Meta-analysis macros for SAS

Oliver Kuss & Armin Koch

Institut für Medizinische Biometrie und Informatik, Abteilung Medizinische Biometrie, Universität Heidelberg, Im Neuenheimer Feld 305, D-69120 Heidelberg, Germany

Summary: This paper presents a collection of macros for The SAS System to perform meta-analyses of clinical trials where the results of a single trial can be displayed in a fourfold-table. These macros provide the diagnostic plots most of which can be found in almost every meta-analysis publication (Funnel-Plot, confidence interval plot, Galbraith-Plot, Sensitivity-Plot by Thompson) as well as the computation of the different estimators of the treatment effect along with their summary estimators. Results from a meta-analysis on the comparison of low-molecular weight heparin and standard heparin for the prophylaxis of thromboembolic events are used for illustrative purposes.

(SSNinCSDA 22, 325-333 (1996))

Keywords: SAS-macros, meta-analysis.

Received: November 1995 Revised: May 1996

I. Introduction

Boissel et al. (1988) have quoted meta-analysis as a quantitative systematic approach for the combination and integration of results from various studies with the aim to gain information that is not available from any of the single studies. Primarily developed in the social and behavioral sciences the analytical method has become extremely popular in the field of medicine during the last decade. The majority of methods necessary to perform meta-analysis in practice are simple and, as there is no principal difference between a meta-analysis and the analysis of a multicentre-trial, available in most standard statistical packages. As meta-analysis often attempts to compare studies that are, at best, similar, heterogeneity analysis gains more importance than it might have in the analysis of multicentre-trials.

Due to the fact that meta-analysis is becoming more common in many disciplines, Normand (1995) states a growing need for meta-analysis software and gives some recommendations what methods should be covered by a meta-analysis software package. Below others she emphasizes the benefit of graphical methods and the random effects framework.

Although programmable systems, like The SAS System, provide the basic tools to do simple calculations and to plot graphs, usually a considerable amount of time doing the same things over and over again is needed. We therefore want to present a set of macros, most of which can be readily used, to prepare the graphs typically found in almost every meta-analysis publication. These are accompanied by a macro that provides computations for the general fixed and general random effects approaches. For illustrative purposes a meta-analysis on the comparison of low-molecular weight heparin (LMWH) and standard unfractionated heparin (UFH) for the prophylaxis of thromboembolic events will be used. See Table 1 for the original studies that were included into the meta-analysis.

II. The example

Major surgical interventions, anaesthesia and the often necessary longer confinement to bed lead to an increased risk of the development of deep vein thrombosis (DVT) in surgical patients. Some of the thrombi might move to the lung and lead to pulmonary embolism which is an event of high fatality. As an enormous number of surgical int-

Table 1: Double blind clinical trials in general abdominal and orthopedic surgery for the comparison of LMWH and UFH for the prevention of thromboembolic events after major surgical interventions: **Dose**: daily administered dose in the LMWH-group ≤/> 3400 IU (low / high); **DVT**: end point deep vein thrombosis in the two treatment groups (LMWH and UFH) **WH:** end point wound haematoma in the two treatment groups (LMWH and UFH). (* Three-armed studies were treated as two two-armed studies with the same UFH-group)

				DVT		WH	
Author	**Year**	**Type of surgery**	**Dose**	**LMWH**	**UFH**	**LMWH**	**UFH**
Schmitz-Hübner	1984	general	high	3/40	0/39	7/40	8/39
Schmitz-Hübner*	1984	general	high	0/41	-----	1/41	-----
Kakkar	1985	general	low	5/196	15/199	11/196	14/199
Maroske	1985	orthop.	low	7/40	7/40	10/40	12/40
Bergquist	1986	general	high	13/215	9/217	13/215	7/217
Koller	1986	general	low	2/70	1/68	3/74	2/72
Sasahara	1986	general	low	14/134	13/126	4/137	3/132
Voigt	1986	general	low	1/103	1/97	29/103	24/97
Onarheim	1986	general	high	1/25	0/27	0/25	1/27
Haas	1987	orthop.	low	15/73	15/73	21/80	21/80
Bergquist	1988	general	high	28/505	41/497	36/505	47/497
Caen	1988	general	low	6/190	7/195	35/195	38/190
Lassen	1988	orthop.	low	35/107	34/112	0/107	0/112
Adolf	1989	general	low	25/202	24/202	14/202	19/202
Baumgartner	1989	general	low	6/87	7/89	1/87	0/89
Heilmann	1989	general	low	2/150	6/150	19/150	27/150
Kakkar	1989	general	low	8/88	10/91	2/88	1/91
Monreal	1989	orthop.	high	14/32	6/30	2/46	2/44
Pini	1989	orthop.	high	5/25	7/24	2/25	1/24
Reilmann	1989	orthop.	low	8/63	6/58	15/63	17/58
Lassen	1989	orthop.	low	14/53	23/54	9/68	14/71
Hartl	1990	general	low	5/112	5/115	1/112	1/115
Boneu	1990	general	low	30/648	29/663	55/665	80/677
Koppenhagen	1990	general	low	4/51	7/53	0/51	0/53
Planes	1990	orthop.	high	15/120	27/106	8/124	5/113
Koppenhagen	1990	general	low	2/36	4/41	6/42	2/41
Eriksson	1991	orthop.	high	19/63	25/59	0/67	0/69
Freick	1991	orthop.	low	5/52	12/48	1/55	1/55
Leizorovicz	1991	general	low	16/431	7/429	5/431	28/429
Leizorovicz*	1991	general	low	7/430	-----	21/430	-----
Levine	1991	orthop.	high	50/258	61/263	17/333	31/332
GHAT	1992	orthop.	high	45/136	47/137	32/167	36/169
Hoffmann	1992	general	low	1/298	3/296	10/298	7/296
Koppenhagen	1992	general	low	27/323	27/330	2/323	0/330
Kakkar	1993	general	low	6/1894	6/1915	27/1894	52/1915

terventions is performed every day, prophylaxis against thromboembolism is a major health problem. Benefit from a prophylaxis with UFH is beyond any doubt. Recently, so-called low-molecular weight heparins (a fraction of standard heparin) were developed in the hope to achieve an increased efficacy and safety (heparin prophylaxis increases the risk of bleeding complications). A thorough summary on the topic can be found in a meta-analysis of Leizorowicz et al. (1992). Up to now, the question whether low molecular weight heparin is superior to standard heparin is not completely answered. This example is used to demonstrate what can be learned from careful heterogeneity analysis.

III. Results and macros

3.1 %metafunn

The first step in performing a meta-analysis, unless the number of studies is less than about 15, should be a Funnel-plot, where sample size is plotted against an arbitrary estimator for the treatment effect for each study in the meta-analysis. This indicates completeness and unbiasedness of the literature-search, which is the most important prerequisite for a meaningful interpretation of computed summary estimates. Many scientists are under the impression that non-significant results might be less likely to be accepted by the journals or that less effort might be made to publish them due to the incorrect feeling that non significant results are not important (Dickersin et al.(1992); Easterbrook et al.(1991)). This "publication bias" results in a distortion of the findings of the meta-analysis by generating "over-optimistic" statements and the rise of overstated hopes in new treatments.

Of course, looking at a Funnel-plot is a little subjective, but it can give a helpful first impression of the situation and show a possible bias.

In the present example there is no direct evidence for "publication bias": due to random fluctuation one would expect, especially in small trials, to see an approximately symmetric distribution of observed effects around the expected true effect, generating the funnel (see Fig. 1). Moreover, the fact that estimates with both signs are found indicates that in the present example study results were published even if they contradicted the hypothesis that lead to the initiation of the trial.

3.2 %metacalc

Although meta-analysis can be done for any type of variable for which stratified analyses are available, we prefer binary responses that are simpler to assess, especially in the field of publication based meta-analysis. We mention that this macro, after some preliminary calculations, could also be used for ordinal data or serve as a starting point for developing a macro that could handle normally distributed data (Whitehead and Whitehead (1991)), but this was not primarily intended.

In the situation of a clinical trial comparing two treatments where the results can be displayed in a fourfold-table, risk difference, (Log-)risk ratio or (Log-)odds ratio are all well-known estimators for the unknown treatment effect, none of which is a priori superior; the choice is rather a matter of taste or justified by interpretation.

Two different methods exist for the combination of the effects estimated in the single trials into the overall effect of treatment: the general fixed effects approach (GFA) and the general random effects approach (GRA).

Under the assumption of homogeneity, i.e. the same treatment effect is estimated in each study, the weighted average

$$\hat{\Theta} = \Sigma \hat{\Theta}_i w_i / \Sigma w_i$$

is an unbiased estimator (Cox, 1982) for the overall treatment effect Θ in the GFA (Θ_i estimates the treatment effect in the i-th study and $w_i = 1/v_i$ where v_i is the estimated variance of Θ_i, i=1,....., k and k denotes number of included studies). Further, a 95%-confidence interval is given by

$$\hat{\Theta} \pm 1.96\sqrt{1/\Sigma w_i}\ .$$

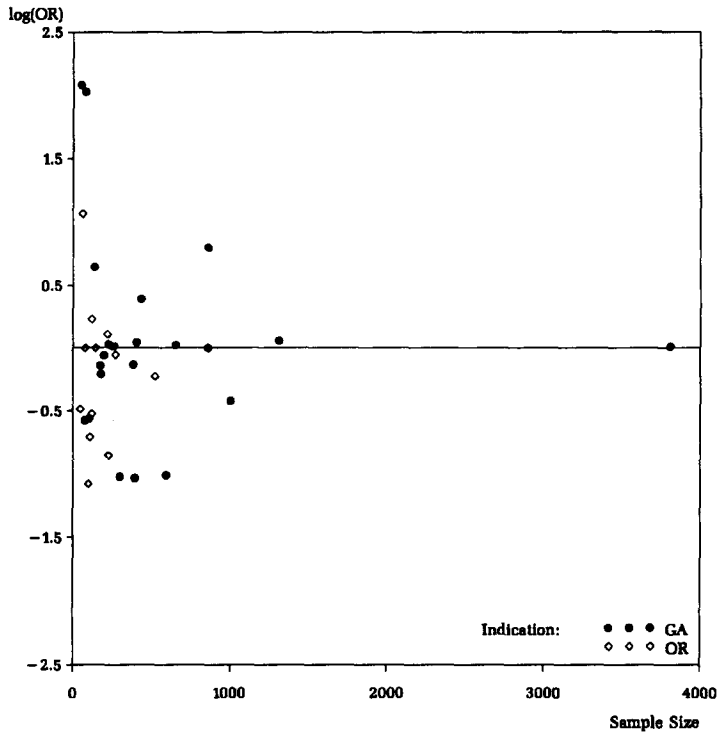

Fig. 1: Funnel-plot for all double blind clinical trials in general abdominal and orthopedic surgery (GA: general (abdominal), OR: orthopedic), Treatment effect is estimated by the Log odds ratio (Yusuf/Peto method). Log(OR)<0 if treatment is superior to control.

For this estimate of Θ the greatest weight is given to the most informative studies (those with the smallest variance, in general those enrolling the largest number of included patients), which seems quite sensible.

A hypothesis test of H_0:$\Theta_1 = \ldots\ldots = \Theta_k = 0$ can be performed by comparing

$$U = (\Sigma \hat{\Theta}_i w_i)^2 / \Sigma w_i$$

to a χ^2-distribution with one degree of freedom, and H_0 is rejected if U is significantly large. An overall test of the homogeneity assumption (H_0:$\Theta_1 = \ldots\ldots = \Theta_k = \Theta$) is provided by comparing

$$Q = \Sigma (\hat{\Theta}_i - \bar{\Theta})^2 w_i$$

to a χ^2-distribution with k-1 degrees of freedom.

A number of slightly different versions of the overall effect estimator were proposed. Yusuf/Peto's method (Yusuf et al. (1985)), for example, is based on comparing observed and expected numbers of events under the hypothesis of no treatment effect. The Mantel-Haenszel method in general leads to slightly different values of the overall effect estimator due to modified formulas for the estimation of the variance of treatment effect in the single studies. For an overview on methods see Fleiss (1993) or Petitti (1994).

In the opinion of some authors the assumption of homogeneity of effects across all studies seems to be a strong or even unrealistic limitation. Therefore an alternative model is proposed: the single studies are regarded as a random sample from a bigger population of studies. In each study, it is not a common treatment effect Θ for all studies that is estimated, but a single study-specific Θ_i. These true effects Θ_i are assumed to be normally distributed around the true overall treatment effect Θ with variance τ^2. In this model the estimator for Θ is given by

$$\hat{\Theta} = \Sigma w_i^* \hat{\Theta}_i / \Sigma w_i^*$$

Table 2: Computational results (GFA and GRA) for the various estimators and tests for the overall effects for the subgroup of low-dose studies in general surgery for the variable "safety" calculated with the %metacalc-macro. (DELTA: risk difference, LogOR(D): Log odds ratio, LogOR(M): Log odds ratio according to the Mantel-Haenszel-method, LogOR(P): Log odds ratio after Yusuf/Peto, LogRR: Log risk ratio)

```
***** METACALC:        tests and estimates for overall effects for the fixed
                       effects model
```

OBS	EFFECT	N	ESTIMATE	KIL95	KIU95	T_EFFEKT	P_EFFEKT
1	DELTA	17	-0.00444	-0.00963	0.00075	2.8155	0.093356
2	LogOR(D)	15	-0.31331	-0.49578	-0.13084	11.3262	0.000764
3	LogOR(M)	15	-0.32034	-0.49799	-0.14268	12.4903	0.000409
4	LogOR(P)	17	-0.31789	-0.49363	-0.14215	12.5697	0.000392
5	LogRR	15	-0.25994	-0.42065	-0.09923	10.0502	0.001523

OBS	T_HETGEN	P_HETGEN	EXP_EFF	EXP_KIL	EXP_KIU
1	37.0642	0.00205	.	.	.
2	20.9946	0.10177	0.73102	0.60910	0.87736
3	.	.	0.72590	0.60775	0.86703
4	26.8455	0.04324	0.72768	0.61041	0.86749
5	21.8427	0.08190	0.77110	0.65662	0.90553

```
***** METACALC: tests and estimates for overall effects
               for the random effects model
```

OBS	EFFECT	N	ESTIMATE	KIL95	KIU95	T_EFFEKT	P_EFFEKT
1	DELTA	17	-0.00673	-0.01768	0.004212	1.45377	0.22792
2	LogOR(D)	15	-0.27612	-0.52705	-0.02518	4.65145	0.03103
3	LogOR(P)	17	-0.24111	-0.50483	0.022609	3.21114	0.07314
4	LogRR	15	-0.24171	-0.46889	-0.01453	4.34883	0.03703

OBS	T_HETGEN	P_HETGEN	EXP_EFF	EXP_KIL	EXP_KIU	BS_VAR
1	16.3405	0.42946	.	.	.	0.00019
2	15.5792	0.33974	0.75872	0.59034	0.97513	0.07003
3	17.5934	0.34824	0.78576	0.60361	1.02287	0.09999
4	15.7835	0.32677	0.78528	0.62570	0.98557	0.06077

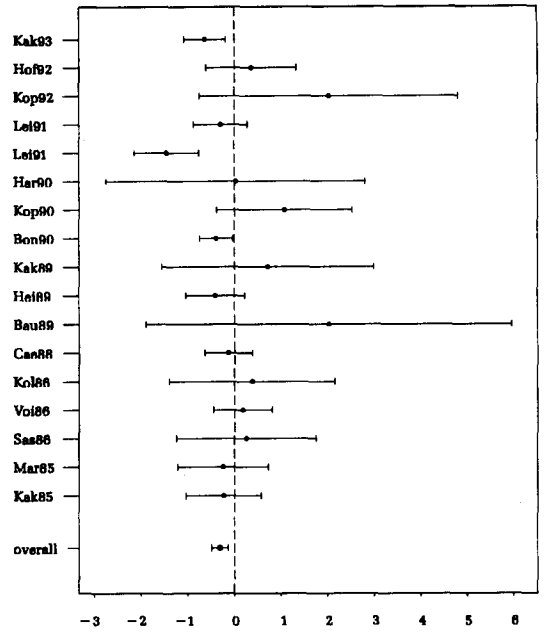

Fig. 2: Confidence interval plot for the subgroup of low-dose studies in general surgery for the variable "safety" plotted with the %metaki-macro. Treatment effect is estimated by the Log odds ratio (Yusuf/Peto method).

with the same $\hat{\Theta}_i$ as in the GFA but modified weights $w_i^* = (w_i^{-1} + \tau^2)^{-1}$. The between-study variance τ^2 has to be estimated from the data by

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

based on the statistics Q of the homogeneity test in the GFA (DerSimonian&Laird (1986)).

In our meta-analysis dose and indication were identified among others as possible sources of heterogeneity, therefore analyses in several subgroups were carried out. Table 2 presents the results for the subgroup of low-dose studies in general surgery for the variable "safety" calculated with the %metacalc-macro. The results indicate that treatment is superior to control, independent of the applied method. The various approaches for the odds-ratio lead to similar outcomes, whereas formal conclusions based on the risk difference approach would differ. Further, it can be seen how heterogeneity affects the estimators for the overall effect. In the GFA we get lower p-values for the test of a zero-effect compared to the GRA.

3.3 %metaci

The classical method for visualizing the results of a meta-analysis is to plot the estimated treatment effect $\hat{\Theta}_i$ with its 95%-confidence interval for each single study and adding the overall effect estimator with its confidence interval into the graph. A disadvantage is that the visual impression is dominated by small studies with large confidence intervals, while larger studies with smaller confidence intervals seem to be less important. As a remedy some authors (e.g. Thompson (1993)) propose to combine the five smallest studies into one trial.

Fig. 2 shows the confidence interval plot for the subgroup introduced in 3.2 visualizing the analytically assessed heterogeneity.

3.4 %metagalb

Radial-Plots (Galbraith (1988)) are bivariate scatter plots of

$$y = \hat{\Theta}_i \sqrt{w_i} = \hat{\Theta}_i / s.e.(\hat{\Theta}_i),$$

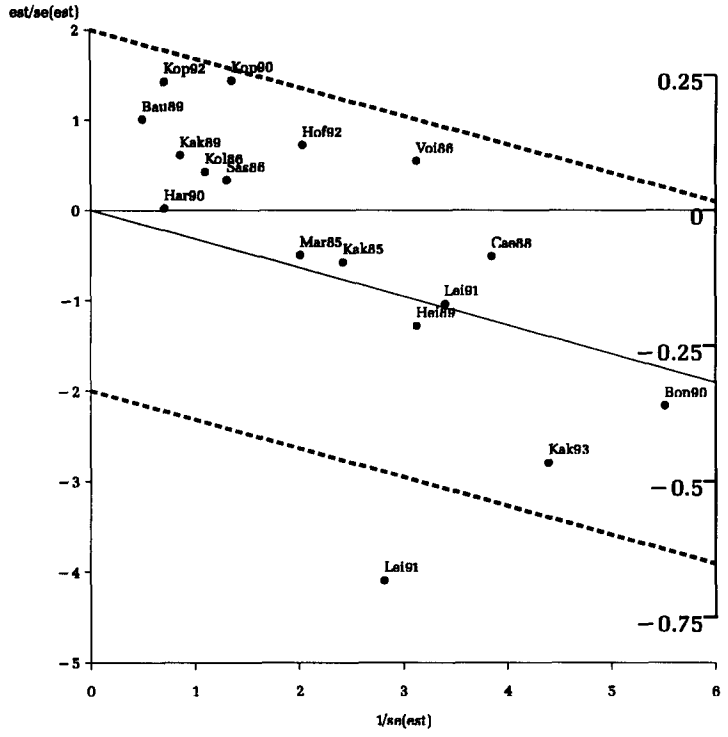

Fig. 3: Galbraith-plot for the subgroup of low-dose studies in general surgery for the variable "safety" plotted with the %metagalb-macro. Treatment effect is estimated by the Log odds ratio (Yusuf/Peto method).

$$y = \hat{\Theta}_i \sqrt{w_i} = \hat{\Theta}_i / s.e.(\hat{\Theta}_i),$$

the standardized estimate of the single treatment effect against

$$x = \sqrt{w_i} = 1/s.e.(\hat{\Theta}_i),$$

the "precision" of the estimate. The radial axis is plotted on the right-hand side of the plot. These displays have some useful and instructive features:

Small trials correspond to points close to the y-axis, while large trials provide influential points on the right hand side. The estimate Θ_i of a single trial is the slope of the line from the origin through the corresponding point (x_i, y_i) and therefore readily available from the radial axis at the point of intersection with this line.

Under the assumption of homogeneity $E(y_i) = \Theta x_i$, and Θ is the slope of the regression line through the origin in a formal regression of y on x. Because the least-squares estimator of the slope equals to $\hat{\Theta}$, in consequence the estimator for the overall treatment effect (which is the GFA-estimator) can also be read from the radial axis, at the point of intersection with the regression line through the origin.

From the contributions to the homogeneity statistics Q we see that the residuals of the regression $y_i - \Theta x_i = (\Theta_i - \Theta)\sqrt{w_i}$ follow roughly a standard normal distribution. Because $(\Theta_i - \Theta)\sqrt{w_i}$ can be interpreted as the contribution to heterogeneity for every single study, a plot of the two lines $y = \Theta x \pm 2$ provides an approximate 95%-confidence band with studies lying outside of it contributing somehow significantly to heterogeneity.

Due to the fact that $y = \hat{\Theta}_i / s.e.(\hat{\Theta}_i)$ and $\Theta_i / s.e.(\Theta_i) \sim N(0,1)$, single studies with significant treatment effect at the 95%-confidence level can be readily detected by having y-values outside the interval [-1.96,1.96].

Fig. 3 shows the radial-plot for the above mentioned subgroup plotted with the %metagalb-macro.

3.5 %metasens

Additional insight about the relation between the GFA- and the GRA-estimator can be gained through a diagnostic plot proposed by Thompson (1993): by investigating the weights of individual studies one can see that moving from a fixed effects to a random effects model is simply a process of equalizing the weights given to each study.

This idea was extended by Thompson by making this process continuous, determining the overall effect estimator as a function of the between-study variance τ^2, ranging from $\tau^2 = 0$ (GFA) and $\tau^2 = \hat{\tau}^2$ (GRA) to $\tau^2 = \infty$ (giving equal weight to each study) and plotting the overall estimator vs. the between-study variance.

Fig. 4 shows this sensitivity-plot for the subgroup of low-dose studies in general surgery, scaled that x=0 corresponds to GFA, x=0.5 to GRA and x=1 to equal weighting. A 95%-confidence interval for every overall estimator is added. It can be seen that the estimator for the overall Log odds ratio is highly sensitive to the weights given to the single studies, even up to the point of reversing the sign and thus claiming inferiority for the new treatment in the extreme case of equal weighting.

IV. Conclusions

Careful investigation of heterogeneity and its potential sources is one of the principal tasks in meta-analysis and a prerequisite for a meaningful interpretation of the results. To perform this, graphical methods and intermediate computational results, for example the contribution each trial makes to the test statistic of homogeneity, are needed, but not directly available in SAS. We hope that the SAS-Macros presented here promote a sensible way to deal with heterogeneity in meta-analysis and help to

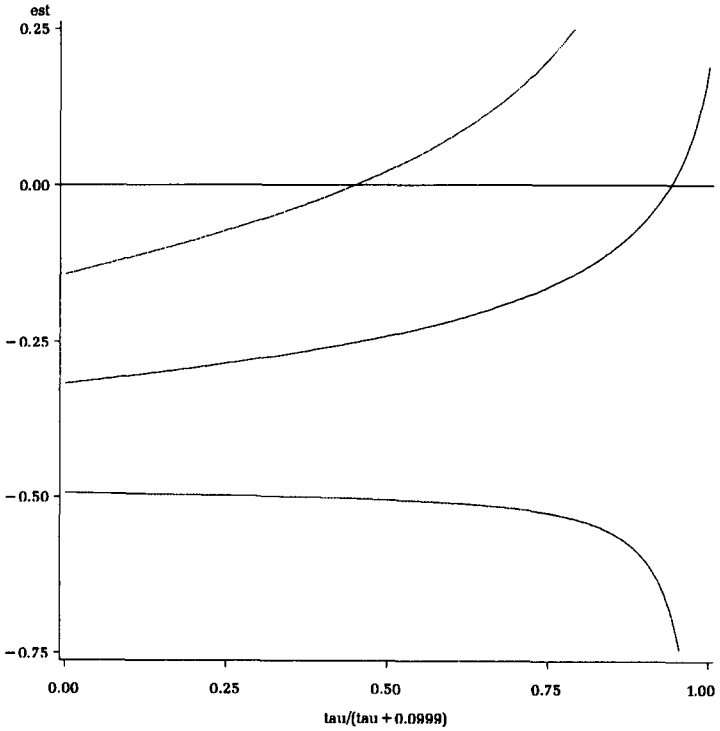

Fig. 4: Sensitivity-plot by Thompson for the subgroup of low-dose studies with indication "GA" for the variable "safety" plotted with the %metasens-macro. Treatment effect is estimated by the Log odds ratio (Yusuf/Peto method). (0.0999 is the estimated value for the between-study variance, see Table 2).

overcome the problem that this analyses are not executed due to non-availability of methods in conventional statistical packages. In addition we hope that these macros can serve as a starting point for further developments.

The SAS-Macros that realized the plots and calculations in this paper are available from the authors free of charge (Please send a letter or a short e-mail to armin@imbi.uni-heidelberg.de). They have been tested under SAS for Windows and the SAS System on a HP 9000/700 (Release 6.09).

Acknowledgment

This work has been sponsored in part by the Deutsche Forschungsgemeinschaft (DFG grant Vi 107/4-1).

References

Boissel, J. P. et al. (1988): Meta-analysis of clinical trials: summary of an international conference. *Eur. J. Clin. Pharmacol.*, 34, 535-538.

Cox, D. R. (1982): Combination of data. In: Kotz, S. and Johnson, N. L. (eds.): Encyclopedia of statistical sciences. New York, 45-53.

Dickersin, K. et al. (1992): Factors influencing publication of research results. *JAMA*, 267, 374-378.

DerSimonian, R. and Laird, N. (1986): Meta-analysis in clinical trials. *Controlled clinical trials*, 7, 177-188.

Easterbrook, P. J. et al. (1991): Publication bias in clinical research. *Lancet*, 337, 867-872.

Fleiss, J. L. (1993): The statistical basis of meta-analysis. *Statist. Meth. in Med. Res.*, 2, 121-145.

Galbraith, R. F. (1988): A note on graphical representation of estimated odds ratios from several clinical trials. *Statist. Med.*, 7, 889-894.

Leizorowicz, A. et al. (1992): Low molecular weight heparin in prevention of preoperative thrombosis. *Brit. Med. J.*, 305, 913-920.

Normand, S. L. (1995): Meta-Analysis software: A comparative review. *The American Statistician*, 49, 3, 298-309.

Petitti, D. B. (1994): Meta-analysis, decision analysis and cost-effectiveness analysis. New York, Oxford.

Thompson, S. G. (1993): Controversies in meta-analysis: the case of the trials of serum cholesterol reduction. *Statist. Meth. in Med. Res.*, 2, 173-192.

Whitehead, A. and Whitehead, J. (1991): A general parametric approach to the meta-analysis of randomized trials. *Statist. Med.*, 10, 1665-1677.

Yusuf, S., Peto, R., Lewis, J., Collins, R. and Sleight, P. (1985): Beta blockade during and after myocardial infarction: An overview of the randomized trials. *Progress in cardiovascular diseases*, 27, 335-371.