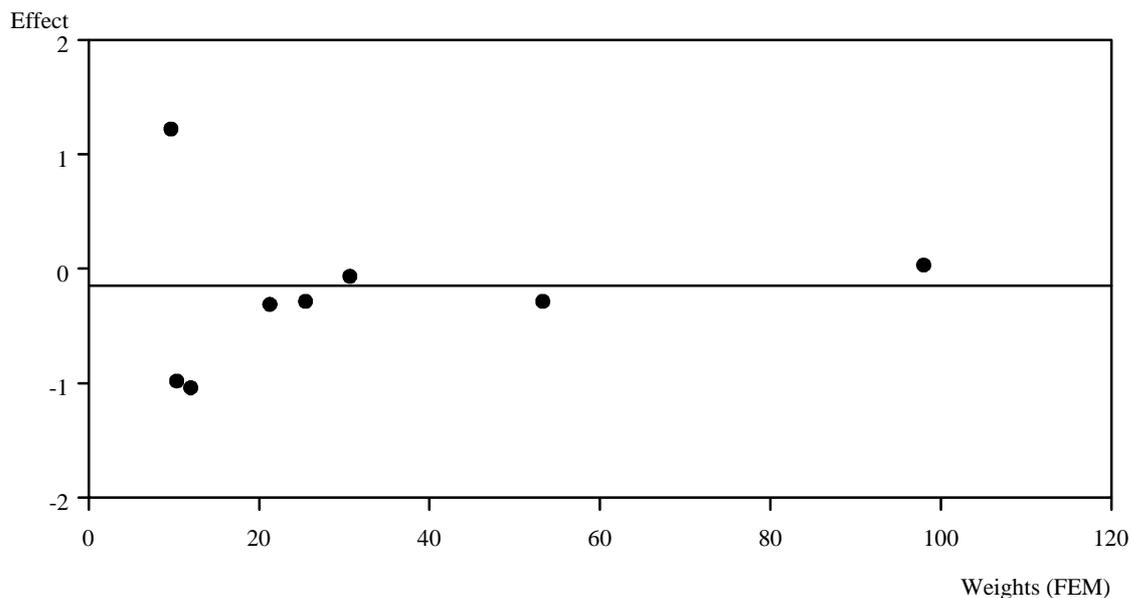


Abbildung 1: Funnelplot

Fügt man außerdem die Parameter label = refid und lpos = lpos hinzu werden labels an die einzelnen Punkte geschrieben, um eine Identifikation zu ermöglichen. lpos ist eine spezielle Variable mit den Ziffern {1, ...,9}, die die Positionierung angibt. Die Verwendung ist im Makro beschrieben. Im Beispiel haben CEC1980: lpos = "1", CAP1985 und PEL1985: lpos = "9" und alle übrigen lpos = "3".

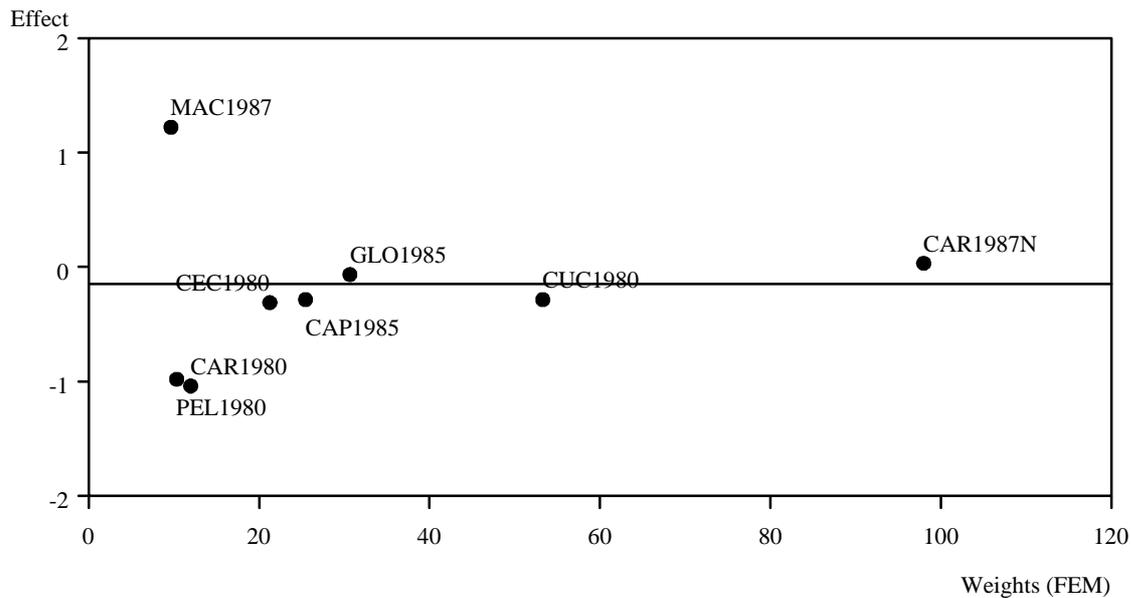


Abbildung 2: Funnelplot mit labels

Der Forestplot ist die gebräuchlichste Darstellung für die Ergebnisse von MA. Dabei werden die Effekte und Konfidenzintervalle der Einzelstudien und der MA untereinander dargestellt, was manchmal an einem Nadelbaum erinnert (daher der Name „Forest-Plot“). Von Wang & Bushman [10] wird die Verwendung von proc timeplot empfohlen. Das Ergebnis ist allerdings nicht zufriedenstellend.

```
proc timeplot;
```

```
  plot lower="[" theta="*" upper="]" / overlay hiloc ref=0 refchar="0";
```

```
  id refid;
```

```
run;
```

Die Ausgabe erfolgt über eine ASCII-Graphik und sieht folgendermaßen aus.

```
refid      lower  theta  upper      min                                max
          -1.604994284                                1.8520280156
*-----*-----*
CAP1985   -0.67  -0.28  0.10      [-----*-----0-]
CAR1980   -1.60  -1.04  -0.47      [-----*-----]
CAR1987N  -0.16   0.03  0.23      [---0*---]
CEC1980   -0.73  -0.31  0.12      [-----*-----0-]
CUC1980   -0.55  -0.28  -0.02      [---*-----]
GLO1985   -0.42  -0.06  0.29      [-----*0-----]
MAC1987   0.59   1.22  1.85      [-----*-----]
PEL1980  -1.59  -0.98  -0.37      [-----*-----]
FEM       -0.27  -0.15  -0.03      [---*---]
REM       -0.53  -0.21  0.11      [-----*-----0-]
*-----*-----*

```

Abbildung 3: Forestplot mit proc timeplot

Eine andere Möglichkeit ist die Verwendung des SAS-Makros %metaki. Außer dem Datensatznamen und den Variablenbezeichnungen definiert der parameter name= die Variable der Studienidentifikation, das Format namefmt= die Beschriftung auf der linken Seite der Graphik und der Parameter by= spezifiziert die Gruppen, hier ist group=1 für alle Studien und group=2 für die MA-Schätzer im FEM und REM. Der Aufruf des Makros ist %metaki(inset = metaki, est = theta, kilist = lower, kiuest = upper, by = group, name = refid, namefmt = \$refid., mrad = 0.2); wobei mrad= die Größe des Punktschätzers in der Graphik vorgibt und metaki die beiden Datensätze, die %metacont erzeugt hat enthält.

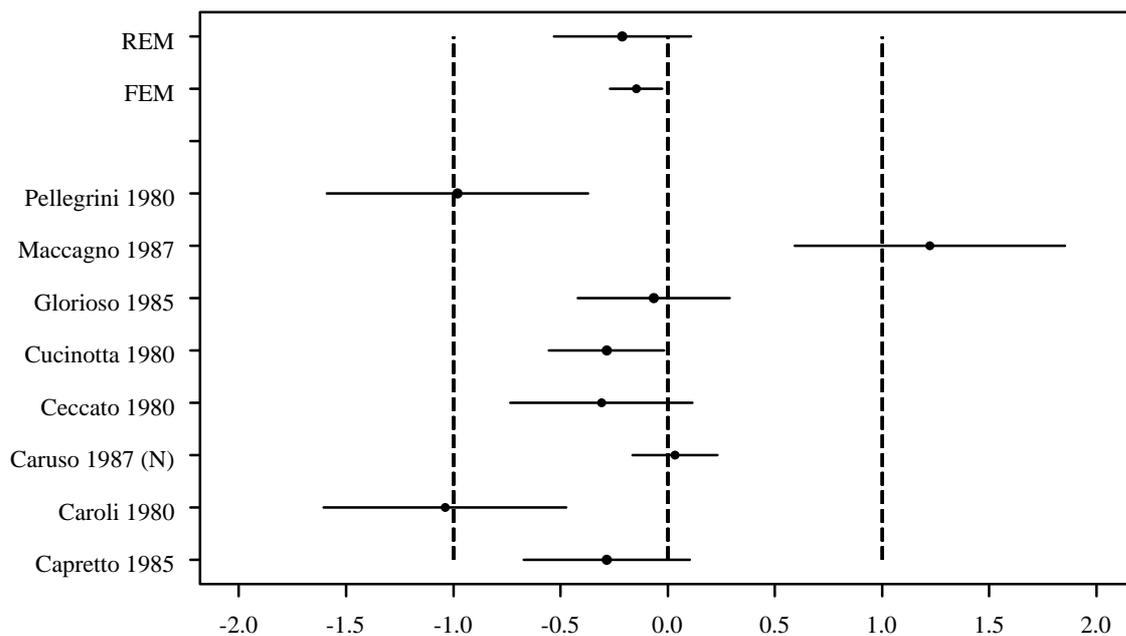


Abbildung 4: Forestplot mit %metaki

## 4. Diskussion

Die Verwendung der Makros %metacont zur Analyse von Meta-Analysen mit stetigen Zielgrößen sowie %metafun und %metaki zur graphischen Darstellung sind sehr hilfreich und einfach zu verwenden, zum anderen können beide Makros leicht an eigene Erfordernisse angepasst werden.

Makros und ein Beispielprogramm können beim Autor angefordert werden.

### Danksagung

Das Beispiel stammt aus einer Meta-Analyse, die von der Arzneimittelkommission der deutschen Ärzteschaft unterstützt wurde.

## 5. Literatur

- [1] DerSimonian R, Laird NM, Meta-Analysis in clinical trials: Controlled Clinical Trials 1986, 7, 177-188
- [2] EpiMeta: <http://www.cdc.gov/epo/dpram/epimeta/epimeta.htm>
- [3] Hartung J, Makambi KH, Simple t-distribution based tests for meta-analysis: to appear
- [4] Hedges LV, Olkin I, Statistical Methods for Meta-Analysis, Academic Press 1985
- [5] Houwelingen HC, Arends LR, Stijnen T, Tutorial in Biostatistics, Advanced Methods in Meta-Analysis: Multivariate Approach and Meta-Regression, to appear
- [6] Kuß O, Koch A, Meta-analysis macros for SAS, The Statistical Software Newsletter, 1996, 22, 325-333
- [7] MetaAnalysis: <http://www.meta-analysis.com>,  
[http://www.yorku.ca/faculty/academic/schwarze/meta\\_e.htm](http://www.yorku.ca/faculty/academic/schwarze/meta_e.htm)
- [8] Petitti DB, Meta-Analysis, Decision Analysis, and Cost-Effectiveness Analysis, Oxford University Press 1994
- [9] RevMan: <http://www.cochrane.de/cc/cochrane/revman.htm>
- [10] Wang MC, Bushman BJ, Integrating Results through Meta-Analytic Review Using SAS® - Software: SAS Institute Inc. 1999, Cary, NC, USA
- [11] Whitehead A, Whitehead J, A General Parametric Approach to the Meta-Analysis of Randomized Clinical Trials: StatMed 10 (1991), 1665-1677