# Protection against bias in randomised clinical trials and meta-analyses.

## The role of allocation concealment and other methodological design measures to protect against bias.

**Julie Pildal, MD**

**PhD Thesis**
**Faculty of Health Sciences**
**University of Copenhagen**
**2005**

**The Nordic Cochrane Centre, H:S Rigshospitalet**
**DK-2100 Copenhagen**
**Denmark**

Cover illustration reproduced courtesy of Dorte Holbæk.

# Protection against bias in randomised clinical trials and meta-analyses.

**The role of allocation concealment and other methodological design measures to protect against bias.**

**PhD Thesis by Julie Pildal, MD**

---

**CONTENTS**

## Acknowledgements

The research contained in this thesis was done during my employment at the Nordic Cochrane Centre, the Copenhagen Hospital Corporation, in the period 2001 to 2004.

First and foremost I would like to thank my primary supervisor Peter Gøtzsche for the opportunity of doing a PhD-thesis and for providing first-rate working conditions. Peter Gøtzsche has given highly qualified and detailed feedback, often at very short notice, and has provided valuable ideas, guidance and inspiration, while also permitting me to pursue my own ideas. The few responsibilities I have had in addition to doing research have all been interesting and relevant to my further career.

Through Peter Gøtzsche's international collaboration, I established contact with Doug Altman and An-Wen Chan, and Peter has supported the harvest of the full potential of the extensive material collected for study III by encouraging me to collaborate in the future with other researchers who have performed similar work in a project funded by a research grant from the UK National Health Service.

I am also indebted to my co-supervisor Jørgen Hilden, who has been there for me at various critical moments and given guidance both on statistical and design issues and on manuscripts, as well as with regard to wider perspectives such as whether and how to go abroad. He has also managed to deal efficiently with my sometimes desperately tight schedules.

Doug Altman gave me the opportunity of experiencing the wonderful people and the intellectual life in his inspiring research group during my three-month stay in Oxford. It was well worth lining up with the other researchers in front of his office to seek his advice, but it was also good to get out of the office again while I could still recall some of the many suggestions he had given me. Among Doug Altman's staff, I am particularly grateful for Mike Bradburn's help in sorting out and replicating the most esoteric meta-analyses that I encountered in study III.

During my employment at the Nordic Cochrane Centre I had the fortune to have Asbjørn Hróbjartsson, Karsten Juhl Jørgensen, and Elisabeth Forfang as fun, flexible and highly competent colleagues and co-authors. An-Wen Chan visited the centre in 2003 and has been an inspiring and hard-working colleague to collaborate with.

I am much indebted to the IT development group at the Nordic Cochrane Centre: Jacob Riis, Monica Kjeldstrøm, Rasmus Moustgaard and Peter Friis, who have been very creative, fast and helpful in solving all kinds of IT and layout related problems that I have encountered.

Finally, I would like to thank the secretaries of the Nordic Cochrane Centre, Jannie Hedegaard and Marian Pandal. I have been completely dependent on Jannie Hedegaard's dedicated help of high quality in the retrieval of approximately 600 references for study III. It has also been a pleasure to work with Marian Pandal, even with the most frustrating tasks such as trying to sort out my budget.

I would like to thank the Scientific and Ethical Committees for Copenhagen and Frederiksberg, who provided access to the trial protocols and offered administrative support. Without this, study II would not have been possible.

I also thank the Medicinsk Bibliotek at Rigshospitalet and The Danish National Library of Science and Medicine for extraordinary help with the retrieval of the literature; the people at Odense University who are responsible for the STATA-newcomers list, in particular Lars Korsholm, for quick and valuable responses; and the PhD-board at the Faculty of Health Sciences, Copenhagen University, for financial support of my participation in several courses abroad and for financing my three-month stay in Oxford.

Lastly, I thank my husband, Dan Zahavi, for all his support and encouragement.

*Julie Pildal*

## Summary

The overall theme of this thesis is the degree of bias protection in randomised clinical trials (RCTs) and how this is dealt with in systematic reviews. The main focus is on allocation concealment, which serves to avoid selection bias by masking what treatment the next patient to be enrolled will receive until the patient has been irreversibly included. On average the lack of reported adequate measures of allocation concealment in the trial reports is associated with approximately 20% exaggeration of the paucity of undesirable outcomes of the experimental treatment when measured as ratio of odds ratios. As the effect of many of the interventions provided by health-care systems is within this range, it is often unknown whether our health care interventions can be relied upon to convey a true effect rather than an apparent effect reflecting bias. On a worldwide perspective the financial implications are overwhelming.

In a systematic review of RCTs of whether intravenous administration of polyclonal immunoglobulin for adjuvant treatment of sepsis reduces all cause mortality, we found that relative risk of death with immunoglobulin treatment was 0.77 (95% confidence interval, 0.68–0.88) if all trials were included, whereas the trials with a high level of reported bias protection, which comprised almost half of the total evidence, failed to show a reduction in mortality: relative risk of 1.02 (95% CI, 0.84–1.24).

In the second paper we compared 102 pairs of published trial reports and their protocols and found that most trials with unclear allocation concealment in the trial reports also had unclear allocation concealment in the protocols. This suggests that drug-regulatory authorities and the science ethics committees should be given the resources and incentive to sanction only trial protocols with descriptions of adequate methods of allocation concealment.

In the last paper we estimated the percentage of "positive" conclusions drawn from meta-analyses that remained supported at a 5% significance level when only trials with adequate allocation concealment were included in the re-analyses. Two-thirds of conclusions lost their support, partly due to loss of statistical power, and partly due to a shift in the overall point estimate towards no effect.

In conclusion, we have shown an example of the importance of adequate bias protection, and have estimated the prevalence and characteristics of bias protection measures at protocol, trial and review level, and thereby disclosed that these issues need greater attention.

## Resumé

Det overordnede emne for afhandlingen er graden af beskyttelse mod bias i randomiserede kliniske studier og hvordan dette håndteres i oversigtsartikler. Der fokuseres hovedsageligt på maskering af patient allokeringen. Denne forholdsregel har til formål at undgå selektion-bias, dvs. at undgå, at patienter med god prognose fortrinsvis kanaliseres ind i den behandlingsarm, der foretrækkes af den person, som står for inklusion af patienter og tildeling af, hvilken behandling der skal gives. Det opnås ved at maskere, hvilken behandling den næste patient, som står overfor at blive inkluderet i et klinisk studie, vil blive allokeret til indtil vedkommende er uigenkaldeligt inkluderet. I gennemsnit er manglende beskrivelse af denne forholdsregel i publikationer af kliniske randomiserede studier associeret med en overdrivelse af den eksperimentelle behandling med ca. 20% (målt som ratio af odds ratioer). Eftersom effekten af mange af de interventioner, der anvendes i sundhedsvæsnet er af denne størrelsesorden er det ofte uvist, hvorvidt man forsøger at tilvejebringe en tilsyneladende effekt generet af bias eller en reel effekt. I et internationalt perspektiv er det således et spørgsmål om potentielt spild af overvældende ressourcer.

I en systematisk oversigtsartikel over randomiserede kliniske studier af hvorvidt supplerende behandling af sepsis med immunoglobuliner reducerer dødeligheden, fandt vi, at den relative risiko for død indenfor 30 dage var 0.77 (95% usikkerheds-interval, 0.68–0.88) ved behandling med immunoglobuliner, hvis alle studierne indgik. Derimod viste de studier, der havde et højt niveau af beskyttelse mod bias, og som udgjorde ca. halvdelen af den statistiske information, ikke nogen reduktion af den relative risiko for død: 1.02 (0.84–1.24).

I den anden publikation sammenlignede vi 102 par af samhørende forsøgsprotokoller og publicerede rapporteringer af randomiserede kliniske studier. Vi fandt, at de fleste studier, som havde uklar maskering af allokeringen, også havde uklar maskering af allokeringen i følge

protokollen. Dvs. at politikere på internationalt plan bør give lægemiddelstyrelser og etiske komiteer ressourcer og incitament til kun at godkende protokoller, der eksplicit beskriver tilstrækkelige metoder til at sikre maskering af allokering.

I det sidste manuskript estimerede vi hvilken andel af de konklusioner, der drages af meta-analyser, der fortsat understøttes af statistisk signifikante resultater, hvis der kun inddrages studier med tilstrækkelig maskering af allokering. Hovedparten af konklusioner mistede deres støtte, dels p.gr.a. tab af statistisk styrke, og dels fordi de tilbageblevne studier viste en mindre effekt af den eksperimentelle behandling.

Samlet set har vi i form af en systematisk oversigtsartikel givet et eksempel på vigtigheden af tilstrækkelig beskyttelse mod bias. Desuden har vi vurderet prævalensen og karakteren af beskyttelse mod bias i randomiserede kliniske interventions-studier og disses protokoller, samt i oversigtsartikler. Vi  har dermed påvist et stort behov for, at dem som godkender protokoller, er opmærksomme på at forhindre gennemførslen af studier, der kan give skævvredne resultater. Vi har endvidere afdækket et tilsvarende stort behov for, at forfattere, som opsummerer randomiserede studier i oversigtsartikler, forhindrer implementering af interventioner, der af samme grund ikke er tilstrækkeligt velfunderede.

## Introduction

The overall theme of this thesis concerns the level of protection against bias in randomised clinical trials (RCTs) and how this is dealt with in research synthesis at review level. Bias is a systematic distortion of the results in such a way that they do not exclusively reflect the difference of the effect of the interventions under investigation, but also (or exclusively) the impact of flawed research methods. The sources of bias are multiple; here we focus on 1) selection bias, 2) performance and ascertainment bias, 3) attrition bias and 4) publication bias.

Selection bias occurs if patients with a better prognosis are preferentially channelled into one of the treatment arms. If this happens, the results of the trials will to some extent reflect the difference in prognosis. Selection bias may be caused by more or less conscious manipulation by the persons involved in enrolling patients into a trial. If, for instance, patients are allocated to two treatments on the basis of a computer-generated simple randomisation scheme, then this may seem to be a plausible guarantee of adequate randomisation, but if the randomisation plan is then posted on the bulletin board it will be quite transparent to the persons involved in the recruitment of patients, what treatment the next patient to be enrolled will get. This would allow the investigators to select the sequence of patients to be enrolled, e.g., if the next eligible patient has a good prognosis, but would be randomised into the treatment that the investigator does not wish to promote, then he or she could pretend to be too busy to enrol patients at that particular time or describe the trial in terms that would cause the patient to decline the participation request. To prevent the persons in charge of enrolment from introducing selection bias, the sequence of upcoming treatment allocations has to remain concealed until the patients have been irreversibly enrolled. Schulz et al. coined the term “allocation concealment” in 1995 to differentiate the concealment of the allocation sequence up until intervention assignment, which can always be done,

from the blinding with respect to the nature of the intervention to be implemented, which cannot always be done.[1] Allocation concealment can be achieved in a several ways, for instance by requiring that the clinician first enrols the patient into the trial and then contacts a remote central randomisation centre to obtain the patient's treatment assignment.

According to authoritative sources double-blinding serves to prevent ascertainment bias (differential reporting on and detection of symptoms and treatment effects by patients and investigators) and performance bias (differential administration of the intervention or co-interventions by care provider or patients). It is less recognised that it also serves to retain patients in trials and to improve their compliance with the control treatment [2-4] Most often double-blinding entails that the patient and investigators are blinded, and frequently that the investigators also serve as outcome assessors. However, in general, the term double-blinding is inconsistently used and defined. [5-8] Finally, the success of double-blinding should not be taken for granted, as, for example, side-effects may undermine it. Albeit, tests of the success of blinding are rare in published trial reports. [9]

Attrition bias may occur when patients drop out of, or are excluded from, a study. This will entail bias if the pattern of these patients' prognoses differs between the treatment arms. In fact, it will undermine the randomisation. To avoid violation of the randomisation, the trial results can be analysed by "intention to treat analysis", which means that the analysis compares the results of patients according to the groups they were originally allocated to. The term "intention-to-treat analysis" is inconsistently and sometimes overtly incorrectly used.[10] It can refer to that some or all of the following randomised patient categories were included in the analysis: 1) False inclusions (enrolled patients not fulfilling the inclusion criteria) 2) Patients who failed to start the intervention. 3) Deviation from randomisation due to non-compliance of the patients. However, full application of intention to treat analysis is possible only when complete outcome data are available for all

randomised subjects. Thus, a differential loss to follow-up within or between arms, remains problematic, because one can only guess at what happened to these patients, leading to an increased level of uncertainty about what the true effect of the treatment under investigation, is. Yet it is, of course, important that investigators honestly report such losses and discuss their potential effect. [10]

Finally, as to publication bias, Study I illustrates an example. Publication bias means that the scientific literature does not fairly reflect the body of research. In particular, trials with statistically significant results are more likely to be published [11-13] than trials with negative results. As a consequence, they are more likely to become identified and included in the research synthesis of review articles. [14] Thus, this is not a problem at the design or analysis level of the individual trial, but rather at the publication and subsequent research synthesis level. However, the prevention of publication bias or at least detection of it will be at the protocol level, in that only if trial protocols become publicly registered, will researchers be able to reliably detect publication bias.

All of these issues of bias protection measures have something to do with how to prevent investigators from distorting the results of their own research. Preoccupation with such matters may seem excessively misanthropic. However, as will be described in the background section, substantial empirical evidence suggests that investigators do have inclinations that tend to undermine the validity of their own research, although this may not be deliberate or even come to the attention of the investigators themselves. Regardless of whether these inclinations are caused by strong beliefs, or noble or questionable intentions, it is important to be aware of them and try to prevent their influence.

## Background

The shared background of all the three papers of this thesis is the accumulating empirical evidence of the importance of bias prevention. This evidence primarily concerns the following components of bias protection: allocation concealment, double-blinding and intention to treat analysis at the level of randomised controlled trials (RCTs). When the results of RCTs are to be summarised at review level their susceptibility to bias has to be taken into account in conjunction with the potential problem of publication bias. The most extensive documentation available regards the impact of unclear or inadequate allocation concealment in published reports of trials. Accordingly, this lack of reported adequate allocation concealment receives the primary attention in this thesis.

At the planning stage of the thesis, four studies had consistently (formal test for statistical heterogeneity <0.10) shown that lack of reported adequate allocation concealment was associated with an exaggeration of the experimental treatment of approximately 30% when measured as ratio of odds ratios. [1 15-17]. That is, the ratio of odds ratios of trials with *lack of* reported adequate allocation concealment to the odds ratios of trials *with* reported adequate allocation concealment was 0.70 (95% confidence interval 0.62 to o.80), thus on average the relative odds ratios of trials with lack of reported adequate allocation concealment were 30% lower (more beneficial, since the events were undesirable). What this corresponds to in terms of absolute reduction of risk depends on the baseline event rate and the magnitude of the treatment effect, but, in general, it is within the magnitude of the treatment effect most RCTs seek to detect [18]. Thus, it is crucial to investigate when the results of a trial can be ascribed to a true treatment effect rather than bias. Schulz et al. were the first to quantify an exaggeration associated with lack of reported adequate allocation concealment based on meta-analyses. The merit of using meta-analyses for this purpose is that it allows comparisons of the treatment effects in similar trials with and without the bias protection

component of interest, because meta-analyses are based on the assumption that any variation in the interventions and conditions will not influence the treatment effect in a systematic way. Thus, the estimated influence of the bias protection component will not be confounded by the type of intervention or medical condition, provided that this assumption is not violated. Schulz et al. found that on average the ratio of odds ratios of trials with unclear vs. reported adequate allocation concealment was 0.70 (95% confidence interval 0.62 to 0.79). In this analysis other components of bias protection i.e. double-blinding, no exclusions of patients after randomisation, and adequate generation of allocation sequence, were controlled for in a logistic regression analysis. The ratio was determined as an statistical interaction between treatment arm (experimental vs. control) and status of allocation concealment, and a fixed effect within and between meta-analyses was assumed.[1] The meta-analyses included in this study were all of RCTs in obstetrics,[1] but the study was followed-up by Moher et. al., Kjaergard et al and Jüni et al., who found similar results in other specialties; the latter two studies used slightly different criteria when operationalising the definition of adequate allocation, other inclusion criteria for meta-analyses and different strategies for analysing data. Thus, the results appeared to be widely generalisable, and the sum of them was a ratio of odds ratios of 0.70 (95% CI: 0.62 to 0.80) when trials with lack of reported adequate concealment (that is, trials with inadequate and unclear allocation concealment) were compared to trials with reported adequate allocation concealment.[17] Subsequently, additional studies were published [19 20]. The characteristics and the implications of these studies [19 20] are outlined and discussed in the third paper as well as in the Discussion chapter of this overview.

Schulz has provided a collection of personal accounts of how subversion of the random sequence was easily achieved when inadequate methods for allocation concealment were employed, for instance by altering the sequence of patients to be enrolled when an open list with the allocation sequence was posted on the bulletin board. Examples of subversion of measures that usually

provide adequate concealment, e.g. central randomisation, are also given, for instance: " trial investigators related ringing the central number and asking for several assignments all at once".[21] Hence, allocation concealment can be achieved with a certain level of assurance, but probably never with a 100% guarantee.

Although reported adequate allocation concealment seems so important for the validity of the trials' results, it is not particularly frequently reported. The prevalence of unclear or inadequate allocation concealment ranges from 39% in recent trial publications in high impact general medical journals that endorse the CONSORT statement[22] to a much higher prevalence in less recent publications in lower impact specialty journals; for instance, 93% of RCT in a dermatology journal from 1976 to1997 [23] and 97% of RCTs published in Intensive Care Medicine from 1975 to 2000.[24] Other examples based on RCTs published more recently or published in specialty journals with an intermediate impact factor find a prevalence of unclear or inadequate allocation concealment somewhere in between. [25-30] [16] [31] [32].

This suboptimal reporting is targeted in the CONSORT statement (Consolidated Standards of Reporting Trials), which is a guideline under continuous development that seeks to help authors and editors to improve reporting of intervention trials. Allocation concealment is just one of the CONSORT group's concerns; the CONSORT statement provides other recommendations to enhance accurate reporting of essential bias protection measures such as, whether and how double-blinding or analysis by intention to treat was applied.[22] It was first published in 1996,[33] revised and accompanied by elaborations and explanations in 2001[34] and extended in 2004 with a focus on reports of harm[35]. Two studies have demonstrated that of endorsement of the CONSORT statement improves the quality of reporting on allocation concealment and other bias protection measures in general medical journals,[22] as well as in specialty journals,[36] although there is still room for strengthened enforcement of these recommendations. Today, 175 journals endorse CONSORT

(www.consort-statement.org accessed 9 March 2005). Influential editorial groups have also adopted it, including the International Committee of Medical Journal Editors (ICMJE, also known as the Vancouver group),[37] the council of Science Editors (CSE), and the World Association of Medical Editors (WAME).[38]

The impact of lack of double-blinding has been estimated in the same studies mentioned above [1 15-17]. The pooled estimate of the impact was a ratio of odds ratios of 0.86 (95% CI: 0.74 to 0.99). This implies that, on average, trials without double-blinding show a 14% more beneficial effect compared with similar trials with double-blinding on an odds ratio scale. [17] This estimate of 14% is to be interpreted with the same precautions as detailed above for the corresponding estimate of the impact of lack of reported adequate concealment. The four studies each contributed with point estimates that were more heterogeneous than was the case when the impact of allocation concealment was assessed. However, there was no demonstrable statistical heterogeneity, partly because the two most aberrant results ROR = 1.11 [15] and ROR = 0.56 [16] were derived from relatively small samples, but also because the sensitivity of the test for heterogeneity is low [39]. Lack of double-blinding may be associated with different size of bias depending the type of outcome assessed; all cause mortality will be less susceptible to ascertainment bias, whereas more subjective outcomes such as pain are more susceptible to bias. [5] Nevertheless, trials with lack of blinding may yield more conservative estimates of the effect of the experimental treatment effect, because of a more extensive contamination, than trials with successful blinding. Contamination occurs if patients assigned to what they believe is the inferior treatment, seek to compensate for this by obtaining a more efficient treatment e.g. from their general practitioner. [7] The studies assessing the impact of lack of double-blinding defined double-blinding by the following phrases: "purported to be so according to the trial report", [1 15] as "identical placebo tablets or similar", [16] or "described as double-blind or assessor blind".[40] In Study I to III of this thesis, a trial is categorized as double-

blind if described in the trial report or protocol as double-blind; or patient and provider/physician were said to be blinded. If the phrase "placebo-controlled" was used without any indication that the treatments might be distinguishable or that the care provider might have become un-blinded before the onset of the treatment, then this is also categorized as double-blinded. Patient and assessor blinding is not categorized as double-blinding.

Attrition bias and publication bias are relevant in Study I, but not in Study II and III. Several studies have attempted to estimate the influence of lack vs. presence of an intention to treat analysis, [1 16 19] but found little evidence of an influence on apparent treatment effects. However, the methods employed to assess the risk of attrition bias in these studies were problematic. For instance, Schulz et al compared trials that reported exclusions with those that either did not report exclusions or reported that no exclusions had occurred. The underlying assumption that exclusions did not occur if they were not reported was questioned by the authors themselves, who argued that published information on exclusions may be of little value in assessing the risk of attrition bias.[1] However, a recent study overcame the problem of relying on published data by using individual patient data meta-analyses.[41] They found that on average, original analyses compared with analyses including all randomised patients tended to favour the experimental treatment.

Publication bias is also associated with an overestimation of the effect the experimental treatment,[40 42] with a combined average ratio between effect estimates derived from unpublished trials vs. published trials of 1.11, where a ratio above 1 indicated a less beneficial effect of the unpublished trials (95% CI: 1.03 to 1.18).[43] That is, the results of trials that remain unpublished tend not to be statistically significant, and the treatment effects are also less beneficial than for comparable published trials.

## Aims of the studies

Prompted by the accumulating empirical evidence of the importance of bias prevention in randomised controlled trials and in meta-analyses, we ventured to investigate the following primary aims:

- In Study I: RCTs and systematic reviews have shown conflicting results as to whether polycloncal immunoglobulin as adjunctive treatment of bacterial sepsis reduces mortality. The level of bias protection in the RCTs that contributed to the reviews was highly variable. We hypothesised that bias susceptibility of the individual trials and how this was dealt with at review level might explain the discrepant findings. Thus we aimed to determine whether, polycloncal immunoglobulin as adjunctive treatment of bacterial sepsis reduces mortality if only RCTs with a high level of bias protection are relied upon.
- In Study II: Allocation concealment is very frequently unclear in published reports of RCTs. As this bias protection measure appears to be one of the most important, it is crucial to gain knowledge on whether and how allocation concealment can be assumed to have occurred in such trials. Among RCTs with unclear allocation concealment in the published trial report, we aimed to determine the fraction that represents unclear reports of adequate measures to ensure allocation concealment as judged by their protocol.
- In Study III: On an overall average RCTs with unclear or inadequate allocation concealment appear to report exaggerated estimates of treatment effects. We wished to quantify how much this influenced clinical decision-making. To do so, we determined the percentage of "positive" conclusions based on a meta-analysis result that remains supported if only trials with reported adequate allocation concealment are included.

MAJOR ARTICLE

# Polyclonal Immunoglobulin for Treatment of Bacterial Sepsis: A Systematic Review

**Julie Pildal and Peter C. Gøtzsche**

The Nordic Cochrane Centre, Rigshospitalet, Copenhagen, Denmark

**Randomized trials of adjunctive treatment of bacterial sepsis with polyclonal immunoglobulin show conflicting results. We performed a systematic review and a meta-analysis of the results of randomized trials that compared reductions in mortality rates in patient groups treated with polyclonal immunoglobulin versus either placebo or no treatment in addition to conventional treatment. High-quality trials had adequate concealment of allocation, were double-blinded and placebo-controlled, and made data available for intention-to-treat analyses. Twenty trials were included. Meta-analysis of all trials showed a relative risk of death with immunoglobulin treatment of 0.77 (95% confidence interval [CI], 0.68–0.88). High-quality trials (involving a total of 763 patients, 255 of whom died) showed a relative risk of 1.02 (95% CI, 0.84–1.24), whereas other trials (involving a total of 948 patients, 292 of whom died) showed a relative risk of 0.61 (95% CI, 0.50–0.73). Because high-quality trials failed to demonstrate a reduction in mortality, polyclonal immunoglobulin should not be used for treatment of sepsis except in randomized clinical trials.**

Randomized trials of polyclonal immunoglobulin for treatment of sepsis have yielded conflicting results [1, 2]. Systematic reviews have also come to different conclusions. Alejandria et al. [1] found that polyclonal immunoglobulin reduced mortality substantially and significantly among adults (relative risk [RR], 0.62; 95% CI, 0.49–0.79), but not among neonates (RR, 0.70; 95% CI, 0.42–1.18). A review by Ohlsson and Lacy [2] reported a marginally statistically significant reduction in mortality among neonates with suspected sepsis (RR, 0.63; 95% CI, 0.40–1.00).

At our hospital, immunoglobulin constitutes the second largest drug cost. That expenditure may be justified if it saves lives. Most of the evidence supporting its use is provided by small trials (which have a large random error) with methodological shortcomings (including increased risk of systematic error [i.e., bias]). Thus, we decided to perform an independent systematic review, with emphasis on the methodological quality of the studies.

## PATIENTS AND METHODS

***Study selection and search strategy.*** We selected clinical trials described as randomized by the investigators, comparing reductions of mortality in any patient group with suspected or proved sepsis or septic shock treated with polyclonal immunoglobulin versus in those receiving placebo or no treatment in addition to conventional treatment. Studies focusing solely on prevention of sepsis were excluded. A free text literature search of all records in the databases of PubMed, Embase, and the Cochrane Library was last updated 21 January 2004. The search strategy included bacterial infection, to allow identification of studies containing results derived from subgroups with sepsis. The following groups of terms were searched: (1) "sepsis OR septicemia OR septicaemia OR shock-septic OR bacteriemia OR bacteraemia OR bacteremia," (2) "bacterial infections OR bacterial infection OR bacterial-infections," (3) "immunoglobulin OR immunoglobulins OR antibodies OR antibody OR polyclonal," (4) "randomi* OR controlled

Received 5 October 2003; accepted 4 February 2004; electronically published 1 June 2004.

Reprints or correspondence: Dr. Peter Gøtzsche, The Nordic Cochrane Centre, Rigshospitalet, Dept. 7112, Blegdamsvej 9, DK-2100 Copenhagen Ø, Denmark (p.c.gotzsche@cochrane.dk).

**Clinical Infectious Diseases 2004; 39:38–46**



1058-4838/2004/3901-0008$15.00

OR blind* OR placebo OR "controlled ? trial," and (5) the combination of the terms listed in (3), (4), and either (1) or (2).

The database-specific indexing term is one of the synonyms in each of the first 3 search strings. No restrictions were applied. Decisions on which of the retrieved trials to include were made independently by the 2 reviewers. The first authors of the included trials were asked if they were aware of any unpublished trials. Reference lists were scanned for additional trials.

***Outcomes.*** According to the protocol for the review, the primary aim was to assess whether treatment with immunoglobulin reduced total 30-day mortality in patients with suspected or proved sepsis. Secondary outcomes were number of days in hospital (if separate data for survivors and nonsurvivors were available, because pooled data can be misleading), complications to the infection, and adverse effects of immunoglobulin treatment.

The following sensitivity analyses were planned according to the protocol: High- versus lower-quality trials (a priori primary subgroup analysis); sepsis due to gram-negative organisms versus sepsis due to gram-positive organisms; neonates versus nonneonatal patients; immunocompetent versus nonimmunocompetent patients; underlying diseases; and albumin as placebo versus other placebos or no placebo (because albumin has been implied to increase mortality in seriously ill patients) [3].

***Quality assessment.*** Trials were considered high quality if they (1) had adequate concealment of allocation, (2) were double-blinded and placebo controlled, and (3) applied an intention-to-treat analysis or data were available that allowed an intention-to-treat analysis [4]. Trials failing to meet $\geqslant 1$ of these criteria were considered lower quality. We restrict the use of the term "quality" to refer to these criteria.

We considered concealment of allocation adequate if there was central randomization; serially numbered, opaque, sealed envelopes; sequentially numbered but otherwise identical vehicles, including their contents; or other descriptions of convincing concealment of allocation. Concealment was inadequate if there was alteration; reference to case record numbers or date of birth; an open table of random numbers (unless the vehicles were correspondingly numbered and the blinding impeccable). Unclear concealment meant that there was no description of the method or that the description did not allow a clear distinction.

***Data extraction.*** The 2 investigators independently extracted the data. Disagreements were rare and were the result of simple errors. All first authors of the included trials were contacted and asked for additional information on trial quality.

***Data analysis.*** RRs were combined in a meta-analysis by the Mantel-Haenszel method with use of RevMan software, version 4.2.3 (Cochrane; available from http://www.cochrane.org) [5]. A fixed-effect model was used, which assumes that the true effect of the intervention is the same in all of the included trials, differences between study results being ascribed to sampling error. Variation in study results not ascribable to sampling error were referred to as heterogeneity. Large studies with high event rates received the most weight in the meta-analysis. The a priori primary hypothesis for exploring sources of heterogeneity was the influence of methodological quality, followed by the other sensitivity analyses.

According to the protocol, tests for heterogeneity were to be performed with use of the method of DerSimonian and Laird [6] and a test for interaction [7]. The former method [6] was replaced by a more sensitive test ($I^2$) [8] that became available during the preparation of our article. Post hoc analyses to explore alternative explanations of heterogeneity included a random-effects model (assuming that the true effect varies around an overall average treatment effect) and a stepwise backward random-effects metaregression of the logarithm of the RR on quality, small-studies effect, age group, baseline risk, immunoglobulin preparation, and total dose provided within a week.

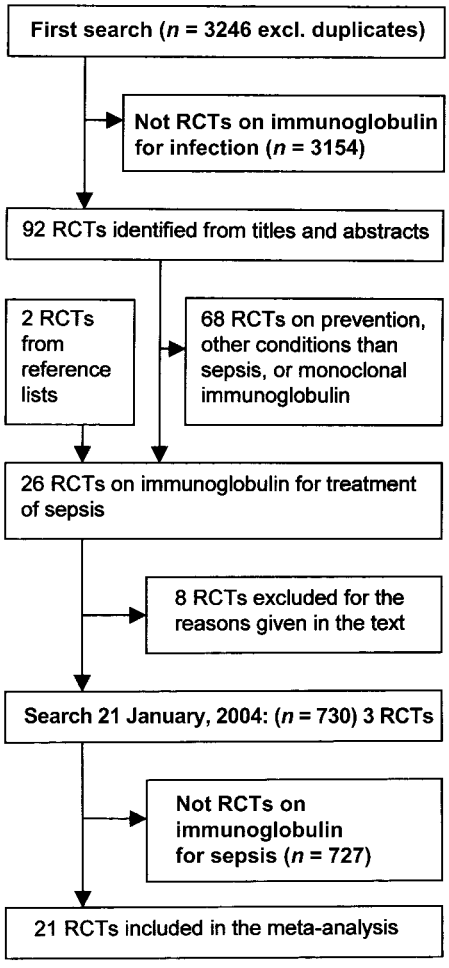



**Figure 1.** Study selection flow diagram. RCT, randomized controlled trial

**Table 1. Summary of studies involving polyclonal immunoglobulin treatment for bacterial sepsis.**

| Study | Dosing regimens, duration | Total dose[a] | Disease | No. of patients randomized | Concealment of allocation category (description)[b] | Patient and care provider blinded? | Intention-to treat-data available | Length of follow-up |
|---|---|---|---|---|---|---|---|---|
| Behre et al. [11] | Pentaglobin (IgGMA) vs. 5% albumin, 3 days | 70,000[c] | Hematological malignancy and sepsis | 52 | B (NA) | Yes (placebo controlled) | No | 28 days |
| Burns et al. [13] | Sandoglobulin (IgG) vs. albumin, 3 days | 1200 | Septic thrombo cytopenia | 38 | A (sealed, opaque, and sequentially numbered envelopes) | Yes (placebo controlled; anonymously labeled identical infusion sets) | Yes | 9 days |
| Chen [14] | Intraglobin (Class NA) vs. NaCl, single dose | 500 | Sepsis in neonates (36% preterm) | 141 | B (NA) | Uncertain (placebo controlled; difference in opacity expected; no described precautions to conceal this) | No | Until 6 weeks after discharge from hospital |
| Darenberg et al. [38] | Endobulin (IgG) vs. albumin 1%, 3 days | 2000 | Streptococcal toxic shock syndrome | 21 | A (centralized randomization) | Yes (placebo controlled) | Yes | Primary outcome mortality data: 28 days. |
| de Simone et al. [16] | Sandoglobulin (IgG) vs. no treatment, 2 days + 1 extra dose on day 5 (if necessary) | 800–1200 | Sepsis after surgery | 24 | B (sealed envelopes; unknown whether they were opaque and sequentially numbered) | No | Yes | 9 days |
| Dominioni et al. [18] | Sandoglobulin (IgG) vs. albumin in 5% dextrose and saline, doses on days 0, 1, and 5 | 1000 | Surgery or trauma complicated with sepsis | 117 | B (NA) | Yes (placebo controlled) | Yes | Until death or discharge |
| Erdem et al. [19] | Pentaglobin (IgGMA) vs. no treatment, 3 days | 750 | Sepsis in preterm neonates | 44 | C (alternation)[d] | No | No | NA |
| Grundmann and Hornung [20] | Intraglobin (Class NA) vs. no treatment, 2 days | 500 | Surgery or trauma complicated with sepsis | 46 | B (randomized envelope technique; unknown whether envelopes were opaque and sequentially numbered) | No | Yes | Until death or discharge (minimum, 30 days) |
| Haque et al. [21] | Pentaglobin (IgGMA) vs. 10% dextrose, 4 days | 1000 | Sepsis in pre-term neonates | 60 | A (sealed, opaque, and sequentially numbered envelopes) | No (placebo controlled but pentaglobin was opaque, dextrose was not; no precautions taken to conceal this) | Yes | NA |
| Just et al. [22] | Immunoglobulin preparation from Biotest (Frankfurt) (IgGMA) vs. no treatment, 1.5 days | 20,000[c] | Sepsis in a subgroup of adults with severe infection | 104 (sepsis group, 35) | B (NA) | No | No | Until death or discharge |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| Karatzas et al. [39] | Pentaglobin (IgGMA) vs. no treatment, 3 days | 750 | Sepsis | 82 | A (computer-generated randomization sequence kept centralized apart from clinical center) | No | Yes | 28 days + 6 months |
| Lindquist et al. [23] | Gamma-Venin (IgG, pepsin treated) vs. no treatment, 3 days | 450 | Sepsis in a subgroup of adults with severe infection | 177 (sepsis group, 67) | A (sealed, opaque, and sequentially numbered envelopes) | No | No | NA |
| Mancilla-Ramirez et al. [24] | Gamimune N (IgG) vs. maltose 10%, single dose | 500 | Sepsis in a subgroup of neonates | 80 (sepsis group, 37) | A (sealed, opaque, and sequentially numbered envelopes) | Yes (placebo controlled; different opacity concealed by amber-colored bottles; infusion lines covered with aluminum wraps) | Yes | 30 days or until death or discharge |
| Samatha et al. [26] | Pentaglobin (IgGMA) vs. no treatment, 3 days | 750 | Sepsis in neonates (73% preterm) | 60 | B (picking up lots) | No | No | NA |
| Schedel et al. [27] | Pentaglobin (IgGMA) vs. no treatment, 3 days | 60,000[c] | Sepsis | 69 | A (sealed, opaque, and sequentially numbered envelopes; centralized randomization procedure) | No | Yes | 6 weeks |
| Shenoi et al. [29] | Sandoglobulin (IgG) vs. NaCl and dextrose, 3 days | 3000 | Sepsis in neonates | 51 | A (sealed, opaque, and sequentially numbered envelopes) | No (placebo controlled but bottles nonidentical containing fluid of different opacity) | Yes | Until death or discharge |
| Sidiropoulos et al. [30] | Sandoglobulin (class NA) vs. no treatment, 6 days | 6000 in term and 3000 in preterm neonates[c] | Suspected sepsis in neonates (term and preterm) | 82 | C (alternation) | No | No | 1–4 years |
| Tugrul et al. [37] | Pentaglobin (IgGMA) vs. no treatment, 3 days | 750 | Sepsis | 42 | C (open table of random numbers at allocation site) | No | Yes | 28 days |
| Wesoly et al. [33] | Pentoglobin (IgGMA) vs. no treatment, 3 days | 750 | Subgroup with sepsis after surgery | 100 (sepsis group, 35) | C (alternation) | No | Yes | Until death or discharge |
| Werdan and Pilz Gat [34] | Polyglobin N (IgG) vs. 0.1% albumin, 2 days | 900 | Severe sepsis | 653 | A (serially numbered and coded identical bottles containing identical-appearing fluid) | Yes (placebo controlled; bottles and fluid indistinguishable) | Yes[e] | 28 days |
| Yakut et al. [35] | Gamümine N (IgG) vs. albumin, 6 days | 1400 | Sepsis after surgery | 40 | B (NA) | Yes (placebo controlled; identical bottles with indistinguishable fluid) | Yes | NA |

**NOTE.** IgA, immunoglobulin A; IgG, immunoglobulin G; IgGMA, IgG preparations enriched with immunoglobulin M and IgA; NA, not available.

[a] Total dose is expressed in milligrams per kilogram of body weight, unless otherwise indicated.

[b] Concealment of allocation was categorized as follows: A, adequate; B, unclear; C, inadequate. For a description of the criteria for categorization, see Methods.

[c] Total dose is expressed in milligrams, irrespective of body weight.

[d] This information was not available in the trial report [19], but according to Ohlsson and Lacy [2]), who corresponded with Erdem et al. [19], alternation was used.

[e] Primary end point data are available for 624 of 653 included patients.

The small-studies effect was present if the effect estimate varied with smaller study size (which may occur, for example, as a result of publication bias) [5]. Baseline risk is the underlying risk at trial entry. Because few trial reports provided this information as baseline sepsis score, we used the control group event rate instead (although this will tend to overestimate the association with treatment effect, because the control group event rate itself enters into the treatment effect estimate). High quality was coded as 1 and lower quality as 0; small-studies effect was modeled as the standard error of the logarithm of the RR; age groups were defined as neonates versus nonneonates (i.e., adults, except for very few school-age children) and were coded as 0 and 1, respectively; IgG preparations were coded as 1, and IgG preparations enriched with IgM and IgA (IgGMA) were coded as 0. The total dose was expressed as milligrams per kilogram of body weight. The metaregression was performed with use of Stata software, version 8 (StataCorp) [9, 10].

## RESULTS

***Description of studies.*** Twenty-nine trial reports were identified (figure 1) [11–39]. Eight reports were excluded for the following reasons: no mortality data available in the subgroup of septic patients [25, 36]; unclear whether the deaths among septic patients occurred in the intervention or the control group [12, 15]; fundamental design problems [32]; an interim analysis of a later full trial report [17]; and duplicate publications [28, 31].

The 21 included trials comprised 1711 patients and 547 deaths. Thirteen of the 21 corresponding authors answered our questions (see Acknowledgments), and 4 studies [13, 24, 34, 38] were reclassified from lower quality to high quality as a consequence of these responses. The characteristics of the trials are shown in table 1. One large trial by Werdan et al. [34] involved 624 patients and 239 deaths, and it provided 38% of the weight in the meta-analysis. The mortality data from this trial have previously only been reported qualitatively ("the 28-day mortality was not reduced" [34]) in an abstract. However, the authors have provided us with quantitative intention-to-treat data, and the trial was performed according to a detailed, published protocol, so the quality of the trial could be assessed [40]. Seven of the trials comprised neonates [14, 19, 21, 24, 26, 29, 30], and 14 of the trials comprised nonneonates (i.e.,

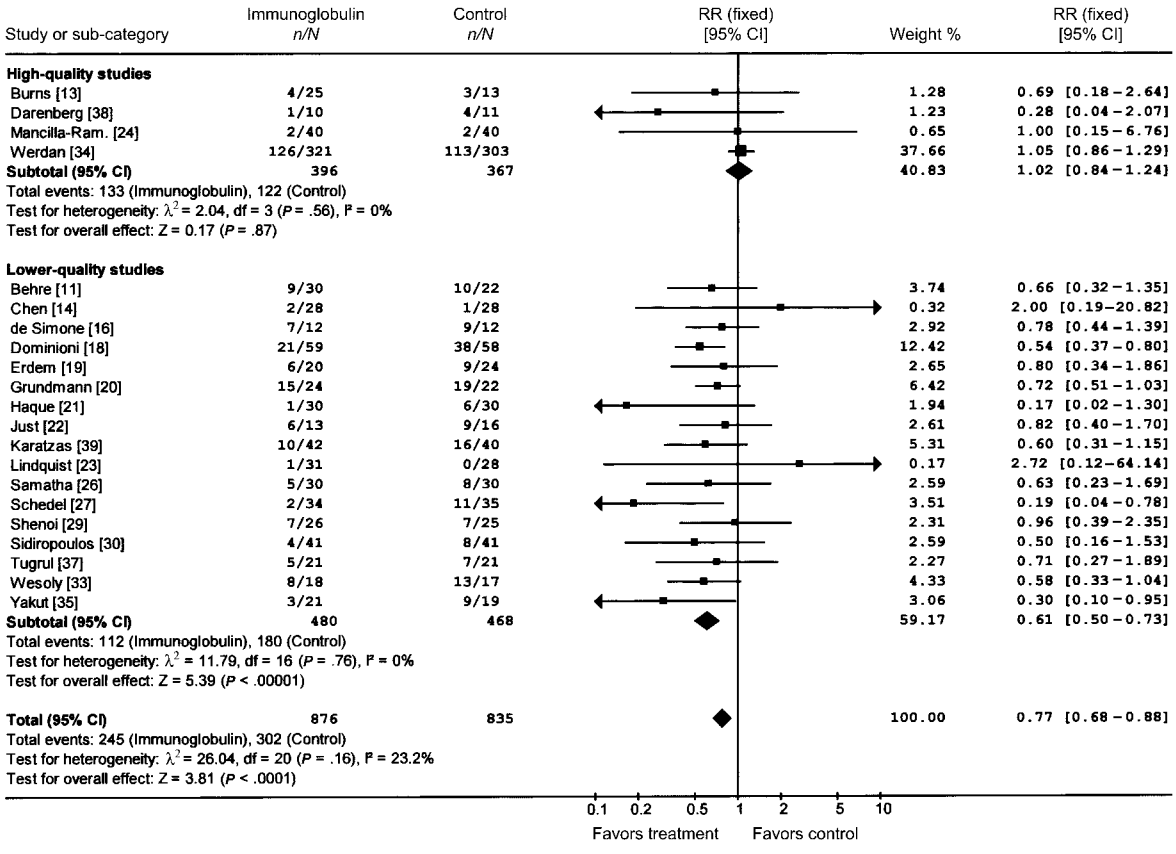



**Figure 2.** Meta-analysis of relative risk of all-cause mortality comparing patients with sepsis treated with polyclonal immunoglobulin (Immunoglobulin) with patients with receiving placebo or no additional treatment for sepsis (Control). Subtotals designate the subgroup analysis of trials of high quality and lower quality. *Bars*, 95% CI; *n,* number of deaths; *N,* number of patients; RR, relative risk; Fixed, fixed-effect model. $I^2$ quantifies the percentage of variation between study results that is not ascribable to sampling error.

**Table 2. Sensitivity analysis of the loss to follow-up in the trial by Werdan et al. [34].**

| Characteristic | Immunoglobulin group | Placebo group | Relative risk (95% CI) |
|---|---|---|---|
| Status, no. of patients | | | |
| Dead | 126 | 113 | … |
| Alive | 195 | 190 | … |
| Unknown[a] | 14 | 14 | … |
| Death rates, % (*n*/*N*) | | | |
| Complete case analysis | 39.3 (126/321) | 37.3 (113/303) | 1.05 (0.86–1.29) |
| Assuming all lost patients died | 41.8 (140/335) | 40.1 (127/317) | 1.04 (0.87–1.25) |
| Assuming all lost patients survived | 37.6 (126/335) | 35.6 (113/317) | 1.06 (0.86–1.29) |
| Extreme case, favoring immunoglobulin[b] | 37.6 (126/335) | 40.1 (127/317) | 0.94 (0.77–1.14) |
| Extreme case, favoring placebo[c] | 41.8 (140/335) | 35.6 (113/317) | 1.17 (0.97–1.42) |

[a] Equal distribution of the loss is assumed.
[b] Assumes that all patients lost to follow-up in the immunoglobulin group survived and all patients lost to follow-up in the placebo group died.
[c] Assumes that all patients lost to follow-up in the placebo group survived and all patients lost to follow-up in the immunoglobulin group died.

adults, except for very few school-age children) [11, 13, 16, 18, 20, 22, 23, 27, 33–35, 37–39].

The methodological quality of the studies was highly variable (table 1), and only 4 of the studies met all 3 quality criteria and were categorized as high quality [13, 24, 34, 38]. Nine studies had adequate concealment of allocation [13, 21, 23, 24, 27, 29, 34, 38, 39], 8 had unclear concealment of allocation [11, 14, 16, 18, 20, 22, 26, 35], and 4 were inadequately concealed [19, 30, 33, 37]. Thirteen were not double-blinded [16, 19–22, 26, 27, 29, 30, 33, 37–39], and 7 did not make data available for intention-to-treat analysis [11, 14, 19, 22, 23, 26, 30].

Four studies reported follow-up until death or discharge [18, 22, 29, 33]. In 5 studies, the length of follow-up was not available [19, 21, 23, 26, 35], and in the remaining studies, it varied and was often imprecisely reported. Thus, we report mortality data at the length of follow-up provided by the authors (table 1) and did not include length of follow-up in the metaregression.

***Mortality.*** When data from all trials were pooled, there appeared to be a beneficial effect of immunoglobulin treatment on the RR of death of 0.77 (95% CI, 0.68–0.88; $P = .0001$). However, 23.2% of the variability between the study results could not be ascribed to sampling error ($I^2$, 23.2%; figure 2). When the trials were analyzed in separate subgroups of high and lower quality, heterogeneity was no longer detectable ($I^2$, 0%). The pooled RR for the 4 high-quality trials was 1.02 (95% CI, 0.84–1.24; $P = .87$). In contrast, the 17 lower-quality trials had a pooled RR of 0.61 (95% CI, 0.50–0.73; $P < .00001$) (figure 2). The difference between the estimates from the trials of high methodological quality versus those from the trials of lower methodological quality was highly statistically significant ($P = .0002$).

The large study had a loss to follow-up of 4.3% of patients, but even extreme-case scenarios in favor of immunoglobulin treatment did not alter the finding that high-quality trials did not show a statistically significant effect on mortality (table 2). Loss to follow-up was not reported in other studies.

***Sensitivity analyses of mortality.*** The results were similar if a random-effect model was applied. The overall estimate of RR was 0.70 (95% CI, 0.59–0.85), the RR for high-quality trials only was 1.03 (95% CI, 0.85–1.25), and the RR for lower-quality trials only was 0.64 (95% CI, 0.54–0 .76). Levels of heterogeneity were unaltered.

If the quality criteria for high-quality trials were reduced to require that only the most important criterion of methodological quality (i.e., concealment of allocation [41]) be fulfilled, the results would be as follows: 9 trials with adequately concealed allocation (RR, 0.91; 95% CI, 0.76–1.09; $I^2$, 40%) versus 12 trials with unclear or inadequate concealment of allocation (RR, 0.63; 95% CI, 0.52–0.77; $I^2$, 0%). The introduction of 40% heterogeneity in the high-quality trial group indicates that the lack of double blinding in the 5 reclassified trials that had adequate concealment made an important difference.

**Table 3. Exploratory stepwise backward random effects metaregression.**

| Variable | Coefficient | Standard error | *P* |
|---|---|---|---|
| Methodological quality | 0.48 | 0.13 | <.001 |
| Small-studies effect | −0.57 | 0.43 | .18 |
| Age group | −0.38 | 0.28 | .18 |
| Immunoglobulin preparation | −0.12 | 0.22 | .59 |
| Total dose in mg/kg $\times 10^3$ | 0.17 | 0.29 | .53 |
| Baseline risk | −0.55 | 1.12 | .62 |

**NOTE.** Each line states the test result for the individual covariate in the last step, where it is included in the model along with the variables above.

Most trials comprised a mixture of patients with sepsis due to gram-negative organisms and patients with sepsis due to gram-positive organisms, as well as immunocompetent and immunoincompetent patients with different underlying diseases. In general, separate mortality data were not provided for any of these subgroups, which precluded the planned sensitivity analyses. In only 2 studies was albumin used as placebo in a concentration within the range that has been suggested to increase mortality [11, 38].

The strong association between study quality and the RR of death would confound the planned subgroup analysis. Instead, we did an exploratory stepwise backward random-effects metaregression (table 3). It confirmed the strong association between study quality and effect, but it found no evidence for an association of the effect with age groups, baseline risk, immunoglobulin preparation, or total immunoglobulin dose. When the covariables were modeled alone, the only other covariable apart from quality with a *P* value suggestive of an association with the effect was the small-studies effect ($P = .032$). When both variables were included in the model, the *P* value for the regression coefficient for quality was .01; it was .18 for the small-studies effect. This reflects that many of the small studies also had lower quality, and after controlling for the quality of the study, the association of small study size with larger effect estimates was no longer significant. Thus, trial quality was the only variable that explained a statistically significant amount of variation in the outcomes of the included trials.

***Length of hospital stay.*** For nonsurvivors, patients in the immunoglobulin group died 2.7 days earlier than others (95% CI, 0.2–5.3). For survivors, there was no statistically significant difference in length of hospital stay between groups (3.8 days, 95% CI, −2.3 to 9.9) [18, 35].

***Complications and adverse effects.*** The information on complications and adverse effects was too scarce to be combined in a meta-analysis.

## DISCUSSION

***Major findings and possible explanations.*** Our most reliable estimate of the effect of treatment with intravenous, polyvalent immunoglobulin on mortality in patients with sepsis relied on the 41% of the statistical information that came from the high-quality trials designed to minimize bias. This estimate was a RR of 1.02 (95% CI, 0.84–1.24), which is compatible with a 16% reduction in mortality as well as with a 24% increase.

The overall pooled estimate, based on all of the trials, showed a large and significant reduction in mortality with immunoglobulin treatment, but one-fourth of the variation between the study results could not be ascribed to sampling error; this unexplained variability disappeared when high-quality and lower-quality studies were analyzed in subgroups. The difference in the results from these subgroups is large, but it is consistent with the expected influence of methodological quality. Trials with inadequate or unclear concealment of allocation exaggerated the effect of the experimental intervention by ~30%, on average (when measured as a ratio of ORs) in 4 out of the 5 empirical studies of bias [41, 42]. Furthermore, the difference is highly statistically significant, and it is the result our primary subgroup analysis as defined a priori. Thus, it is likely to reflect a true difference between high-quality and lower-quality trials. This result remained robust to the metaregression that explored whether differences in other trial characteristics (including age group, type of immunoglobulin preparation, etc.) were better explanations for the heterogeneity between the results of the individual trials.

Placebo treatment to ensure blinding of patients and care providers may seem unimportant when the outcome is mortality. However, if lack of blinding concurs with a lack of intention-to-treat analysis or with a lack of predetermined stopping rules, the risk of bias is obvious. Lack of intention-to-treat data may imply that the patients who did not receive the full intervention were not accounted for in the published report [43]. But some patients in the intervention group may not have received full treatment because of rapid deterioration and subsequent death, whereas similar patients in the control group are not excluded because no placebo intervention was required. Hence, differential exclusions could lead to bias in favor of the intervention group. In 7 studies, the number of patients withdrawn or excluded from analysis was not available, and there was no statement that there were no exclusions [11, 14, 19, 22, 23, 26, 30]. Four of these studies did not apply placebo treatment [19, 22, 23, 26].

Lack of predetermined stopping rules increases the risk of spurious findings because of multiple looks at the data. If there is no blinding, the number of informal interim analyses can be large. The trial that reported the largest statistically significant effect was unblinded and prematurely terminated, and it stated that 12 interim analyses had been performed [27]. Predetermined stopping rules were not mentioned in 8 of the 11 trials without double blinding.

Two unblinded studies had predetermined goals of samples sizes [26, 29], but the sponsoring company (Sandoz India) withdrew support while the trials were ongoing and caused their premature termination. In one of the studies, the sponsor also made blinding impossible by refusing to provide identical vials with placebo [29].

***Previous systematic reviews of immunoglobulin for treatment of sepsis.*** Alejandria et al. [1] find that polyclonal immunoglobulin significantly reduces mortality, both when all studies (including those involving adults and neonates) are pooled (RR, 0.64; 95% CI, 0.51–0.80) and when only high-quality studies are considered (RR, 0.30; 95% CI, 0.09–0.99) [21, 32]. The discrepancy with our findings can be explained by their less sensitive search strategy, their less rigorous appli-

cation of quality assessment, and their retrieval of less information from the authors of the trials.

Ohlsson and Lacy [2] report results of trials from 2 settings. The first addresses mortality in neonates with clinically suspected sepsis; there is a borderline statistically significant reduction in mortality, as mentioned above. In the other setting (neonates with subsequently proven sepsis), they find a markedly reduced RR of 0.55 (95% CI, 0.31–0.98). A trial with a fundamental error in study design (in which inclusion of patients was dependent on the effect of the treatment) is included in the second setting [32]. If it were excluded, as in our analysis, then the combined result would no longer be statistically significant. Ohlsson and Lacy [2] do not present sensitivity analyses of the influence of the quality of the trials, but they cautiously conclude that there are insufficient data to support routine use of immunoglobulin for treatment of sepsis in neonates.

A recent review (without a meta-analysis) mentions the negative finding of the large high-quality trial and some of the methodological shortcomings of 6 of the smaller trials included here [44]. What our study adds to this is the presentation of 15 additional randomized controlled trials, with more detail on the methodological quality of these trials and a quantitative analysis of the sum of the evidence.

***Strengths and limitations of our study.*** Our review demonstrated that the overall effect estimate of immunoglobulin on mortality among septic patients not only hinges on the precision provided by the largest trial, but also on the methodological quality of the trials. The intermediate publication status of the large study by Werdan et al. [34] entails some uncertainty, because we cannot know why it has not been fully published yet.

The classification of trials as lower quality did not indicate that they were necessarily all of low quality. Some trials classified as lower quality may even have been high-quality but failed to report the measures taken to ensure this. Further, lack of guarding against bias did not prove that bias occurred, just that it may have occurred. But with different results derived from well-guarded versus uncertainly or less well-guarded trials, we recommend trusting the former.

Combining trials that occurred in different settings and involved different severities of sepsis may seem counterintuitive, but the results of a sepsis trial are likely to be extrapolated beyond the particular inclusion criteria. In addition, if there is no effect of the treatment, then any trial result deviating from no effect would be ascribable to sampling error or bias, and then it would be legitimate to combine all trials according to their susceptibility to bias. We explored whether there was any evidence against this assumption in the metaregression and found none.

The metaregression could be used to gauge whether there were obvious alternative explanations (other than quality) for the observed heterogeneity, but we could not exclude an effect of immunoglobulin treatment in defined patient subgroups. This would require large studies, such as the one currently being conducted by Brocklehurst et al. [45], who plan to include 5000 neonates.

***Implications for practice.*** Most of the immunoglobulin used in the United States is used off-label [46] and could be spurred by undue emphasis on results found in subgroups of the trials included here. However, the present review should serve to avoid this undue emphasis. For a common condition like sepsis, the burden of proof should be statistically and clinically significant treatment effects derived from high-quality randomized trials. Such evidence is not available, and we therefore suggest that polyvalent immunoglobulin for treatment of sepsis is not recommended for clinical practice. Exceptions could exist for rare conditions like streptococcal toxic shock syndrome, but guidelines will have to rest on a comprehensive analysis of the totality of the relevant data, including safety issues such as the risk of acute renal failure [47].

## Acknowledgments

We thank the following corresponding authors for providing additional information on their trials: E. R. Burns, A. Norrby-Teglund (for Darenberg et al. [38]), R. Grundmann, K. N. Haque, L. Lindquist, J. Mancilla-Ramirez, I. Schedel, A. Shenoi, C. Wesoly, K. Werdan, M. Yakut, S. Tugrul, and S. Karatzas.

***Financial support.*** The Nordic Cochrane Centre is financed by the Copenhagen Hospital Corporation. The views expressed in this article represent those of the authors and are not necessarily the views of other members of the Cochrane Collaboration.

***Conflict of interest.*** All authors: No conflict.

## References

1. Alejandria MM, Lansang MA, Dans LF, et al. Intravenous immunoglobulin for treating sepsis and septic shock. Cochrane Database Syst Rev **2002**; 4:CD001090.
2. Ohlsson A, Lacy JB. Intravenous immunoglobulin for suspected or subsequently proven infection in neonates. Cochrane Database Syst Rev **2004**; 1:CD001239.
3. Alderson P, Bunn F, Lefebvre C, et al. Human albumin solution for resuscitation and volume expansion in critically ill patients. Cochrane Database Syst Rev **2002**; 4:CD001208.
4. Hollis S, Campbell F. What is meant by intention to treat analysis?: survey of published randomised controlled trials. BMJ **1999**; 319:670–4.
5. Egger M, Davey Smith G, Altman D. Systematic reviews in health care: meta-analysis in context. London: BMJ Books, **2001**.
6. DerSimonian R, Laird N. Meta-analysis in clinical trials. Control Clin Trials **1986**; 7:177–88.
7. Altman DG, Bland JM. Interaction revisited:the difference between two estimates. BMJ **2003**; 326:219.
8. Higgins JP, Thompson SG, Deeks JJ, Altman DG. Measuring inconsistency in meta-analyses. BMJ **2003**; 327:557–60.
9. Bradburn MJ, Deeks JJ, Altman DG. sbe24: metan—an alternative meta-analysis command. Stata Techn Bull **1998**; 44:4–15.
10. Sharp S. sbe23: meta-analysis regression. Stata Techn Bull **1998**; 42: 16–22.
11. Behre G, Ostermann H, Schedel I, et al. Endotoxin concentrations and

therapy with polyclonal IgM-enriched immunoglobulins in neutropenic cancer patients with sepsis syndrome: pilot study and interim analysis of a randomized trial. Antiinfect Drugs Chemother **1995**; 13: 129–34.
12. Berg A, Laguenie G, Goutail F, Couturier C, Saint-Maurice C. [Venoglobulins in pediatric surgical resuscitation]. Cah Anesthesiol **1989**; 37:511–5.
13. Burns ER, Lee V, Rubinstein A. Treatment of septic thrombocytopenia with immune globulin. J Clin Immunol **1991**; 11:363–8.
14. Chen JY. Intravenous immunoglobulin in the treatment of full-term and premature newborns with sepsis. J Formos Med Assoc **1996**; 95:839–44.
15. Christensen RD, Brown MS, Hall DC, Lassiter HA, Hill HR. Effect on neutrophil kinetics and serum opsonic capacity of intravenous administration of immune globulin to neonates with clinical signs of early-onset sepsis. J Pediatr **1991**; 118:606–14.
16. De Simone C, Delogu G, Corbetta G. Intravenous immunoglobulins in association with antibiotics:a therapeutic trial in septic intensive care unit patients. Crit Care Med **1988**; 16:23–6.
17. Dominioni L, Dionigi R, Zanello M, et al. Effects of high-dose IgG on survival of surgical patients with sepsis scores of 20 or greater. Arch Surg **1991**; 126:236–40.
18. Dominioni L, Bianchi V, Imperatori A, et al. High-dose intravenous IgG for treatment of severe surgical infections. Dig Surg **1996**; 13:430–4.
19. Erdem G, Yurdakok M, Tekinalp G, Ersoy F. The use of IgM-enriched intravenous immunoglobulin for the treatment of neonatal sepsis in preterm infants. Turk J Pediatr **1993**; 35:277–81.
20. Grundmann R, Hornung M. Immunoglobulin therapy in patients with endotoxemia and postoperative sepsis—a prospective randomized study. Prog Clin Biol Res **1988**; 272:339–49.
21. Haque KN, Zaidi MH, Bahakim H. IgM-enriched intravenous immunoglobulin therapy in neonatal sepsis. Am J Dis Child **1988**; 142:1293–6.
22. Just HM, Metzger M, Vogel W, Pelka RB. [Effect of adjuvant immunoglobulin therapy on infections in patients in an surgical intensive care unit: results of a randomized controlled study]. Klin Wochenschr **1986**; 64:245–56.
23. Lindquist L, Lundbergh P, Maasing R. Pepsin-treated human gamma globulin in bacterial infections: a randomized study in patients with septicaemia and pneumonia. Vox Sang **1981**; 40:329–37.
24. Mancilla-Ramirez J, Gonzalez-Yunes R, Castellanos-Cruz C, et al. Intravenous immunoglobulin in the treatment of neonatal septicemia. Bol Med Hosp Infant Mex **1992**; 49:4–11.
25. Masaoka T, Hasegawa H, Takaku F, et al. The efficacy of intravenous immunoglobulin in combination therapy with antibiotics for severe infections. Jpn J Chemother **2000**; 48:199–217.
26. Samatha S, Jalalu MP, Hegde RK, et al. Role of IgM enriched intravenous immunoglobulin as an adjuvant to antibiotics in neonatal sepsis. Karnataka Pediatr J **1997**; 11:1–6.
27. Schedel I, Dreikhausen U, Nentwig B, et al. Treatment of gram-negative septic shock with an immunoglobulin preparation: a prospective, randomized clinical trial. Crit Care Med **1991**; 19:1104–13.
28. Schedel I, Dreichhausen U. The therapy of gram-negative septicotoxic diseases with pentaglobin, an immunoglobulin with an elevated IgM content [in Russian]. Anesteziol Reanimatol **1996**; 3:4–9.
29. Shenoi A, Nagesh NK, Maiya PP, Bhat SR, Subba Rao SD. Multicenter randomized placebo controlled trial of therapy with intravenous immunoglobulin in decreasing mortality due to neonatal sepsis. Indian Pediatr **1999**; 36:1113–8.
30. Sidiropoulos D, Bohme U, von Muralt G, et al. Immunoglobulin substitution in the treatment of neonatal septicemia [in German]. Schweiz Med Wochenschr **1981**; 111:1649–55.
31. Sidiropoulos D, Boehme U, Von Muralt G, et al. Immunoglobulin supplementation in prevention or treatment of neonatal sepsis. Pediatr Infect Dis **1986**; 5:S193–4.
32. Weisman LE, Stoll BJ, Kueser TJ, et al. Intravenous immune globulin therapy for early-onset sepsis in premature neonates. J Pediatr **1992**; 121:434–43.
33. Wesoly C, Kipping N, Grundmann R. Immunoglobulin therapy of postoperative sepsis [in German]. Z Exp Chir Transplant Kunstliche Organe **1990**; 23:213–6.
34. Werdan K, Pilz Gat SSG. Polyvalent immune globulins [abstract 18]. Shock **1997**; 7(Suppl):5.
35. Yakut M, Cetiner S, Akin A, et al. Effects of immunuglobulin G on surgical sepsis and septic shock. Bull Gulhane Mil Med Acad **1998**; 40:76–81.
36. Gokalp AS, Toksoy HB, Turkay S, Bakici MZ, Kaya R. Intravenous immunoglobulin in the treatment of *Salmonella typhimurium* infections in preterm neonates. Clin Pediatr Phila **1994**; 33:349–52.
37. Tugrul S, Ozcan PE, Akinci O, et al. The effects of IgM-enriched immunoglobulin preparations in patients with severe sepsis. Crit Care **2002**; 6:357–62.
38. Darenberg J, Ihendyane N, Sjolin J, et al. Intravenous immunoglobulin G therapy in streptococcal toxic shock syndrome:a European randomized, double-blind, placebo-controlled trial. Clin Infect Dis **2003**; 37: 333–40.
39. Karatzas S, Boutzouka E, Venetsanou K, et al. The effects of IgM-enriched immunoglobulin preparations in patients with severe sepsis: another point of view [letter]. Crit Care **2002**; 6:543–5.
40. Pilz G, Fateh Moghadam S, Viell B, et al. Supplemental immunoglobulin therapy in sepsis and septic shock—comparison of mortality under treatment with polyvalent iv immunoglobulin versus placebo: protocol of a multicenter, randomized, prospective, double-blind trial. Theor Surg **1993**; 8:61–83.
41. Juni P, Altman DG, Egger M. Systematic reviews in health care: assessing the quality of controlled clinical trials. BMJ **2001**; 323:42–6.
42. Balk EM, Bonis PA, Moskowitz H, et al. Correlation of quality measures with estimates of treatment effect in meta-analyses of randomized controlled trials. JAMA **2002**; 287:2973–82.
43. Moher D, Schulz KF, Altman DG, Lepage L. The CONSORT statement: revised recommendations for improving the quality of reports of parallel-group randomised trials. Lancet **2001**; 357:1191–4.
44. Laupland KB. Polyclonal intravenous immunoglobulin for the prophylaxis and treatment of infection in critically ill adults. Can J Infect Dis **2002**; 13:100–6.
45. Brocklehurst P, Tarnow-Mordi W, Haque K, et al. International Neonatal Immunotherapy Study (INIS). Funded by Medical Research Council (U.K.). Trial end date, 2007. Available at: http://controlled-trials.com/mrct/trial/INIS/1042/22315.html. Accessed 3 October **2003**.
46. Chen C, Danekas LH, Ratko TA, Vlasses PH, Matuszewski KA. A multicenter drug use surveillance of intravenous immunoglobulin utilization in US academic health centers. Ann Pharmacother **2000**; 34:295–9.
47. Pierce LR, Jain N. Risks associated with the use of intravenous immunoglobulin. Transfus Med Rev **2003**; 17:241–51.

## Study II

BMJ Paper; pilj203505

type=paper in-section=ppr id=pilj203505 in-journal=bmj doi=10.1136/bmj.38414.422650.8F role=papweb

# Comparison of descriptions of allocation concealment in trial protocols and the published reports: cohort study

Julie Pildal, An-Wen Chan, Asbjørn Hróbjartsson, Elisabeth Forfang, Douglas G Altman, Peter C Gøtzsche

**Nordic Cochrane Centre, H:S Rigshospitalet, DK-2100 Copenhagen, Denmark**

Julie Pildal
*research fellow*

Asbjørn Hróbjartsson
*senior research fellow*

Elisabeth Forfang
*physician*

Peter C Gøtzsche
*director*

**Department of Medicine, University of Toronto**

An-Wen Chan
*resident physician*

**Cancer Research UK Medical Statistics Group, Centre for Statistics in Medicine, Oxford OX3 7LF**

Douglas G Altman
*director*
Correspondence to: J Pildal jpl@dadlnet.dk

(Accepted: 2 March 2005)

P+ Further details are on bmj.com

## Abstract

Objectives To compare how allocation concealment is described in publications of randomised clinical trials and corresponding protocols, and to estimate how often trial publications with unclear allocation concealment have adequate concealment according to the protocol.
Design Cohort study of 102 sets of trial protocols and corresponding publications.
Setting Protocols of randomised trials approved by the scientific and ethical committees for Copenhagen and Frederiksberg, 1994 and 1995.
Main outcome measures Frequency of adequate, unclear, and inadequate allocation concealment and sequence generation in trial publications compared with protocols, and the proportion of protocols where methods were reported to be adequate but descriptions were unclear in the trial publications.
Results 96 of the 102 trials had unclear allocation concealment according to the trial publication. According to the protocols, 15 of these 96 trials had adequate allocation concealment (16%, 95%

confidence interval 9% to 24%), 80 had unclear concealment (83%, 74% to 90%), and one had inadequate concealment. When retrospectively defined loose criteria for concealment were applied, 83 of the 102 trial publications had unclear concealment. According to their protocol, 33 of these 83 trials had adequate allocation concealment (40%, 29% to 51%), 49 had unclear concealment (59%, 48% to 70%), and one had inadequate concealment.
Conclusions Most randomised clinical trials have unclear allocation concealment on the basis of the trial publication alone. Most of these trials also have unclear allocation concealment according to their protocol.
*166;171*

**Introduction**

Selection bias occurs in randomised clinical trials if patients with a better prognosis are preferentially allocated to one of the treatment arms. The results of the trial will then to some degree reflect this difference in prognosis rather than just a difference in the effects of the compared treatments.

The purpose of randomisation is to avoid selection bias, as patients with known and unknown differences in prognosis will tend to be equally distributed between the treatment groups. To ensure true randomisation, however, the random allocation sequence should not only be sequentially and irreversibly administered but should also be concealed to the individuals in charge of enrolment and treatment allocation. Otherwise, knowledge of the upcoming allocation will permit selective assignment of patients by manipulation of either the sequence of treatments to be allocated or the sequence of patients to be enrolled.[1]

Surveys have shown that 44% to 93% of publications of randomised controlled trials lack a clear description of allocation concealment.[2] Empirical studies have shown that publications of trials in which allocation concealment is unclear or inadequate are associated with, on average, a 20-30% exaggeration of the treatment effect (measured as a ratio of odds ratios) compared with trials of the same interventions with adequate concealment.[3-9] Generation of a truly random sequence is an interrelated issue for which there is suggestive empirical evidence of an associated inflation of the treatment effect if the trial publication does not document adequate procedures.[4-6] Both issues are of major concern because the effect of many treatments is less than these average biases. We compared how allocation concealment is described in publications of randomised clinical trials and corresponding protocols, and we estimated how often trial publications with unclear allocation concealment have adequate concealment according to the protocol.

**Methods**

Our cohort consisted of all published randomised trials (apart from trials in dentistry) whose protocols were approved by the scientific and ethical committees for Copenhagen and Frederiksberg

in 1994 and 1995. We identified trial publications by contacting the principal investigators and by searching PubMed, Embase, and the Cochrane Central Register of Controlled Trials (final search in May 2003; median publication year 1999, range 1995-2003).[10] In total, 102 protocols were published in 122 trial reports. When there was more than one publication of a trial, we examined all publications for information on allocation concealment.

**Outcomes and data extraction**

Our outcome measures were frequency of adequate, unclear, and inadequate allocation concealment and sequence generation in trial publications compared with protocols; the proportion of protocols where methods were reported to be adequate when the trial publications gave unclear descriptions; the type and frequency of methods used for allocation concealment; and the prevalence of other trial characteristics that might undermine concealment (for example, stating the block size in the protocol).

Two authors extracted data from the first half of the published reports and the second half of the protocols while another pair of authors extracted data from the rest. Disagreements were resolved within each pair, thus avoiding assessment of both trial publication and protocol for the same trial.

**Assessment of adequacy of allocation concealment**

We considered the following methods for allocation concealment as adequate:[3-6] [11-13] central randomisation; numbered coded vehicles; opaque, sealed, and sequentially numbered envelopes; and other methods containing convincing means of concealment. Inadequate methods concerned open or predictable sequences of allocation (for example, alternation), date of birth, case record number or similar, and open tables of random numbers. We categorised studies as unclear that did not fall into one of these categories or that provided no information.

To ensure consistency and transparency and to capture how strict application compared with loose application of our criteria might influence our results, we operationalised our interpretations of authors' descriptions of allocation concealment (see table A on bmj.com). The strict criteria are those recommended for Cochrane reviews,[11] except for an elaboration on central randomisation, as specified in table A.[12] [13] The loose criteria, which we defined retrospectively, comprised the most liberal criteria used in any of the previous empirical studies of bias associated with unclear or inadequate allocation concealment.[3-9] For instance, in a study by Schulz et al[4] envelopes had to be opaque, sealed, and sequentially numbered to qualify as adequate concealment, whereas in a study by Kjaergard et al[6] use of sealed envelopes without further details qualified as adequate (see bmj.com for examples of how the criteria were applied on our sample (Box 1).

#### Assessment of adequacy of sequence generation

Adequate methods of sequence generation included computer generated random numbers, tables of random numbers, or drawing lots or envelopes. Inadequate methods could be related to prognosis such as date of birth or year of admission. Unclear methods were methods not falling into one of these two categories or where the methods were not described.

#### Statistical analysis

We calculated 95% confidence intervals using the exact binomial method in Stata version 8.

## Results

#### Allocation concealment

Using the strict criteria, 96 of the 102 trials (94%, 95% confidence interval 88% to 98%) had an unclear allocation concealment according to their publications. According to their protocols, 15 of these 96 trials (16%, 9% to 24%) had adequate allocation concealment and one had inadequate concealment, whereas most (80 of 96; 83%, 74% to 90%) had unclear concealment (table 1[t1]).

*Table 1 - pilj203505.t1*

Using the loose criteria, 83 of the 102 trials had unclear allocation concealment (81%, 72% to 88%). According to the protocols, 33 of these 83 publications (40%, 29% to 51%) had adequate allocation concealment, one had inadequate concealment, and 49 (59%; 48% to 70%) had unclear concealment (see table 1).

According to the strict criteria, 20 of the 102 studies (five publications and 19 protocols; see table 1) described adequate allocation concealment. When the loose criteria were applied, however, 51 studies (18 publications and 45 protocols) described adequate concealment.

#### Sequence generation

Eighty one of the 102 trial publications gave no information on how the allocation sequence was generated; 16 of these 81 trials (20%; 12% to 30%) described adequate sequence generation in the protocol. No protocols or trial publications reported inadequate methods of sequence generation.

#### Methods used for allocation concealment

Table 2[t2] lists the methods used to achieve allocation concealment. Numbered coded vehicles was the most frequently applied method according to the protocols (26 of 102) but had the lowest rate of appearance in the trial publications (three of 26). None of the 17 trials using central randomisation fulfilled the strict criteria, as none described concealment of the randomisation sequence from the central staff, only four described irreversibility of the treatment assignment, and none described that prognostic data irrelevant to stratification must not be revealed to the central office (in three trials such data were positively requested). In 39 of the 102 trials neither the

protocols nor the publications provided any information on attempts to conceal the allocation. In four trials, the protocol and the publication gave conflicting information on which method was used.

*Table 2 - pilj203505.t2*

**Trial characteristics that might weaken an otherwise adequate allocation concealment regimen**

*Block randomisation*

In 14 trials, block randomisation could partly have compromised allocation concealment because the block size was explicitly stated in the protocol. This is problematic since a known block size enables qualified guesswork to predict upcoming allocations towards the end of the block. This can weaken allocation concealment even in multicentre studies if they are stratified per centre and in double blind studies if the blinding becomes compromised—for example, because of adverse effects.

*Tasks that should not be carried out by the same party*

The preparation of envelopes for concealment was described in the passive tense in nine of the 13 studies using the envelope method for allocation concealment (see table 2). Thus it is unknown whether the same person prepared the envelopes, enrolled the patients, and administered the envelopes, particularly as seven of the nine studies were single centre studies. An example of lack of separation of functions for central randomisation was when the same party had information on the prognosis of the next patient to be enrolled and was involved in concealing the sequence and in administering it (see bmj.com).

*Code envelopes*

In 42 of the 55 double blind studies, a security system for emergency code breaking was described in the protocol but mentioned in only one publication. Overall, 90% (38 of 42) of these protocols specified that envelopes or a similar system would be present at the clinical location. Deciphering the contents of such envelopes, for instance by holding them against strong light, might have revealed the allocation for the next patient; yet only one of the 38 protocols (3%) described the envelopes as opaque. Although such code envelopes are a theoretical threat to the allocation concealment, it is unknown whether their presence on the clinical location is associated with exaggerated effect estimates. Consequently, our criteria for assessment of allocation concealment by the means of envelopes did not include assessment of code envelopes.

## Discussion

Most trial publications provided unclear information on allocation concealment. When we applied strict criteria the corresponding protocols clarified that 16% had adequate concealment compared with 40% when we applied loose criteria. Thus, regardless of the criteria applied, most of

the protocols also provided unclear information or gave rise to additional concern that the allocation concealment might have been compromised (for example, by disclosing the block size). A similar pattern of insufficient reporting was found for sequence generation. The lack of clarity in the protocols is consistent with, but does not prove, the notion that unclear reporting of allocation concealment in trial publications often reflects inadequate safeguards against selection bias.[4]

Our results make it reasonable to assume that the empirical surveys, which show a 20-30% exaggeration of the treatment effect for trial publications with unclear or inadequate allocation concealment, included some trials with allocation concealment that was adequately carried out but insufficiently reported.[3-6] This implies that if inadequate concealment with ensuing selection bias is to explain the observed exaggeration in the previous studies,[3-6] then an even larger exaggeration would be expected for those trials where neither the publication nor the protocol indicated adequate concealment.

**Strengths and limitations of study**

The strength of our study is that it is the first account of how allocation concealment is described in a representative cohort of trial protocols and subsequent publications of trials. The detailed data extraction allowed for sensitivity analysis of the strictness of the applied criteria and for finding additional elements that could compromise allocation concealment.

One limitation is that even in the cases where the protocols provided explicit descriptions of allocation concealment, the assumption that the trials were conducted according to the protocol, might not always be true.[10] However, only four of 102 trials gave conflicting information when the publications were compared with their protocols. Another limitation is that it is still unresolved as to what extent the exaggeration associated with unclear allocation concealment in trial publications can be explained by inadequate concealment and ensuing selection bias, as opposed to unclear concealment being a marker of other sources of bias.[4]

**Relation of our findings to those of other studies**

Our strict criteria might have been too stringent, and four related studies used criteria with a stringency somewhere between our strict and loose criteria.[14] [15] [17] [18]

Three studies indicated that trial publications with unclear allocation concealment reflect poor reporting of adequate methods, rather than poor methods.[14] [15] [17]

In a retrospective questionnaire survey of investigators by Hill et al, 78% of 32 trials with unclear allocation concealment in trial publications were adequately concealed according to the primary investigators.[14] The finding, however, centred on a small sample, on the reliability and

memory of the investigators, and on assumptions of what the 20% of non-responders would have replied.

Devereaux et al. found that 54 of 56 trials with unclear allocation concealment in the trial publication were adequately concealed according to a pre-announced telephone interview of the investigators.[15] These trials were published in journals with higher impact factors than ours and might be of higher methodological quality. Or maybe some of the protocols in our cohort failed to adequately to detail all the bias-protection procedures to be adopted. Devereaux et al. argue that since investigators were willing to report lack of blinding of some parties, they would probably answer reliably regarding lack of allocation concealment. However, while lack of blinding may be a question of feasibility, lack of allocation concealment is inexcusable and hence potentially less likely to be admitted. The reliability of surveyed trial investigators has previously been reported on in two surveys where 86% (42/49) and 80% (28/35) of investigators denied the existence of unreported outcomes, although there was evidence to the contrary in their study protocols. [10] [16]

Another survey was done on trials carried out within the framework of the Radiation Therapy Oncology Group, where all trial protocols undergo a rigorous six step peer review process.[17] Although all studies had adequate allocation concealment (central randomisation) only 42% reported adequate concealment in the trial publication. However, as the authors pointed out, their result has limited generalisability since few trial protocols undergo such rigorous peer review and, as documented in our broad cohort, central randomisation is not the most commonly used method across medical specialties.

Finally, Liberati et al[18] reported results similar to ours; among 47 trials with unclear allocation concealment in the publications, 11 (23%) used adequate randomisation methods (defined as central randomisation) according to a subsequent telephone interview of all but one investigator. The discrepancy with the findings of Hill et al and Devereaux et al might reflect the difference in response rate, criteria for adequate concealment, recentness of the included trials, or the strategies for contacting and phrasing the questions to the investigators.

**Implications for clinicians and policy makers**

It is prudent to assume that a notable fraction of the overestimation of the treatment effect associated with unclear allocation concealment is caused by selection bias. This fraction can be reduced through several mechanisms. Journals should endorse and enforce the consolidated standards of reporting trials statement (www.consort-statement.org), which recommends explicit description of the allocation procedures in publications of trials, and the gatekeepers who sanction

protocols for funding and approval should demand that adequate methods are described in protocols and implemented in trials. Furthermore, our study adds to the argument that protocols should be made publicly available,[10] [19] [20] because public access would increase the reliability of critical appraisal of the fraction of trials where the protocol does describe methods for allocation concealment. Such access would most likely require international legislation and implementation by drug regulatory authorities for trials on pharmaceutical interventions, and research ethics committees for trials on non-pharmaceutical interventions. Both necessitates appropriate investment because these institutions are already pressured to review too much, too quickly.[21] [22]

**What is already known on this topic**

In most trial publications, allocation concealment is unclear or inadequate

Unclear or inadequate concealment in publications is associated with an exaggeration of the treatment effect by 20-30%, on average

**What this study adds**

Most often allocation concealment is also unclear in the protocol

Gatekeepers who sanction protocols should require that adequate methods of allocation concealment be described and used

Protocols should be publicly accessible to enhance critical appraisal of trials

We thank the scientific and ethical committees for Copenhagen and Frederiksberg, who provided access to the trial protocols and offered administrative support; the researchers who responded to the questionnaires; and J Hedegaard for secretarial support.

Contributors: All authors conceived and designed the study, analysed and interpreted the data, critically revised the manuscript and approved the final version of the manuscript. JP, AH, EF, and PCG acquired the data. PGC provided administrative support for the study. JP drafted the manuscript. She is guarantor for the paper and accepts full responsibility for the conduct of the study, had access to the data, and controlled the decision to publish.

Funding: JP, AH, EF, and PCG are funded by Copenhagen Hospital Corporation. AWC was supported by the Rhodes Trust. DGA is supported by Cancer Research UK. The funding organisations had no role in any aspect of the study, the manuscript, or the decision to publish.

Competing interests: None declared.

Ethical approval: Not required.

1. Schulz KF, Grimes DA. Allocation concealment in randomised trials: defending against deciphering. *Lancet* 2002;359:614-8.
2. Altman DG, Schulz KF, Moher D, Egger M, Davidoff F, Elbourne D, et al. The revised CONSORT statement for reporting randomized trials: explanation and elaboration. *Ann Intern Med* 2001;134:663-94.

3. Jüni P, Altman DG, Egger M. Systematic reviews in health care: assessing the quality of controlled clinical trials. *BMJ* 2001;323:42-6.
4. Schulz KF, Chalmers I, Hayes RJ, Altman DG. Empirical evidence of bias. Dimensions of methodological quality associated with estimates of treatment effects in controlled trials. *JAMA* 1995;273:408-12.
5. Moher D, Pham B, Jones A, Cook DJ, Jadad AR, Moher M, et al. Does quality of reports of randomised trials affect estimates of intervention efficacy reported in meta-analyses? *Lancet* 1998;352:609-13.
6. Kjaergard LL, Villumsen J, Gluud C. Reported methodologic quality and discrepancies between large and small randomized trials in meta-analyses. *Ann Intern Med* 2001;135:982-9.
7. Balk EM, Bonis PA, Moskowitz H, Schmid CH, Ioannidis JP, Wang C, et al. Correlation of quality measures with estimates of treatment effect in meta-analyses of randomized controlled trials. *JAMA* 2002;287:2973-82.
8. Jüni P, Egger M. Allocation concealment in clinical trials. *JAMA* 2002;288:2407-8.
9. Schulz KF, Altman DG, Moher D. Allocation concealment in clinical trials. *JAMA* 2002;288:2406-7.
10. Chan AW, Hróbjartsson A, Haahr MT, Gøtzsche PC, Altman DG. Empirical evidence for selective reporting of outcomes in randomized trials: comparison of protocols to published articles. *JAMA* 2004;291:2457-65.
11. Alderson P, Green S, Higgins JPT, ed. Assessment of study quality. *Cochrane reviewers' handbook 4.2.2* [Updated Mar 2004]; sect 6. In: Cochrane Library, Issue 2. Chichester: Wiley, 2004.
12. Meinert C. *Clinical trials. Design, conduct and analysis*. Oxford: Oxford University Press, 1986.
13. Berger VW, Bears JD. When can a clinical trial be called 'randomized'? *Vaccine* 2003;21:468-72.
14. Hill CL, LaValley MP, Felson DT. Discrepancy between published report and actual conduct of randomized clinical trials. *J Clin Epidemiol* 2002;55:783-6.
15. Devereaux PJ, Choi PT, El-Dika S, Bhandari M, Montori VM, Schunemann HJ et al. An observational study found that authors of randomized controlled trials frequently use

concealment of randomization and blinding, despite the failure to report these methods. *J Clin Epidemiol* 2004;57:1232-6.

16. Chan AW, Krleza-Jeric K, Schmid I, Altman DG. Outcome reporting bias in randomized trials funded by the Canadian Institutes of Health Research. *CMAJ* 2004;171:735-40.
17. Soares HP, Daniels S, Kumar A, Clarke M, Scott C, Swann S, et al. Bad reporting does not mean bad methods for randomised trials: observational study of randomised controlled trials performed by the Radiation Therapy Oncology Group. *BMJ* 2004;328:22-4.
18. Liberati A, Himel HN, Chalmers TC. A quality assessment of randomized control trials of primary treatment of breast cancer. *J Clin Oncol* 1986;4:942-51.
19. Jüni P, Rutjes AW, Dieppe PA. Are selective COX 2 inhibitors superior to traditional non steroidal anti-inflammatory drugs? *BMJ* 2002;324:1287-8.
20. Chalmers I, Altman DG. How can medical journals help prevent poor medical research? Some opportunities presented by electronic publishing. *Lancet* 1999;353:490-3.
21. Ghersi D, Campbell EG, Pentz R, Cox Macpherson C. The future of institutional review boards. *Lancet Oncol* 2004;5:325-9.
22. Abraham J. The pharmaceutical industry as a political player. *Lancet* 2002;360:1498-502.

Box 1. Examples from our cohort of how we assessed allocation concealment for the three most common methods.

*1. Central randomisation*

In an open trial clinicians contacted the central randomisation office to obtain the treatment allocation of the next patient. The clinicians provided more information on the patient's prognosis than necessary for stratification. *Comment:* We had no guarantee that the central staff person did not use this additional information to alter an unconcealed allocation sequence to "help" the trial to show the desired result; unclear by the strict criteria, adequate by the loose.

*2. Numbered coded vehicles*

A study used numbered coded vehicles containing treatment and control treatment in random order, but there was no information on whether the numbering or administration of the vehicles was sequential. *Comment:* Here the numbers could have been a random sequence of two numbers i.e. 22212111, meaning that if the blinding was broken for just one patient it would be broken for all. Or if there was no demand for sequential delivery of them to the patients, then known or decipherable block sizes or security envelopes could allow for informed adjustment of the sequence of their delivery; unclear by the strict criteria, adequate by the loose.

*3. Envelope method*

An open trial used sealed envelopes to allocate patients to each group. *Comment:* No information on whether the envelopes were transparent if held up against strong light allowing selection of the next patient to be enrolled to have a prognosis that would make the preferred treatment appear superior; or if the sequence of patients was difficult to alter, then to change the sequence of envelopes, as these were not described as pre-numbered; unclear by the strict criteria; but adequate by the loose.

Table A. Definitions of methods for allocation concealment and strict vs. loose criteria for assessing them

| | **Strict criteria** (predefined) | **Loose criteria** (post hoc) |
|---|---|---|
| **Central randomisation** defined as: the clinician enrolling the participants contacts a remote centre and obtains the treatment assignment | | |
| Adequate | Not falling into the category of unclear measures described below | All trials where treatment allocation was obtained by contacting a remote centre |
| Unclear | *If* disclosure of participants' prognostic data* to the central office staff was possible before the clinician obtained the treatment assignment, operationalised as: no negation of this possibility<br>*And* no precautions were reported to avoid central selection bias, operationalised as: no information on whether the allocation sequence was concealed to the central staff until the participant is irreversibly registered *and* no assurance that the sequence is strictly sequentially administered** | If not explicitly inadequate then trials were classified as having adequate allocation concealment |
| **Envelopes method** defined as: Treatment allocated either by sequential administration of envelopes containing the treatment assignment or by drawing them at random | | |
| Adequate | *If* the envelopes were opaque *and* sealed *and* serially numbered | If envelopes were sealed |
| Unclear | If one or more of the abovementioned criteria were not met | If envelopes were not described as sealed |
| **Numbered coded vehicles** defined as: Numbered coded vehicles (implicitly or explicitly described as containing the treatment in random order) when no other means of allocation concealment was implied | | |
| Adequate | *If* vehicles were indistinguishable, sequentially numbered, and sequentially administered<br>*And* no implication that the investigator allocating them to patients had any foreknowledge of their contents | Vehicles were indistinguishable<br>*And* no implication that the investigator allocating them to patients had any foreknowledge of their contents |
| Unclear | If no information on whether the vehicles were sequentially administered | Not relevant, since failing one of the adequacy criteria above would imply inadequacy*** |
| **Measures that were considered inadequate according to both strict and loose criteria**: allocation by alternation, date of birth, case record number or open table of random numbers*** | | |
| **Other measures of convincing allocation concealment would be classified as such according to both strict and loose criteria** | | |
| **Studies with information on concealment that did not fall into one of the categories defined above or did not provide any information at all were classified as unclear** | | |

* Prognostic data for a pre hoc stratification and minimization not included
** Minimisation was interpreted as being inherently strictly sequentially administered, as the allocation of a patient will be uniquely determined by how the patients' baseline data correspond to the factors for which minimization is employed
*** Two studies, that supposedly used numbered coded vehicles for concealment, explicitly described the vehicles or their content as distinguishable. These two studies were classified as having inadequate allocation concealment (see table 2)

Table 1. Comparison of adequacy of allocation concealment as described in pairs of protocols and corresponding trial publications in the 102 trials according to strict (top half) and loose criteria (bottom half).

| | | **Protocols** | | | |
|---|---|---|---|---|---|
| | | Adequate | Unclear | Inadequate | Total |
| **Trial publications** | Adequate | 4 | 1 | 0 | 5 |
| | *Unclear* | *15* | *80* | *1* | *96* |
| | Inadequate | 0 | 0 | 1 | 1 |
| | Total | 19 | 81 | 2 | 102 |
| | | | | | |
| | | **Protocols** | | | |
| | | Adequate | Unclear | Inadequate | Total |
| **Trial publications** | Adequate | 12 | 6 | 0 | 18 |
| | *Unclear* | *33* | *49* | *1* | *83* |
| | Inadequate | 0 | 0 | 1 | 1 |
| | Total | 45 | 55 | 2 | 102 |

Table 2. Allocation concealment methods in pairs of protocols and corresponding trial publications

| | | **Protocols** | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | | Centralized | Envelopes | Numbered coded vehicles | Other | Type of method uncertain | No information available | Total |
| **Trial Publications** | Centralized | 3 | | 1 | | | 2 | 6 |
| | Envelopes | | 7 | 1 | | | 1 | 9 |
| | Numbered coded vehicles | | | 3 | | 1 | | 4 |
| | Other | | 1 | 1 | | | 1 | 3 |
| | Type of method uncertain | | 1 | 3 | | | 1 | 5 |
| | No information available | 11 | 2 | 17 | | 6 | 39 | 75 |
| | Total | 14 | 11 | 26 | | 7 | 44 | 102 |

## Study III

# Impact of allocation concealment on conclusions drawn from meta-analyses of randomised trials

Pildal, J (MD)[1], Hróbjartsson, A (MD, Ph.D)[1], Jørgensen KJ (MD)[1]: Hilden J (MD)[2], Altman DG (DSc)[2], Gøtzsche PC (MD, DMSc)[1]:

*[1] The Nordic Cochrane Centre, Rigshospitalet, DK, [2] Department of Biostatistics, University of Copenhagen, DK. [3] Centre for Statistics in Medicine, Oxford, UK.*

**Keywords:** allocation concealment, bias, methodological quality, bias protection, randomised controlled trials, meta-analysis.

**Corresponding author**

Julie Pildal, MD

The Nordic Cochrane Centre, H:S Rigshospitalet, Dept. 7112

Blegdamsvej 9, DK-2100 København Ø, Denmark

jpl@dadlnet.dk

Word count: 3.416

**Abstract**

**Aim** Among conclusions holding an intervention preference, we estimated the percentage that remains supported when only trials with adequate allocation concealment are maintained in the analysis.

**Material** Random sample of 38 reviews from PubMed accessed in 2002 and 32 reviews from The Cochrane Library 2003, issue 2. Eligible reviews presented a binary effect estimate from a meta-analysis of randomised controlled trials (RCTs) as the first statistically significant result quoted in support of the conclusion.

**Methods** The methods sections of the RCTs of each included meta-analysis were assessed for adequacy of allocation concealment blinded to the results section. We replicated each meta-analysis using the authors' method of meta-analysis except that we only included those trials that had adequate allocation concealment. Conclusions were defined as insufficiently supported if our result was no longer statistically significant.

**Results** Thirty-four meta-analyses contained a mixture of trials with unclear or inadequate allocation concealment as well as trials with adequate allocation concealment, whereas four exclusively contained trials with adequate concealment, and 32 meta-analyses exclusively contained trials with lack of adequate concealment. The reduced sample of RCTs with adequate allocation concealment comprised 51% of the patients in all included meta-analyses. Twenty-two of the 70 reviews' conclusions (31% (95% CI: 21% to 44%)) remained supported. This reflected loss of power as well as a shift in the point estimate towards a less beneficial effect.

**Conclusion** Two-thirds of conclusions holding an intervention preference based on a meta-analysis are no longer supported if only trials with adequate allocation concealment are considered.

## Introduction

Concealment of treatment allocation is one of the most important safeguards against bias in randomised controlled trials. It means that persons who recruit patients must be unaware of which group the next patient will be allocated to, if enrolled. Otherwise that person might channel patients with a better prognosis into his or her preferred treatment. Empirical studies show that the effects of the experimental interventions are exaggerated by approximately 20% on average when measured on the odds ratios scale, if allocation concealment is unclear or inadequate.[1-6] That is, the ratio of odds ratios of trials with *unclear or inadequate* allocation concealment to the odds ratios of trials with *adequate* allocation concealment, abbr. ROR, is around 0.80; thus on average, the odds ratios of trials with unclear or inadequate adequate allocation concealment drop by 20% of their value, suggesting fewer undesirable clinical events on the experimental regimen. What this corresponds to in terms of absolute reduction of risk depends on the baseline event rate and the magnitude of the treatment effect, but in general it is within the magnitude of the treatment effect most RCTs seek to detect.[7] It is therefore of interest to estimate what fraction of our health care interventions can be relied upon to convey a true effect as opposed to the fraction where the effect is uncertain, since the interventions were introduced based on potentially biased trials. Here we aimed to explore the prevalence and consequence of failure to consider concealment of allocation in reviews. Parallel investigations of another component of protection against bias, viz. double blinding, will also be reported.

**Material and methods**

*Primary outcome:* The percentage of reviews where the conclusions were no longer supported if only trials with adequate allocation concealment were included. A sensitivity analysis was planned to estimate the percentage of reviews where the conclusions were no longer supported if the estimates from trials with unclear or inadequate allocation were adjusted and weighted according to the expected size, direction and variance of bias.

*Secondary outcome:* A pooled estimate of the ratio of the treatment effect estimates from trials with unclear or inadequate allocation concealment compared with those from trials with adequate allocation concealment in the included meta-analyses.

*Identification and selection of reviews*

We aimed to retrieve 70 reviews; half from the Cochrane Library and half from PubMed (hereinafter referred to as Cochrane reviews and PubMed reviews). The first samples to pass all four steps of the selection process as outlined in Fig. 1.a and 1.b were too small; hence, the sample of 181 Cochrane reviews and 128 PubMed reviews reflect our best guess at the necessary number needed to be sifted in an additional round of selection. Thirty-eight Cochrane reviews and 32 PubMed reviews were included.

Step 1: Identifying samples of reviews to be assessed

PubMed and Cochrane reviews were identified by the following search strategies. The PubMed database was searched for the years 2001 and 2002:

#1 randomi* OR controlled OR blind* OR placebo OR "controlled ? trial"

#2 meta-analysis OR metaanalysis Field: All Fields, Limits: Meta-Analysis

#3 #1 AND #2

The first 128 matches in the order of date of publication were assessed for eligibility.

The Cochrane Library: An IT specialist not otherwise involved in the project identified all reviews containing at least one meta-analysis with a binary outcome in Cochrane Library 2003, issue 2 and numbered them randomly. The first 181 reviews were assessed for eligibility in the assigned random order.

Steps 2-4: Applying selection criteria for reviews and meta-analyses

Meta-analyses of randomized trials of prophylactic or therapeutic interventions were eligible. Only one meta-analysis from each review was included and was identified as follows: a) The first statistically significant meta-analysis result presented in the results section of the abstract identified the potentially eligible meta-analysis. b) If this analysis concerned a binary outcome and was supported by a preference stated in the conclusion then it was included. Otherwise the review was excluded. Other exclusion criteria were:

- Substantial uncertainty concerning what the authors of the review perceived as experimental and conventional treatment
- More than 40 trials in the index-analysis
- Arithmetical average (instead of a weighted average of test statistics)
- The abstract of the review explicitly stated that it was partly based on non-randomized trials

Steps 3 and 4 involved some subjective judgement and were therefore carried out independently and in duplicate by two authors; disagreements led to exclusion.

*Retrieval of trials*

All trials of the included meta-analyses were retrieved. If the authors of a review referred to several reports of one trial, then all these reports were retrieved and formed the basis of the assessment of allocation concealment. If the methods section in a trial report referred to another report for further details, then this was also obtained.

*Assessment of adequacy of allocation concealment and double-blinding*

A student not otherwise involved in the project copied the methods sections of the RCTs of the included meta-analyses. Hence, pairs of authors assessed the two bias protection components, adequacy of concealment of allocation and double blinding, independently, in duplicate, and blinded to the results sections. Allocation concealment was assessed according to the same criteria as those employed by Schulz et al.[1]:

Trials with adequate concealment employed central randomisation including pharmacy-controlled randomisation (where a pharmacy remote from the clinical ward allocates the treatment); numbered or coded bottles or containers; serially numbered, opaque, sealed envelopes; or the trialists presented other descriptions that implied convincing concealment. Methods were deemed to provide inadequate concealment if it was obvious to which treatment the next patient would be allocated if enrolled; for instance alternation; reference to case record numbers or to dates of birth. Trials with unclear concealment did not report an allocation concealment approach at all or reported an approach that did not fall into one of the above categories.

Trials were categorized as double-blind if described as double-blind; or "patient and provider" blinded. Trials reported to be "placebo-controlled" without any indication that the treatments might be distinguishable or that any investigators might have become un-blinded before

the onset of the treatment were also categorized as double-blinded. "Patient and assessor"-blinding was not categorized double-blinded.

*Extraction of 2 x 2 data for each meta-analysis*

The 2 x 2 data of the trials in the individual meta-analyses were reported in all Cochrane reviews and in a minority of the PubMed reviews. When lacking in the review, data were sought in the trial publications instead. However, the methods sections in nine reviews were not detailed enough to allow unequivocal identification of the exact numbers that had been entered into the meta-analysis.[8-15] In those cases the authors of the reviews were contacted. If they supplied the data, we used the data for the re-analysis using the authors' method of meta-analysis.[13] Otherwise the point estimate and its standard error were extracted from the meta-analyses and a generic inverse variance analysis was used instead (here "generic" designates that it is not one of the more specific types of meta-analyses such as Peto's odds ratio method). [16] In one case neither was possible; the authors were willing to provide the data, but no longer had access to them, and the company (Novartis), which had provided the data and sponsored the trial, declined our request for the data, and it was not possible to replicate the meta-analysis.[12] Since this meta-analysis exclusively comprised trials with unclear or inadequate allocation concealment, it contributed to our primary outcome but not to the secondary.

*Data analysis and statistics*

First we checked whether we could replicate the meta-analyses using the review authors' method of analysis while retaining all trials in the analysis. The summary statistics in the included meta-analyses were odds ratio, relative risk, risk difference and hazard ratio; and a range of different methods for meta-analysis was applied. When a generic inverse variance analysis had to be employed instead, we used a fixed effect or random effects model depending on what was closest to

the original method. We did not correct any errors of analysis but replicated the original meta-analyses as performed by the review authors, because we wished to isolate the impact of insisting on reported adequate allocation concealment in a representative sample of meta-analyses. The term "reported adequate allocation concealment" is sometimes used in following text, because the absence vs. presence of it denotes unclear or inadequate vs. adequate allocation concealment, whereas "absence" vs. "presence of adequate concealment" would signify inadequate vs. adequate allocation concealment.

For the primary outcome, we redid each meta-analysis using the review authors' method of analysis when possible, but included only those trials that had reported adequate allocation concealment. Conclusions were considered insufficiently supported by the data if our estimates were not statistically significant at a two-sided 5% significance level.

For the secondary outcome, we used an approach described by Sterne et al.[17] Briefly, we estimated the ROR comparing trials with absence vs. presence of reported adequate bias protection by performing a random effects meta-regression analysis on each meta-analysis. To ensure consistency, we re-calculated the effect estimates where necessary, so that all results were expressed as undesirable events (e.g. presence of symptoms, not absence of symptoms). The odds in the numerator of the odds ratio of each trial were the experimental treatment. These ROR estimates were then combined in a random effects generic inverse variance meta-analysis. This entailed that reviews where the 2 x 2 data could not be obtained, or where all trials either lacked or reported adequate bias protection (concealment or blinding, as the case may be), did not contribute to the analysis.

## Results

*Characteristics of included meta-analyses*

The selection process for Cochrane reviews and PubMed reviews is outlined in Fig 1.a and 1.b, respectively. The most noticeable difference between Cochrane and PubMed reviews was the larger fraction of Cochrane reviews excluded because no preference was stated in the conclusion: 59/109 (54%; CI: 44% to 64%) vs. 20/73 (27%; 17% to 39%) in PubMed reviews. Thirty-eight Cochrane reviews were included and comprised a total of 202 trials and a median of 4 trials in each meta-analysis (10-90 percentile: 2-12).[18-55] The 32 included PubMed reviews comprised 297 trials with a median of 7 trials included in each meta-analysis (10-90 percentiles: 4-17).[56-87]Three trials appeared in 2 reviews based on 6 and 17 trials respectively. As the overlap was small and the reviews addressed different outcomes none of them were excluded.

The sample size (number of patients) was reduced by 48.6% when trials with unclear or inadequate allocation concealment were excluded. The distribution of trials with adequate, unclear, inadequate and no randomisation are given in Table 1. Although all reviews purportedly included randomised trials only, our re-assessments revealed that seven studies were not randomised. However, as this is a pragmatic study the reviews containing these trials were not excluded, and the few non-randomised trials were treated in the analyses as having inadequate allocation concealment. The meta-analyses were reproduced exactly everywhere, or to within one or two rounding units (0.01 when an odds ratio scale was used); there were 2 exceptions but even here the discrepancy was never > a tenth of the associated standard error, which in turn seemed everywhere correctly calculated. The co-occurrence of allocation concealment with double-blinding is outlined in Table 2.

*Impact of insisting on reported adequate allocation concealment*

In total, 34 meta-analyses comprised a mixture of trials with and without reported adequate allocation concealment, whereas 32 meta-analyses exclusively contained trials with unclear/inadequate concealment, and the last 4 meta-analyses exclusively contained trials with adequate concealment. Overall, 22 of the 70 reviews' conclusions remained supported when only trials with adequate concealment were included, i.e., 31% (95% CI: 21% to 44%); 14 of 38 (37%) conclusions in Cochrane reviews, and 8 of 32 (25%) in PubMed reviews remained supported.

*Magnitude of overestimation of treatment benefit*

The pooled estimate of the ratio of the treatment effect estimates from trials with unclear or inadequate allocation concealment compared with those from trials with adequate allocation concealment in the included meta-analyses was a ratio of odds ratios (ROR) of 0.92 (95% CI: 0.82 to 1.03) (Fig. 2). The corresponding ROR of trials without double-blinding compared to those with double-blinding was similar: ROR 0.93 (95% CI: 0.80 to 1.09). Hence, there was a trend towards a seemingly more beneficial effect of the experimental treatment in the trials without allocation concealment. But absence of adequate bias protection was not statistically significantly associated with an overall inflation of the effect of the experimental treatment for any of the two components.

We explored how our result adds to the current evidence of the impact of lack of reported allocation concealment on treatment effect estimates. This was done in a random effects generic inverse variance meta-analysis of our result and those of other similar studies [1-6]. The overall estimate was a ROR of 0.82 (0.72 to 0.93), implying a relative overestimation of the treatment effect of 18% in trials without reported allocation concealment. However, the level of heterogeneity was highly statistically significant (Fig. 3), and we consequently abstained from the planned sensitivity analysis of the primary outcome.

## Discussion

### *Major findings*

Two-thirds of the review conclusions stating a preference were no longer supported if only trials with adequate allocation concealment were included. The loss of support is partly due to loss of statistical power since the excluded trials comprised approximately half of all the patients, and partly due to the tendency for estimates of treatment effects derived from trials reporting adequate concealment to be less beneficial than those of the excluded trials.

### *Strengths and limitations of this study*

We aimed to achieve consistent and valid assessments of the bias protection components by retrieving all the original trial publications and re-assessing the adequacy of these components in independent duplication blinded to the results sections.

Our finding, that most conclusions are no longer supported if only trials with adequate allocation concealment are relied upon, could have limited generalisability. This would be the case if meta-analyses were not representative of the evidence that support current health care interventions. Meta-analyses are often performed when a clinical question has not been satisfactorily answered in a single large trial, and as large trials may tend to have adequate allocation concealment, a fraction of current interventions will be derived from a single large trial with a high level of bias protection. However, that fraction is probably small since most RCTs have low statistical power and a level of bias protection comparable to those included in the meta-analyses of the present study,[88-91].

We pragmatically chose the 5% significance level as a cut-off point for when a conclusion was to be considered supported by the evidence. In practice, one will no doubt encounter exceptions

where a higher *p*-value would be sufficient to support a preference. Conversely, there are reviews in which the conclusions we analyzed were based on other findings than the primary outcome (i.e., the first statistically significant meta-analysis result presented in the results section of the abstract in support of the reviewer's conclusion was not the authors' primary outcome). When this situation arises, it may be appropriate to require a smaller *p*-value in order to allay concerns about multiple testing.

The estimated impact of lack of reported adequate bias protection on the overall effect estimates was protected against confounding by disease area and type of intervention because it was based on meta-analyses. But other confounders may have been important; for instance, the two bias protection components might well have been mutual confounders. We found that trials with adequate allocation concealment were more often also double-blinded than trials with unclear or inadequate allocation concealment (Table 2). Furthermore, a statistical interaction between allocation concealment and double-blinding may exist. However, the meta-analyses were too few and too small to permit exploration of these possibilities.

*Relation of our finding to those of other studies*

The proportion of interventions in internal medicine supported by RCTs has previously been estimated.[92] Our study adds what fraction of these RCT-supported interventions remains supported if reported adequate allocation concealment is required, and to our knowledge it is the first to do so. That only 31% of conclusions remained supported is consistent with the high prevalence of trial reports with unclear or inadequate allocation concealment.[93] But this does not imply that our result was predictable; especially since review authors might have been more cautious when drawing conclusions based on trials with an uncertain or high level of bias susceptibility.

Our estimate of the impact of unclear or inadequate allocation concealment was less than what we had expected. Our finding of a ROR of 0.92 when trials with absence vs. presence of reported adequate allocation concealment were compared was not statistically significant and it was closer to 1 than any of the corresponding estimates reported in the first four similar studies. [1-4] A fifth study, by Balk et al,[5] found a ROR of 0.95 (0.83 to 1.09) (to facilitate comparisons all RORs in the present paper are re-expressed so that the denominator represents trials with adequate bias protection). However, the latter result was questioned because one of the inclusion criteria was statistically significant heterogeneity between the estimated treatments effect reported in the trials of each meta-analyses,[94] which would probably introduce too much noise to allow detection of the full effect of lack of reported adequate concealment. While this might be true for the Balk study, it does not apply to our study or to a sixth study, which found a ROR of 1.02 (0.90 to 1.16).[6] A test for statistical heterogeneity between these first four studies was not statistically significant ($p$ = 0.13),[1-4] but when all the studies mentioned above, including our own, were pooled in a generic inverse variance meta-analysis, the test for heterogeneity was highly significant ( < 0.00001) (Fig. 3). Several explanations of this heterogeneity could be offered. Firstly, confounders may have differentially influenced the results of the studies. Other confounders could have had a differential impact in the individual studies. For instance, the individual bias protection components might have been correlated to each other to a different extent in the different studies. Whether (and how) this was taken into account by the authors of the individual studies varied. Secondly, the apparent impact of absence of a bias protection component might differ according to subgroups, which might be differentially represented in the different meta-epidemiological studies. For example, the impact of unclear allocation concealment might be less in a cohort where drug trials with double-blinding comprise a large subgroup, because an adequate method for allocation concealment (numbered coded vehicles) is very frequently employed in these trials, but often not explicitly described in the

trial report.[95] Thirdly, the studies used slightly different criteria for adequate allocation concealment and different strategies for statistical analysis.

*Implications for research, clinicians and policy makers*

The majority of "positive" conclusions drawn from meta-analyses would be disregarded if trials with unclear or inadequate allocation concealment were excluded. This may seem too radical, especially since the bias associated with these trials appears to be smaller and less consistent than previously thought. Nevertheless, results of meta-analyses should always be accompanied by sensitivity analyses presenting the results with and without the trials with unclear or inadequate bias protection components. While such sensitivity analyses will inform the reader, a decision has to be made on whether or not the investigated intervention should be implemented. Currently, we do not know enough about the size and direction of different types bias under different circumstances to satisfactorily guide this decision. Thus, most importantly bias and uncertainty regarding the risk of bias should be prevented. This could be achieved in several ways: firstly, the gatekeepers of trials protocols (primarily drug-regulatory authorities and science ethics committees) should insist on description and ensure implementation of adequate methods to guard against bias. Secondly, trial protocols should be publicly available to facilitate critical appraisal of trials; and thirdly, the CONSORT statement, which requires explicit and appropriate reporting on bias protection components, should be broadly enforced.

However, we also need guidance as to how to interpret the vast majority of the available evidence for clinical interventions where the level of bias protection is unclear. The reasons for the heterogeneity between the studies of the impact of unclear bias protection have to be unravelled to inform such interpretation. Collaboration has been established to provide reassessment and pooling of raw data from the studies and to conduct further studies on new material.

**Acknowledgements**

J. Hedegaard is thanked for excellent secretarial support. Rasmus Moustgaard is thanked for generating the random list of eligible Cochrane Reviews.

**Funding**

JP, AH, KJJ and PCG are funded by Copenhagen Hospital Corporation. Cancer Research UK supports DGA. The funding organizations had no role in any aspect of the study, the manuscript, or the decision to publish.

**Conflict of interest**

None known to any of the authors



**Individual contributions from involved participants**

JP is the guarantor of the study, had full access to all the data in the study, and takes responsibility for the integrity of the data and the accuracy of the data analysis. All authors have approved the final version of the manuscript.

Study concept and design: All authors.

Acquisition of data: JP, AH, KJJ, PCG.

Analysis and interpretation of data: All authors.

Drafting of the manuscript: JP.

Critical revision of the manuscript for important intellectual content: All authors.

Administrative, technical or material support: PCG.

**Tables**

**Table 1** Distribution of number of trials according to our assessment of the adequacy of allocation concealment

| | Trials in Cochrane reviews<br>*n* = 202 (%) | Trials in PubMed reviews<br>*n* = 297 (%) | All trials<br>*n* = 499 (%) |
|---|---|---|---|
| Adequate | 51 (25) | 31 (10) | 82 (16) |
| Unclear | 129 (64) | 250 (84) | 379 (76) |
| Inadequate | 19 (9) | 12 (4) | 31 (6) |
| Not randomised | 3 (1) | 4 (1) | 7 (1) |

**Table 2** Number of double-blinded trials related to the level of allocation concealment. N = number of trials

| | Adequate concealment<br>*n* = 82 (%) | Unclear concealment<br>*n* = 379 (%) | Inadequate concealment<br>*n* = 38 (%) |
|---|---|---|---|
| Double-blinding | 56 (68) | 119 (31) | 3 (8) |

## Figures

**Figure 1.a** Flowchart for selection of Cochrane reviews

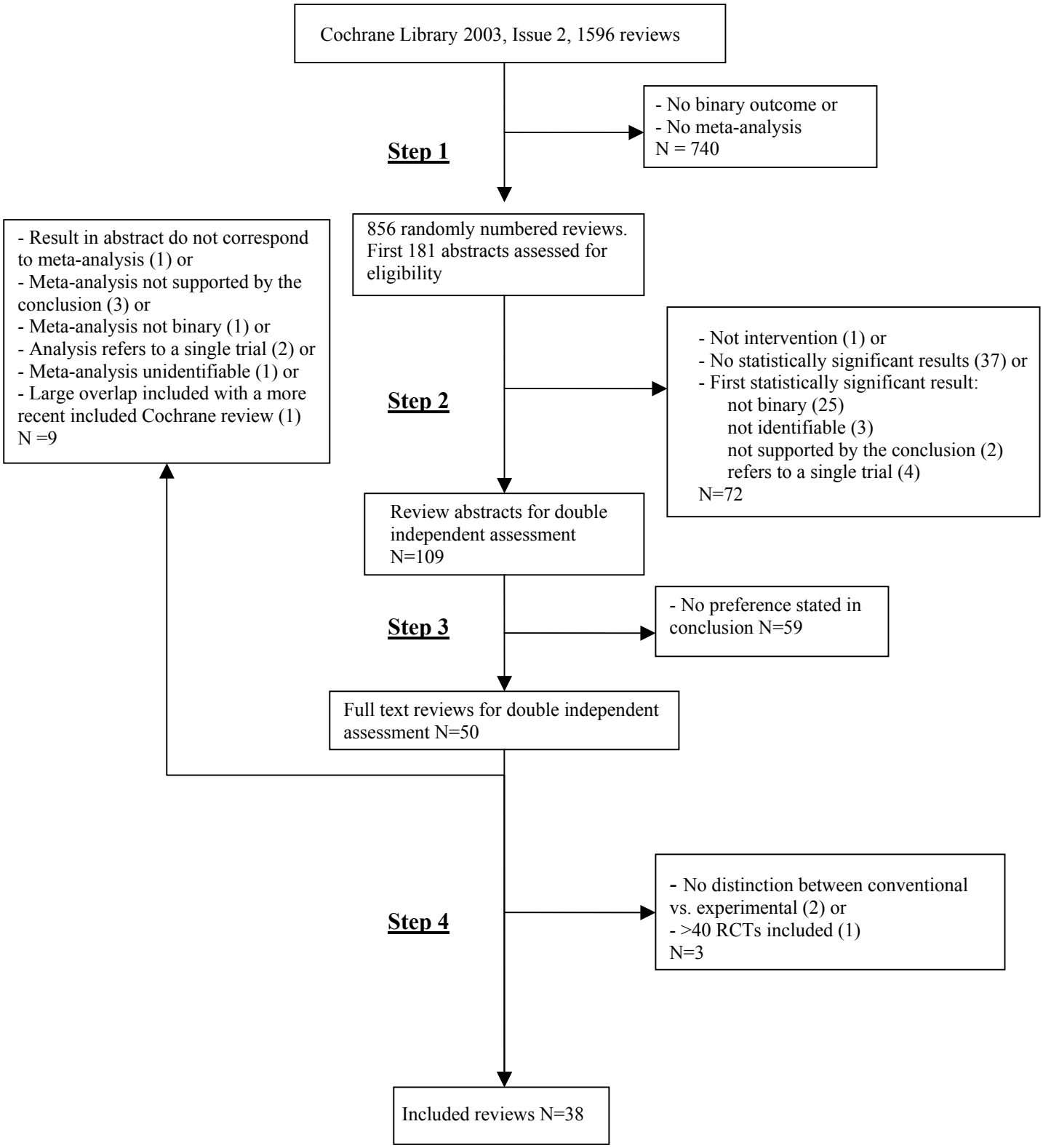


**Figure 1.b** Flowchart for selection of PubMed reviews

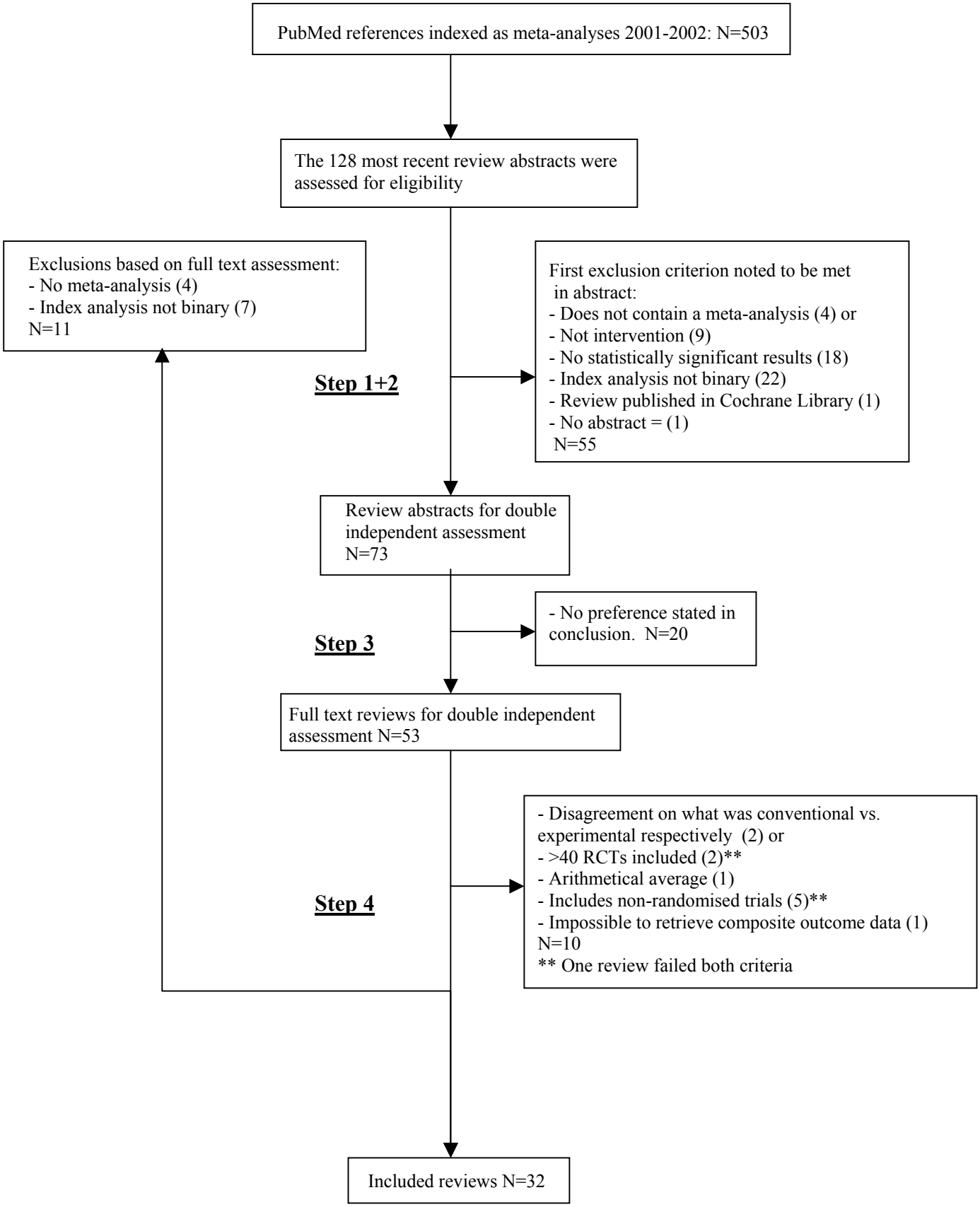


**Figure 2** Impact of lack of allocation concealment: meta-analysis of ratio of odds ratios of absence vs. presence of reported allocation concealment in trials of 29 meta-analyses

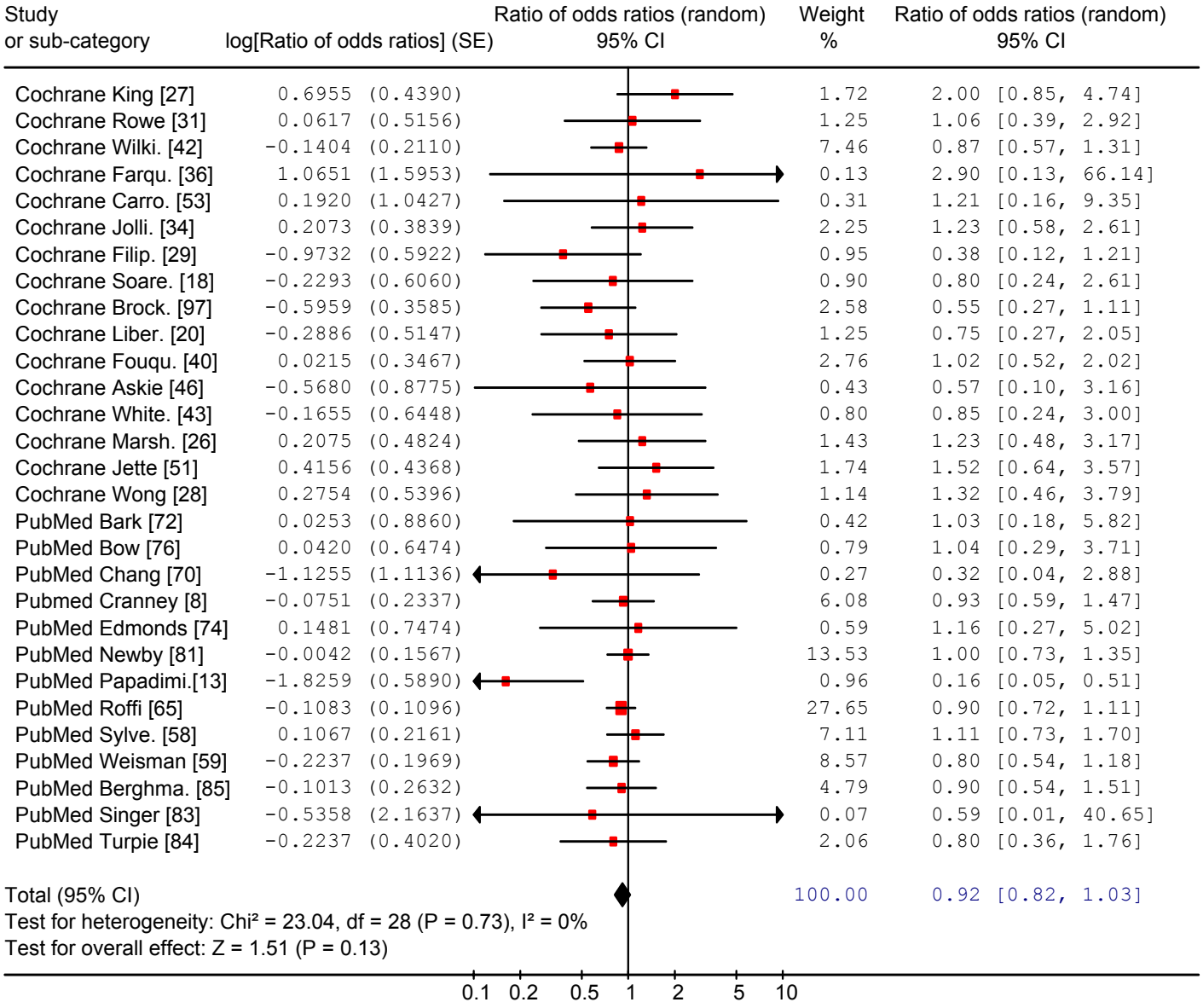
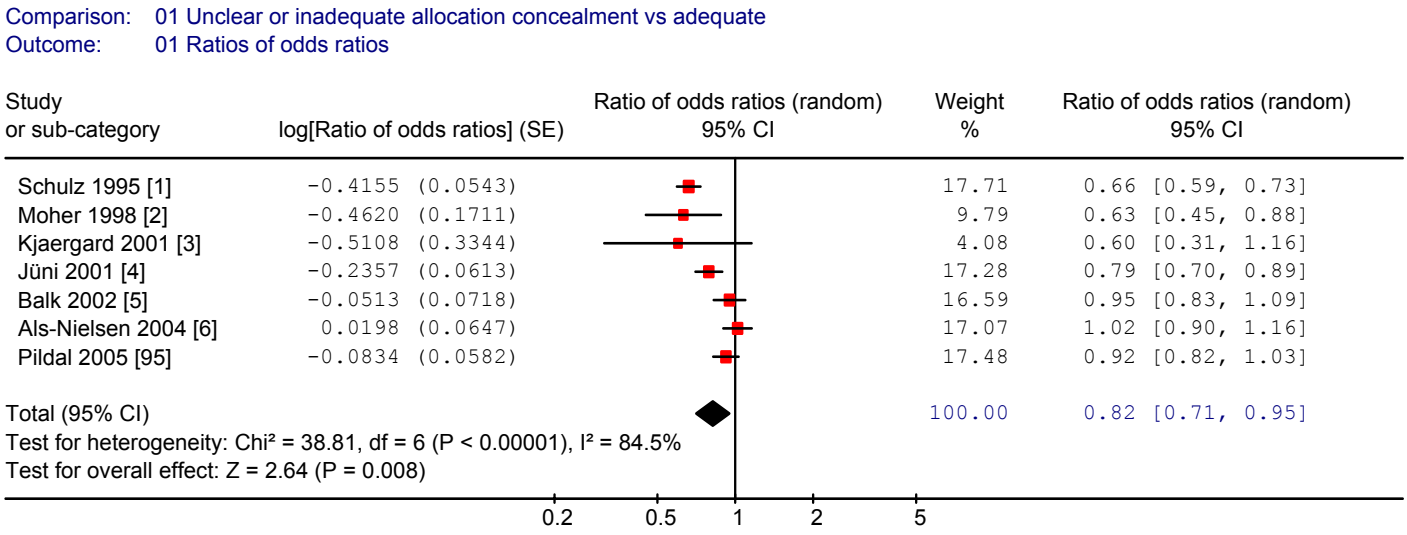


**Figure 3** Meta-analysis of the studies of the impact of unclear or inadequate allocation concealment



**Legends to figures**

**Figure 2**

A comparison of treatment effect estimates on the odds ratio scale, comparing results from trials with unclear or inadequate allocation concealment with those from trials with reported adequate allocation concealment in 29 meta-analyses. The ratios of odds ratios (squares) with 95% confidence intervals are shown. The size of the squares reflects the statistical weight (shown in a separate column) in the overall analysis. A ratio of odds ratios (ROR) below 1 (i.e., left of the centre line) implies that trials with unclear or inadequate allocation concealment show a more beneficial effect as more undesirable events are prevented. Meta-analyses where 2 x 2 data could not be obtained did not contribute to this analysis. Square brackets: references.

**Figure 3**

Meta-analysis of the empirical studies on the impact of unclear or inadequate allocation concealment on the estimates of treatment effects. All studies compared estimates of treatment effect within a number of meta-analyses and calculated the ratio of odds ratios for this purpose. A ratio of odds ratios (ROR) below 1 implies that trials with unclear or inadequate allocation concealment show a more beneficial effect as more undesirable events are prevented. Square brackets: references.

## References


1. Schulz KF, Chalmers I, Hayes RJ, Altman DG. Empirical evidence of bias. Dimensions of methodological quality associated with estimates of treatment effects in controlled trials. JAMA 1995; 273(5):408-12.

2. 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(9128):609-13.

3. Kjaergard LL, Villumsen J, Gluud C. Reported methodologic quality and discrepancies between large and small randomized trials in meta-analyses. Ann Intern Med 2001; 135(11):982-9.

4. Jüni P, Altman DG, Egger M. Systematic reviews in health care: Assessing the quality of controlled clinical trials. BMJ 2001; 323(7303):42-6.

5. Balk EM, Bonis PA, Moskowitz H *et al*. Correlation of quality measures with estimates of treatment effect in meta-analyses of randomized controlled trials. JAMA 2002; 287(22):2973-82.

6. Als-Nielsen B , Chen W, Gluud LL, Sierma V, Hilden J, Gluud C. Are trial size and reported methodological quality associated with treatment effects? Program & Abstract Book, 12th Cochrane Colloquium 2004; 102.

7. Yusuf S, Collins R, Peto R. Why do we need some large, simple randomized trials? Stat Med 1984; 3(4):409-22.

8. Cranney A, Tugwell P, Adachi J *et al*. Meta-analyses of therapies for postmenopausal osteoporosis. III. Meta-analysis of risedronate for the treatment of postmenopausal osteoporosis. Endocr Rev 2002; 23(4):517-23.

9. Cranney A, Wells G, Willan A *et al*. Meta-analyses of therapies for postmenopausal osteoporosis. II. Meta-analysis of alendronate for the treatment of postmenopausal women. Endocr Rev 2002; 23(4):508-16.

10. Crystal E, Connolly SJ, Sleik K, Ginger TJ, Yusuf S. Interventions on prevention of postoperative atrial fibrillation in patients undergoing heart surgery: a meta-analysis. Circulation 2002; 106( 1):75-80.

11. Salas M, Ward A, Caro J. Are proton pump inhibitors the first choice for acute treatment of gastric ulcers? A meta analysis of randomized clinical trials. BMC Gastroenterol 2002; 2(1):17.

12. Haugh M, Helou S, Boissel JP, Cribier BJ. Terbinafine in fungal infections of the nails: a meta-analysis of randomized clinical trials. Br J Dermatol 2002; 147(1):118-21.

13. Papadimitropoulos E, Wells G, Shea B *et al*. Meta-analyses of therapies for postmenopausal osteoporosis. VIII: Meta-analysis of the efficacy of vitamin D treatment in preventing osteoporosis in postmenopausal women. Endocr Rev 2002; 23(4):560-9.

14. Eccleston C, Morley S, Williams A, Yorke L, Mastroyannopoulou K. Systematic review of randomised controlled trials of psychological therapy for chronic pain in children and adolescents, with a subset meta-analysis of pain relief. Pain 2002; 99(1-2):157-65.

15. Fiddian P, Sabin CA, Griffiths PD. Valacyclovir provides optimum acyclovir exposure for prevention of cytomegalovirus and related outcomes after organ transplantation. J Infect Dis 2002; 186 Suppl 1:S110-5.

16. Alderson P, Green S, Higgins JPT, editors. Assessment of study quality. Cochrane Reviewers' Handbook 4.2.2 [Updated, March 2004]; Section 6. In: The Cochrane Library, Issue 2, 2004. Chichester, UK: John Wiley & Sons, Ltd.

17. Sterne JA, Juni P, Schulz KF, Altman DG, Bartlett C, Egger M. Statistical methods for assessing the influence of study characteristics on treatment effects in 'meta-epidemiological' research. Stat Med 2002; 21(11):1513-24.

18. Soares-Weiser K, Brezis M, Tur-Kaspa R, Leibovici L. Antibiotic prophylaxis for cirrhotic patients with gastrointestinal bleeding. In: The Cochrane Library, Issue 2, 2003. Oxford: Update Software.

19. Guidugli F, Castro AA, Atallah AN. Antibiotics for preventing leptospirosis. In: The Cochrane Library, Issue 2, 2003. Oxford: Update Software.

20. Liberati A, D'Amico R, Pifferi S *et al*. Antibiotics for preventing respiratory tract infections in adults receiving intensive care. In: The Cochrane Library, Issue 2, 2003. Oxford: Update Software.

21. Sheikh A, Hurwitz B, Cave J. Antibiotics versus placebo for acute bacterial conjunctivitis. In: The Cochrane Library, Issue 2, 2003. Oxford: Update Software.

22. Sandercock P , Mielke O, Liu M, Counsell C. Anticoagulants for preventing recurrence following presumed non-cardioembolic ischaemic stroke or transient ischaemic attack. In: The Cochrane Library, Issue 2, 2003. Oxford: Update Software.

23. Bacaltchuk J , Hay P, Trefiglio R. Antidepressants versus psychological treatments and their combination for bulimia nervosa. In: The Cochrane Library, Issue 2, 2003. Oxford: Update Software.

24. Sandercock P , Gubitz G, Foley P, Counsell C. Antiplatelet therapy for acute ischaemic stroke. In: The Cochrane Library, Issue 2, 2003. Oxford: Update Software.

25. Brocklehurst P, Volmink J. Antiretrovirals for reducing the risk of mother-to-child transmission of HIV infection. In: The Cochrane Library, Issue 2, 2003. Oxford: Update Software.

26. Marshall M, Lockwood A. Assertive community treatment for people with severe mental disorders. In: The Cochrane Library, Issue 2, 2003. Oxford: Update Software.

27. King JF, Flenady VJ, Papatsonis DNM, Dekker GA, Carbonne B. Calcium channel blockers for inhibiting preterm labour. In: The Cochrane Library, Issue 2, 2003. Oxford: Update Software.

28. Wong R, Malthaner R. Combined chemotherapy and radiotherapy (without surgery) compared with radiotherapy alone in localized carcinoma of the esophagus. In: The Cochrane Library, Issue 2, 2003. Oxford: Update Software.

29. Filippini G, Brusaferri F, Sibley WA *et al*. Corticosteroids or ACTH for acute exacerbations in multiple sclerosis. In: The Cochrane Library, Issue 2, 2003. Oxford: Update Software.

30. Phelps DL, Lakatos L, Watts JL. D-Penicillamine for preventing retinopathy of prematurity in preterm infants. In: The Cochrane Library, Issue 2, 2003. Oxford: Update Software.

31. Rowe BH, Spooner C, Ducharme FM, Bretzlaff JA, Bota GW. Early emergency department treatment of acute asthma with systemic corticosteroids . In: The Cochrane Library, Issue 2, 2003. Oxford: Update Software.

32. Amaragiri SV , Lees TA. Elastic compression stockings for prevention of deep vein thrombosis. In: The Cochrane Library, Issue 2, 2003. Oxford: Update Software.

33. Smith LA, Oldman AD, McQuay HJ, Moore RA. Eletriptan for acute migraine. In: The Cochrane Library, Issue 2, 2003. Oxford: Update Software.

34. Jolliffe JA, Rees K, Taylor RS, Thompson D, Oldridge N, Ebrahim S. Exercise-based rehabilitation for coronary heart disease. In: The Cochrane Library, Issue 2, 2003. Oxford: Update Software.

35. Jeffery GM, Hickey BE, Hider P. Follow-up strategies for patients treated for non-metastatic colorectal cancer. In: The Cochrane Library, Issue 2, 2003. Oxford: Update Software.

36. Farquhar C, Basser R, Hetrick S, Lethaby A, Marjoribanks J. High dose chemotherapy and autologous bone marrow or stem cell transplantation versus conventional chemotherapy for women with metastatic breast cancer. In: The Cochrane Library, Issue 2, 2003. Oxford: Update Software.

37. Myers RP, Regimbeau C, Thevenot T *et al*. Interferon for acute hepatitis C. In: The Cochrane Library, Issue 2, 2003. Oxford: Update Software.

38. Acuin J, Smith A, Mackenzie I. Interventions for chronic suppurative otitis media. In: The Cochrane Library, Issue 2, 2003. Oxford: Update Software.

39. Chaisewikul R, Privitera MD, Hutton JL, Marson AG. Levetiracetam add-on for drug-resistant localization related (partial) epilepsy. In: The Cochrane Library, Issue 2, 2003. Oxford: Update Software.

40. Fouque D, Wang P, Laville M, Boissel JP. Low protein diets for chronic renal failure in non diabetic adults. In: The Cochrane Library, Issue 2, 2003. Oxford: Update Software.

41. Durkan A, Hodson E, Willis N, Craig J. Non-corticosteroid treatment for nephrotic syndrome in children. In: The Cochrane Library, Issue 2, 2003. Oxford: Update Software.

42. Wilkinson D, Ramjee G, Tholandi M, Rutherford G. Nonoxynol-9 for preventing vaginal acquisition of HIV infection by women from men. In: The Cochrane Library, Issue 2, 2003. Oxford: Update Software.

43. Whitelaw A. Postnatal phenobarbitone for the prevention of intraventricular hemorrhage in preterm infants. In: The Cochrane Library, Issue 2, 2003. Oxford: Update Software.

44. Smyth A, Walters S. Prophylactic antibiotics for cystic fibrosis. In: The Cochrane Library, Issue 2, 2003. Oxford: Update Software.

45. McGuire W, Clerihew L, Austin N. Prophylactic intravenous antifungal agents to prevent mortality and morbidity in very low birth weight infants. In: The Cochrane Library, Issue 2, 2003. Oxford: Update Software.

46. Askie LM, Henderson-Smart DJ. Restricted versus liberal oxygen exposure for preventing morbidity and mortality in preterm or low birth weight infants. In: The Cochrane Library, Issue 2, 2003. Oxford: Update Software.

47. Miller RG, Mitchell JD, Lyon M and Moore DH. Riluzole for amyotrophic lateral sclerosis (ALS)/motor neuron disease (MND). In: The Cochrane Library, Issue 2, 2003. Oxford: Update Software.

48. Oldman AD, Smith LA, McQuay HJ, Moore RA. Rizatriptan for acute migraine. In: The Cochrane Library, Issue 2, 2003. Oxford: Update Software.

49. Sinclair JC. Servo-control for maintaining abdominal skin temperature at 36C in low birth weight infants. In: The Cochrane Library, Issue 2, 2003. Oxford: Update Software.

50. Gibson JNA, Grant IC, Waddell G. Surgery for lumbar disc prolapse. In: The Cochrane Library, Issue 2, 2003. Oxford: Update Software.

51. Jette NJ, Marson AG, Hutton JL. Topiramate add-on for drug-resistant partial epilepsy. In: The Cochrane Library, Issue 2, 2003. Oxford: Update Software.

52. McGuire W, McEwan P. Transpyloric versus gastric tube feeding for preterm infants. In: The Cochrane Library, Issue 2, 2003. Oxford: Update Software.

53. Carroli G, Bergel E. Umbilical vein injection for management of retained placenta. In: The Cochrane Library, Issue 2, 2003. Oxford: Update Software.

54. Soares KVS, McGrath JJ. Vitamin E for neuroleptic-induced tardive dyskinesia. In: The Cochrane Library, Issue 2, 2003. Oxford: Update Software.

55. Chadwick DW, Marson AG. Zonisamide add-on for drug-resistant partial epilepsy. In: The Cochrane Library, Issue 2, 2003. Oxford: Update Software.

56. Tournaye H, Verheyen G, Albano C *et al*. Intracytoplasmic sperm injection versus in vitro fertilization: a randomized controlled trial and a meta-analysis of the literature. Fertil Steril 2002; 78(5):1030-7.

57. van der Vyver M, Halpern S, Joseph G. Patient-controlled epidural analgesia versus continuous infusion for labour analgesia: a meta-analysis. Br J Anaesth 2002; 89(3):459-65.

58. Sylvester RJ , van der MEIJDEN AP, Lamm DL. Intravesical bacillus Calmette-Guerin reduces the risk of progression in patients with superficial bladder cancer: a meta-analysis of the published results of randomized clinical trials. J Urol 2002; 168(5):1964-70.

59. Weisman SM, Graham DY. Evaluation of the benefits and risks of low-dose aspirin in the secondary prevention of cardiovascular and cerebrovascular events. Arch Intern Med 2002; 162(19):2197-202.

60. Fiddian P, Sabin CA, Griffiths PD. Valacyclovir provides optimum acyclovir exposure for prevention of cytomegalovirus and related outcomes after organ transplantation. J Infect Dis 2002; 186 Suppl 1:S110-5.

61. Eberhart LH, Morin AM, Kranke P, Geldner G, Wulf H. [Transient neurologic symptoms after spinal anesthesia. A quantitative systematic overview (meta-analysis) of randomized controlled studies]. Anaesthesist 2002; 51(7):539-46.

62. Eccleston C, Morley S, Williams A, Yorke L, Mastroyannopoulou K. Systematic review of randomised controlled trials of psychological therapy for chronic pain in children and adolescents, with a subset meta-analysis of pain relief. Pain 2002; 99(1-2):157-65.

63. Akai M, Kawashima N, Kimura T, Hayashi K. Electrical stimulation as an adjunct to spinal fusion: a meta-analysis of controlled clinical trials. Bioelectromagnetics 2002; 23(7):496-504.

64. Grant AM. Open mesh versus non-mesh repair of groin hernia: meta-analysis of randomised trials based on individual patient data. Hernia 2002; 6(3):130-6.

65. Roffi M, Chew DP, Mukherjee D *et al*. Platelet glycoprotein IIb/IIIa inhibition in acute coronary syndromes. Gradient of benefit related to the revascularization strategy. Eur Heart J 2002; 23(18):1441-8.

66. Papadimitropoulos E, Wells G, Shea B *et al*. Meta-analyses of therapies for postmenopausal osteoporosis. VIII: Meta-analysis of the efficacy of vitamin D treatment in preventing osteoporosis in postmenopausal women. Endocr Rev 2002; 23(4):560-9.

67. Cranney A, Tugwell P, Zytaruk N *et al*. Meta-analyses of therapies for postmenopausal osteoporosis. VI. Meta-analysis of calcitonin for the treatment of postmenopausal osteoporosis. Endocr Rev 2002; 23(4):540-51.

68. Cranney A, Tugwell P, Adachi J *et al*. Meta-analyses of therapies for postmenopausal osteoporosis. III. Meta-analysis of risedronate for the treatment of postmenopausal osteoporosis. Endocr Rev 2002; 23(4):517-23.

69. Cranney A, Wells G, Willan A *et al*. Meta-analyses of therapies for postmenopausal osteoporosis. II. Meta-analysis of alendronate for the treatment of postmenopausal women. Endocr Rev 2002; 23(4):508-16.

70. Chang CH, Chen KY, Lai MY, Chan KA. Meta-analysis: ribavirin-induced haemolytic anaemia in patients with chronic hepatitis C. Aliment Pharmacol Ther 2002; 16(9):1623-32.

71. San Miguel R , Guillen F, Cabases JM, Buti M. Meta-analysis: combination therapy with interferon-alpha 2a/2b and ribavirin for patients with chronic hepatitis C previously non-responsive to interferon. Aliment Pharmacol Ther 2002; 16(9):1611-21.

72. Barker FG 2nd. Efficacy of prophylactic antibiotic therapy in spinal surgery: a meta-analysis. Neurosurgery 2002; 51(2):391-400; discussion 400-1.

73. Kern JW, Shoemaker WC. Meta-analysis of hemodynamic optimization in high-risk patients. Crit Care Med 2002; 30(8):1686-92.

74. Edmonds ML, Camargo CA Jr, Pollack CV Jr, Rowe BH. The effectiveness of inhaled corticosteroids in the emergency department treatment of acute asthma: a meta-analysis. Ann Emerg Med 2002; 40(2):145-54.

75. Salas M, Ward A, Caro J. Are proton pump inhibitors the first choice for acute treatment of gastric ulcers? A meta analysis of randomized clinical trials. BMC Gastroenterol 2002; 2(1):17.

76. Bow EJ, Laverdiere M, Lussier N, Rotstein C, Cheang MS, Ioannou S. Antifungal prophylaxis for severely neutropenic chemotherapy recipients: a meta analysis of randomized-controlled clinical trials. Cancer 2002; 94(12):3230-46.

77. Haugh M, Helou S, Boissel JP, Cribier BJ. Terbinafine in fungal infections of the nails: a meta-analysis of randomized clinical trials. Br J Dermatol 2002; 147(1):118-21.

78. Crystal E, Connolly SJ, Sleik K, Ginger TJ, Yusuf S. Interventions on prevention of postoperative atrial fibrillation in patients undergoing heart surgery: a meta-analysis. Circulation 2002; 106( 1):75-80.

79. Camma C, Schepis F, Orlando A *et al*. Transarterial chemoembolization for unresectable hepatocellular carcinoma: meta-analysis of randomized controlled trials. Radiology 2002; 224 (1):47-54.

80. D'Souza AL, Rajkumar C, Cooke J, Bulpitt CJ. Probiotics in prevention of antibiotic associated diarrhoea: meta-analysis. BMJ 2002; 324(7350):1361.

81. Newby LK, Califf RM, White HD *et al*. The failure of orally administered glycoprotein IIb/IIIa inhibitors to prevent recurrent cardiac events. Am J Med 2002; 112(8):647-58.

82. A meta-analysis using individual patient data of trials comparing artemether with quinine in the treatment of severe falciparum malaria. Trans R Soc Trop Med Hyg 2001; 95(6):637-50.

83. Singer MA, Nelson RL. Primary repair of penetrating colon injuries: a systematic review. Dis Colon Rectum 2002; 45(12):1579-87.

84. Turpie AG, Bauer KA, Eriksson BI, Lassen MR. Fondaparinux vs enoxaparin for the prevention of venous thromboembolism in major orthopedic surgery: a meta-analysis of 4 randomized double-blind studies. Arch Intern Med 2002; 162(16):1833-40.

85. Berghmans T, Paesmans M, Lafitte JJ, Mascaux C, Meert AP, Sculier JP. Role of granulocyte and granulocyte-macrophage colony-stimulating factors in the treatment of small-cell lung cancer: a systematic review of the literature with methodological assessment and meta-analysis. Lung Cancer 2002; 37(2):115-23.

86. Sewnath ME, Karsten TM, Prins MH, Rauws EJ, Obertop H, Gouma DJ. A meta-analysis on the efficacy of preoperative biliary drainage for tumors causing obstructive jaundice. Ann Surg 2002; 236(1):17-27.

87. Chaiyakunapruk N, Veenstra DL, Lipsky BA, Saint S. Chlorhexidine compared with povidone-iodine solution

for vascular catheter-site care: a meta-analysis. Ann Intern Med 2002; 136(11):792-801.

88. Mulward S, Gøtzsche PC. Sample size of randomized double-blind trials 1976-1991. Dan Med Bull 1996; 43(1):96-8.

89. Chan AW, Hróbjartsson A, Haahr MT, Gøtzsche PC, Altman DG. Empirical evidence for selective reporting of outcomes in randomized trials: comparison of protocols to published articles. JAMA 2004; 291(20):2457-65.

90. Kjaergard LL, Gluud C. Funding, disease area, and internal validity of hepatobiliary randomized clinical trials. Am J Gastroenterol 2002; 97(11):2708-13.

91. Kyriakidi M, Ioannidis JP. Design and quality considerations for randomized controlled trials in systemic sclerosis. Arthritis Rheum 2002; 47(1):73-81.

92. Nordin-Johansson A, Asplund K. Randomized controlled trials and consensus as a basis for interventions in internal medicine. J Intern Med 2000; 247(1):94-104.

93. Altman DG, Schulz KF, Moher D *et al*. The revised CONSORT statement for reporting randomized trials: explanation and elaboration. Ann Intern Med 2001; 134(8):663-94.

94. Jüni P, Egger M. Allocation concealment in clinical trials. JAMA 2002; 288(19):2407-8; author reply 2408-9.

95. Pildal J, Chan AW, Hróbjartsson A, Forfang E, Altman DG, Gøtzsche PG. Comparison of descriptions of allocation concealment in trial protocols and the published reports: cohort study. BMJ 2005; In press.

## Discussion of the studies

This section is structured to meet the following formal requirements for the overview to accompany a PhD Thesis submitted to the Faculty of Health Sciences at Copenhagen University: 1) Presentation of the major findings 2) Assessments of the applied methods 3) Critical review of the conclusions drawn 4) Comparisons with and assessments of other similar studies to delineate what the thesis adds to the current knowledge.

These requirements are partly covered in Study I to III as presented in the previous section. However, to preserve coherence some repetition was inevitable, although the present section assumes that the reader has recently read the three studies. Assessments of the applied methods were given a high priority in the discussion of all the studies for two reasons. First, this issue is not covered in depth in the papers or the manuscript. Second, the suitability of the methods is intertwined with the validity and generalisability of the conclusions.

## Discussion of Study I

## Polyclonal Immunoglobulin for Treatment of Bacterial Sepsis: A Systematic Review.

### Major findings

RCTs and systematic reviews have shown conflicting results as to whether polycloncal immunoglobulin as adjunctive treatment of bacterial sepsis reduces mortality. The level of bias protection in the RCTs that contributed to the reviews was highly variable. We hypothesised that bias susceptibility of the individual trials and how this was dealt with at review level might explain the discrepant findings. Thus, we aimed to determine whether polycloncal immunoglobulin as adjunctive treatment of bacterial sepsis reduces mortality if only RCTs with a high level of bias protection are relied upon.

The major finding was that high quality trials, which comprised almost half the evidence, showed no reduction in mortality among patients with sepsis receiving adjunctive treatment with immunoglobulin compared to those who, in addition to standard treatment, received placebo or no treatment. However, if the sum of the evidence derived from RCTs on polyclonal immunoglobulin for treatment of sepsis was pooled disregarding the individual studies' reported level of bias protection, the treatment effect was then large and highly statistically significant. Yet, 23% of the variation between the results of all studies could not be explained by sampling error. This excess of unexplained variation is referred to in statistical terms as "between-trial heterogeneity", or simply heterogeneity. The methodological quality (level of bias protection) was the only covariate that explained a statistically significant amount of the heterogeneity, while other trial characteristics such as age groups (neonates vs. adults), baseline risk, dose, and type of immunoglobulin preparation did not. Thus this confirms our pre-specified primary hypothesis for explaining any between-trial heterogeneity.

**Assessment of methods**

*Feasibility of the comprehensive literature search*

Our search strategy involved no language restriction, and three databases were searched. Substantial effort was invested to retrieve and translate any RCT whether Turkish,[44] French,[45] German,[46 47] or Japanese.[48]

The feasibility of including trials published in non-English languages has been assessed by Jüni and Egger[49] who found that RCTs included in Cochrane meta-analyses and published in non-English language include fewer patients, tend to be of lower methodological quality and report more beneficial effect estimates than similar RCTs included in the same meta-analyses, but published in English. Moher et al. also assessed whether trials in non-English languages were of lower methodological quality, or tended to report more beneficial treatment effects, but detected neither.[50] However, Moher et al compared trials from 13 journals with an intermediary impact factor and used a score to assess the methodological quality of the studies,[51] which does not assess allocation concealment, but contains components unrelated to internal validity. When the two studies [49 50] are pooled, the overall effect estimate of the experimental intervention is then 12 % relatively lower (more beneficial) in trials published in non-English languages (ratio of effect estimate 0.88; 95% CI: 0.80 to 0.99). Thus, our effort to include trials published in non-English languages was likely to yield low quality trials inflating the impact of the experimental treatment and confounding (to an unknown extent) the methods-related bias issue. Of our five non-English RCTs two were excluded. The three others were of lower quality, contributed with 10% of the weight of the analysis, and excluding them moved the overall estimate of the relative risk from 0.77 to the less beneficial of 0.80. [46 44 47] In general, excluding trials in other languages had little impact in the study by Jüni and Egger (change in effect estimate less than 5% in 58% of the meta-analyses

they studied). Thus, we gained little from including trials in non-English languages. Although, this seemed likely as judged from the study by Jüni and Egger, it was not entirely predictable.

Three of the trials included in our systematic review were not indexed in PubMed. [52 53 44] These were also of lower methodological quality, but the sum of their contribution to the total statistical information was 19%, and excluding them moved the overall estimate of the relative risk from 0.77 to the less beneficial 0.83.

*Reliability of obtaining additional unpublished information from trial authors*

We sent questions to all the primary investigators of the included trials regarding additional information on the quality of their trials. The reliability of interrogating trial investigators after the publication of their trial has been published has been assessed in two surveys, where 86% (42/49) and 80% (28/35) of investigators denied the existence of unreported outcomes, although there was evidence to the contrary in their study protocols. [54 55]. If this finding can be extrapolated to the reliability of authors' responses to pivotal questions on trial quality in general, then correspondence with authors might not be worth the effort or might even be misleading. In this study we dichotomised the assessment of the level of bias protection into two groups. High-quality trials reported adequate allocation concealment, double-blinding, and provided data that made intention to treat analysis possible. Lower quality trials failed to meet one or more of these criteria. Four trials were re-classified as a consequence of correspondence with the authors. Three of these comprised 3% of the weight in the overall meta-analysis, thus the main question is whether it was justified to up-classify the largest trial by Werdan et. al., which comprised 38% of the weight in the analysis.[56] The two most important bias protection components, allocation concealment and double-blinding, were adequate as assessed by the pre-published protocol of this trial.[57] Thus, what motivated the up-classification, was the presentation of data that made intention to treat analysis possible. We had

clearly explained in our letter that data on all randomised patients were requested regardless of whether the patients actually fulfilled the inclusion criteria or got the assigned treatment. Therefore there should be no ambiguity regarding the meaning of availability of data for intention to treat analysis. Werdan also provided the number of patients lost to follow up. It seems unlikely that Werdan et al. had anything to gain by claiming that data were as randomised if this was not true or by providing data that would make immunoglobulin seem less effective/more harmful than was the case.

Another concern regarding incorporation of unpublished information on trial quality is that we do not know what the association[1] between unclear or inadequate reported bias protection measures and inflated effect estimates reflects. The association might reflect the mechanism of the biases that these components should protect against; for instance, selection bias when the status of allocation concealment is unclear. Or it might reflect that these particular ways of reporting on bias protection components are more or less consistent markers of other mechanisms generating inflated results. If the latter is the case, then the marker-to-mechanisms association may change if markers are changed in response to correspondence.

*Inclusion of data from a large trial yet to be fully published*

The uncertainty related to the intermediate publication status of the study by Werdan et al. limits the inferences that can be drawn, as we cannot exclude the possibility that problems affecting the validity of the trial occurred. However, other reasons could be, e.g., that the authors did not published because they anticipated rejection by journals because the result was negative. This seems unlikely though, as this is a large presumably high-quality trial on severely ill patients. Thus, another explanation seems more likely, that the sponsor controls the data.[14] There are several high-

[1] Please refer to the background section of this overview

profile cases of companies trying to suppress publication of results unfavourable to their product.[58] [59] [60] The study by Werdan et al. was an industry sponsored trial according to the published protocol.[57] Studies sponsored by pharmaceutical companies are less likely to published than those funded by other sources, and published studies sponsored by pharmaceutical industry are more likely have outcomes favouring the sponsor.[61] Finally, regardless of the cause, there is empirical evidence that unpublished trials on average show less favourable treatment effects.[40] Accordingly, if unpublished studies are excluded from meta-analyses, then the overall effect estimates are likely to be inflated.

*Quality assessment*

We aimed to determine whether polycloncal immunoglobulin as adjunctive treatment of bacterial sepsis reduces mortality if only RCTs with a high level of bias protection are relied upon. Hence, a high level of bias had to be defined. In this study it comprised adequate allocation concealment, double-blinding, and data that made intention to treat analysis possible. Whereas trials where one or more of these requirements were not met were classified as lower quality, that is, less well protected against bias. These components were assessed because of the empirical evidence of their importance. It may seem like a loss of information to dichotomize these into whether or not all three of them were fulfilled. We did this, partly because we believe that a trial is only as strong as its weakest link, and partly to avoid data-dredging by testing different combinations' fulfilment of the criteria without a pre-specification of which combination would be the primary. However, we also presented an exploratory sensitivity analysis (clearly labelled as such) where fewer criteria had to be fulfilled in the "high level of bias protection"-category. Finally, table 1 in the paper depicts the quality components of the individual studies, which allows the reader to perform sensitivity analyses according to other classifications.

*Suitability of using meta-analysis and meta-regression analysis*

The suitability of pooling the trial results in a meta-analysis when trials were conducted in different settings and with different patient categories is briefly discussed in the paper and will not be repeated here. Ideally, we should have obtained individual patient data from each trial, as this would have allowed trial participants to be directly compared with others in the same study, instead of the whole dataset being pooled as though it came from a single, homogeneous study. Obtaining individual patient data would also have permitted use of a time-to-event analysis, as well as more powerful subgroup analyses with less risk of confounding. However, judging from the responses we received to our questions to the trial authors on quality, it seems highly unlikely that we would have receive IPD from more than a couple of principal investigators at most.

An alternative could have been not to pool data in a meta-analysis, but to simply state that there are no high-quality trials to support the use of immunoglobulin for treatment of sepsis and that a large high-quality trial not yet fully published did not find any beneficial effect. But without a meta-analysis we would not have been able to assess statistically whether there were other or better explanations than methodological quality of the heterogeneity between the results of the individual trials.

Typical competing explanations offered in the literature are that the use of certain immunoglobulin preparations offers an efficient treatment of sepsis in certain patients, etc. For instance, that immunoglobulin may well be effective for treatment of sepsis in neonates, but not in adults. We assessed whether such claims offered better explanations of the between-trial heterogeneity than methodological quality in a meta-regression analysis. However, theoretically we could have employed within trial subgroup analyses and pooled these in a subsequent meta-analysis, instead. The latter approach would have permitted comparison within rather than between trials. However, firstly, the trials did not report mortality in separate subgroups (for instance

patients with gram-positive vs. gram-negative sepsis). Secondly, even if such data had been available, we would not have been able to simultaneously take into account other potential confounding covariates. Thirdly, most of the potential explanations of heterogeneity in the treatment effect were only available on between-trial level, *viz*. neonates vs. non-neonates. One limitation of meta-regression in the present context is that comparisons occur between trials rather than within trials. Another limitation is that its power is limited by the number of observations (21 trials). Finally, we might have been accused of data-dredging if we had not pre-defined methodological quality as our primary hypothesis for explaining the between-trial heterogeneity. [62] Regression analysis in general rests on the assumption that the tested variables are additive and that the influence of quantitative covariates is linear. [63] We had no prior suspicion of interactions between the covariates, so additivity was not explored, and the data set was too small with too limited a spread in the variable "total dose per kg" to reliably allow detection of deviations from linearity. The apparent correlation between study size and study quality is briefly discussed in the paper. We used control group event rate to explore whether base-line risk explained a statistically significant amount of heterogeneity. If this had been the case then interpretation would have been difficult, see paper. [64 65]

**Justification of conclusions**

Regardless of whether the assumptions that meta-analysis and meta-regression rest upon were fulfilled, and regardless of whether authors' replies to our questions were reliable, the conclusion still is that there are no large, high quality trials to support use of immunoglobulin for treatment of sepsis outside the setting of randomised clinical trials.

**Contribution to the current knowledge**

We have shown that high-quality RCTs on polyclonal immunoglobulin as adjunctive treatment for bacterial sepsis failed to show a reduction in mortality. Another systematic review that addressed the same question found that high-quality trials did indeed show a reduction in mortality.[66] The reasons for this discrepancy are that we applied a more rigorous assessment of trial quality, applied a more sensitive search strategy and retrieved more information from trial authors. Ohlsson et al. also reported a statistically significant reduction in mortality among neonates, but did not report sensitivity analyses of the impact of methodological quality.[67] While Ohlsson et al. concluded cautiously that there was insufficient data to support routine use of immunoglobulin for treatment of sepsis in neonates, we draw a stronger conclusion and say that it should not be used at all outside the context of clinical trials. Furthermore, we provide a thorough exploration of other explanations of why trials on the effect of immunoglobulins for treatment of sepsis come to different results, and show that currently there is no evidence of other explanations than the trials' susceptibility to bias. This finding should prevent spurring of use of immuglobulin for treatment of sepsis driven by undue emphasis on selected groups of low quality trials. Off-label drug prescription is the prescription of a drug for diseases other than those for which the drug-regulatory authorities have approved it. While this use is legal, no authorities have assessed the evidence for prescribing the drug for other conditions, leaving it to the individual clinician to assess whether the evidence is strong enough to do so. When clinicians are faced with situations such as neonatal sepsis, they will be under pressure to "do something", which might explain why most of the immunoglobulin used in the United States is used off-label.[68] An example of how the undue emphasis on a subgroup of low quality trials might encourage off-label use of immunoglobulin for treatment of sepsis is provided by Jenson and Pollock.[69] They conducted a meta-analysis of three of the trials on neonatal sepsis, which we also consider in our review. [70-72] On the basis of two studies that we had to rate as lower-

quality studies [70] [71] and one of the trials that we had to exclude, because of fundamental design errors,[72] they claim that "the additive benefit of IVIG (intravenous immunoglobulin) given to neonates with sepsis in decreasing acute mortality is clearly unequivocal and substantial. Neonates with sepsis not afforded this therapy suffer a nearly six-fold higher short-term mortality rate."

Interestingly, four of the trials included in our review were published after the publication of the abstract of the study by Werdan et al., but failed to discuss or mention it. So it is important to draw attention to the appalling contrast between the high number of fully published lower-quality trials that show an overall favourable effect and the failure of the large high-quality negative trial to become published in full. This also emphasizes that despite the Helsinki Declaration's insistence on publication of trials involving humans, it currently remains unspecified what level of publication will suffice.[73] Clearly, publication of an abstract (which will not be indexed in any of the major medical databases) is not satisfactory. Werdan wrote in 1996 that the trial would soon be published in full. [74] This is now nine years ago, and this example calls for a revision of the Helsinki Declaration to prevent infinite postponement of publication of trials.

## Discussion of Study II

### Comparison of descriptions of allocation concealment in trial protocols and the published reports: cohort study

### Major findings

This study was a cohort study comparing descriptions of allocation concealment in published reports of randomised controlled trials (RCTs) and their protocols. [75] As allocation concealment appears to be one of the most important measures of protection against bias, [1 15-17 19 76 77] it is crucial to gain knowledge on whether and how allocation concealment is achieved in trials. Trial protocols are of particular interest with regard to the vast majority of trials with unclear allocation concealment in the published report.[34] Among such RCTs, we aimed to determine the fraction that represents unclear reports of adequate allocation concealment as judged by their protocol. When we applied strict operationalisation of the criteria for adequate allocation concealment, the protocols made it clear that only 16% of the trials had adequate concealment, rising to 40% in a sensitivity analysis where we applied a more lax operationalisation of the criteria. Moreover, regardless how laxly the criteria were applied, most of the protocols also provided unclear information or gave rise to additional concern that the allocation concealment might have been compromised (for example, by revealing the randomisation block size).

### Assessment of methods

The primary research question was what fraction of trials with unclear allocation concealment in the trial report had adequate allocation concealment according to their protocols. The implicit research question is slightly different, namely, does unclear allocation concealment in trial publications reflect poor methods or poor reporting of adequate methods?

*Choice of research material*

The merits and limitations of our cohort of pairs of published reports of RCTs and their protocols are addressed below in the discussion of the justification of our conclusions and of what our study adds to the current knowledge.

*Quality assessments: validity of operational criteria*

All the empirical studies of lack of reported adequate allocation concealment operationalised the concealment concept in slightly different ways.[1 15 16 17 19] In turn, authors of Cochrane reviews are requested to apply criteria that are again slightly different if interpreted literally.[78] For instance, vehicles not only have to be numbered and coded to provide adequate concealment but also to be serially administered. Additional theoretical considerations regarding what adequate allocation concealment entails have been suggested by Meinert [3] and Berger.[79] Our strict operationalisation of criteria for adequate allocation concealment corresponded to a literal interpretation of the recommendation for authors of Cochrane review, with an elaboration of the criteria for central randomisation incorporating suggestions by Meinert and Berger. This approach was chosen because all these elements are relevant to ensure adequate allocation concealment, and because we had expected that the protocols would give more detailed descriptions.

However, it became obvious that very few protocols and trial reports met these strict criteria, so despite their theoretical justification, they might not be fully relevant, partly because a certain level of implicit understanding is often assumed in writing, and partly because the empirical studies of the impact of lack of reported allocation concealment had used laxer operationalisation. Therefore, we used the strict criteria for the presentation of the primary outcome because they were pre-specified; however, we also presented a sensitivity analysis applying the laxest criteria adopted

by any of the empirical studies of bias to capture the full range of how operationalisation of the same conceptual set of criteria could influence the results. The exact number of trials fulfilling the criteria for adequate concealment varied substantially depending on which set was applied. This illustrates the relevance of explicit priori operationalisation of the concept of bias protection components: firstly, to inform comparisons between different studies, and secondly, to avoid data dredging by ad hoc or retrospective definitions.

We do not know whether the association between inflated treatment effects and unclear or inadequate allocation concealment reflects selection bias. It might also be that lack of reported adequate allocation concealment is, instead, a marker associated with other unrecognised mechanisms of bias. This distinction would not be as important if the size and direction of the impact of lack of reported adequate concealment was predictable, and if adequacy of allocation concealment was always assessed in the same way, because the impact of lack of reported adequate concealment would then be the same regardless of the mechanism. However, as this is not the case, it is important to be aware that an apparently slight alteration in the criteria might substantially alter how studies are categorised and also alter a possible association with other determinants of bias in an unpredictable way.

*Empirical evidence of selection bias: baseline discrepancies*

We could have assessed whether lack of reported adequate allocation concealment in the trial report or the protocol, or both, was associated with an increased prevalence of baseline imbalances of important prognostic factors. This would have provided an indication of whether these trials were more frequently affected by selection bias than those with adequate concealment according to both sources. However, this presupposes, that reports on baseline imbalances are reliable.

**Justification of conclusions**

The main conclusion, that most trials with unclear allocation concealment in the published trial report also have unclear allocation concealment according the protocol, is straightforward and is robust to both the lax and stringent operationalisation of the concept of allocation concealment.

Our cohort is the most representative of the overall literature of all the other studies that have assessed the fraction of adequately concealed trials among trials with unclear allocation concealment by consulting other sources. Thus, it is reasonable to assume that it is also the most generalisable. Our main conclusion relies on the assumption that the trials were conducted as outlined in the protocol. We believe this is likely since there were very few (four) explicit discrepancies between the methods of allocation concealment stated in the protocols and those reported the trial.

An unanswered question is what to assume when both protocols and trial reports do not provide any clear account of attempts to achieve allocation concealment. It could be that investigators simply failed to detail the bias protection procedures to be adopted in the trial or it could be that they did not have an appropriate understanding of the purpose and means of providing allocation concealment and therefore gave no detail. Such a lack of understanding is implied when, for instance, unvaried and relatively small randomisation block sizes are described in the protocol. This, however, is speculative.

*The implications for practice*

The call for drug regulatory authorities and science ethics committee to condition sanctioning of protocols on described adequate allocation concealment hardly needs further justification than the mere documentation of the present lack of it. The responsibility of legislators and policy makers to demand this and provide the authorities/committees with the resources should not be ignored, since

these institutions are already under pressure to review too much too quickly.[80] Implementing this request will reduce the conducting of trials with selection bias due to lack of understanding of the importance and means of achieving allocation concealment. Currently, the ICH-CGP rules require investigators to account for how the treatment is going to be allocated to patients and to describe how bias will be prevented in the trial protocols.[81] The guideline from the Danish drug-regulatory authority that was valid when the protocols included in the present study were approved gave very similar instructions.[82] So there is no reason to believe that the recent national endorsement of the updated ICH-GCP rules in Europe has already solved the problem.

We conclude that our findings add to the argument for public access to protocols, which seems difficult to contradict. The financial interests of companies can easily be protected. Any descriptions of the pharmaceutical content of drugs or the production of them have no relevance to the conduct of the trial per se, and we did not encounter any such descriptions in our cohort of protocols. The research question itself may be innovative, but once the trial is launched it will, nevertheless, become publicly known. The costs and logistic effort of implementing public access could be minimised by requesting the protocols to be submitted electronically; the approved version could then simply be available on a central public Internet site. Another potential merit of requiring access to protocols is that the enforcement of the CONSORT statement [35] might tempt authors to overstate the quality of their trial, but if the protocol is available, this can then be detected.

**Contribution to the current knowledge**

This study is the first to reveal the lack of description of the most important bias prevention design measure, allocation concealment, in a representative cohort of trial protocols.

Furthermore, it adds to the other studies that have estimated the fraction of trial reports with unclear allocation concealment that have adequate concealment according to other sources.[83-86] All

but one,[86] of these studies found that most trials with unclear allocation concealment according to the trial report had adequate allocation concealment according to other sources. Our study adds to these findings. Firstly, because it does not depend on the reliability of surveying investigators on their published trials' quality, such as the case was for three of the abovementioned studies,[83 84 86] The reliability of surveying investigators on another validity issue (reporting bias) has recently been assessed in two studies, where 86% (42/49) and 80% (28/35) of investigators denied the existence of unreported outcomes, although there was evidence to the contrary in their study protocols.[54 55] Secondly, our cohort is the most representative of the overall literature and consequently the more generalisable of the other four studies.[83 84 86 85] This is particularly relevant with regard to the study by Soares et al., which was also a study on pairs of protocols and trials. Soares et al. found that unclear allocation concealment in the published trial reports invariably reflected adequate allocation concealment according to the trials' protocol; however, the cohort was a highly selected sample. Our own finding, that 16% to 40% of trials with unclear allocation concealment in the trial report had adequate allocation concealment according to their protocols, is limited by the lack of clear descriptions of allocation concealment in the rest of the protocols of these trials (except in one protocol where concealment was inadequate). Although protocols with unclear allocation concealment might be expected when allocation concealment is inadequate, we do not know to what extent this was the case.

In addition, our study is the first to quantify the impact of how the concept of allocation concealment is operationalised. Also, we report the prevalence of, hitherto little heeded, threats to allocation concealment other than those addressed by the commonly used criteria for adequacy of concealment,[1 15-17 19] for instance, the frequent presence of code envelopes (for security purposes) at the trial location.

Finally, our study raises the question of whether central randomisation is overrated when it is assumed to always provide adequate allocation concealment, since none of the trials employing this method fulfilled our strict criteria for adequate concealment. Conversely, we found that only 3 of 26 trials using the adequate method of numbered coded vehicles according the protocol described this method clearly in the trial report. This raises the question of whether allocation concealment in this type of trials might frequently be underrated, since double-blind drug trials often use numbered coded vehicles for concealment.

## Discussion of Study III

## Impact of allocation concealment on conclusions drawn from meta-analyses of randomised trials

### Major findings

Two-thirds of conclusions holding preferences based on a meta-analysis are no longer supported if only trials with reported adequate allocation concealment are relied upon. This is partly due to a substantial loss of statistic power and partly (at least in our data-set) to a tendency for estimates of treatment effects derived from trials reporting adequate concealment to be less beneficial than those that do not.

### Assessment of methods

*Operationalisation of the primary aim of the study*

The conceptual aim was to determine the fraction of RCT-supported clinical interventions that would remain supported if the requirement for the reliability of the evidence were enhanced to also include reported adequate allocation concealment. This question was operationalised as follows. What fraction of conclusions holding a preference based on a statistically significant result of a meta-analysis of RCTs remained supported if only RCTs with reported adequate allocation concealment were included in "re-meta-analyses" using the original method of analysis and summary statistic when possible. Lack of 2 x 2 data and imprecise methods sections in the reviews sometimes made this impossible. In those cases, generic inverse variance analyses based on the point estimates and their standard deviations as reported in the review were used instead. As detailed in the manuscript, we were able to closely replicate the original analyses; accordingly, we trust that the subgroup re-analyses appropriately reflect the desired outcome.

Operationalisation of when a conclusion holds a preference was defined as lending more support to one treatment than another. This could be as vague as "treatment effects were about equal, but one had minor advantages" or as strong as "One of the interventions is highly preferred and should be considered the standard intervention in all similar patients". We excluded reviews where the authors stated a conclusion compatible with the following categories: "There is not enough evidence to justify any preference" or "There is evidence of no clinically relevant difference" or "Whether one prefers one intervention to another depends on values which will determine the trade off between the benefits and drawbacks of the interventions in question". We used double independent assessment of the fulfilment of this criterion and disagreements led to exclusion of the review. This was to minimise any subjective judgement. We believe that this adequately discriminates between reviews that will influence clinicians to generally prefer one intervention to another and those that will not influence clinical practice in one particular direction.

Only one meta-analysis from each review was included and this was identified as follows: a) the first statistically significant meta-analysis result presented in the results section of the abstract identified the potentially eligible meta-analysis. b) If this analysis concerned a binary outcome and supported the preference stated in the conclusion it was then accepted. If not, the review was excluded. This approach was chosen because the our pilot studies had shown that "positive" conclusions do not necessarily refer to the primary outcome (which is often not defined), nor the largest meta-analysis of the review, but rather to what the authors perceived as the most important finding, which tended to be the first reported statistically significant finding. Only three reviews had to be excluded, because the conclusion could not be assumed to rest on the meta-analysis identified in step a.

*Relevance of obtaining all the original reports of the included meta-analyses*

All the original trial reports of RCTs included in meta-analyses were retrieved. Partly because the authors of PubMed reviews rarely reported assessments of the adequacy of allocation concealment and partly because they infrequently reported the 2 x 2 data of the meta-analyses. Most authors of Cochrane reviews had assessed adequacy of allocation concealment and all reported the 2 x 2 data. However, the authors of the Cochrane reviews had a tendency to overrate the adequacy of allocation concealment and generally applied the Cochrane handbook criteria[78] for adequacy inconsistently and sometimes incorrectly (Appendix 1)

*Choice of method for estimating the impact of absence vs. presence of reported adequate bias protection components on the treatment effect*

To estimate the impact of absence vs. presence of reported adequate allocation concealment on the treatment effect we had to define what a treatment effect was in order to achieve consistent comparisons between meta-analyse. We defined the treatment effect as that of the experimental compared to the control treatment as follows: if an active intervention was compared with placebo or with no intervention, the active treatment was then experimental. If the authors of the review referred to one of the interventions as experimental or new, this distinction was then used. Two authors made this assessment independently and any disagreement led to exclusion of the review. The relevance of seeking increased objectivity by requiring independent interrater agreement was illustrated by our exclusion of 4 of 103 reviews at step 4 (Fig 1.a and 1.b in the manuscript).

As noted in the manuscript, several different strategies for estimating the impact of absence vs. presence of reported adequate bias protection components on the treatment effect have been employed in previous comparable studies. For instance, Schulz et al. [1] and Moher et al [15] used a logistic regression model and assumed a fixed effect between and within the meta-analyses, while

the analysis applied in the present study assumed a random effect at both levels. We used a random effects model at the within-meta-analysis level because many of the included meta-analyses originally used this model. This probably led to an overestimation of the between-trial heterogeneity. We abstained from trying to control for confounding between allocation concealment and double-blinding. If more and larger meta-analyses had been available we would have estimated the effect of each component by controlling for confounding using meta-regression in each meta-analysis, and then combined these estimates in two separate meta-analyses the same way as was done for the uncontrolled estimates.

**Justification of conclusions**

That two-thirds of conclusions based on meta-analyses were no longer supported by the data if only trials with reported adequate allocation concealment were included, was predictably partly due to loss of statistical power. We did not try to disentangle how much of the loss of support for conclusions was due to loss of power rather than a less beneficial point-estimated represented by the trials included compared to those excluded. Because it is the compound influence of both mechanisms that comprise the relevant answer to the conceptual question.

Two major inferences are drawn based on this conclusion: one with regard to prevention, the other with regard to the interpretation of the vast majority of available evidence with an unclear level of bias susceptibility. It is obvious that uncertainty with regard to the level of bias protection should be prevented, and the ways in which this can be done have already been detailed in the manuscript. How to deal with and interpret the vast majority of the available evidence is more controversial. We suggest that sensitivity analyses that will inform the reader of the findings, depending on the level of bias protection, should always be presented. Since this could be done in several ways, it will be prone to data-dredging, and a strategy for assessing the impact of bias

susceptibility should be pre-specified in a protocol. Additional explorative analyses could be presented provided that they are labelled as such. But a conclusion has to be drawn as to whether or not to recommend the investigated intervention should be implemented. Here we call for more research to guide this assessment. It is hardly realistic to expect that a number of "correction factors" for simple application can be derived to solve the problem of uncertainty with regard to size, direction and variability associated with unclear reports of bias. But more research will probably increase our understanding of the circumstances that might influence the impact of different sources of bias under different conditions.

**Contribution to the current knowledge**

The primary aim of this study was to provide a documentation of the scale of the problem we are faced with because of the poor reporting on the level of bias protection in RCTs and to do so by illustrating the problem in relation to the clinically relevant situation, namely the clinical decisions. The proportion of interventions in internal medicine supported by RCTs has previously been estimated.[87] Our study further details what fraction of these RCT-supported interventions that remains supported if reported adequate allocation concealment is required, and to our knowledge we are the first to do so. That only 31% of conclusions remained supported is consistent with the high prevalence of trial reports with unclear or inadequate allocation concealment.[34] But this does not imply that our result was predictable, especially since review authors might have been more cautious when drawing conclusions based on trials with an uncertain or high level of bias susceptibility.

Our estimate of the impact of unclear or inadequate allocation concealment was less than we had expected. Our finding of a ratio of odds ratios of 0.92 when trials with absence vs. presence of reported adequate allocation concealment were compared was not statistically significant and it was

closer to 1 than any of the corresponding estimates of the first four similar studies. A more recent study, by Balk et al, found a ROR of 0.95 (0.83 to 1.09) However, this result was questioned because one of the inclusion criteria was statistically significant heterogeneity between the estimated treatment effects reported in the trials of each meta-analyses, which would probably introduce too much noise to allow detection of the full effect of lack of reported adequate concealment.[76] But this explanation does not apply to our study nor to a sixth study published as an abstract, which found a ROR of 1.02 (0.90 to 1.16).[20] As opposed to the study by Als-Nielsen et al. half of the studies that contributed to our ROR estimate were derived from PubMed reviews, which makes our results more broadly representative. In addition, we used exactly the same criteria for adequacy of allocation concealment as Schulz et al., which is important, because subtle differences in operationalisation of the same conceptual criteria can make a large difference as to how trials are categorised, as illustrated in Study II and Appendix 1.

## Conclusions and implications for practice and research

Sound theoretical considerations justify the precautions recommended to avoid bias at clinical trial and at review level.[17] Empirical evidence suggests that a failure to implement these precautions,[1] and a failure to report whether they were employed is associated with bias in favour of the experimental study.[1 15-17] One of the most imperative of these components is allocation concealment. In Study III we showed the importance of reporting this, since two-thirds of the conclusions stating a preference lose support if only trials with reported adequate allocation concealment are relied upon. We also estimated the inflation of the effect of the experimental treatment associated with inadequate or unclear allocation concealment in our data-set. We found less inflation than that of the first four similar studies.[1 15-17] A fifth comparable study became available during Study III.[19] This study by Balk et al. reported results that were consistent with ours, but might be explained as an exceptional finding due to a particular inclusion criterion; however, this does not explain our finding. Thus, the determinants of the size and direction of bias associated with unclear reports have to be further investigated. This would inform our assessment of the likely impact of bias in the individual meta-analyses. Such assessments will involve some subjective judgment. The readers should consequently be presented with sensitivity analyses of the possible impact of bias.

Study II offers three possible explanations of why the studies of the impact of lack of reported adequate allocation concealment reach different results, [1 15-17 19 20] since the following phenomena might have confounded the studies in various degrees. First, a certain fraction of trials will have adequate allocation concealment but fail to report so. This fraction was 16% to 40% according to the corresponding trial protocols. Second, the quoted 16% to 40% represents our quantification of how differently trials may become classified according to a seemingly simple set

of criteria, depending on whether these are perceived as strict ideals that have to be fulfilled in detail, or whether they are perceived as examples of concepts to be applied more laxly. Third, other mechanisms for undermining adequate allocation concealment than those assessed in previous similar studies may be important, for instance the disclosure of small fixed block sizes in the protocol (Additional explanations are offered in the discussion section of Study III).

Study II is the first to disclose that the majority of broadly representative trial protocols lack description of the most important bias protection precaution, allocation concealment. Thus, the regulatory authorities should be given the resources and incentive to prevent conduct of biased trials and to minimise uncertainty of the true effect of the interventions investigated in trials. That is, these authorities should enforce that trial protocols are approved only if they contain explicit description of adequate bias protection precautions. Furthermore, trial protocols should be made publicly available to strengthen critical appraisal the reliability of clinical trials. Another implication of our quantification of how differently trials can be categorised depending on how the criteria for adequate allocation concealment are applied, is to avoid data-dredging. Accordingly, the sensitivity analyses to assess the impact of bias in meta-analyses should be pre-specified in the protocol of the review and any additional exploratory analyses should be clearly labelled as such.

Finally, the risk of introducing bias or uncritically summarising results from trials with an uncertain or low bias protection in review articles also deserves attention. We illustrated this in Study I which is a systematic review including a meta-analysis of whether polyclonal immunoglobulin as adjunctive treatment for sepsis reduces mortality. A pivotal contribution was that of a large trial yet to be fully published, and the overall meta-analysis result of the hinged on the bias susceptibility of individual trials. Evidence from trials with a high level of bias protection failed to show a reduction in mortality. Assessment of the level of bias protection was conducted according to a pre-specified strategy. Additional exploratory analyses demonstrated that there was

no convincing evidence of other explanations of the differences between the results of the included trials, than their susceptibility to bias. Accordingly, any claim that certain subgroups of patients will benefit from the treatment remains to be proven in large high quality trials, and polyclonal immunoglobulin should not be used for treatment of sepsis outside the setting of such trials.

## References


1. Schulz KF, Chalmers I, Hayes RJ, Altman DG. Empirical evidence of bias. Dimensions of methodological quality associated with estimates of treatment effects in controlled trials. JAMA 1995; 273(5):408-12.

2. Pocock SJ. Clinical trials, a practical approach. 1st edition. Chichester: Wiley, 1983.

3. Meinert C, Author. Clinical Trials. Design, Conduct and Analysis. 1st edition. Oxford: Oxford University Press, 1986.

4. Friedman LM, Furberg CD, DeMets DL. 3rd edition. St. Louis: Mosby, 1996.

5. Hróbjartsson A, Gøtzsche PC. Is the placebo powerless? An analysis of clinical trials comparing placebo with no treatment. N Engl J Med 2001; 344(21):1594-602.

6. Devereaux PJ, Manns BJ, Ghali WA *et al*. Physician interpretations and textbook definitions of blinding terminology in randomized controlled trials. JAMA 2001; 285(15):2000-3.

7. Schulz KF, Chalmers I, Altman DG. The landscape and lexicon of blinding in randomized trials. Ann Intern Med 2002; 136(3):254-9.

8. Montori VM, Bhandari M, Devereaux PJ, Manns BJ, Ghali WA, Guyatt GH. In the dark: the reporting of blinding status in randomized controlled trials. J Clin Epidemiol 2002; 55(8):787-90.

9. Fergusson D, Glass KC, Waring D, Shapiro S. Turning a blind eye: the success of blinding reported in a random sample of randomised, placebo controlled trials. BMJ 2004; 328(7437):432.

10. Hollis S, Campbell F. What is meant by intention to treat analysis? Survey of published randomised controlled trials. BMJ 1999; 319(7211):670-4.

11. Easterbrook PJ, Berlin JA, Gopalan R, Matthews DR. Publication bias in clinical research. Lancet 1991; 337(8746):867-72.

12. Dickersin K, Min YI, Meinert CL. Factors influencing publication of research results. Follow-up of applications submitted to two institutional review boards. JAMA 1992; 267(3):374-8.

13. Stern JM, Simes RJ. Publication bias: evidence of delayed publication in a cohort study of clinical research projects. BMJ 1997; 315(7109):640-5.

14. Egger M, Davey Smith G, Altman D, Editors. Systematic reviews in Health Care, Meta-analysis in Context. 2nd edition. London: BMJ Books, 2001.

15. 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(9128):609-13.

16. Kjaergard LL, Villumsen J, Gluud C. Reported methodologic quality and discrepancies between large and small randomized trials in meta-analyses. Ann Intern Med 2001; 135(11):982-9.

17. Jüni P, Altman DG, Egger M. Systematic reviews in health care: Assessing the quality of controlled clinical trials. BMJ 2001; 323(7303):42-6.

18. Yusuf S, Collins R, Peto R. Why do we need some large, simple randomized trials? Stat Med 1984; 3(4):409-22.

19. Balk EM, Bonis PA, Moskowitz H *et al*. Correlation of quality measures with estimates of treatment effect in meta-analyses of randomized controlled trials. JAMA 2002; 287(22):2973-82.

20. Als-Nielsen B, Chen W, Gluud LL, Sierma V, Hilden J, Gluud C. Are trial size and reported methodological quality associated with treatment effects? Program & Abstract Book, 12th Cochrane Colloquium 2004; 102.

21. Schulz KF, Grimes DA. Allocation concealment in randomised trials: defending against deciphering. Lancet 2002; 359(9306):614-8.

22. Moher D, Jones A, Lepage L. Use of the CONSORT statement and quality of reports of randomized trials: a comparative before-and-after evaluation. JAMA 2001; 285(15):1992-5.

23. Adetugbo K, Williams H. How well are randomized controlled trials reported in the dermatology literature? Arch Dermatol 2000; 136(3):381-5.

24. Latronico N, Botteri M, Minelli C, Zanotti C, Bertolini G, Candiani A. Quality of reporting of randomised controlled trials in the intensive care literature. A systematic analysis of papers published in Intensive Care Medicine over 26 years. Intensive Care Med 2002; 28(9):1316-23.

25. Gummesson C, Atroshi I, Ekdahl C. The quality of reporting and outcome measures in randomized clinical trials related to upper-extremity disorders. J Hand Surg [Am] 2004; 29(4):727-34; discussion 735-7.

26. Strippoli GF, Craig JC, Schena FP. The number, quality, and coverage of randomized controlled trials in nephrology. J Am Soc Nephrol 2004; 15(2):411-9.

27. Harrison JE. Clinical trials in orthodontics II: assessment of the quality of reporting of clinical trials published in three orthodontic journals between 1989 and 1998. J Orthod 2003; 30(4):309-15; discussion 297-8.

28. Karassa FB, Tatsioni A, Ioannidis JP. Design, quality, and bias in randomized controlled trials of systemic lupus erythematosus. J Rheumatol 2003; 30(5):979-84.

29. Kyriakidi M, Ioannidis JP. Design and quality considerations for randomized controlled trials in systemic sclerosis. Arthritis Rheum 2002; 47(1):73-81.

30. Kjaergard LL, Frederiksen SL, Gluud C. Validity of randomized clinical trials in gastroenterology from 1964-2000. Gastroenterology 2002; 122(4):1157-60.

31. Kjaergard LL, Nikolova D, Gluud C. Randomized clinical trials in HEPATOLOGY: predictors of quality. Hepatology 1999; 30(5):1134-8.

32. Gluud C, Nikolova D. Quality assessment of reports on clinical trials in the Journal of Hepatology. J Hepatol 1998; 29(2):321-7.

33. Begg C, Cho M, Eastwood S *et al*. Improving the quality of reporting of randomized controlled trials. The CONSORT statement. JAMA 1996; 276(8):637-9.

34. Altman DG, Schulz KF, Moher D *et al*. The revised CONSORT statement for reporting randomized trials: explanation and elaboration. Ann Intern Med 2001; 134(8):663-94.

35. Ioannidis JP, Evans SJ, Gotzsche PC *et al*. Better reporting of harms in randomized trials: an extension of the CONSORT statement. Ann Intern Med 2004; 141(10):781-8.

36. Devereaux PJ, Manns BJ, Ghali WA, Quan H, Guyatt GH. The reporting of methodological factors in randomized controlled trials and the association with a journal policy to promote adherence to the Consolidated Standards of Reporting Trials (CONSORT) checklist. Control Clin Trials 2002; 23(4):380-8.

37. Davidoff F. News from the International Committee of Medical Journal Editors. Ann Intern Med 2000; 133(3):229-31.

38. The consort-statement [Web Page]. Available at http://www.consort-statement.org/. (Accessed 9 March 2005).

39. Hardy RJ, Thompson SG. Detecting and describing heterogeneity in meta-analysis. Stat Med 1998; 17(8):841-56.

40. Egger M, Juni P, Bartlett C, Holenstein F, Sterne J. How important are comprehensive literature searches and the assessment of trial quality in systematic reviews? Empirical study. Health Technol Assess 2003; 7(1):1-76.

41. Tierney JF, Stewart LA. Investigating patient exclusion bias in meta-analysis. Int J Epidemiol 2005; 34(1):79-87.

42. McAuley L, Pham B, Tugwell P, Moher D. Does the inclusion of grey literature influence estimates of intervention effectiveness reported in meta-analyses? Lancet 2000; 356(9237):1228-31.

43. Egger M, Ebrahim S, Smith GD. Where now for meta-analysis? Int J Epidemiol 2002; 31(1):1-5.

44. Yakut M, Cetiner S, Akin A *et al*. Effects of immunuglobulin G on surgical sepsis and septic shock. BULL-GULHANE-MIL-MED-ACAD. Bulletin-of-Gulhane-Military-Medical-Academy. 1998; 40/1 (76-81) 1998.

45. Berg A, Laguenie G, Goutail F, Couturier C, Saint-Maurice C. [Venoglobulins in pediatric surgical resuscitation]. Cah Anesthesiol 1989; 37(7):511-5.

46. Just HM, Metzger M, Vogel W, Pelka RB. [Effect of adjuvant immunoglobulin therapy on infections in patients in an surgical intensive care unit. Results of a randomized controlled study]. Klin Wochenschr 1986; 64(6):245-56.

47. Wesoly C, Kipping N, Grundmann R. [Immunoglobulin therapy of postoperative sepsis]. Z Exp Chir Transplant Kunstliche Organe 1990; 23(4):213-6.

48. Masaoka T. [Combination therapy of antibiotics and intravenous immunoglobulin]. Nippon Rinsho 2001; 59(4):781-4.

49. Juni P, Holenstein F, Sterne J, Bartlett C, Egger M. Direction and impact of language bias in meta-analyses of controlled trials: empirical study. Int J Epidemiol 2002; 31(1):115-23.

50. Moher D, Fortin P, Jadad AR *et al*. Completeness of reporting of trials published in languages other than English: implications for conduct and reporting of systematic reviews. Lancet 1996; 347(8998):363-6.

51. Jadad AR, Moore RA, Carroll D *et al*. Assessing the quality of reports of randomized clinical trials: is blinding necessary? Control Clin Trials 1996; 17(1):1-12.

52. Behre G, Ostermann H, Schedel I *et al*. Endotoxin concentrations and therapy with polyclonal IgM-enriched immunoglobulins in neutropenic cancer patients with sepsis syndrome: Pilot study and interim analysis of a randomized trial. ANTIINFECT-DRUGS-CHEMOTHER. Antiinfective-Drugs-and-Chemotherapy. 1995; 13/2 (129-134) 1995.

53. Dominioni L, Bianchi V, Imperatori A, Minoia G, Dionigi R. High-dose intravenous IgG for treatment of severe surgical infections. DIG-SURG. Digestive-Surgery. 1996; 13/4-5 (430-434) 1996.

54. Chan AW, Hróbjartsson A, Haahr MT, Gøtzsche PC, Altman DG. Empirical evidence for selective reporting of outcomes in randomized trials: comparison of protocols to published articles. JAMA 2004; 291(20):2457-65.

55. Chan AW, Krleza-Jeric K, Schmid I, Altman DG. Outcome reporting bias in randomized trials funded by the Canadian Institutes of Health Research. CMAJ 2004; 171(7):735-40.

56. Werdan K, Pilz GatSSG. Polyvalent immune globulins [abstract]. Shock 1997; 7:Suppl: 5.

57. Pilz G, Fateh Moghadam S, Viell B *et al*. Supplemental immunoglobulin therapy in sepsis and septic shock - comparison of mortality under treatment with polyvalent i.v. immunoglobulin versus placebo: Protocol of a multicenter, randomized, prospective, double-blind trial. THEOR-SURG. Theoretical-Surgery. 1993; 8/2 (61-83) 1993.

58. Kondro W, Sibbald B. Drug company experts advised staff to withhold data about SSRI use in children. CMAJ 2004; 170(5):783.

59. Egger M, Juni P, Bartlett C, Holenstein F, Sterne J. How important are comprehensive literature searches and the assessment of trial quality in systematic reviews? Empirical study. Health Technol Assess 2003; 7(1):1-76.

60. Rennie D. Thyroid storm. JAMA 1997; 277(15):1238-43.

61. Lexchin J, Bero LA, Djulbegovic B, Clark O. Pharmaceutical industry sponsorship and research outcome and quality: systematic review. BMJ 2003; 326(7400):1167-70.

62. Thompson SG, Higgins JP. Treating individuals 4: can meta-analysis help target interventions at individuals most likely to benefit? Lancet 2005; 365(9456):341-6.

63. Altman DG. Practical statistics for medical research. 1st edition. London: Chapman and Hall, 1991.

64. Sharp SJ, Thompson SG, Altman DG. The relation between treatment benefit and underlying risk in meta-analysis. BMJ 1996; 313(7059):735-8.

65. Sharp SJ, Thompson SG. Analysing the relationship between treatment effect and underlying risk in meta-analysis: comparison and development of approaches. Stat Med 2000; 19(23):3251-74.

66. Alejandria MM, Lansang MA, Dans LF, Mantaring JB. Intravenous immunoglobulin for treating sepsis and septic shock (Cochrane Review). In: *The Cochrane Library*, Issue 1, 2002. Oxford: Update Software .

67. Ohlsson A, Lacy JB. Intravenous immunoglobulin for suspected or subsequently proven infection in neonates (Cochrane Review). In: *The Cochrane Library*, Issue 1, 2004. Oxford: Update Software .

68. Pierce LR, Jain N. Risks associated with the use of intravenous immunoglobulin. Transfus Med Rev 2003; 17(4):241-51.

69. Jenson HB, Pollock BH. Meta-analyses of the effectiveness of intravenous immune globulin for prevention and treatment of neonatal sepsis. Pediatrics 1997; 99(2):E2.

70. Haque KN, Zaidi MH, Bahakim H. IgM-enriched intravenous immunoglobulin therapy in neonatal sepsis. Am J Dis Child 1988; 142(12):1293-6.

71. Sidiropoulos D, Bohme U, von Muralt G, Morell A, Barandun S. [Immunoglobulin substitution in the treatment of neonatal septicemia]. [German]. Schweizerische Medizinische Wochenschrift. Journal Suisse De Medecine 1981; 111:1649-55.

72. Weisman LE, Stoll BJ, Kueser TJ *et al*. Intravenous immune globulin therapy for early-onset sepsis in premature neonates. J Pediatr 1992; 121(3):434-43.

73. WORLD MEDICAL ASSOCIATION DECLARATION OF HELSINKI [Web Page]. Available at http://www.wma.net/e/policy/b3.htm. (Accessed 9 March 2005).

74. Werdan K, Pilz G. Supplemental immune globulins in sepsis: a critical appraisal. Clin Exp Immunol 1996; 104 Suppl 1:83-90.

75. Pildal J, Chan AW, Hróbjartsson A, Forfang E, Altman DG, Gøtzsche PG. Comparison of descriptions of allocation concealment in trial protocols and the published reports: cohort study. BMJ 2005; In press.

76. Jüni P, Egger M. Allocation concealment in clinical trials. JAMA 2002; 288(19):2407-8; author reply 2408-9.

77. Schulz KF, Altman DG, Moher D. Allocation concealment in clinical trials. JAMA 2002; 288(19):2406-7; author reply 2408-9.

78. Alderson P, Green S, Higgins JPT, editors. Assessment of study quality. Cochrane Reviewers' Handbook 4.2.2 [Updated, March 2004]; Section 6. In: The Cochrane Library, Issue 2, 2004. Chichester, UK: John Wiley & Sons, Ltd.

79. Berger VW, Bears JD. When can a clinical trial be called 'randomized'? Vaccine 2003; 21(5-6):468-72.

80. Ghersi D, Campbell EG, Pentz R, Cox Macpherson C. The future of institutional review boards. Lancet Oncol 2004; 5(5):325-9.

81. The ICH guidelines. The International Conference on Harmonisation of Technical Requirements for Registration of Pharmaceuticals for Human Use (ICH). [Web Page]. Available at http://www.ich.org/UrlGrpServer.jser?@_ID=276&@_TEMPLATE=254. (Accessed 9 May 2004).

82. Sundhedsstyrelsen. Kliniske afprøvninger af lægemidler. Vejledning om anmeldelse efter lov om lægemidler. Sundhedsstyrelsen, 1992.

83. Hill CL, LaValley MP, Felson DT. Discrepancy between published report and actual conduct of randomized clinical trials. J Clin Epidemiol 2002; 55(8):783-6.

84. Devereaux PJ, Choi PT, El-Dika S *et al*. An observational study found that authors of randomized controlled trials frequently use concealment of randomization and blinding, despite the failure to report these methods. J Clin Epidemiol 2004; 57(12):1232-6.

85. Soares HP, Daniels S, Kumar A *et al*. Bad reporting does not mean bad methods for randomised trials: observational study of randomised controlled trials performed by the Radiation Therapy Oncology Group. BMJ 2004; 328(7430):22-4.

86. Liberati A, Himel HN, Chalmers TC. A quality assessment of randomized control trials of primary treatment of breast cancer. J Clin Oncol 1986; 4(6):942-51.

87. Nordin-Johansson A, Asplund K. Randomized controlled trials and consensus as a basis for interventions in internal medicine. J Intern Med 2000; 247(1):94-104.

## Appendix 1

Summary of the 253 reports of 202 trials included in 38 the Cochrane reviews. The trial reports were retrieved and reassessed in duplication and blinded to the results section. Exact details on the methods employed for allocation concealment were extracted; for instance, it was extracted whether envelopes were 1) sealed, 2) opaque and 3) sequentially numbered.
The adequacy of allocation concealment described in the trial reports was categorized according to the Cochrane Handbook and our categorizations were compared to those of the review authors.

51 of the 253 trial reports were co-publications (None of them contained any additional information on allocation concealment compared to the main publication).

Among the remaining 202 trials, we categorized 96 trials differently than the reviewers for the following reasons:

1) Envelopes method: Review authors classified trials as reporting adequately concealment although not fulfilling all the criteria: opaque and sealed and sequentially numbered. ($n$ = 11)

2) Review authors classified trials as reporting adequate concealment, where we classified them as unclear. The difference was not due to review authors obtaining unpublished information or incorporating information from co-publications. Review authors did not give any specification of the methods, which could support their classification. ($n$ =15)

3) Review authors rated trials as reporting adequate concealment (where we classified them as unclear) while having obtained unpublished information from the trial authors, but without specifying what methods were employed. (*n* =6)

4) Review authors classified "Numbered coded vehicles" as adequate concealment, although administration of the vehicles was not described as sequential. *(n* = 11 trials) [2]

5) Review authors did not assess adequacy of allocation concealment, where our assessment were:
    a. Inadequate: (*n* = 5)
    b. Unclear: (*n* = 12)
    c. No information available: (*n* = 11)
    d. Adequate: (*n* = 4)

6) Review authors classified as unclear allocation concealment, although the trial reports did not even claim or even allude to randomisation (*n* = 3)
7) Review authors classified as unclear allocation concealment, although allocation concealment was inadequate (*n* = 3)
8) Review authors had obtained unpublished information and described the methods in sufficient detail to justify that their classification was different than ours (*n* = 10)
9) Different other reasons (*n* = 5)

[2] The criterion for "numbered coded vehicles", as specified in the Cochrane Handbook includes that the administration of the vehicles has to be sequential. Schulz et al. did not require explicit reports of sequential administration. The classification we used for study III was pre-specified to be that of Schulz et al. but we extracted detailed information that permitted classification according to both sets of criteria.