The unpredictability paradox: review of empirical comparisons of randomised and non-randomised clinical trials

Regina Kunz, Andrew D Oxman

Abstract

Objective To summarise comparisons of randomised clinical trials and non-randomised clinical trials, trials with adequately concealed random allocation versus inadequately concealed random allocation, and high quality trials versus low quality trials where the effect of randomisation could not be separated from the effects of other methodological manoeuvres.
Design Systematic review.
Selection criteria Cohorts or meta-analyses of clinical trials that included an empirical assessment of the relation between randomisation and estimates of effect.
Data sources Cochrane Review Methodology Database, Medline, SciSearch, bibliographies, hand searching of journals, personal communication with methodologists, and the reference lists of relevant articles.
Main outcome measures Relation between randomisation and estimates of effect.
Results Eleven studies that compared randomised controlled trials with non-randomised controlled trials (eight for evaluations of the same intervention and three across different interventions), two studies that compared trials with adequately concealed random allocation and inadequately concealed random allocation, and five studies that assessed the relation between quality scores and estimates of treatment effects, were identified. Failure to use random allocation and concealment of allocation were associated with relative increases in estimates of effects of 150% or more, relative decreases of up to 90%, inversion of the estimated effect and, in some cases, no difference. On average, failure to use randomisation or adequate concealment of allocation resulted in larger estimates of effect due to a poorer prognosis in non-randomly selected control groups compared with randomly selected control groups.
Conclusions Failure to use adequately concealed random allocation can distort the apparent effects of care in either direction, causing the effects to seem either larger or smaller than they really are. The size of these distortions can be as large as or larger than the size of the effects that are to be detected.

Introduction

> Observational evidence is clearly better than opinion, but it is thoroughly unsatisfactory. All research on the effectiveness of therapy was in this unfortunate state until the early 1950s. The only exceptions were the drugs whose effect on immediate mortality were so obvious that no trials were necessary, such as insulin, sulphonamide, and penicillin.[1]

"The basic idea, like most good things, is very simple."[1] Randomisation is the only means of controlling for unknown and unmeasured differences between comparison groups as well as those that are known and measured. Random assignment removes the potential of bias in the assignment of patients to one intervention or another by introducing unpredictability. When alternation or any other preset plan (such as time of admission) is used, it is possible to arrange to enter a patient into a study at an opportune moment. With randomisation, however, each patient's treatment is assigned according to the play of chance. It is a paradox that unpredictability is introduced into the design of clinical trials by using random allocation to protect against the unpredictability of the extent of bias in the results of non-randomised clinical trials.

Despite this simple logic, and many examples of harm being done because of delays in conducting randomised trials, there are limitations to the use of randomised trials, both real and imagined, and scepticism about the value of randomisation.[2–5] We believe this scepticism is healthy. It is important to question assumptions about research methods, and to test these assumptions empirically, just as it is important to test assumptions about the effects of health care. In this paper we have attempted systematically to summarise empirical studies of the relation between randomisation and estimates of effect.

Methods

We included four types of comparisons in our review: randomised clinical trials versus non-randomised clinical trials of the same intervention, randomised clinical trials versus non-randomised clinical trials across different interventions, adequately concealed random allocation versus inadequately concealed random allocation in trials, and high quality trials versus low quality trials in which the specific effect of randomisation or allocation concealment could not be separated from the effect of other methodological manoeuvres such as double blinding. Both descriptive and analytical assessments of the relation between the use of random allocation and estimates of effect are included, based on cohorts or meta-analyses of clinical trials.

We identified studies from the Cochrane Review Methodology Database,[6] other methodological bibliographies, Medline, and SciSearch, and by hand searching journals, personal communication with methodologists, and checking the reference lists of relevant articles. These searches were conducted up to July 1998. Potentially relevant citations were retrieved and assessed for inclusion independently by both authors. Disagreements were resolved by discussion.

We used the following criteria to appraise the methodological quality of included studies: Were explicit criteria used to select the trials? Did two or more investigators agree regarding the selection of

Department of Nephrology, Charité, Berlin, Germany
Regina Kunz, *registrar*
Health Services Research Unit, National Institute of Public Health, Oslo, Norway
Andrew D Oxman, *director*
Correspondence to: Dr Oxman
andrew.oxman@labmed.uio.no

BMJ 1998;317:1185–90

BMJ: first published as 10.1136/bmj.317.7167.1185 on 31 October 1998. Downloaded from http://www.bmj.com/ on 28 March 2024 by guest. Protected by copyright.

trials? Was there a consecutive or complete sample of clinical trials? Did the study control for other methodological differences such as double blinding and complete follow up? Did the study control for clinical differences in the participants and interventions in the included trials? Were similar outcome measures used in the included trials? The overall quality of each study was summarised as: no important flaws, possibly important flaws, or major flaws.

For each study one of us (RK) extracted information about the sample of clinical trials, the comparison that was made, the type of analysis undertaken, and the results, and the other checked the extracted data against the published article. The reported relation between randomisation and estimates of effect was recorded and, if possible, converted to the relative overestimation or underestimation of the relative risk reduction. We prepared tables for each type of comparison to facilitate a qualitative analysis of the extent to which the included studies yielded similar results, and heterogeneity in the included studies was explored both within and across comparisons.

In summarising the results we have assumed that evidence from randomised trials is the reference standard to which estimates from non-randomised trials are compared. However, as with other gold standards, randomised trials are not without flaws, and this assumption is not intended to imply that the true effect is known, or that estimates derived from randomised trials are always closer to the truth than estimates from non-randomised trials.

Results

We have identified 18 cohorts or meta-analyses that met our inclusion criteria, totalling 1211 clinical trials.[7–24] Efforts to develop an efficient electronic search strategy using Medline have thus far not been successful due to poor indexing. Searches for studies that cited Colditz and colleagues,[15] Miller and colleagues,[16] Chalmers and colleagues,[18] or Schulz and colleagues[19] using SciSearch yielded seven additional studies. Searches using SciSearch for studies that cited the other studies meeting our inclusion criteria did not yield any other additional studies. Exploratory hand searching of three methodological journals (*Controlled Clinical Trials*, *Statistics in Medicine*, and the *Journal of Clinical Epidemiology*) for four years (1970, 1980, 1990, and 1995) yielded a single relevant study published in 1990. The 18 included studies were published in 14 different journals. The majority of studies were identified through personal communication with methodologists and through bibliographies and reference lists.

Table 1 Randomised controlled trials (RCTs) compared with non-randomised controlled trials (non-RCTs) of the same intervention

Study	Sample (search strategy)	Comparison	Results	Direction of bias
Chalmers 1977[7]	32 controlled studies of anticoagulation in acute myocardial infarction (systematic)	RCTs with CCTs and HCTs on case fatality rate, rate of thromboembolism, and haemorrhages	Relative risk reduction for mortality overestimated by 35% in HCTs and 6% in CCTs compared with RCTs. Case fatality rate highest in HCTs (38.3%) compared with RCTs (19.6%) and CCTs (29.2%). Similar pattern for thromboembolism	Overestimation of effect
Sacks 1982[8]	Sample of 50 RCTs and 56 HCTs, assessing 6 interventions (treatment of oesophageal varices, coronary artery surgery, anticoagulation in myocardial infarction, chemotherapy for colon cancer and melanoma, and diethylstilboestrol for recurrent miscarriage) (at hand)	RCTs with HCTs on frequency of detecting statistically significant results ($P \leq 0.05$) of primary outcome and reduction of mortality	20% of the RCTs found a statistically significant benefit from the new treatment compared with 79% of the HCTs. Relative risk reduction of mortality in HCTs *v* RCTs was 0.49/0.27 (1.8) for cirrhosis, 0.68/0.26 (2.6) for coronary artery surgery at 3 years, 0.49/0.22 (2.2) for anticoagulation in myocardial infarction, and 0.67/–0.02 for diethylstilboestrol in recurrent miscarriage. Outcomes in treatment groups were similar in both designs, but outcomes in control groups were worse among historical controls	Overestimation of effect
Diehl 1986[9]	19 RCTs and 17 HCTs for 6 types of cancer (breast, colon, stomach, lung cancer, melanoma, soft tissue sarcoma) (reference lists of two textbooks)	Matching of randomised and historical controls for disease, stage, and follow up, and comparison on survival and relapse free survival	18 of 43 matched control groups (42%) varied by >10% (absolute difference in either outcome), 9 (21%) by >20%, and 2 (5%) by >30%. Survival or relapse free survival was better in RCTs compared with HCTs in 17/18 matches	Overestimation of effect
Reimold 1992[10]	6 RCTs and 6 CCTs of chinidine in atrial fibrillation (systematic)	RCTs and CCTs on maintenance of sinus rhythm 3, 6, and 12 months after cardioversion	At 3 months, beneficial effect of maintaining sinus rhythm with chinidine was 54% less in non-RCTs compared with RCTs, and was 76% less at 12 months	Underestimation of effect
Recurrent Miscarriage Immunotherapy Trialists Group 1994[11]	9 RCTs and 6 CCTs (with self selected treatment) of allogenic leucocyte immunotherapy for recurrent miscarriage (systematic)	RCTs and CCTs on live birth rate	Beneficial effect of immunotherapy on birth rate among pregnant women was 9% larger in CCTs compared with RCTs, but was 63% lower in CCTs when all women were considered	Underestimation of effect when all women considered, similar effect for pregnant women
Watson 1994[12]	4 RCTs and 6 CCTs/HCTs of oil soluble contrast media during hysterosalpingography in infertile couples (systematic)	RCTs and CCTs/HCTs on pregnancy rate	RCTs and CCTs/HCTs detected similar increases in pregnancy rates: odds ratio for RCTs 1.92 (95% CI, 1.33 to 2.68) and for CCTs/HCTs 1.92 (1.55 to 2.38)	Similar effect
Pyörälä 1995[13]	11 RCTs and 22 (not further specified) non-RCTs on hormonal therapy in cryptorchidism (systematic)	RCTs and non-RCTs on the descent of testes after therapy with luteinising hormone releasing hormone or human chorionic gonadotrophin	Success rate of descent of testes after therapy with luteinising hormone releasing hormone was 2.3 times larger in non-RCTs than in RCTs and 1.7 times larger after therapy with human chorionic gonadotrophin	Overestimation of effect
Carroll 1996[14]	17 RCTs and 19 non-RCTs (including HCTs or trials with inadequate randomisation procedures) on transcutaneous electrical nerve stimulation (systematic)	RCTs and non-RCTs on control of postoperative pain	Transcutaneous electrical nerve stimulation judged ineffective at improving postoperative pain in 85% of RCTs, while 89% of non-RCTs concluded that it did improve postoperative pain	Overestimation of effect

CCT=concurrently controlled trial; HCT=historically controlled trial.

BMJ: first published as 10.1136/bmj.317.7167.1185 on 31 October 1998. Downloaded from http://www.bmj.com/ on 28 March 2024 by guest. Protected by copyright.

Table 2 Randomised controlled trials (RCTs) compared with non-randomised controlled trials (non-RCTs) across different interventions

Study	Sample (search strategy)	Comparison	Results	Direction of bias
Colditz 1989[15]	113 studies published in 1980 comparing new interventions with old, identified in leading cardiology, neurology, psychiatry, and respiratory journals (systematic)	36 parallel RCTs, 29 randomised COTs, 46 non-randomised COTs, 3 CCTs, 5 ECTs, 9 observational studies compared for "treatment gain" (Mann-Whitney statistic), and relation between quality score and "treatment gain" assessed	All but one design achieved similar "treatment gains" (0.56-0.65). Overall, 89% of new treatments were rated as improvements, but only non-randomised COTs detected a significantly higher "treatment gain" from the new treatment compared with RCTs (P=0.004). Within RCTs, there was no correlation between quality score and "treatment gain" (P=0.18)	Inconclusive
Miller 1989[16]	188 studies comparing new surgical interventions with old, published in 1983 and identified in leading surgical journals (systematic)	81 RCTs, 15 CCTs, 27 HCTs, 91 observational studies, 7 BASs compared on "treatment gain" (Mann-Whitney), and association between treatment success and study design and the relation between quality score and treatment gains assessed	Non-significant trend towards larger "treatment gains" for new treatments on the principal disease in non-RCTs (0.56 to 0.78) than in RCTs (0.56). For treatment of complications the "treatment gain" was similar across all study designs (0.54 to 0.55) except in BASs (0.90). Within RCTs, there was no correlation between quality scores and treatment gains (P=0.7)	Inconclusive
Ottenbacher 1992[17]	Sample of 30 RCTs and 30 trials with non-random process of allocation, eg matching or HCTs (systematic search of *N Engl J Med* and *JAMA* across several medical specialties)	RCTs and non-RCTs on treatment effects as measured by standardised mean differences	No difference in treatment effect found between non-RCTs (0.23) and RCTs (0.21)	Similar effects

COT=Crossover trial; CCT=concurrently controlled trial; ECT=external control study; BAS=before and after study; HCT=historically controlled trial.

Randomised trials versus non-randomised trials of the same intervention

Table 1 summarises the eight studies comparing randomised clinical trials and non-randomised clinical trials of the same intervention. In five of the eight studies, estimates of effect were larger in non-randomised trials. Outcomes in the randomised treatment groups and non-randomised treatment groups were frequently similar, but worse outcomes among historical controls spuriously increased the estimated treatment effects. One study found comparable results for both allocation procedures, and two studies reported smaller treatment effects in non-randomised studies. In one study the smaller estimate of effect was due to a poorer prognosis for patients in the non-randomised treatment groups. The deviation of the estimates of effect for non-randomised trials compared with randomised trials ranged from an underestimation of effect of 76% to an overestimation of effect of 160%.

Randomised trials versus non-randomised trials across different interventions

The evidence from comparisons across different interventions and various study designs (randomised controlled trials and non-randomised controlled trials, crossover designs, and observational studies) is less clear (table 2). In all three studies several study designs and clinical conditions were combined and their diverse outcomes converted to a standardised effect size. There was substantial clinical heterogeneity, and there were many other factors that could distort or mask a possible association between randomisation and estimates of effect. No consistent relation between study design or quality and the magnitude of the estimates of effect was detected.

Adequately concealed allocation versus inadequately concealed allocation

Concealed random allocation to treatment—that is, blinding of the randomisation schedule to prevent subversion by the investigators or trial participants—should ensure protection against biased allocation. Chalmers and colleagues found that within randomised controlled trials failure adequately to conceal allocation was associated with larger imbalances in prognostic factors and larger treatment effects (table 3).[18] They reported a more than sevenfold overestimation of the treatment effect in trials with inadequately concealed allocation. They did not, however, control for other methodological factors in their descriptive analysis.[18] Schulz and colleagues conducted a multivariate analysis that controlled for blinding and completeness of follow up, which yielded similar results.[19] They found that inadequately concealed random allocation (for example, alternation) compared with adequately concealed random allocation (for example, assignment by a central office) resulted in

Table 3 Trials with adequately concealed allocation compared with inadequately concealed allocation

Study	Sample (search strategy)	Comparison	Results	Direction of bias
Chalmers 1983[18]	145 controlled trials of treatment for acute myocardial infarction (systematic)	Studies with different allocation schemes (non-random, non-concealed random, and concealed random allocation) on maldistribution of prognostic variables, frequency of significant outcomes, and case fatality rates	In non-RCTs, non-concealed RCTs, and RCTs with concealed allocation, the maldistribution of prognostic factors was 34%, 7%, and 3.5% respectively, frequency of significant outcomes was 25%, 11%, and 5% respectively, average relative risk reduction for mortality was 33%, 23%, and 3% respectively. Case fatality rate for control groups was 32%, 23%, and 16% and for treatment groups was 21%, 18%, and 16% respectively	Overestimation of effect
Schulz 1995[19]	250 RCTs from 33 meta-analyses (Cochrane Pregnancy and Childbirth Database)	Association between methodological features of controlled trials (allocation concealment, double blinding, and follow up), and treatment effect (odds ratio)	Treatment effect overestimated by 41% in RCTs with inadequate concealment and by 30% in RCTs with unclear adequacy of concealment compared with those with adequate concealment (P≤0.001) after adjustment for other methodological features. Studies with no double blinding overestimated treatment effect by 17% compared with double blinded studies (P =0.01). Lack of complete follow up had no influence on treatment effect (7%, P=0.32)	Overestimation of effect

RCT=Randomised controlled trial.

BMJ: first published as 10.1136/bmj.317.7167.1185 on 31 October 1998. Downloaded from http://www.bmj.com/ on 28 March 2024 by guest. Protected by copyright.

Table 4 Studies of high quality trials compared with low quality trials

Study	Sample (search strategy)	Comparison	Results	Direction of bias
Emerson 1990[20]	Sample of 7 meta-analyses with 107 primary studies where full information about quality scores was available (at hand)	Assessment of relation between quality score and (a) observed treatment difference and (b) variation of observed treatment difference	No correlation detected between either quality score and treatment difference or variation of treatment difference within each meta-analysis or in combined analysis (P=0.29)	Similar effects
Imperiale 1990[21]	Meta-analysis of 11 RCTs of steroids in alcoholic hepatitis (systematic)	Short term mortality in studies with high and low methodological quality	In studies with low quality, relative risk reduction on mortality was 86% smaller than the reduction observed in high quality studies. In studies with low quality and hepatic encephalopathy no effect was observed, while the relative risk reduction of mortality in high quality studies was 55%	Underestimation of effect
Nurmohamed 1992[22]	Meta-analysis of 35 surgical and orthopaedic RCTs on low molecular weight heparin as thromboprophylaxis (systematic)	Relative risk reduction for deep vein thrombosis and pulmonary embolism in studies of high and low methodological quality	In studies with low quality, relative risk reduction for venous thrombosis in surgical trials was 2.6 times larger, and in orthopaedic trials 1.4 times larger, than studies with high quality. Relative risk reduction for pulmonary embolus in surgical trials was 1.7 times larger, and in orthopaedic trials 2.8 times larger, than studies with high quality	Overestimation of effect
Khan 1996[23]	Meta-analysis of 9 RCTs (parallel or crossover design) evaluating the effect of anti-oestrogen treatment in male infertility (systematic)	Pregancy rates in studies with high and low methodological quality	In studies of low quality, pregnancy rate increased under treatment (odds ratio 2.6), but declined under treatment in high quality studies (0.5)	Reversal of effect
Ortiz 1998[24]	Meta-analysis of 7 RCTs on the effect of folic or folinic acid *v* placebo (systematic)	Frequency of gastrointestinal side effects in studies with high and low methodological quality	In studies with low quality there was a 43% reduction in the odds ratio of side effects (0.57) compared with a 70% reduction in studies with high quality (0.3)	Underestimation of effect

RCT=Randomised controlled trial.

estimates of effect (odds ratios) that were on average 40% larger.

High quality trials versus low quality trials

Considerable differences in the observed treatment effect were detected when the results of high quality studies were compared with those of low quality studies in the context of systematic reviews of specific health care (table 4). In these studies the estimates of effect were distorted in both directions and even caused the alarming situation of a harmful intervention associated with a reduction in pregnancies (odds ratio 0.5, on the basis of high quality studies) seeming beneficial in low quality studies (odds ratio 2.6, on the basis of low quality studies). In two meta-analyses, low quality studies consistently underestimated the beneficial effect of the intervention being evaluated by 27% to 100%, and an effective treatment could have been discarded based on the results of low quality studies.

Methodological quality

The methodological quality of the studies included in this review varied. Four studies met all of our criteria.[19 21–23] Three of these assessed the impact of bias on the effect of a specific healthcare intervention as part of a systematic review, and the analysis was performed as part of a subgroup analysis to test the robustness of the overall finding.[21–23] The other 14 studies had one or more methodological flaws including not controlling for other methodological manoeuvres[16 18 22 27] or clinical differences.[7 13–17 20 24]

Discussion

It has proved difficult to develop efficient search strategies for locating empirical methodological studies such as the ones included in this review. Although we believe it is unlikely that there are many published methodological studies such as the ones by Sacks and colleagues,[8] Schulz and colleagues,[19] Chalmers and colleagues,[18] and Emerson and colleagues[20] that we have not identified, there may be unpublished or ongoing studies like these that we have not identified, and it is likely that there are many meta-analyses that meet the inclusion criteria for this review that we have not identified. The Cochrane Library contains 428 completed reviews and 397 protocols, and there are over 1700 entries in the database of abstracts of reviews of effectiveness.[26] We have not systematically gone through all of these meta-analyses. An expanded version of this review will be published in the Cochrane Library and kept up to date through the Cochrane Empirical Methodological Studies Methods Group.[27] Additional studies will be added to the review, and any errors that are identified will be corrected.

We have not included comparisons between randomised controlled trials and cohort studies,[28] case-control studies,[29 30] or evaluations of effectiveness using large healthcare administrative databases,[3] although some of the studies in this review included observational studies. Observational studies often provide valuable information that is complementary to the results of clinical trials. For example, case-control studies may be the best available study design for evaluating rare adverse effects, and large database studies may provide important information about the extent to which effects that are expected based on randomised clinical trials are achieved in routine practice. However, it is important to remember that it is only possible to control for confounders that are known and measured in observational studies, and we should be wary of hubris and its consequences in assuming that we know all there is to know about any disease.

As with any review the quality of the data is limited by the quality of the studies that we have reviewed. Most of the studies included in the review had one or more methodological flaws. In many of the included comparisons, particularly those between randomised controlled trials and historically controlled trials, methodological differences other than randomisation may account for some of the observed differences in estimates of effect.[7–9 13 18]

Four of the studies met all of our criteria for assessing methodological quality,[19 21–23] and one study in par-

BMJ: first published as 10.1136/bmj.317.7167.1185 on 31 October 1998. Downloaded from http://www.bmj.com/ on 28 March 2024 by guest. Protected by copyright.

ticular provided strong support for the conclusion that clinical trials that lack adequately concealed random allocation produce estimates of effect that are on average 40% larger than clinical trials with adequately concealed random allocation, but that the degree and the direction of this bias varies widely.[19] This study also shows the potential contribution that systematic reviews, and notably the Cochrane Database of Systematic Reviews, can make towards developing an empirical basis for methodological decisions in evaluations of health care. Currently this empirical basis is lacking, and many methodological debates rely more on logic or rhetoric than evidence. Analyses such as the one undertaken by Schulz and colleagues, in which methodological comparisons are made among trials of the same intervention, are likely to yield more reliable results than comparisons that are made across different interventions which, not surprisingly, tend to be inconclusive.[15–17]

We have assumed that, in general, differences between randomised trials and non-randomised trials or between trials with adequately concealed random allocation and inadequately concealed random allocation are best explained by bias in the non-randomised controlled trials and inadequately concealed trials. This assumption is supported by findings of large imbalances in prognostic factors as well. However, it is possible that randomised controlled trials can sometimes underestimate the effectiveness of an intervention in routine practice by forcing healthcare professionals and patients to acknowledge their uncertainty and thereby reduce the strength of placebo effects.[4 25 31] It is also possible that publication bias can partly explain some of the differences in results observed in studies such as the one by Sacks and colleagues.[8] This would be the case if randomised trials are more likely to be published regardless of the effect size, than historically controlled trials. However, we are not aware of any evidence that supports this hypothesis, and the available evidence shows consistently that randomised trials, like other research, are also more likely to be published if they have results that are considered significant.[32–35]

Several explanations for discrepancies between estimates of effect derived from randomised trials and non-randomised trials are possible. For example, it can be argued that estimates of effect might be larger in randomised trials if the care provided in the context of trials is better than that in routine practice, assuming this is the case for the treatment group and not the control group. Similarly, strict eligibility criteria might select people with a higher capacity to benefit from a treatment, resulting in larger estimates of effect in randomised trials than non-randomised trials with less strict eligibility criteria. If, for some reason, patients with a poor prognosis were more likely to be allocated to the treatment group in non-randomised trials then this would also result in larger estimates of effect in randomised trials. Conversely, if patients with a poor prognosis were more likely to be allocated to the control group in non-randomised trials, as often seems to be the case based on the results of this review, this would result in larger estimates of effect in the non-randomised trials.

Key messages

- Empirical studies support using random allocation in clinical trials and ensuring that the allocation process is concealed—that is, that assignment is impervious to any influence by the people making the allocation
- The effect of not using concealed random allocation can be as large or larger than the effects of worthwhile interventions
- On average, failure to use concealed random allocation results in overestimates of effect due to a poorer prognosis in non-randomly selected control groups compared with randomly selected control groups, but it can result in underestimates of effect, reverse the direction of effect, mask an effect, or give similar estimates of effect
- The adequacy of allocation concealment may be a more sensitive measure of bias in clinical trials than scales used to assess the quality of clinical trials
- It is a paradox that the unpredictability of randomisation is the best protection against the unpredictability of the extent and direction of bias in clinical trials that are not properly randomised

Conclusion

Overall, this review supports using random allocation in clinical trials and ensuring that the randomisation schedule is adequately concealed. The effect of not using random allocation with adequate concealment can be as large or larger than the effects of worthwhile interventions. On average, non-randomised trials and randomised trials with inadequately concealed allocation result in overestimates of effect. This bias, however, can go in either direction, can reverse the direction of effect, or can mask an effect.

For those undertaking clinical trials this review provides support for using randomisation to assemble comparison groups.[25] For those undertaking systematic reviews of clinical trials, this review provides support for considering sensitivity analyses based on the adequacy of allocation concealment in addition to or instead of on the basis of overall quality scores, which may be less sensitive measures of bias.

As Cochrane stated: "The [randomised controlled trial] is a very beautiful technique, of wide applicability, but as with everything else there are snags."[1] Those making decisions on the basis of clinical trials need to be cautious of small trials (even when they are properly randomised) and systematic reviews of small trials both because of chance effects and the risk of biased reporting.[36 37] It is also possible to introduce bias into a trial despite allocation concealment.[19 38] Finally, even when the risk of error due to either bias or chance is small, judgments must be made about the applicability of the results to individual patients[39 40] and about the relative value of the probable benefits, harms, and costs.[41 42]

We thank Alex Jadad, Steve Halpern, and David Cowan for help in locating studies, Dave Sackett and Iain Chalmers for encouragement and advice, Mike Clarke for reviewing the manuscript,

BMJ: first published as 10.1136/bmj.317.7167.1185 on 31 October 1998. Downloaded from http://www.bmj.com/ on 28 March 2024 by guest. Protected by copyright.

Annie Britton and other colleagues for provision of their bibliographies on research methodology, and the investigators who conducted the studies we reviewed.

Contributors: RK and ADO contributed to the preparation of the protocol and the final manuscript and assessed the relevance and methodological quality of retrieved reports. RK prepared the first drafts of the protocol and the paper, undertook the majority of the searches with help from David Cowan, Steve Halpern, Alex Jadad, and collected data from the included studies. ADO checked the collected data against the original reports. Both authors will act as guarantors for the paper.

Funding: Norwegian Ministry of Health and Social Affairs.

Competing interests: None declared.

1 Cochrane AL. *Effectiveness and efficiency: random reflections on health services*. London: Nuffield Provincial Hospitals Trust, 1972:20-5.
2 Committee for Evaluating Medical Technologies in Clinical Use. *Assessing medical technologies*. Washington DC: National Academy Press, 1985:76-8.
3 US Congress, Office of Technology Assessment. *Identifying health technologies that work: searching for evidence*, OTA-H-608. Washington DC: US Government Printing Office, 1994:41-51.
4 Black N. Why we need observational studies to evaluate the effectiveness of health care. *BMJ* 1996;312:1215-8.
5 Weiss CH. Evaluation. *Methods for studying programs and policies*. 2nd ed. Upper Saddle River: Prentice Hall, 1998:229-33.
6 Clarke M, Carling C, Oxman AD, eds. Cochrane Review Methodology Database. *The Cochrane Library*. Oxford: Update Software, 1998. Issue 3.
7 Chalmers TC, Matta RJ, Smith H Jr, Kunzler AM. Evidence favoring the use of anticoagulants in the hospital phase of acute myocardial infarction. *N Engl J Med* 1977:297:1091-6.
8 Sacks H, Chalmers TC, Smith H Jr. Randomized versus historical controls for clinical trials. *Am J Med* 1982;72:233-40.
9 Diehl LF, Perry DJ. A comparison of randomized concurrent control groups with matched historical control groups: are historical controls valid? *J Clin Oncol* 1986;4:1114-20.
10 Reimold SC, Chalmers TC, Berlin JA, Antman EM. Assessment of the efficacy and safety of antiarrhythmic therapy for chronic atrial fibrillation: observations on the role of trial design and implications of drug related mortality. *Am Heart J* 1992;124:924-32.
11 Recurrent Miscarriage Immunotherapy Trialists Group. Worldwide collaborative observational study and meta analysis on allogenic leukocyte immunotherapy for recurrent spontaneous abortion. *Am J Reprod Immunol* 1994;32:55-72.
12 Watson A, Vandekerckhove P, Lilford R, Vail A, Brosens I, Hughes E. A meta-analysis of the therapeutic role of oil soluble contrast media at hysterosalpingography: a surprising result? *Fertil Steril* 1994;61:470-7.
13 Pyorala S, Huttunen NP, Uhari M. A review and meta-analysis of hormonal treatment of cryptorchidism. *J Clin Endocrinol Metab* 1995;80:2795-9.
14 Carroll D, Tramer M, McQuay H, Nye B, Moore A. Randomization is important in studies with pain outcomes: systematic review of transcutaneous electrical nerve stimulation in acute postoperative pain. *Br J Anaesth* 1996;77:798-803.
15 Colditz GA, Miller JN, Mosteller F. How study design affects outcomes in comparisons of therapy. I: medical. *Stat Med* 1989;8:441-54.
16 Miller JN, Colditz GA, Mosteller F. How study design affects outcomes in comparisons of therapy. II: surgical. *Stat Med* 1989;8:455-66.
17 Ottenbacher K. Impact of random assignment on study outcome: an empirical examination. *Control Clin Trials* 1992;13:50-61.
18 Chalmers TC, Celano P, Sacks HS, Smith H Jr. Bias in treatment assignment in controlled clinical trials. *N Engl J Med* 1983;309:1358-61.
19 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.
20 Emerson JD, Burdick E, Hoaglin DC, Mosteller F, Chalmers TC. An empirical study of the possible relation of treatment differences to quality scores in controlled randomized clinical trials. *Control Clin Trials* 1990;11:339-52.
21 Imperiale TF, McCullough AJ. Do corticosteroids reduce mortality from alcoholic hepatitis? A meta analysis of the randomized trials. *Ann Intern Med* 1990;113:299-307.
22 Nurmohamed MT, Rosendaal FR, Buller HR, Dekker E, Hommes DW, Vandenbroucke JP, et al. Low molecular weight heparin versus standard heparin in general and orthopaedic surgery: a meta-analysis. *Lancet* 1992;340:152-6.
23 Khan KS, Daya S, Jadad A. The importance of quality of primary studies in producing unbiased systematic reviews. *Arch Intern Med* 1996;156:661-6.
24 Ortiz Z, Shea B, Suarez Almazor ME, Moher D, Wells GA, Tugwell P. The efficacy of folic acid and folinic acid in reducing methotrexate gastrointestinal toxicity in rheumatoid arthritis. A meta-analysis of randomized controlled trials. *J Rheumatol* 1998;25:36-43.
25 Chalmers I. Assembling comparison groups to assess the effects of health care. *J R Soc Med* 1997;90:379-86.
26 NHS Centre for Reviews and Dissemination. Database of abstracts of reviews of effectiveness. *The Cochrane Library*. Oxford: Update Software, 1998. Issue 3.
27 Cochrane Empirical Methodological Studies Methods Group. *The Cochrane Library*. Oxford: Update Software, 1998. Issue 3.
28 Forgie MA, Wells PS, Laupacis A, Fergusson D. Preoperative autologous donation decreases allogeneic transfusion but increases exposure to all red blood cell transfusion: results of a meta-analysis. *Arch Intern Med* 1998;158:610-6.
29 Colditz GA, Brewer TF, Berkey CS, Wilson ME, Burdick E, Fineberg HV, et al. Efficacy of BCG vaccine in the prevention of tuberculosis. Meta analysis of the published literature. *JAMA* 1994;271:698-702.
30 Stieb D, Frayha HH, Oxman AD, Shannon HS, Hutchison BG, Crombie F. The effectiveness and usefulness of Haemophilus influenzae type b vaccines: a systematic overview (meta-analysis). *Can Med Assoc J 1990;142:719-32.*
31 Kleijnen J, Gøtzsche P, Kunz RH, Oxman AD, Chalmers I. So what's so special about randomisation? In: Maynard A, Chalmers I, eds. *Non-random reflections on health services research: on the 25th anniversary of Archie Cochrane's effectiveness and efficiency*. London: BMJ Publishing Group, 1997:93-106.
32 Dicksersin K, Min YI. NIH clinical trials and publication bias. *Online J Curr Clin Trials* [serial online] 1993; document No 50.
33 Dickersin K. How important is publication bias? A synthesis of available data. *AIDS Education and Prevention* 1997;9(suppl A):15-21.
34 Stern JM, Simes RJ. Publication bias: evidence of delayed publication of clinical research projects. *BMJ* 1997;315:640-5.
35 Ioannidis JPA. Effect of the statistical significance of results on the time to completion and publication of randomized efficacy trials. *JAMA* 1998;279:281-6.
36 Counsell CE, Clarke MJ, Slattery J, Sandercock PAG. The miracle of DICE therapy for acute stroke: fact or fictional product of subgroup analysis? *BMJ* 1994;309:1677-81.
37 Egger M, Davey SG, Schneider M, Minder C. Bias in meta-analysis detected by a simple, graphical test. *BMJ* 1997;315:629-34.
38 Guyatt GH, Sackett DL, Cook DJ, for the Evidence-Based Working Group. Users' guides to the medical literature, II: how to use an article about therapy or prevention, A: are the results of the study valid? *JAMA* 1993;270:2598-601.
39 Dans AL, Dans LF, Guyatt GH, Richardson S. Users' guides to the medical literature: XIV. How to decide on the applicability of clinical trial results to your patient. *JAMA* 1998;279:545-9.
40 Cochrane Methods Working Group on Applicability and Recommendations. *The Cochrane Library*. Oxford: Update Software, 1998. Issue 3.
41 Guyatt GH, Sackett DL, Cook DJ, for the Evidence-Based Working Group. Users' guides to the medical literature, II: how to use an article about therapy or prevention, B: what were the results and will they help me in caring for my patients? *JAMA* 1994;270:59-63.
42 Oxman AD, Flottorp S. An overview of strategies to promote implementation of evidence based health care. In: Silagy C, Haines A, eds. *Evidence based practice*. London: BMJ Books, 1998:91-109.

(Accepted 29 September 1998)

A memorable patient

Never underestimate a placebo

When I became a clinical student in 1952 medical wards had many patients with incurable chronic rheumatic heart disease. Mitral valv(ul)otomy had just started, and at the Central Middlesex Hospital Dr Keith Ball presented an outpatient who had been operated on. "It was a miracle for me, and once a cripple I can now breathe and walk adequately," and she then showed us what she could do. After she had gone Dr Ball explained that when the surgeon put his finger in the left atrium the mitral valve was so tight and hard that the stenosis was uncorrectable.

I learnt three lessons which helped to make me a competent clinician. Firstly, accept calmly praise (or blame) from a patient, but know in your conscience that you may not have been responsible, because, secondly, although the public may, you must not confuse sequence and consequence—post hoc, non ergo propter hoc. Thirdly, never underestimate or denigrate the tremendous efficacy of a placebo, especially a dramatic procedure such as an operation. Fortunately in 1952 I was taught also by Richard Doll and Francis Avery Jones, who were putting controlled clinical trials into mainstream British medicine and providing evidence to replace "in my experience."

J H Baron, *honorary professorial lecturer, Mount Sinai School of Medicine, New York*

BMJ: first published as 10.1136/bmj.317.7167.1185 on 31 October 1998. Downloaded from http://www.bmj.com/ on 28 March 2024 by guest. Protected by copyright.