Maastricht University

The influence of methodologic quality on the conclusion of a landmark metaanalysis on thrombolytic therapy.

Document status and date:
Published: 01/01/2002

Document Version:
Publisher's PDF, also known as Version of record

Please check the document version of this publication:

• A submitted manuscript is the version of the article upon submission and before peer-review. There can be important differences between the submitted version and the official published version of record. People interested in the research are advised to contact the author for the final version of the publication, or visit the DOI to the publisher's website.
• The final author version and the galley proof are versions of the publication after peer review.
• The final published version features the final layout of the paper including the volume, issue and page numbers.

Link to publication

THE INFLUENCE OF METHODOLOGIC QUALITY ON THE CONCLUSION OF A LANDMARK META-ANALYSIS ON THROMBOLYTIC THERAPY

Arianne P. Verhagen

Maastricht University and Erasmus University Rotterdam

Henrica C. W. de Vet

Maastricht University and Vrije Universiteit

Frank Vermeer

Maastricht University Hospital

Jos W. M. G. Widdershoven

Twee Steden Hospital

Robert A. de Bie

Maastricht University

Alphons G. H. Kessels

Maastricht University and Maastricht University Hospital

Maarten Boers

Vrije Universiteit

Piet A. van den Brandt

Maastricht University

Abstract

Objective: We studied the influence of the methodologic quality of individual trials on the outcome of a landmark meta-analysis on thrombolytic therapy in acute myocardial infarction. From each study we extracted the number of patients in both groups who died in hospital or during follow-up. Methodologic quality was assessed using the Delphi list. We first recalculated pooled odds ratios (ORs) and 95% confidence intervals (CIs), on the studies found and compared them with the original results of Yusuf et al. Next we incorporated the results of quality assessment in five different ways in the calculation of the pooled ORs: a) component analysis; b) visual plot; c) quality score as a threshold score; d) quality score as a weighting factor; and e) cumulative pooling.

Results and conclusion: No correlation between quality scores and ORs was found. Studies with a proper description of the different quality components provided an estimate close to the true treatment effect. No major differences were found between the results of the five different methods of incorporating the quality scores into the final conclusion.

Verhagen et al.

Scientific guidelines for reviewing the literature often include assessment of the methodologic quality of the trials (20;21). The value of the conclusion of a meta-analysis not only depends on the quality of the review process itself, but also on the methodologic quality of the randomized clinical trials (RCTs) included (10). A leading paradigm in empirical research is that clinical trials that do not meet certain design criteria, such as concealed randomization and double blinding, will be usually biased in favor of the intervention and are therefore more likely to produce larger treatment effects (18;19;22;23).

Assessment of the quality of clinical trials by criteria lists provides an estimation of the possibility of biased results of a trial. One approach in assessing quality is to focus on components such as randomization and blinding in trial reports (16;18). Furthermore, a criteria list can provide a quality score as an estimation of the overall methodologic quality of the design and conduct of the trial (2). The included studies can be ordered hierarchically according to quality scores, with higher scores indicating studies with a better methodologic quality (4). Quality scores can be used as a threshold score for inclusion of the article in a review, as a weighting factor in the statistical analysis (6;12;23), or as the input sequence in a cumulative meta-analysis (14;15;23). Finally, a visual plot of the effect size against a quality score can be presented (14;15;23).

Historically, the effectiveness of thrombolytic therapy for acute myocardial infarction (AMI) was long disputed. Before 1980 intravenous streptokinase (SK) was tested in RCTs, but the results were not unequivocally in favor of this therapy. Later, intracoronary application of SK became in use because of the angiographically documented recanalization of the occluded coronary artery. SK became licensed for use in AMI after positive results of a meta-analysis in 1985 of a study by Yusuf et al. (29) and two very large trials (9;11) in which the benefit of intravenous thrombolysis in AMI was confirmed. However, cumulative meta-analysis showed (in retrospect) that there already was a clear evidence of the benefit in 1973 (1;7). Yusuf et al. (29) did not carry out a formal quality assessment in their meta-analysis. They closed their discussion with a comment on the general validity of their overview because some trials were "undoubtedly less well executed than others."

We chose the meta-analysis of the study by Yusuf et al. (29) because the results had great impact on health care. Its results appeared to be valid, because large trials studying the same intervention (9;11) confirmed their conclusions, and in a study of Egger et al. (8) the funnelplot derived from the meta-analysis of Yusuf et al. (29) was not skewed. Therefore, this meta-analysis offered the possibility of determining a kind of gold standard effect estimate to study the influence of design characteristics on outcome. In the literature, statements were made about the direction of bias from design characteristics (19;22), but a reference effect estimate was never used.

In this research, quality will be measured by the Delphi list (26), which has been developed by expert consensus, thereby providing some validity (27). We set out to investigate whether and in which way quality can affect the overall conclusions of a meta-analysis. In this study we used the meta-analysis of Yusuf et al. (29), and the conclusions of the authors are regarded as the gold standard.

METHODS

Selection of Studies

All full reports presenting mortality data included in the meta-analysis of Yusuf et al. (29) are included in this study. Where they used an abstract or personal communication, we

Table 1. The Delphi List for Quality Assessment

Items	Answer option
1. Treatment allocation	
a. Was a method of randomization performed?	Yes/No/Don't know
b. Was the treatment allocation concealed?	Yes/No/Don't know
2. Were the groups similar at baseline regarding the most important prognostic indicators?	Yes/No/Don't know
3. Were the eligibility criteria specified?	Yes/No/Don't know
4. Was the outcome assessor blinded?	Yes/No/Don't know
5. Was the care provider blinded?	Yes/No/Don't know
6. Was the patient blinded?	Yes/No/Don't know
7. Were point estimates and measures of variability presented for the primary outcome measures?	Yes/No/Don't know
8. Did the analysis include an intention-to-treat analysis?	Yes/No/Don't know

required full reports. Therefore, we searched in MEDLINE and EMBASE and consulted leading cardiologists in the thrombolytic field. Studies only available as abstracts or personal communications were excluded, because quality assessment could not be performed.

Quality Assessment

For the quality assessment of individual studies, we applied the Delphi list (26). The Delphi list is a systematically developed criteria list including nine items, measuring three dimensions of quality: internal validity, external validity, and statistical considerations (Table 1). It is not designed as a composite scale; items can be used as separate components. The assessment of the studies was performed independently by two epidemiologists and two cardiologists. The assessors reached a final score during a consensus meeting resulting in an overall quality score (QS). The QS consists of the number of items satisfied and ranges from 0–9. We studied the relationship between QS and effect size first by analyzing the effects of the main components of the QS (component analysis). We then studied the relationship between the overall QS and effect sizes in several ways: visual plot, using QS as threshold or weight, or as an input sequence in cumulative pooling.

For the component analysis we divided the studies into three categories of *randomization*: method concealed, method appropriate but not concealed, and method unknown. Concealed randomization implies that a random (unpredictable) assignment sequence is generated by an independent person not responsible for determining eligibility of the patients, and this sequence is concealed until allocation occurs. By appropriate we mean that reports present additional information about a randomization procedure, which is considered appropriate in being at random and unpredictable in preventing allocation bias. For *blinding* we divided the studies into two categories: blinding reported or not reported. Concerning *withdrawals*, we divided the studies into three categories: a) no withdrawals or a withdrawal rate not leading to bias; b) withdrawal rate unknown; and c) a withdrawal rate possibly leading to bias (meaning > 5% dropouts [=withdrawal during the intervention period] > 20% loss to follow-up [=withdrawal during follow-up]). These cut-off points of 5% and 20% were set according to other frequently used criteria lists (5;24). For the *statistical analysis* items, we divided the studies into two categories: performance of an intention-to-treat (ITT) analysis or not.

Data Extraction

For the main outcome measure, we extracted from each report the number of patients in the treated and control groups who died in hospital or during follow-up, as mentioned in the report.

Analysis

We first recalculated pooled ORs and their 95% confidence intervals (CIs) using a Peto fixed effects model (all our methods are comparable with the method used by Yusuf et al.), and compared it with the results of Yusuf et al. (29). Also, a funnelplot was made according to Egger et al. (7) to evaluate possible publication/selection bias of the studies found. We calculated intraclass correlation coefficient (ICC) between the four different quality assessors, and a Spearman rank correlation coefficient to evaluate the relationship between quality and effect estimate. Next, the QS were incorporated in the pooling in the five different ways mentioned above.

For the component analysis, we chose to perform a sensitivity analysis and refrained from meta-regression techniques because of the small number of studies included. For the overall analysis, we first construct a scatterplot of the QS against the individual ORs. We also made a scatterplot, including the studies we were unable to find or did not include. These studies all received a QS of 2. To use the QS as a threshold score, we followed a suggestion by Chalmers et al. (3) and did a restricted analysis on studies receiving a QS similar or higher than the mean QS. Further, we also used the QS as a weight: we weighted each individual study estimate by their achieved Delphi QS, thereby deriving more impact from higher quality studies on the overall pooled results (6). Finally, to achieve cumulative pooling, we started the pooling with the study with the highest QS and subsequently added the others, rank-ordered by decreasing QS.

RESULTS

Trials Included

In the original meta-analysis of Yusuf et al. (29), 33 trials were included. Of four studies, two reports (short-term and long-term results) of the same trial were available (32;33) (39;40) (45;46) (56;62). For quality assessment, we chose the report in which the method of research was most clearly described (32;40;45;56). We identified 27 of the original 33 references (30;31;32;34;35;36;37;38;40;41;42;43;44;45;47;48;49;50;51;52;54;55;56;58;61;63). The language of most publications was in English, three in German (34;36;55), one in French (63), and one in Spanish (49). The Spanish report was translated into English to facilitate quality assessment. Of the original 33 references, two were based on personal communications: Hugenholtz PG, Serruys PW, Simoons ML, et al. Randomized trial of intracoronary thrombolysis in acute myocardial infarction. Personal communication 1985; Theroux P, personal communication to Furberg C. A trial of intracoronary streptokinase in acute myocardial infarction. We were able to trace one of them (Hugenholtz et al.) as a full report (60), and this paper was also included.

After detailed reading, 2 of the 28 references did not meet the selection criteria (i.e., RCT, full report, presenting mortality data). One appeared to be a case series instead of an RCT (42). Although the paper used the terms RCT and control group, it reported the results of a case series (n = 23), later on compared with a control group (n = 11). Another report focused on the complications of SK treatment based on data derived from an RCT, but no data about mortality were presented (52). Both studies were excluded from the analysis.

Of the remaining 26 references, four were abstracts (48;54;58;61). We found two full reports of the same group of authors, about the same trial (53;59) and included them in this research. The main characteristics of the remaining 24 trials and their Delphi QS are presented in Table 2. In our calculations we used the data of the longest follow-up period found in the reports.

Table 2. Characteristics of the Studies

Study	Participants	Method	Intervention	Odds ratio (95% CI)	Delphi QS
Gørmson, 1973 (43)	AMI < 24 hours; age < 80, n = 28	Concealed R; blinding of patient, therapist and observer	UK vs placebo	0.61 (0.09–4.37)	8
Schreiber, 1986 (59)	AMI < 6 hours; age < 76, n = 38	Unknown R; blinding of patient and therapist	SK + heparin vs placebo + heparin	0.20 (0.02–2.07)	7
Italian (Dioguardi), 1971 (37)	AMI < 12 hours; no age limits, n = 321	Concealed R; no blinding; multicenter	SK + anticoag. vs glucose + anticoag	1.01 (0.51–2.01)	6
2nd Frankfurt (Breddin), 1973 (36)	AMI < 12 hours; age < 70, n = 206	Unknown R; blinding of patient and therapist; multicenter	SK + heparin vs placebo + heparin	0.38 (0.18–0.77)	6
Lippschutz, 1965 (51)	AMI < 48 hours; no age limits, n = 84	Concealed R; blinding of patient and observer	UK + heparin vs placebo + heparin	0.79 (0.24–2.58)	6
Rentrop, 1984 (57)	AMI < 12 hours; age < 72, n = 124	Unknown R; blinding of observer	IC-SK or IC-SK + NTG vs NTG or control	2.38 (0.84–6.75)	6
Kennedy, 1985 (46)	AMI < 12 hours; age < 75, n = 250	Concealed R; no blinding; multicenter	IC-SK + heparin vs heparin	0.52 (0.23–1.16)	6
2nd European, 1971 (41)	AMI < 24 hours; no age limits, n = 730	Concealed R; no blinding; multicenter	SK vs heparin	0.64 (0.45–0.9)	5
Australian (Bett), 1973 (35)	AMI < 24 hours; age < 65, n = 517	Adequate R; no blinding; multicenter	SK + heparin + anticoag. vs heparin + anticoag	0.75 (0.43–1.31)	5
UK Collab. (Aber), 1976 (30)	AMI < 24 hours; no age limits, n = 595	Concealed R; no blinding; multicenter	SK vs control	0.88 (0.57–1.35)	5
3rd European, 1981 (40)	AMI < 12 hours; age < 80, n = 315	Concealed R; no blinding; multicenter	SK + coumarin vs glucose + coumarin	0.41 (0.24–0.72)	5
Olson, 1986 (53)	AMI < 12 hours; no age limits; males only, n = 52	Unknown R; no blinding	SK + heparin vs saline + heparin	0.83 (0.20–3.29)	5
European Cooperative, 1975 (38)	AMI < 12 hours; age < 80, n = 341	Concealed R; no blinding; multicenter	UK + heparin + anticoag. vs glucose + heparin + anticoag.	1.09 (0.64–1.87)	5

(*Continued*)

Table 2. (*Continued*)

Study	Participants	Method	Intervention	Odds ratio (95% CI)	Delphi QS
Khaja, 1983 (47)	AMI <6 hours; no age limits, n = 40	Adequate R; no blinding	IC-SK vs dextrose placebo	0.21 (0.02–2.08)	5
Anderson, 1984 (33)	AMI <4 hours; age < 80, n = 50	Unknown R; Probably blinding of observer	IC-SK + heparin vs heparin	0.38 (0.06–2.19)	5
Simoons, 1985 (60)	AMI <4 hours; age < 70, n = 533	Adequate R; no blinding; Zeelen design	IC-SK + conventional vs conventional treatment	0.51 (0.30–0.88)	5
2nd German (Poliwoda), 1977 (55)	AMI <12 hours; no age limits, n = 492	Concealed R; no blinding; multicenter	SK + heparin + marcoumar vs heparin + standard	1.28 (0.84–1.94)	5
1st European (Amery), 1969 (31)	AMI <72 hours; no age limits, n = 167	Adequate R; no blinding	SK + coumarin vs heparin + coumarin	1.46 (0.69–3.1)	4
Heikinheimo, 1971 (44)	AMI <72 hours; no age limits, n = 426	Adequate R; no blinding; multicenter	SK + anticoag. vs glucose + antocoag.	1.25 (0.64–2.42)	4
Witchitz, 1977 (63)	AMI <24 hours; age < 75, n = 58	Adequate R; no blinding	SK vs heparin	0.77 (0.20–3.04)	4
Leiboff, 1984 (50)	AMI <4 hours; age < 75, n = 40	Unknown R; no blinding	IC-SK + heparin vs NTG + heparin	1.78 (0.29–11.04)	4
Raizner, 1985 (56)	AMI <6 hours; age < 70, n = 64	Concealed R; no blinding	IC-SK + NTG vs NTG or control	2.8 (0.48–16.5)	4
Austrian (Benda), 1977 (34)	AMI <12 hours; no age limits, n = 728	Adequate R; no blinding; multicenter	SK vs control	0.56 (0.36–0.87)	3
Lasierra, 1977 (49)	AMI <48 hours; no age limits, n = 24	Adequate R; no blinding	SK + heparin vs heparin	0.22 (0.02–2.53)	2

Definition of abbreviations: SK = streptokinase; UK = urokinase; R = randomization procedure; IC-SK = intracoronary streptokinase; NTG = nytroglycerin.
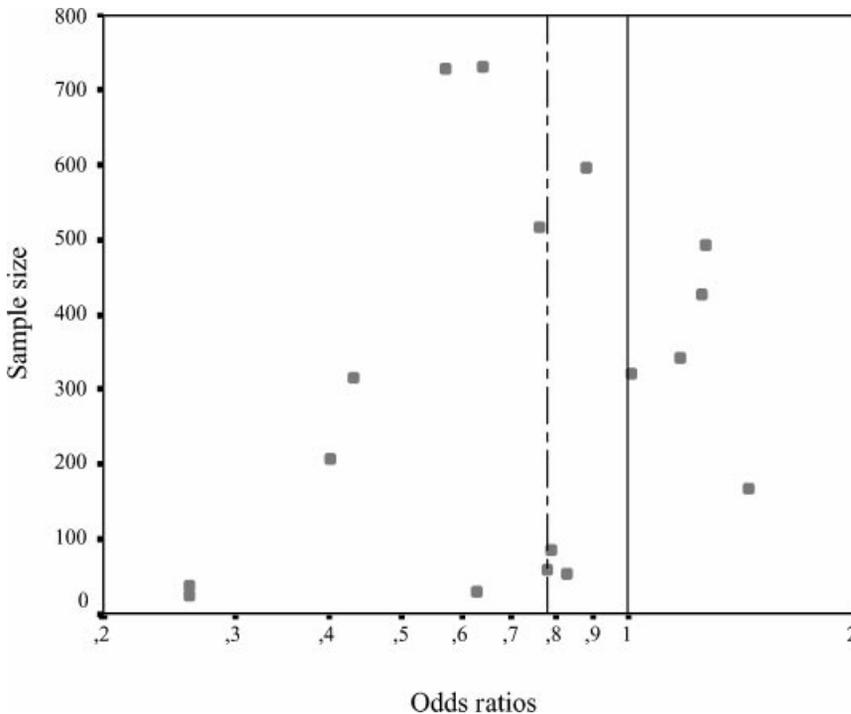
Figure 1. Funnelplot of the studies concerning intravenous treatment of MI.

Pooled OR

Following the approach of Yusuf et al., we divided the reports into two categories: intravenous (SK or urokinase) and intracoronary (SK) treatment. Regarding intravenous treatment, we included 17 of the original 24 references. Our pooled OR was 0.78 (0.67–0.90), equal to that of Yusuf et al., 0.78 (0.68–0.89).

Regarding intracoronary treatment, we included seven of the nine original references of Yusuf et al. (29). We calculated an OR of 0.68 (0.48–0.98). Contrary to Yusuf et al., we found a significant pooled OR, suggesting that intracoronary treatment with SK is beneficial for patients with AMI. The pooled OR found by Yusuf et al. was only presented in a figure (OR $\approx$ 0.8; 95% CI $\approx$ 0.55–1.1).

To study the relationship between quality of an RCT and outcome, we restricted ourselves to the studies concerning intravenous treatment, because we reproduced the same pooled OR as Yusuf et al. did with the studies found. The funnelplot we made of these 17 studies (Figure 1) showed no apparent publication/selection bias.

Quality Assessment

The QS of the Delphi list is presented in Table 2. The mean QS was 5 (max = 9). The ICC between the four quality assessors was 0.54. The Spearman rank correlation coefficient of the QS versus the OR was −0.21.

INCORPORATION OF QUALITY

Component Analysis

Of the included 17 studies, nine (30;35;37;38;39;41;43;51;55) mentioned a concealed randomization, in five studies (31;34;44;49;63) the method was appropriate but not concealed,
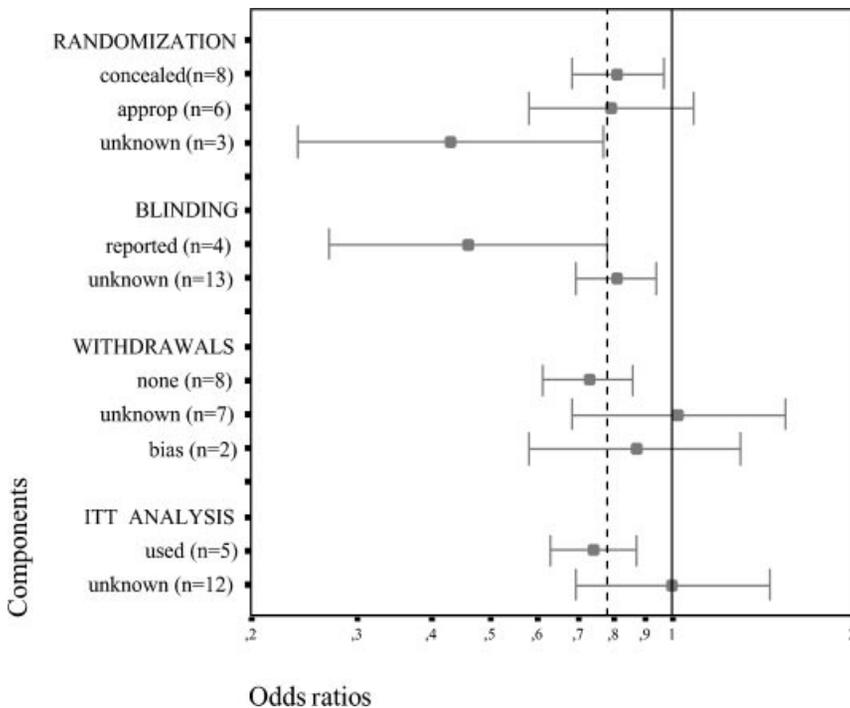
Figure 2. Pooled ORs (95% CI) for subgroup of studies according to the individual components.

and in three (36;53;59) the method of randomization was unknown. Four studies (36;43;51; 59) mentioned a form of blinding, and all four used the term "double blind." In eight studies (34;35;36;37;38;40;41;55) no withdrawals or a withdrawal rate not leading to bias was found, in seven the withdrawal rate was unknown (31;43;44;49;53;59;63), and in two studies (30;51) a withdrawal rate "possibly leading to bias" was found. In five studies (37;38;53;59;63) an ITT analysis was performed. In Figure 2 we present the pooled ORs and 95% CIs for each component in the different categories.

When the randomization method is unknown or when blinding is reported, the pooled OR greatly differs from the overall average of true effects (the gold standard pooled OR). Overall, there is a tendency to underestimate the effect in the categories unknown except for randomization. When the different components (except blinding) are properly described, the pooled estimate of effect is close to the gold standard OR.

Visual Plot. To get insight into the relationship between overall quality and effect sizes, we constructed a scatterplot of the QS against the individual ORs. Figure 3A shows the Delphi QS (vertically) against the effect size (horizontally) of all 17 full reports. Figure 3B presents the QS and the ORs of all 24 original studies; the studies that we were unable to find received a QS of 2. Both plots show no correlation between overall quality score and effect.

Threshold Score. We found a mean quality score of 5 and included studies with at least a QS of 5. The pooled OR was 0.77 (0.65–0.91). Using a threshold of the median score (median = 5) or 50% of the maximum available score (4.5), all provide the same results.

Weighting Factor. We weighted each individual study estimate by their achieved Delphi QS (Table 1). The pooled estimate was 0.78 (0.73–0.83). There is no difference in pooled OR using quality scores as a threshold or as a weight. The confidence interval here using this weighted analysis is smaller because all studies received a weight above 1.
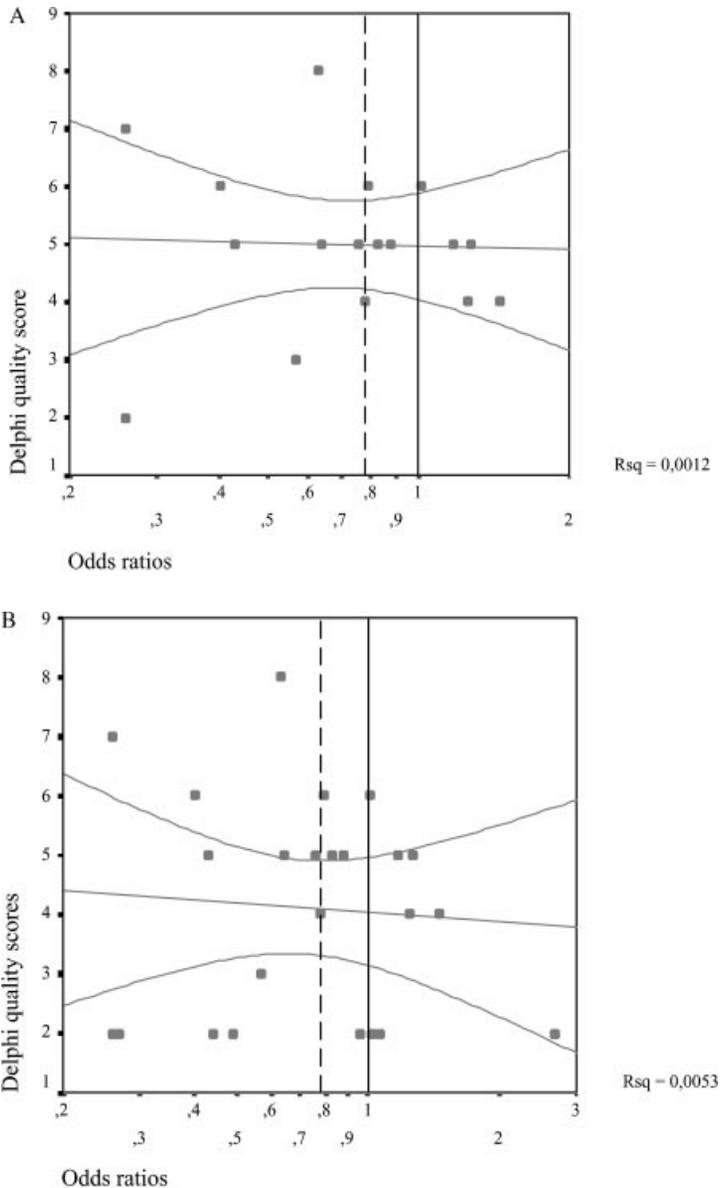
Figure 3. Scatterplot of the QS against the individual ORs, with (A) showing the ORs of the 17 full reports, and (B) showing the ORs of all 24 original studies.

Cumulative Pooling. We started the cumulative pooling with the highest scoring study, and subsequently added the ones with lower QS. When 100% of the studies is included, the original pooled estimate of effect (OR = 0.78) is reached. Figure 4 shows the cumulative pooling. The number of pooled studies in the category of high-quality studies is small and the confidence intervals are wide. When the best 40% of the studies is included, the pooled OR becomes significantly lower than 1.0.

DISCUSSION

We found no apparent influence of methodologic quality on the conclusion of a landmark meta-analysis on thrombolytic therapy. The results were consistent over the five different
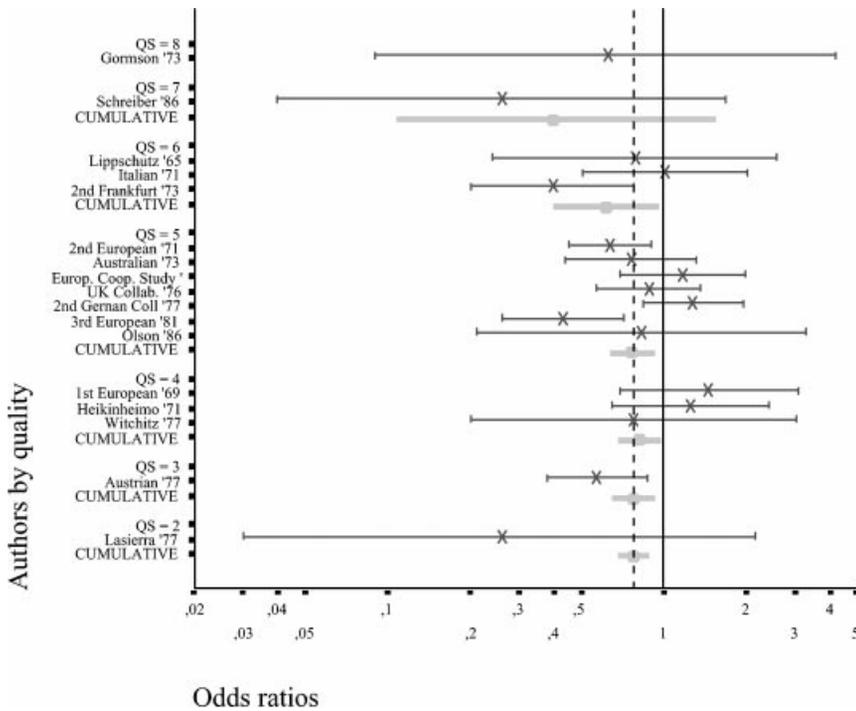
Figure 4. Cumulative odds ratios.

methods of incorporating the quality into a final conclusion. Studies with a proper description of different quality components seem to provide a good estimate of the overall average of true treatment effects.

We decided to accept the pooled OR as found by Yusuf et al. (29) as our gold standard, providing a true treatment effect. The main reason for this is that after 15 years, its results still hold. An alternative for setting a gold standard would have been to use the results of the best qualitative studies. As there were only a few of these, we would have introduced an unreliable gold standard, so in this study the Yusuf pooled OR is an overall average score. If the leading paradigm (19;20;22;23) in empirical research is true, meaning that clinical trials with a poor methodologic quality will usually be biased in favor of the intervention, we expected to find a pooled OR of high-quality studies closer to 1, and a pooled OR of the lower quality studies lower than the average Yusuf OR of 0.78. However, our research does not confirm this paradigm in empirical research. We must acknowledge that the number of studies is small and only a few are high-quality studies. The high-powered studies (i.e., large sample size) were only of low to moderate quality. Thus, our study is not strong enough to confidently reject the paradigm, and more empirical studies are needed.

When a reviewer performs a systematic review or meta-analysis and quality assessment is a part of the review process, two decisions are important. First, the choice of criteria list should be used as a valid and reliable measuring instrument. A large number of criteria lists are available (13;18), and we regard the Delphi list as a relatively valid one because of the way it is developed (27). Second, the reviewer has to decide how quality will be incorporated into the final conclusion. In the literature five different ways of incorporating the quality are described (6;12;15;16;23). Overall, our results of incorporating quality into a final conclusion is consistent, but component analysis on blinding provides a strange result we cannot explain.

Unfortunately, our data set was too small to be able to draw firm conclusions or to use meta-regression techniques for analyzing possible sources of heterogeneity. Because the studies we used were all published before the rise of empirical evidence of the importance of some design characteristics, our data set might be biased one way or the other. It is possible that authors may not have provided information needed to assess the quality, resulting in a lower quality score. On the other hand, authors probably did not provide information in order to be regarded as a high-quality study when it is not.

In conclusion, the use of quality scales and checklists is increasingly criticized. Nevertheless, it is also frequently used because it provides insight into the methodologic quality in a simple manner. Other studies using similar approaches support our view that effect of quality on outcome is unclear (13;25;27;28). Quality assessment is seen as an important part of a meta-analysis, but the influence of quality on outcome is still unclear and needs further research (13;16;17;25;27;28). In this research, the pooled effect estimate is reached irrespective of the way quality is incorporated into the final conclusion.

REFERENCES

1. Antman EM, Lau J, Kupelnick B, Mosteller F, Chalmers TC. A comparison of results of meta-analyses of randomized control trials and recomendations of clinical experts. *JAMA.* 1992;268:240-248.
2. Assendelft WJJ, Koes BW, van der Heijden GJMG, Bouter LM. Efficacy of chiropractic manipulation for back pain: Blinded review of relevant randomized clinical trials. *JMPT.* 1992;15:487-494.
3. Chalmers TC, Smith H, Blackburn B, et al. A method for assessing the quality of a randomized control trial. *Control Clin Trials.* 1981;2:31-49.
4. de Bie RA. Methodology of systematic reviews: An introduction. *Phys Ther Rev.* 1996;1:47-51.
5. de Vet HCW, de Bie RA, van der Heijden GJMG, et al. Systematic reviews on the basis of methodological criteria. *Physiotherapy.* 1997;83:284-289.
6. Detsky AS, Naylor CD, O'Rourke K, McGeer AJ, L'Abbé KA. Incorporating variations in the quality of individual randomized trials into meta-analysis. *J Clin Epidemiol.* 1992;45:255-265.
7. Egger M, Smith GD, Phillips AN. Meta-analysis: Principles and procedures. *BMJ.* 1997; 315:1533-1537.
8. Egger M, Smith GD, Sneider M, Minder C. Bias in meta-analysis detected by a simple graphical test. *BMJ.* 1997;315:629-634.
9. GISSI-1 (Gruppo Italiano per lo Studio della Streptochinasi nell "Infarcto miocardico"). Effectiveness of intravenous thrombolytic treatment in acute myocardial infarction. *Lancet.* 1986;1:397-401.
10. Haynes RB. Clinical review articles should be as scientific as the articles they review. *BMJ.* 1992;304:330-331.
11. ISIS-2 (Second International Study of Infarct Survival) Collaborative Group. Randomized trial of intravenous streptokinase, oral aspirin, both or neither among 17,187 cases of suspected acute myocardial infarction. *Lancet.* 1988;2:349-360.
12. Jenicek M. Meta-analysis in medicine: Where we are and where we want to go. *J Clin Epidemiol.* 1989;42:35-44.
13. Jüni P, Witschi A, Bloch R, Egger M. The hazards of scoring the quality of clinical trials for meta-analysis. *JAMA.* 1999;282:1054-1060.
14. Koes BW, Assendelft WJJ, van der Heijden GJMG, Bouter LM. Spinal manipulation for low back pain: An updated systematic review of randomized clinical trials. *Spine.* 1996;221:2860-2873.
15. Koes BW, van Tulder MW, van der Windt DAWM, Bouter LM. The efficacy of back schools: A review of randomized clinical trials. *J Clin Epidemiol.* 1994;47:851-862.
16. Kunz R, Oxman AD. The unpredictability paradox: Review of empirical comparisons of randomised and non-randomised clinical trials. *BMJ.* 1998;317:1185-1190.
17. Linde K, Scholz M, Ramirez G, et al. Impact of study quality on outcome in placebo-controlled trials of homeopathy. *J Clin Epidemiol.* 1999;52:631-636.
18. Moher D, Jadad AR, Tugwell P. Assessing the quality of randomized controlled trials: Current issues and future directions. *Int J Technol Assess Health Care* 1996;12;195-208.

19. Moher D, Pham B, Jones A, et al. Does quality of reports of randomised trials affect estimates of intervention efficacy reported in meta-analyses?. *Lancet.* 1998;352:609-613.

20. Mulrow CD. The medical review article: State of the science. *Ann Intern Med.* 1987;106:485-488.

21. Oxman A, ed. *Preparing and maintaining systematic reviews: The Cochrane Collaboration Toolkit 1996.* Available at: http://hiru.mcmaster.ca/cochrane.

22. Schultz KF, Chalmers I, Hayes R, Altman DG. Empirical evidence of bias: Dimensions of methodological quality associated with estimates of treatment effects in controlled trials. *JAMA.* 1995;273:408-412.

23. Schultz KF, Grimes DA, Altman DG, Hayes RJ. Blinding and exclusions after allocation in randomised controlled trials: Survey of published parallel group trials in obstetrics and gynaecology. *BMJ.* 1996;312:742-744.

24. van Tulder MW, Assendelft WJJ, Koes BW, Bouter LM, and the Editorial Board of the Cochrane Collaboration. Method guidelines for systematic reviews in the Cochrane Collaboration: Back Review Group for Spinal Disorders. *Spine.* 1997;22:2323-2330.

25. Verhagen AP, de Bie RA, Lenssen AF, et al. Impact of quality items on study outcome: Treatments in acute lateral ankle sprains. *Int J Technol Assess Health Care.* 2000;16:1136-1146.

26. Verhagen AP, de Vet HCW, de Bie RA, et al. The Delphi list: A criteria list for quality assessment of randomised clinical trials for conducting systematic reviews developed by Delphi consensus. *J Clin Epidemiol.* 1998;51:1235-1241.

27. Verhagen AP, de Vet HCW, de Bie RA, Boers M, van der Brandt PA. The art of quality assessment of RCTs for systematic reviews. *J Clin Epidemiol.* 2001;54:651-654.

28. Wahlbeck K, Tuunainen A, Gilbody S, Adams CE. Influence of methodology on outcomes of randomised clozapine trials. *Pharmacopsychiatry.* 2000;33:54-59.

29. Yusuf S, Collins R, Peto R, et al. Intravenous and intracoronary therapy in acute myocardial infarction: Overview of results on mortality, reinfarction and side-effects from 33 randomized controlled trials. *Eur Heart J.* 1985;6:556-585.

References of the Meta-Analysis

30. Aber CP, Bass NM, Berry CL, et al. Streptokinase in acute myocardial infarction: A controlled multicentre study in the United Kingdom. *BMJ.* 1976;2:1100-1104. (UK Collaborative.)

31. Amery A, Roeber G, Vermeulen HJ, Verstraete M. Single-blind randomised multicentre trial comparing heparin and streptokinase treatment in recent myocardial infarction. *Acta Med Scand Suppl.* 1969;505:5-35. (1st European.)

32. Anderson JL, Marshall HW, Bray BE, et al. A randomized trial of intracoronary streptokinase in the treatment of acute myocardial infarction. *N Engl J Med.* 1983;308:1312-1318.

33. Anderson JL, McIlvaine PM, Marshall HW, et al. Long-term follow-up after intracoronary streptokinase for myocardial infarction: A randomized controlled study. *Am Heart J.* 1984;108:1402-1408.

34. Benda L, Haider M, Ambrosch F. Results of the Austrian myocardial infarction study on the effects of streptokinase (in German). *Wein Klin Wochenschr.* 1977;89:779-783.

35. Bett JHN, Biggs JC, Castaldi PA, et al. Australian multicentre trial of streptokinase in acute myocardial infarction. *Lancet.* 1973;1:57-60.

36. Breddin K, Ehrly AM, Fechler L, et al. Short-term fibrinolytic treatment in acute myocardial infarction (in German). *Dtsch Med Wochenschr.* 1973;98:861-873.

37. Dioguardi N, Mannicci PM, Lotto A. Controlled trial of streptokinase and heparin in acute myocardial infarction. *Lancet.* 1971;2:891-895.

38. European Cooperative Study. Controlled trial of urokinase in myocardial infarction. *Lancet.* 1975;2:624-626.

39. European Cooperative Study Group for Streptokinase Treatment in Acute Myocardial Infarction. Streptokinase in acute myocardial infarction. *N Engl J Med.* 1979;301:797-802.

40. The European Cooperative Study Group. Streptokinase in acute myocardial infarction extended report of the European Cooperative Trial. *Acta Med Scand Suppl.* 1981;648:7-57.

41. European Working Party. Streptokinase in recent myocardial infarction: A controlled multicentre trial. *BMJ.* 1971;3:325-331.

42. Fletcher AP, Sherry S, Alkjaersig N, Smyrniotis FE, Jick S. The maintainance of a sustained thrombolytic state in man, II: Clinical observations on patients with myocardial infarction and other thromboembolic disorders. *J Clin Invest.* 1959;38:1111-1119.

43. Gørmsen J, Tidstrom B, Fedderson C, Ploug J. Biochemical evaluation of low dose of urokinase in acute myocardial infarction: A double blind study. *Acta Med Scand.* 1973;194:191-198.

44. Heikinheimo R, Ahrenberg P, Honkapohja H, et al. Fibrinolytic treatment in myocardial infarction. *Acta Med Scand.* 1971;189:7-13.

45. Kennedy JW, Ritchie JL, Davis KB, Fritz JF. Western Washington randomized trial of intracoronary streptokinase in acute myocardial infarction. *N Engl J Med.* 1983;309:1477-1482.

46. Kennedy JW, Ritchie JL, Davis KB, et al. Western Washington randomized trial of intracoronary streptokinase in acute myocardial infarction: A 12-month follow-up report. *N Engl J Med.* 1985;312:1073-1078.

47. Khaja F, Walton JA, Brymer JF, et al. Intracoronary fibrinolytic therapy in acute myocardial infarction: Report of a prospective randomized trial. *N Engl J Med.* 1983;308:1305-1311.

48. Kolibash AJ, Magorien RD, Bashore TM, Bush CA. Does acute thrombolytic recanalization favorably alter segmental left ventricular function? *Circulation.* 1984;70(suppl 2):258. Abstract.

49. Lasierra C, Vilades J, Fernandez C, et al. Estreptoquinasa en el infarcto agudo de miocardio. *Revista Clinica Espaniola.* 1977;144:251-7.

50. Leiboff RH, Katz RJ, Wasserman AG, et al. A randomized, angiographically controlled trial of intracoronary streptokinase in acute myocardial infarction. *Am J Cardiol.* 1984;53:404-407.

51. Lippschutz EJ, Ambrus JL, Ambrus CM, et al. Controlled study on the treatment of coronary occlusion with urokinase-activated human plasmin. *Am J Cardiol.* 1965;16:93-98.

52. Ness PM, Simon TL, Cole C, Walston A. A pilot study of streptokinase therapy in acute myocardial infarction: Observations on complications and relation to trial design. *Am Heart J.* 1974;88:705-712.

53. Olson HG, Butman SM, Piters KM, et al. A randomized controlled trial of intravenous streptokinase in evolving acute myocardial infarction. *Am Heart J.* 1986;111:1021-1029.

54. Olson HG, Lyons KP, Butman S, et al. A randomized controlled trial of intravenous streptokinase in acute myocardial infarction. *Circulation.* 1984;70(suppl 2):155. Abstract.

55. Poliwoda H, Schneider B, Avenarius HJ. Investigations of the clinical course of acute myocardial infarction, I: The fibrinolytic treatment of acute myocardial infarction with streptokinase (in German). *Med Klin.* 1977;72:451-463.

56. Raizner AE, Tortoledo FA, Verani MS, et al. Intracoronary thrombolytic therapy in acute myocardial infarction: A prospective randomized trial. *Am J Cardiol.* 1985;55:301-305.

57. Rentrop KP, Feit F, Blanke H, et al. Effects of intracoronary streptokinase and intracoronary nitroglycerin infusion on coronary angiographic patterns and mortality in patients with acute myocardial infarction. *N Engl J Med.* 1984;311:1457-1463.

58. Schreiber TL, Miller DH, Borer JS, et al. Efficacy of intravenous heparin vs streptokinase in acute myocardial infarction. *Circulation.* 1984;70(suppl 2):27. Abstract.

59. Schreiber TL, Miller DH, Silvasi DA, et al. Randomized double blind trial of intravenous streptokinase for acute myocardial infarction. *Am J Cardiol.* 1986;58:47-52.

60. Simoons ML, Serruys PW, van der Brand M, et al. Improved survival after early thrombolysis in acute myocardial infarction: A randomized trial by the Interuniversity Cardiology Institute in the Netherlands. *Lancet.* 1985;i:578-581.

61. Valère PE, Guérot C, Castillio-Fenoy A, et al. L'infarctus myocardique: Traitement randomisé par la streptokinase. *La Nouvelle Presse Médicale.* 1975;4:190. Abstract.

62. Verani MS, Tortoledo FA, van Reet RE, et al. Intracoronary thrombolysis in acute myocardial infarction: Effects on function of myocardium at risk. Preliminary findings of a randomized trial. *J Am Coll Cardiol.* 1983;1:592. Abstract.

63. Witchitz S, Kolsky H, Moisson P, Chiche P. Streptokinase et infarctus du myocarde aigu. La fibrinolyse peut-elle limiter la nécrose? *Ann Cardiol Angélol.* 1977;26:53-56.