Bias in Clinical Intervention Research

Methodological studies of systematic errors in randomised trials and observational studies

Doctoral Dissertation

Lise Lotte Gluud

Faculty of Health Sciences

University of Copenhagen

2005

Denne afhandling er af Det Sundhedsvidenskabelige Fakultet ved Københavns Universitet antaget til offentligt at forsvares for den medicinske doktorgrad. København, den 19/8 2005.

Forsvaret finder sted kl 14^{30} på Rigshospitalet, Auditorium 93, Juliane Maries Vej 22, 2100 Kbh Ø.

ISBN 87-990924-0-9

Bias in Clinical Intervention Research

Methodological studies of systematic errors in randomised trials and observational studies

Doctoral Dissertation

Lise Lotte Gluud

Cochrane Hepato-Biliary Group, Copenhagen Trial Unit, Centre for Clinical Intervention Research, H:S Rigshospitalet, Copenhagen University Hospital, DK-2100 Copenhagen, Denmark.

Members of the assessment committee Professor, Head of Department Jørgen Rask-Madsen, MD, DrMedSci (Chairman), Associate Professor, Head of Department Susanne Keiding, MD, DrMedSci, and Professor, Head of Department, Christian Torp-Pedersen, MD, DrMedSci.

"Do not believe anything

- simply because it is spoken and rumoured by many,

- simply because it is found written in your religious books,

- merely on the authority of your teachers and elders.

But after observation and analysis, when you find that

anything agrees with reason and is conducive, then accept it."

Gautama Buddha (566-486 BC)

Contents

Acknowledgements

I am pleased to thank the following people who made this thesis possible. For his never-ending inspiring enthusiasm, continuous work efforts, and countless inspiring discussions, I thank my husband Christian Gluud. I also thank my co-authors of included papers Bodil Als- Nielsen, Wendong Chen, Sarah Klingenberg, Kim Krogsgaard, Jianping Liu, Dimitrinka Nikolova, and John Villumsen. I thank Peter Gøtzsche at the Nordic Cochrane Centre for his invaluable help. The staff at Copenhagen Trial Unit, Centre for Clinical Intervention Research, and the members of the Cochrane Hepato-Biliary Group, Editorial Team are thanked for their assistance. For help in tracking down references and IT support I thank Ninna Frydendal and Nader Salas. For financial support I thank The Danish Medical Research Council, The 1991 Pharmacy Foundation, The Copenhagen Hospital Corporation Medical Research Council, Danish Centre for Evaluation and Health Technology Assessment, and The Copenhagen Trial Unit, Centre for Clinical Intervention Research.

This thesis is based on the following nine papers

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

Kjaergard LL, Krogsgaard K, Gluud C. Interferon alfa with or without ribavirin for chronic hepatitis C: systematic review of randomised trials. BMJ 2001;323:1151-5.

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.

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

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

Kjaergard LL, Gluud C. Citation bias of hepato-biliary randomized clinical trials. J Clin Epidemiol 2002;55:407-10.

Kjaergard LL, Als-Nielsen B. Association between competing interests and authors' conclusions: epidemiological study of randomised clinical trials published in the BMJ. BMJ 2002;325:249-52.

Kjaergard LL, Liu J, Als-Nielsen B, Gluud C. Artificial and bioartificial support systems for acute and acute-on-chronic liver failure: a systematic review. JAMA 2003;289:217-22.

Als-Nielsen B, Chen W, Gluud C, Kjaergard LL. Association of funding and conclusions in randomized drug trials: a reflection of treatment effect or adverse events? JAMA 2003;290:921-8.

Summary

The objective of the present thesis is to summarise evidence on factors that may lead to bias in clinical intervention research. The thesis is primarily based on nine previously published cohort studies and systematic reviews of observational studies and randomised trials. Observational studies tend to exaggerate intervention benefits compared to randomised trials. Small trials without adequate randomisation or double blinding tend to overestimate intervention benefits compared to large gold standard trials. This is not the case for small trials with adequate randomisation or double blinding. Unfortunately, only about 37% of randomised trials report adequate generation of allocation sequence and about 25% report adequate allocation concealment. Several trials were performed without double blinding although this would have been feasible. Furthermore, most randomised trials are small and lack sample size calculations. The small size suggests a considerable risk of false negative or false positive results. Financial competing interests may also be associated with bias. On average, conclusions in randomised trials tend to be significantly more favourable towards experimental interventions if trials received funding from a forprofit organisation. The quantitative estimates of intervention benefits and the occurrence of adverse events do not seem to explain the association between funding and conclusions. None of the factors that may lead to bias can predict the extent or direction of bias in individual trials. Different research questions therefore warrant individual evaluations. The combined evidence suggests that several interventions need to be re-evaluated. Large, high quality trials and systematic reviews of randomised trials seem warranted.

Introduction

Descartes (1596-1650) stated that much of what he had learned during his formal education turned out to be wrong. His predecessors made confident claims based on beliefs rather than reliable evidence. It is likely that some of the claims we make today will turn out to be wrong.

Traditionally treatment recommendations were based on clinical experience. Today, evidence-based medicine, which integrates clinical experience with reliable evidence, is replacing the traditional experience-based practice.[1-4] A hierarchy of evidence has been suggested based on the risk of bias associated with different research designs.[5,6]

Uncontrolled experience and experimental models

Uncontrolled clinical experience and experimental models are placed at the lowest levels in the hierarchy (figure). Uncontrolled clinical experience may constitute sufficiently reliable evidence if interventions have dramatic effects in line with insulin for diabetic ketoacidosis. Generally, the reliability of clinical experience is not sufficient because interventions have moderate effects. The shortcomings of human processing, unsystematic data collection, and the fluctuating nature of most diseases can be confused with intervention effects. Evidence from experimental models can be equally misleading, due to the necessary extrapolation. Treatment with β-blockers turned out to reduce mortality in congestive heart failure in spite of their negative inotropic effect.[7] Thalidomide had species specific effects that were not detected in the experimental models.[8] These examples suggest that extrapolation from experimental models to humans can result in both false negative and false positive conclusions.

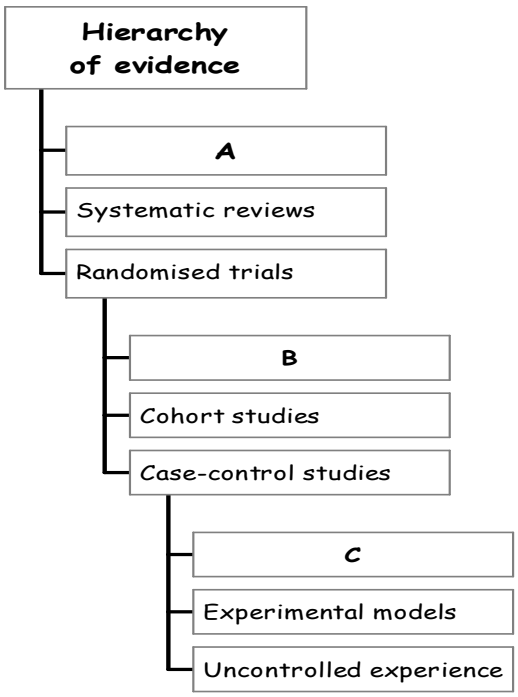

Figure The hierarchy of evidence on intervention effects based on reliability of different research designs.

Observational studies and randomised trials

In the hierarchy of evidence, observational studies have a higher ranking than experimental models and clinical experience because they involve humans and controlled collection of data (figure). The classical observational designs are cohort and case-control studies.[6,9,10] Cohort studies follow persons who are exposed (or not exposed) to something that may influence the probability of a disease. Case-control studies compare characteristics of cases (with the disease) and controls (without the disease). Observational studies are necessary to evaluate rare adverse events, prognostic variables, and behaviour.[11-13] The retrospective study design in some observational studies increases the risk of bias due to factors that change with time, recall bias, and differential measurement errors.[9,10,14]

In randomised trials, patients are randomly allocated to intervention and control groups. The purpose with randomisation is to create groups that are comparable with regard to known and unknown prognostic factors. In observational studies, prognostic factors determine whether patients are allocated to intervention or control groups. This is known as confounding by indication, which may lead to systematic differences between comparison groups.[15,16] When systematic differences exist, intervention benefits may be overestimated. We compared the results of observational studies and randomised trials in a systematic review of artificial support systems for liver failure.[17,18] Included patients allocated to the control groups received standard medical regimens. The control group mortality rates were significantly higher in observational studies than in the randomised trials (76% compared to 56%; $p<0.0001$). The observational studies found that the intervention reduced mortality significantly whereas the randomised trials found no significant effect. This suggests that observational studies may overestimate intervention benefits possibly due to a skewed allocation of patients with the worst prognosis to control groups. The results concur with a systematic review, which found a significant association between the observational study design and intervention benefits.[15] The strength of the association warrants separate evaluations in different situations, although observational studies generally are more susceptible to bias than randomised trials (figure).

Randomised trials and systematic reviews

It is debatable whether large randomised trials or systematic reviews provide the best evidence for intervention comparisons (figure).[19-21] Large randomised trials are generally considered one of our most reliable sources of intervention comparisons.[22-24] However, most trials are small and have inadequate bias control.[4,25-28] The statistical power of the individual trials may be too small to identify significant intervention benefits.[17,29,30] In systematic reviews, the results of individual trials may be combined in a meta-analysis thereby increasing statistical power.[24,31,32] The combination of several trials may also increase the extent to which results can be generalised.[33] Further, systematic reviews may facilitate evaluations of the impact of inadequate quality, publication bias, and other biases.[34] The main disadvantage of systematic reviews is related to their observational design. Subgroup analyses in systematic reviews generally require prospective evaluation.[17,35] Careful methods for identification and selection of trials are necessary to

avoid bias.[32,36-38] Bias may also occur in systematic reviews if the results of the included trials are biased, e.g., due to inadequate quality.[24,39-41] Some systematic reviews remain inconclusive because only low quality trials are included.[42-45] Although systematic reviews and randomised are placed at the top of the hierarchy of evidence, both research designs seem to have advantages as well as disadvantages.

Table 1 Bias in randomised trials

- Liberal versus restrictive use of episiotomy was evaluated in a randomised trial.[54] Some of the investigators who viewed episiotomy favourably failed to include eligible patients with certain characteristics in the trial.
- A randomised trial evaluated the effect of nicotine gum on smoking cessation.[68] Wrigley's chewing gum was used as placebo. It is likely that participants correctly guessed whether they were in the intervention or the control group.
- The effect of ascorbic acid for common cold was evaluated in a 'double-blind' randomised trial.[69,70] Participants were employees of the National Institutes of Health. Lactulose was used as placebo. The results of the trial turned out to be questionable because many participants tasted their capsules and guessed which group they were in.
- The effect of surgery with or without chemotherapy for metastatic colorectal cancer was evaluated in a randomised trial.[75] Chemotherapy was given through a device, which was inserted during operation unless the surgeon found that the prognosis was too poor. Patients who received chemotherapy therefore had a better prognosis. This disrupted the baseline comparability that was established through randomisation. Accordingly, per protocol analyses suggested a significant benefit of adjunctive chemotherapy, whereas intention-to-treat analyses found no significant effect.

The quality of randomised trials

Most researchers agree that quality is important, but few agree on how it should be measured. For clinical trials, quality depends on the control of bias.[46] Adequate bias control means high quality and vice versa. Different strategies are suggested for

incorporating quality in meta-analysis. These include threshold inclusion criteria, use of quality as a weight, plots of effect size against quality, and combination of trials stratified by quality.[34] To incorporate quality in meta-analyses, we have to know the effect of quality on intervention estimates. At least 25 quality scales exist, but few are validated using established criteria.[47,48] Some scales suggest that high quality trials tend to produce more conservative estimates of intervention benefits than low quality trials.[24,40,48] Other scales suggest that low quality trials tend to produce the most conservative estimates.[48] Further, potential overlap between the effects of the individual components in a scale and the overall scale score may also be problematic.[24] A number of cohort studies of randomised trials have evaluated the association between factors that may be related to quality and intervention effects.[24,39,41,47-49] The studies have focused on components relating to basic research methods including randomisation, blinding, and follow up.

Randomisation

Random allocation means that all patients have a known chance of being allocated to one of the intervention groups and that the allocation of the next patient is unpredictable.[50-53] To keep the allocation of patients unpredictable both adequate generation of an allocation sequence and adequate concealment of allocation are necessary. The allocation sequence may consist of random numbers generated by computers or tables. Allocation concealment may consist of randomisation through independent centres or serially numbered identical sealed packages. If the next assignment is known, enrolment of certain patients may be prevented or delayed to ensure that they receive the treatment that is believed to be superior (table 1).[54] Theoretically, serially numbered sealed envelopes may provide adequate allocation concealment, although there is some evidence suggesting that this may not be the case. Some envelopes have been opened before or after patients were excluded.[55] Other envelopes have been transilluminated.[56] The adequacy of using serially numbered sealed envelopes therefore seems debatable.

Many trials are described as randomised, but do not report randomisation methods (table 2).[4,25,57-62] Cohort studies of randomised trials suggest that the proportion of trials with adequate allocation sequence generation ranges from 1% to 52% (median 37%). The

proportion with adequate allocation concealment ranges from 2% to 39% (median 25%). Some of the variation may depend on the evaluated disease areas or different classifications of randomisation methods in the cohort studies.[4,25,27] The proportion of high quality trials may also depend on funding or number of clinical sites. Trials seem to report adequate randomisation methods more often if external funding is declared or several clinical sites are involved.[25,27]

Table 2 Cohort studies on reported randomisation methods in randomised trials from different disease areas

Disease area	Number of trials	Proportion (%) of trials with adequate allocation sequence generation	Proportion (%) of trials with adequate allocation concealment
Various[58]	80	48%	26%
Gynaecology/obstetrics[57]	206	32%	23%
Hepatology[59]	166	28%	37%
Hepatology[25]	235	52%	34%
Dermatology[60]	68	1%	7%
Intensive care medicine[61]	173	27%	6%
Gastroenterology[4]	383	42%	39%
Orthodontics[62]	155	50%	2%

The association between reported randomisation methods and intervention effects was evaluated in a cohort study of 250 trials from meta-analyses on obstetrics.[39] The study compared estimated intervention benefits in trials with or without adequate randomisation methods. The results suggested that intervention benefits were about 30% larger if trials did not report adequate allocation concealment. A similar cohort study of 127 trials from meta-analyses in four disease areas found that intervention benefits were about 37% larger in trials without adequate allocation concealment.[40] Neither of the studies found significant associations between generation of allocation sequence and intervention effects.[39,40] The demonstrated associations between

randomisation and intervention effects may be confounded by selective publication of low quality trials with favourable outcomes.[32,63] Further, without a reference group we do not know whether low quality trials exaggerate or high quality trials underestimate intervention benefits. We addressed this question in a cohort study with 190 randomised trials from eight disease areas.[24] The trials were included in 14 meta-analyses that included at least one large gold standard trial (with >1000 participants), which was used as a reference group. Our results suggested that small trials overestimated intervention benefits if the allocation sequence generation or allocation concealment were inadequate, but not if either of these methods were adequate.[24] The association between randomisation and intervention effects was subsequently evaluated in a cohort study of 276 randomised trials from four medical areas.[49] The trials were included in meta-analyses with significant heterogeneity.[32] The study found no significant association between allocation concealment and intervention effects. The effect of adequate allocation sequence generation was not evaluated. The reasons for the discrepancy between this and previous studies may reflect different definitions of adequate allocation concealment, the selection of trials[64] or random error. The results of the study and four similar cohort studies were combined in a meta-analysis.[24,39-41,49,65,65] Overall, the combined evidence suggests that trials with inadequate allocation concealment tend to overestimate intervention benefits by about 25%. Although the overall estimate is statistically significant, the evidence suggests that there is some variation in the extent of the association between allocation concealment and intervention effects. Separate evaluations in samples of trials therefore seem warranted.

Blinding

In randomised trials, the term blinding refers to keeping participants, health care providers, data collectors, outcome assessors, or data analysts unaware of the assigned intervention.[66] Trials in which patients and investigators are unaware of the assigned intervention are classified as double blind.[67] Sometimes the nature of the intervention precludes double blinding, but blinded outcome assessment and data analyses are usually possible. To ensure adequate double blinding, the compared interventions must be similar. If interventions are compared to no intervention, an identical placebo must be used. Any difference in taste, smell, or appearance can destroy blinding (table 1).[68-70]

In a cohort study with 616 hepato-biliary randomised trials published during 1985–1996, we found that only 34% were double blind.[27] The proportion of double blind trials was significantly associated with the disease area. The variation may reflect that some interventions were difficult to blind or that trials in certain disease areas tend to be performed without double blinding although blinding would have been feasible.[4,25,26,71,71]

Blinding prevents bias associated with patients and investigators expectations.[41] We compared estimated intervention benefits in small trials without blinding and large gold standard randomised trials.[24] The results suggested that small trials without double blinding tend to overestimate intervention benefits significantly. Five cohort studies have evaluated the association between double blinding and intervention effects.[24,39-41,49,65 24,40,41,49,65]A meta-analysis of these studies showed that lack of double blinding was associated with 12% overestimation of intervention benefits.[65] The association between blinding and intervention effects was significant, but there was some variation with regard to the extent of the association in individual studies. Separate evaluations in samples of trials therefore seem warranted.

Follow up

Clinical trials usually have missing data due to losses to follow up.[72] Protocol deviations are often related to prognostic factors and may therefore lead to attrition bias.[73-75] Attempts to obtain data on patients who are lost to follow up and clear descriptions follow up are important.[76-78] Thirty percent of 235 randomised trials published in the journal 'Hepatology' during 1981–1998 did not describe the numbers or reasons for dropouts and withdrawals.[25] In a cohort study of randomised trials, intervention effects did not seem to differ significantly in trials with losses to follow up and trials with complete (explicit or assumed) follow up.[39] In a similar study, we found no significant difference between intervention effects in small randomised trials with unclear follow up reports and large gold standard trials.[24] The lack of association between follow up and intervention effects may reflect discrepancies between reporting and number of dropouts and withdrawals. At present, the association between the number of losses to follow up and intervention effects remains to be established.

Several analytical strategies for dealing with missing data are proposed.[79] Intention-to-treat analyses include all patients whereas per protocol analyses exclude data from patients with protocol deviations.[79,80] Per protocol analyses reduce statistical power when several patients are excluded. Further, if an intervention has adverse effects that lead to losses to follow up, per protocol analyses overestimate the intervention benefit.[29,30] Per protocol analyses also tend to overestimate intervention benefits when prognostic factors are related to treatment withdrawals.[75] Intention-to-treat analyses generally seem to be the most reliable analytical strategy in systematic reviews and randomised trials.

Competing interests

The effect of financial competing interests is debated.[81] Funding is associated with adequate methodological quality.[25,27] On the other hand, financial interests may affect the interpretation of trial results. In a cohort of 159 randomised trials, we found that conclusions were significantly more favourable towards experimental interventions if trials received funding from forprofit organisations.[82] We evaluated the effects of potential confounding factors including quality and statistical power, but none appeared to explain the association between forprofit funding and conclusions. Two systematic reviews have found similar associations between industry sponsorship and pro-industry conclusions.[83,84] The reason for the association between funding and conclusions could be that forprofit organisations tend to evaluate beneficial and safe interventions. We therefore evaluated the effect of intervention benefits and adverse events on the association between funding and conclusions in a cohort of 370 randomised drug trials.[85] The results confirmed the significant association between funding and conclusions. The quantitative estimates of intervention benefits were significantly associated with conclusions, but the occurrence of adverse events was not. Adjusted analyses indicated that neither of these factors explained the association between funding and conclusions. The reason for the potential effect of funding remains to be established.

Sample size and statistical power

The probability of random errors follows a symmetrical bell curve. Errors are equally likely to occur in any direction and may lead to false positive (type I error) or false

negative results (type II error). In randomised trials, the risk of random error depends on the sample size and the size of the intervention effect. The larger the sample size and the larger the intervention effect, the smaller the risk of random error. Small trials on interventions with moderate effects have a substantial risk of producing false positive or negative conclusions.

Table 3 Cohort studies on sample size and statistical power of randomised trials in different disease areas

Disease area	Number of trials	Median sample size (IQR)*	Power to detect 60% to 40% difference
Hepatology[25]	235	52 (28-88)	45
Dermatology[60]	68	46 (30-80)	40
Gastroenterology[4]	385	54 (24-110)	47
Sclerosis[88]	73	28 (17-43)	23
Intensive care medicine[61]	173	30 (20-64)	25
Radiology[89]	130	61 (27-104)	53

*Inter quartile range

Sample size calculations are required in randomised trials because inadequate statistical power can lead to false negative results.[86] The calculations should account for the minimum relevant treatment difference, acceptable probabilities of type I and II error, and losses to follow up.[50,72] The first parameter is adjustable and sensitive. If you reduce the relevant difference by half, four times as many patients are needed. The risks of type I error (α) is usually set to 5%. The risk of type II error (β) errors is usually set to 10% or 20%. The corresponding power (1-β), which indicates the risk of overlooking intervention effects, is 90% or 80%. In cohort studies of randomised trials, sample size calculations were only reported in 8-38% of the included trials.[25,59,87-90] When the pre-set sample size is not reported, it becomes difficult to evaluate whether the planned sample size was reached or whether the trial was extended beyond the planned size or was terminated at an arbitrary point in time.

The power of a trial reflects the risk of overlooking intervention effects. Suppose that you want to perform a trial on a drug that reduces mortality from 40% to 20%. You set α to 5% and include 90 patients in each treatment arm. Such a trial will have 80% power to detect the true treatment effect.[50] If you repeat the trial 100 times, 20 trials will overlook the treatment effect. Now suppose that you evaluate the same drug, but include 45 patients in each treatment arm. This sample size corresponds to a power of 55%. If you repeat the trial 100 times, 45 of you trials will overlook the true treatment effect. The entire sample of trials must therefore be evaluated. In practice, this may be done through a systematic review with a meta-analysis of the trials.

Two studies evaluated statistical power in trials without statistically significant outcomes.[91,92] Both studies found that most trials had insufficient power to detect clinically relevant treatment effects. The relatively small sample size of randomised trials suggests that few have the recommended statistical power (table 3).[4,25,50,60,61,72,88,89] This suggests that several effective interventions may have been disregarded. In many small trials, lack of significant effects may reflect inadequate statistical power or that the intervention does not work. Absence of evidence is not the same as evidence of absence (table 4). Concluding that an intervention is not effective or that two interventions are equally effective is therefore problematic.[93]

Table 4 The Fermi paradox

One day in the early 1940s, nuclear physicists at Los Alamos Laboratories talked about the possibility of other intelligent life in the galaxy. The physicist Enrico Fermi asked: "so if they exist - where are they?" Considering the age of the universe, the galaxy should now be fully colonised. Timothy Ferris answered the question by a simple experiment with a lobster dinner. After setting up the table, he prepared all the fixings for lobster, opened the door, and waited for the lobster to appear. The experiment was stopped after four hours. Ferris had to conclude that lobsters do not exist because none had turned up. This conclusion is obviously wrong. As Ferris pointed out "absence of evidence is not evidence of absence."

Traditional reviews often count the number of supportive trials and choose the view receiving the most votes.[38,94] This may lead to false negative conclusions if trials are

under-powered. Statistical power may be increased if the results of the individual trials are combined in a meta-analysis. We identified 12 randomised trials for a systematic review on interferon with or without ribavirin for non-responders with hepatitis C.[29,30] The primary outcome was sustained clearance of the hepatitis C virus. The sample size of the trials suggested that none had sufficient statistical power. Only three trials found a significant effect of combination therapy, but a meta-analysis of the trials found a significant effect of combination therapy.

Publication bias and related biases

In systematic reviews, unbiased inclusion of trials is essential.[32,95,96] Factors that affect the availability of trials may affect the conclusions of systematic reviews. Selective publication of trials with favourable outcomes may lead to type 1 errors.[97] Several studies suggest that such publication bias exist. The proportion of positive trials indicates the extent of publication bias in the published literature. In a cohort study with 530 published trials, we found that 71% had statistically significant outcomes.[98] The sample size of the included trials suggests that about half had a 60% risk of overlooking a 60% to 30% reduction in mortality. Similar studies in other disease areas found similar high proportions of positive trials.[96] The existence of publication bias has also been evaluated in prospective studies, which found that significant results increased the chances of publication significantly.[63,99-102] None of the prospective studies have evaluated the effect of publication bias, but clinical examples suggest that publication bias may lead to implementations of potentially ineffective or harmful interventions (table 5).[103-115]

Citation habits may also affect the dissemination of trials.[96,116] We evaluated the association between statistical significance and citation frequencies in a cohort of 530 randomised trials.[98] All trials dealt with hepatological diseases. The study suggested that trials were cited significantly more often if their results were statistically significant. Other studies have found similar associations in cardiovascular diseases and rheumatology.[117,118] A similar study of vaccine trials found the direct opposite.[119] In a study of articles that were submitted to an emergency medicine meeting, citation

frequencies did not seem to depend on the statistical significance of the study outcome. The extent and direction of citation bias therefore remains unclear.[96]

Table 5 Dissemination bias

- Selective serotonin reuptake inhibitors (SSRI) have been used for depression since the 1980s.[103] The first trials were initiated in the late 1980s, but several remained unpublished more than 10 years later.[104,105] During subsequent years, the press and medical journals reported cases of suicidal behaviour among children on SSRI.[103] In 2003, a review of data from published and unpublished clinical trials prompted the Food and Drug Administration (FDA) and the Committee on Safety of Medicines (CSM) to warn against the use of SSRI for children.[106-109]
- The Vioxx Gastrointestinal Outcomes Trial (VIGOR) suggested that rofecoxib was a safe alternative to anti-inflammatory drugs for pain relief.[110] Independent researchers were concerned about the potential cardiovascular effects.[111,112] The published report of the VIGOR trial did not describe this question. A review of unpublished safety data showed that rofecoxib increased the risk of serious cardiovascular thrombotic events.[113,114]
- In 1995 the Department of Health commissioned an individual patient meta-analysis of studies of primrose oil supplementation for atopic dermatitis.[115] For unknown reasons, the authors were not allowed to publish their work. In fact, Searle (the company then responsible for marketing evening primrose oil) asked the authors to sign a written statement to indicate that the contents of the report were not leaked.

Conclusions

Bias may occur in randomised trials and systematic reviews although they have the highest rank in the hierarchy of evidence. Lack of adequate randomisation, blinding, and follow-up may bias estimates of intervention effects.[17,24,39,40] The size of the effect of bias is about 25% for inadequate randomisation and 12% for lack of double blinding. The size of most intervention effects is less than or equal to the potential effects of bias. The sample size of trials also seems important because small trials may produce false positive or false negative results due to inadequate statistical power and random

error.[4,25] Many randomised trials may have generated the wrong conclusions regarding intervention benefits because they are small and have unclear bias control.[4,4,25-27,59,120] Selective publication of trials with positive results and financial competing interests may also affect conclusions about intervention benefits.[27,82,85,96,98] In theory, interventions should only be used if they have been shown to be effective in well-designed trials.[121] The evidence presented in this review suggests that several interventions may need re-evaluation in large, high quality randomised trials and systematic reviews of randomised trials. The Cochrane Collaboration and similar initiatives may facilitate this goal.[122]

Additional research is needed for a more detailed evaluation of the effects of bias in randomised trials and systematic reviews. The reliability of sealed envelopes compared to central randomisation, the effect of blinded outcome assessments, the effect of dropouts and withdrawals, and methods for dealing with attrition bias all seem important. Methods for improving recruitment rates and completeness of follow up may improve the reliability of many randomised trials.[123,124] The different factors that are related to bias may be characterised as amalgamated characteristics. We are only able to see the pattern if larger groups of trials are evaluated. None of the evaluated factors that seem to reflect bias control can accurately predict the extent or direction of bias in the individual trials. Different research questions therefore warrant individual evaluations.

References

1 Sackett D, Rosenberg WM, Gray JA, Haynes RB, Richardson WS. Evidence based medicine: what it is and what it isn't. BMJ 1996;**312**:71-2.

2 Guyatt GH, Haynes RB, Jaeschke RZ, Cook D, Green L, Naylor CD, et al. Users' guides to the medical Literature: XXV. Evidence-based medicine: principles for applying The users' guides to patient care. Evidence-Based Medicine Working Group. JAMA 2000;**284**:1290-6.

3 Kjaergard LL, Gluud C. [Health technology assessment]. In: Wang J, Gluud C, editors. [Evidence based medicine and clinical practise]. Shanghai: Science Publishing house; 2001. p. 229-31.

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

5 Guyatt G, Sinclair J, Cook D, Jaeschke R, Schünemann H, Pauker S. Grading recommendations - a qualitative approach. In: Hayward, Robert, electronic editor. Users' guides interactive. Chicago(IL): JAMA Publishing Group; 2002 [cited 2002 11 27]. Available from: http://www.usersguides.org.

6 Grimes DA, Schulz KF. An overview of clinical research: the lay of the land. Lancet 2002;**359**:57-61.

7 MERIT-HF Study Group. Effect of metoprolol CR/XL in chronic heart failure: Metoprolol CR/XL Randomised Intervention Trial in Congestive Heart Failure (MERIT-HF). Lancet 1999;**353**:2001-6.

8 Stern L. In vivo assessment of the teratogenic potential of drugs in humans. Obstet Gynecol 1981;**58**:3S-8S.

9 White E, Hunt JR, Casso O. Exposure measurement in cohort studies: the challenges of prospective data collection. Epidemiol Rev 1998;**20**:43-56.

10 Wacholder S. Design issues in case-control studies. Stat Methods Med Res 1995;**4**:293-309.

11 Nuesch R, Schroeder K, Dieterle T, Martina B, Battegay E. Relation between insufficient response to antihypertensive treatment and poor compliance with treatment: a prospective case-control study. BMJ 2001;**323**:142-6.

12 Black N. Why we need observational studies to evaluate the effectiveness of health care. BMJ 1996;**312**:1215-8.

13 Owen CG, Whincup PH, Gilg JA, Cook DG. Effect of breast feeding in infancy on blood pressure in later life: systematic review and meta-analysis. BMJ 2003;**327**:1189-95.

14 Sacks H, Chalmers TC, Smith H, Jr. Randomized versus historical controls for clinical trials. Am J Med 1982;**72**:233-40.

15 Kunz R, Oxman AD. Randomisation to protect against selection bias in healthcare trials (Cochrane Methodology Review). In: The Cochrane Library. Chichester, UK: John Wiley & Sons, Ltd, Issue 4, 2003.

16 Johnston SC. Combining ecological and individual variables to reduce confounding by indication: case study-subarachnoid hemorrhage treatment. J Clin Epidemiol 2000;**53**:1236-41.

17 Kjaergard LL, Liu JP, Als-Nielsen B, Gluud C. Artificial and bioartificial support systems for acute and acute-on-chronic liver failure: a systematic review. JAMA 2003;**289**:217-22.

18 Liu JP, Kjaergard LL, Als-Nielsen B, Gluud C. Artificial and bioartificial support systems for liver failure (Cochrane Review). In: The Cochrane Library. Chichester, UK: John Wiley & Sons, Ltd, Issue 1, 2004.

19 Villar J, Carroli G, Belizan JM. Predictive ability of meta-analyses of randomised controlled trials. Lancet 1995;**345**:772-6.

20 Cappelleri JC, Ioannidis JP, Schmid CH, de Ferranti SD, Aubert M, Chalmers TC, et al. Large trials vs meta-analysis of smaller trials: how do their results compare? JAMA 1996;**276**:1332-8.

21 LeLorier J, Gregoire G, Benhaddad A, Lapierre J, Derderian F. Discrepancies between meta-analyses and subsequent large randomized, controlled trials. New Engl J Med 1997;**337**:536-42.

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

23 Peto R, Baigent C. Trials: the next 50 years. Large scale randomised evidence of moderate benefits. BMJ 1998;**317**:1170-1.

24 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.

25 Kjaergard LL, Nikolova D, Gluud C. Randomized clinical trials in Hepatology: predictors of quality. Hepatology 1999;**30**:1134-8.

26 Gluud C, Kjaergard LL. Quality of randomised clinical trials in portal hypertension and other fields of hepatology. In: Franchis R, editor. Portal Hypertension III. Proceedings of the Third Baveno International Consensus Workshop on Definitions, Methodology, and Therapeutic Strategies. Oxford: Blackwell Science; 2001.

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

28 McDonald S, Westby M, Clarke M, Lefebvre C, Cochrane Centres' Working Group on 50 Years of Randomized Trials. Number and size of randomized trials reported in general health care journals from 1948 to 1997. Int J Epidemiol 2002;**31**:125-7.

29 Kjaergard LL, Krogsgaard K, Gluud C. Interferon alfa with or without ribavirin for chronic hepatitis C: systematic review of randomised trials. BMJ 2001;**323**:1151-5.

30 Kjaergard LL, Krogsgaard K, Gluud C. Ribavirin with or without alpha interferon versus no intervention, placebo or alpha interferon for chronic hepatitis C (Cochrane Review). In: The Cochrane Library. Oxford: Update Software, Issue 1, 2002.

31 Guyatt GH, Sinclair J, Cook DJ, Glasziou P. Users' guides to the medical literature: XVI. How to use a treatment recommendation. Evidence-Based Medicine Working Group and the Cochrane Applicability Methods Working Group. JAMA 1999;**281**:1836-43.

32 Alderson P, Green S, Higgins JPT, editors. Cochrane Reviewers' Handbook 4.2.1 [updated December 2003]. In: The Cochrane Library. Chichester, UK: John Wiley & Sons, Ltd., Issue 1, 2004.

33 Gøtzsche P. Why we need a broad perspective on meta-analysis. It may be crucially important for patients. BMJ 2000;**321**:585-6.

34 Detsky AS, Naylor CD, O'Rourke K, McGeer AJ, Abbe KA. Incorporating variations in the quality of individual randomized trials into meta-analysis. J Clin Epidemiol 1992;**45** :255-65.

35 Oxman AD, Guyatt GH. A consumer's guide to subgroup analyses. Ann Intern Med 1992;**116**:78-84.

36 Egger M, Smith GD, Phillips AN. Meta-analysis: principles and procedures. BMJ 1997;**315**:1533-7.

37 Smith D, Egger M. Meta-analysis. Unresolved issues and future developments. BMJ 2000;**316**:221-5.

38 Egger M, Smith GD. Meta-analysis. Potentials and promise. BMJ 1997;**315**:1371-4.

39 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.

40 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.

41 Jüni P, Altman DG, Egger M. Assessing the quality of controlled clinical trials. BMJ 2001;**323**:42-6.

42 Als-Nielsen B, Koretz RL, Kjaergard LL, Gluud C. Branched-chain amino acids for hepatic encephalopathy (Cochrane Review). In: The Cochrane Library. Chichester, UK: John Wiley & Sons, Ltd, Issue 4, 2003.

43 Als-Nielsen B, Kjaergard LL, Gluud C. Benzodiazepine receptor antagonists for acute and chronic hepatic encephalopathy (Cochrane Review). In: The Cochrane Library. Chichester, UK: John Wiley & Sons, Ltd, Issue 4, 2003.

44 Als-Nielsen B, Gluud LL, Gluud C. Non-absorbable disaccharides for hepatic encephalopathy: systematic review of randomised trials. BMJ 2004;**328**;1046-50.

45 Als-Nielsen B, Gluud LL, Gluud C. Non-absorbable disaccharides for hepatic encephalopathy (Cochrane Review). In: The Cochrane Library. Chichester, UK: John Wiley & Sons, Ltd, Issue 1, 2004.

46 Oxman AD, Guyatt GH, Singer J, Goldsmith CH, Hutchison BG, Milner RA, et al. Agreement among reviewers of review articles. J Clin Epidemiol 1991;**44**:91-8.

47 Moher D, Jadad AR, Nichol G, Penman M, Tugwell P, Walsh S. Assessing the quality of randomized controlled trials: an annotated bibliography of scales and checklists. Control Clin Trials 1995;**16**:62-73.

48 Jüni P, Witschi A, Bloch R, Egger M. The hazards of scoring the quality of clinical trials for meta-analysis. JAMA 1999;**282**:1054-60.

49 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.

50 Pocock SJ. Clinical trials - a practical approach. Chichester: John Wiley and Sons; 1996.

51 Altman DG, Bland JM. Statistics notes. Treatment allocation in controlled trials: why randomise? BMJ 1999;**318**:1209.

52 Altman DG, Bland JM. How to randomise. BMJ 1999;**319**:703-4.

53 Schulz KF, Grimes DA. Allocation concealment in randomised trials: defending against deciphering. Lancet 2002;**359**:614-8.

54 Klein MC, Kaczorowski J, Robbins JM, Gauthier JP, Jorgensen SH, Joshi AK. Physicians' beliefs and behaviour during a randomized controlled trial of episiotomy: consequences for women in their care. CMAJ 1995;**153**:769-79.

55 Swingler GH, Zwarenstein M. An effectiveness trial of a diagnostic test in a busy outpatients department in a developing country: issues around allocation concealment and envelope randomization. J Clin Epidemiol 2000;**53**:702-6.

56 Schulz KF. Subverting randomization in controlled trials. JAMA 1995;**274**:1456-8.

57 Schulz KF, Chalmers I, Altman DG, Grimes DA, Dore CJ. The methodologic quality of randomization as assessed from reports of trials in specialist and general medical journals. *Online* J Curr Clin Trials 1995;Doc No 197:81.

58 Altman DG, Dore CJ. Randomisation and baseline comparisons in clinical trials. Lancet 1990;**335**:149-53.

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

60 Adetugbo K, Williams K. How well are randomized controlled trials reported in the dermatology literature? Arch Dermatol 2000;**136**:381-5.

61 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. Int Care Med 2002;**28**:1316-23.

62 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**:309-15.

63 Easterbrook PJ, Berlin JA, Gopalan R, Matthews DR. Publication bias in clinical research. Lancet 1991;**337**:867-72.

64 Schulz KF, Altman DG, Moher D. Allocation concealment in clinical trials. JAMA 2002;**288**:2408-9.

65 Jüni P, Egger M. Allocation concealment in clinical trials. JAMA 2002;**288**:2407-8.

66 Deveraux PJ, Manns BJ, Ghali WA, Quan H, Lacchetti C, Montori VM, et al. Physician interpretations and textbook definitions of blinding terminology in randomized controlled trials. JAMA 2001;**285**:2000-3.

67 Schulz KF, Grimes DA. Blinding in randomised trials: hiding who got what. Lancet 2002;**359**:696-700.

68 Campbell IA, Lyons E, Prescott RJ. Stopping smoking. Do nicotine chewing-gum and postal encouragement add to doctors' advice. Practitioner 1987;**231**:114-7.

69 Karlowski TR, Chalmers TC, Frenkel LD, Kapikian AZ, Lewis TL, Lynch JM. Ascorbic acid for the common cold. A prophylactic and therapeutic trial. JAMA 1975;**231**:1038-42.

70 Chalmers TC. Effects of ascorbic acid on the common cold. An evaluation of the evidence. Am J Med 1975;**58**:532-6.

71 Schulz KF, Grimes DA, Altman DG, Hayes RJ. Blinding and exclusions after allocation in randomised controlled trials: survey of published parallel group trials in obstetrics and gynaecology. BMJ 1996;**312**:742-4.

72 International conference on harmonisation expert working group. International conference on harmonisation of technical requirements for registration of pharmaceuticals for human use. ICH harmonised tripartite guideline. Guideline for good clinical practice. 1997 CFR & ICH Guidelines. 1ed. PA 19063-2043, USA: Barnett International/PAREXEL; 1997.

73 Corrigan JD, Harrison-Felix C, Bogner J, Dijkers M, Terrill MS, Whiteneck G. Systematic bias in traumatic brain injury outcome studies because of loss to follow-up. Arch Phys Med Rehabil 2003;**84**:153-60.

74 DiFranceisco W, Kelly JA, Sikkema KJ, Somlai AM, Murphy DA, Stevenson LY. Differences between completers and early dropouts from 2 HIV intervention trials: a health belief approach to understanding prevention program attrition. Am J Public Health 1998;**88**:1068-73.

75 Kemeny MM, Adak S, Gray B, Macdonald JS, Smith T, Lipsitz S, et al. Combined-modality treatment for resectable metastatic colorectal carcinoma to the liver: surgical resection of hepatic metastases in combination with continuous infusion of chemotherapy - an intergroup study. J Clin Oncol 2002;**20**:1499-505.

76 Egger M, Jüni P, Bartlett C, CONSORT Group (Consolidated Standards of Reporting of Trials). Value of flow diagrams in reports of randomized controlled trials. JAMA 2001;**285**:1996-9.

77 Schulz KF, Grimes DA. Sample size slippages in randomised trials: exclusions and the lost and wayward. Lancet 2002;**359**:781-5.

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

79 Montori VM, Guyatt GH. Intention-to-treat principle. CMAJ 2001;**165**:1339-41.

80 Millis SR. Emerging standards in statistical practice: implications for clinical trials in rehabilitation medicine. Am J Phys Med Rehabil 2003;**82**:S32-S37.

81 Smith R. Beyond conflict of interest. Transparency is the key. BMJ 1998;**317**:291-2.

82 Kjaergard LL, Als-Nielsen B. Association between competing interests and authors' conclusions: epidemiological study of randomised clinical trials published in the BMJ. BMJ 2002;**325**:249-52.

83 Bekelman JE, Li Y, Gross CP. Scope and impact of financial conflicts of interest in biomedical research: a systematic review. JAMA 2003;**289**:454-65.

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

85 Als-Nielsen B, Chen W, Gluud C, Kjaergard LL. Association of funding and conclusions in randomized drug trials: a reflection of treatment effect or adverse events? JAMA 2003;**290**:921-8.

86 Altman DG, Bland JM. Absence of evidence is not evidence of absence. BMJ 1995;**311**:485.

87 Maggard MA, O'Connell JB, Liu JH, Etzioni DA, Ko CY. Sample size calculations in surgery: are they done correctly? Surgery 2003;**134**:275-9.

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

89 Huang W, LaBerge JM, Lu Y, Glidden DV. Research publications in vascular and interventional radiology: research topics, study designs, and statistical methods. J Vasc Interv Radiol 2002;**13**:247-55.

90 Skovlund E. A critical review of papers from clinical cancer research. Acta Oncol 1998;**37**:339-45.

91 Freiman JA, Chalmers TC, Smith H, Kuebler RR. The importance of beta, the type II error and sample size in the design and interpretation of the randomized control trial. Survey of 71 "negative" trials. New Engl J Med 1978;**299**:690-4.

92 Moher D, Dulberg CS, Wells GA. Statistical power, sample size, and their reporting in randomized controlled trials. JAMA 1994;**272**:122-4.

93 Alderson P, Chalmers I. Survey of claims of no effect in abstracts of Cochrane reviews. BMJ 2003;**326**:475.

94 Koretz RL. Methods of meta-analysis: an analysis. Curr Opin Metab Care 2002;**5**:467-74.

95 Thase M. Comparing the methods used to compare antidepressants. Psychopharmacol Bull 2002;**36**:1-17.

96 Song F, Eastwood S, Gilbody S, Duley L, Sutton AJ. Publication and related biases. Health Technol Assess 2000;**4**:1-115.

97 Gluud C. "Negative trials" are positive! J Hepatol 1998;**28**:731-3.

98 Kjaergard LL, Gluud C. Citation bias of hepato-biliary randomized clinical trials. J Clin Epidemiol 2002;**55**:407-10.

99 Dickersin K, Chan S, Chalmers TC, Sacks HS, Smith H. Publication bias and clinical trials. Control Clin Trials 1987;**8**:343-53.

100 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**:374-8.

101 Dickersin K, Min YI. NIH clinical trials and publication bias. Online J Curr Clin Trials 1993;**Doc No 50**.

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

103 Interim report of the Committee on Safety of Medicines' Expert Working Group on Selective Serotonin Reuptake Inhibitors. www seroxatusergroup org uk [cited 2004 Jan 05].

104 Khan A, Khan S, Kolts R, Brown WA. Suicide rates in clinical trials of SSRIs, other antidepressants, and placebo: analysis of FDA reports. Am J Psychiatry 2003;**160**:790-2.

105 Khan A, Warner HA, Brown WA. Symptom reduction and suicide risk in patients treated with placebo in antidepressant clinical trials: an analysis of the Food and Drug Administration database. Arch Gen Psychiatry 2000;**57**:311-7.

106 Abbott A. British panel bans use of antidepressant to treat children. Nature 2003;**423**:792.

107 Mitka M. FDA alert on antidepressants for youth. JAMA 2003;**290**:2534.

108 Hirschfeld RMA. Suicide and antidepressant treatment. Arch Gen Psychiatry 2000;**57**:325-6.

109 Oranski I. FDA questions antidepressant safety for children. Lancet 2003;**862**:1558.

110 Bombadier C, Laine L, Reicin A, Shapiro D, Burgos-Vargas R, Davis B, et al. Comparison of upper gastrointestinal toxicity of rofecoxib and naproxen in patients with rheumatoid arthritis. VIGOR Study Group. N Engl J Med 2003;**343**:1520-8.

111 Mukherjee D, Nissen SE, Topol EJ. Risk of cardiovascular events associated with selective COX-2 inhibitors. JAMA 2001;**286**:954-9.

112 Ray WA, Stein CM, Daugherty JR, Hall K, Arbogast PG, Griffin MR. COX-2 selective non-steroidal anti-inflammatory drugs and risk of serious coronary heart disease. Lancet 2002;**360**:1071-3.

113 FDA Advisory Committee Briefing Document. www fda gov/cder/drug/advisory/mdd htm [cited 2003 Dec 12].

114 McCormack J, Rangno R. Digging for data from the COX-2 trials. CMAJ 2002;**166**:1649-50.

115 Williams HC. Evening primrose oil for atopic dermatitis. BMJ 2004;**327**:1358-9.

116 Gluud LL, Sørensen TIA, Gøtzsche P, Gluud C. The Journal Impact Factor as a Predictor of Trial Quality and Outcomes: Cohort Study of Hepato-Biliary Randomized Clinical Trials. Am J Gastroenterol 2005 [In Press].

117 Gøtzsche P. Reference bias in reports of drug trials. BMJ 1987;**195**:654-46.

118 Ravnskov U. Cholesterol lowering trials in coronary heart disease: frequency of citation and outcome. BMJ 1992;**305**:15-9.

119 Hutchison BG, Lloyd S. Comprehensiveness and bias in reporting clinical trials. Study of reviews of pneumococcal vaccine effectiveness. Can Fam Physician 1995;**41**:1356-60.

120 Gluud C. Evidence based medicine in Liver. Liver 1999;**19**:1-2.

121 Cochrane A. Effectiveness and efficiency. Random reflections on health services. London: Nuffield Provincial Hospital Trust, 1972.

122 Chalmers I, Sackett D, Silagy C. Non-random reflections on Health Services Research: On the 25th Anniversary of Archie Cochrane´s Effectiveness and Efficiency. In: Maynard A., Chalmers I, editors. London: BMJ Publishing Group; 1997. p. 231-49.

123 Kjaergard LL, Kruse AY, Krogsgaard K, Gluud C, Mortensen EL, Gottschau A, et al. Outpatients' knowledge about and attitude towards clinical research and randomised clinical trials. Dan Med Bull 1998;**45**:439-43.

124 Kruse AY, Kjaergard LL, Krogsgaard K, Gluud C, Mortensen EL, Gottschau A, et al. A randomized trial assessing the impact of written information on outpatients' knowledge about and attitude toward randomized clinical trials. Control Clin Trials 2000;**21**:223-40.

Danish Summary

Formålet med denne afhandling er at gennemgå evidensen for faktorer, der kan medføre bias i klinisk interventions forskning. Afhandlingen er primært baseret på ni tidligere publicerede kohorte studier og systematiske litteratur oversigter over observationelle studier og randomiserede forsøg. Flere potentielle kilder til bias identificeres. Observationelle studier har en tendens til at overvurdere interventions effekter sammenlignet med randomiserede forsøg. Små forsøg uden adækvat randomisering eller dobbelt blinding har en tendens til at overvurdere interventions effekter sammenlignet med store 'guld-standard' randomiserede forsøg. Dette er ikke tilfældet for små forsøg der har adækvat randomisering eller dobbeltblinding. Desværre rapporterer kun omkring 37% af randomiserede forsøg adækvat generering af allokerings sekvens og omkring 25% adækvat skjult allokering. Flere forsøg gennemføres uden dobbelt blinding selvom dette ville have været muligt. Endvidere er de fleste randomiserede forsøg små og mangler materialestørrelses beregninger. Den lille størrelse indikerer at de har en væsentlig risiko for at generere falsk negative eller falsk positive resultater. Financielle interesser er muligvis også associeret med bias. Konklusioner i randomiserede forsøg har en tendens til at være mere positive overfor eksperimentelle interventioner, hvis forsøget er finansieret af en profit organisation. De kvantitative resultater og hyppigheden af utilsigtede hændelser kan ikke forklare sammenhængen mellem finansiering og konklusioner. Ingen af de potentielle kilder til bias kan forudsige graden eller retningen af bias i enkelte forsøg. Forskellige forskningsområder bør derfor evalueres individuelt. Den samlede evidens tyder på at flere interventioner bør evalueres i store randomiserede forsøg med god kvalitet og i systematiske litteraturoversigter.

ISBN 87-990924-0-9