Blinding and exclusions after allocation in randomised controlled trials: survey of published parallel group trials in obstetrics and gynaecology

Kenneth F Schulz, David A Grimes, Douglas G Altman, Richard J Hayes

Abstract

Objective—To assess the methodological quality of approaches to blinding and to handling of exclusions as reported in randomised trials from one medical specialty.

Design—Survey of published, parallel group randomised controlled trials.

Data sources—A random sample of 110 reports in which allocation was described as randomised from the 1990 and 1991 volumes of four journals of obstetrics and gynaecology.

Main outcome measures—The adequacy of the descriptions of double blinding and exclusions after randomisation.

Results—Though 31 trials reported being double blind, about twice as many could have been. Of the 31 trials only eight (26%) provided information on the protection of the allocation schedule and only five (16%) provided some written assurance of successful implementation of double blinding. Of 38 trials in which the authors provided sufficient information for readers to infer that no exclusions after randomisation had occurred, six (16%) reported adequate allocation concealment and none stated that an intention to treat analysis had been performed. That compared with 14 (27%) and six (12%), respectively, for the 52 trials that reported exclusions.

Conclusions—Investigators could have double blinded more often. When they did double blind, they reported poorly and rarely evaluated it. Paradoxically, trials that reported exclusions seemed generally of a higher methodological standard than those that had no apparent exclusions. Exclusions from analysis may have been made in some of the trials in which no exclusions were reported. Editors and readers of reports of randomised trials should understand that flawed reporting of exclusions may often provide a misleading impression of the quality of the trial.

Introduction

Randomisation, avoidance of exclusions after trial entry, and double blinding (or masking) may represent the most important methodological components for reducing bias in controlled trials.¹⁻³ Indeed, a recent analysis of 250 trials from 33 meta-analyses showed that lack of allocation concealment and double blinding were associated with exaggerated estimates of effects of treatment.⁴ Those associations provide empirical evidence of bias. The analysis did not, however, find an association between reporting exclusions after randomisation and exaggerated estimates.⁴ That implies that some authors of the trials may not have accurately reported those exclusions that had actually taken place.⁴

Though investigations have detailed the poor quality of reporting of randomisation in recently published randomised controlled trials,⁵⁻⁸ few studies have considered other methodological components. We therefore assessed the level of reporting on double blinding and handling of exclusions by evaluating 110 reports of

randomised controlled trials published during 1990-1 in four journals of obstetrics and gynaecology. We also examined the possibility that some reports of trials ostensibly without exclusions may have actually had exclusions.

Methods

We evaluated reports from four journals of obstetrics and gynaecology: the *American Journal of Obstetrics and Gynecology*, the *British Journal of Obstetrics and Gynaecology*, *Obstetrics and Gynecology*, and the *Journal of Obstetrics and Gynaecology*. We identified all 206 reports of randomised trials published in the 1990 and 1991 volumes of those journals.⁹ We included articles in which authors purported to allocate individual subjects randomly to parallel (uncrossed) groups.⁵ From those 206 reports, we identified 110 for evaluation by selecting all 20 from the *Journal of Obstetrics and Gynaecology* and by taking a random sample (computer random number generator¹⁰) of 30 from each of the three other journals.

We developed a data collection instrument to assess blinding and exclusions. For consistency of measurement one investigator (KFS) did all of the initial assessments. To examine the reproducibility of items on our questionnaire another investigator (DAG) unaware of the initial assessments independently assessed a random sample of 10 trials. For this study we analysed only items that exhibited consistency between assessors. We linked the information on allocation concealment from the earlier research⁵ to the records from this study.

An "allocation concealment" process seeks to prevent selection bias, protects the assignment sequence before and until allocation, and can always be implemented.⁵ In contrast with allocation concealment, double blinding seeks to prevent ascertainment bias, protects the sequence after allocation, and cannot always be implemented.⁵ A "double blind" trial shields participants, caregivers, and outcome evaluators from knowledge of assignment to particular treatments.¹⁰⁻¹¹ Exclusions after randomisation can occur because of eventual discovery of ineligibility of a participant, deviations from protocol, withdrawals, or losses to follow up. The valid, unbiased approach analyses all randomised participants in the originally assigned groups, regardless of compliance with protocol ("intention to treat analysis").¹⁰⁻¹⁸

Results

Most of the trials in this study (65%; 72/110) entailed pharmaceutical interventions. Overall, authors reported 31 (28%) trials as double blind, 15 (14%) as having blinded outcome assessments, one (1%) as having blinded the participants or caregivers, and 63 (57%) as not having any form of blinding. We judged double blinding to be feasible in 65 trials. Thus, the authors reported double blinding in only 48% of trials that could have been double blinded. In these 31 double blind trials, authors specified the intervention medium

Division of Sexually Transmitted Diseases Prevention, Centers for Disease Control and Prevention, Atlanta, GA 30333, USA
Kenneth F Schulz, *assistant director*

Department of Obstetrics, Gynecology, and Reproductive Sciences, University of California, San Francisco, CA 94110, USA
David A Grimes, *professor*

Medical Statistics Laboratory, Imperial Cancer Research Fund Medical Statistics Group, Centre for Statistics in Medicine, Oxford OX3 7LF
Douglas G Altman, *head*

Tropical Health Epidemiology Unit, London School of Hygiene and Tropical Medicine, London WC1E 9HT
Richard J Hayes, *head*

Correspondence to:
Dr Schulz.

BMJ 1996;312:742-4

as capsules in three (10%), tablets in 11 (35%), injection or intravenous administration in 11 (35%), double dummy¹¹ in two (6%), and various other approaches—for example, sprays, creams—in four (13%). Only 14 (45%) reports described similarity of the treatment and control regimens—for example, appearance, taste, administration.

Only eight of the 31 reports (26%) provided information on the protection of the allocation schedule. Those statements included keeping the code in a secure location or not breaking the code until the end of the study. Only five (16%) provided statements that blinding had been implemented successfully. Investigators reported testing the efficacy of blinding in only two of the 31 trials. Both found substantial unblinding of the assignments.

Authors provided insufficient information on the number of participants randomised and analysed even to infer whether exclusions had taken place in nine (8%) reports. In 52 (47%) the authors indicated exclusion of at least one participant: of these, 29 reports indicated exclusions of more than 10% of randomised participants. Although all 52 reports provided reasons for exclusions, only 34 provided reasons by assignment to treatment group. Four reports stipulated that the exclusions occurred before the assignment code was broken.

In 49 trials (45%) authors reported no apparent exclusions (table 1). Only 11 of these, however, explicitly stated that no exclusions had occurred after randomisation. Authors in the 38 other trials simply reported analysing the same number of participants as randomised, thereby implying that no exclusions had occurred. In these 38 reports with no apparent exclusions authors less commonly reported an intention to treat analysis ($P=0.037$; two tailed Fisher's exact test⁹) and less commonly reported adequate allocation concealment (relative risk 0.59; Taylor series 95% confidence interval⁹ 0.25 to 1.39) than did authors in the 52 trials with exclusions. Though the 95% confidence interval for concealment overlaps 1.0, it does not overlap the reverse results we expected from trials without exclusions. Indeed, the 38 trials with no apparent exclusions less commonly indicated adequate allocation concealment than the 11 trials that explicitly stated no exclusions (relative risk 0.35; 0.13 to 0.92).

Discussion

Only about half of the trials that could have been double blinded were done double blind. Because double blinding reduces bias⁴ these investigators should have used it more often. Double blinding, however, may not always be appropriate. In certain trials, particularly those with objective end points such as death, the anticipated gain in reduction of bias may not always be worth the additional difficulty and cost. Furthermore, in other trials—for example, surgical trials—double blinding becomes difficult or impossible. Nevertheless, even when double blinding is not feasible, blinding of assessment is advisable and often possible.

For double blinding to be successful, investigators must use and report effective methodological pro-

cedures.³ Because of scanty reporting we question the effectiveness of many of these attempts at double blinding. Indeed, when investigators tested blinding in two trials they found failures. Yet a third trial reported assigning participants by hospital number and giving two drugs to one group and only one drug to the other. Such deficient approaches thwart successful blinding. Thus, blinding seems compromised in at least three trials, although we suspect that such problems existed more widely.

Our results are consistent with other findings. Authors have often just reported their study as "double blind" and not provided much further clarifying information.¹⁹⁻²¹ Though investigators have reported compromised blinding,^{22,23} such candid reporting seems rare. We believe that authors should provide the following information on double blind trials: (a) mechanism (for example, capsules, tablets); (b) similarity of characteristics of treatment (for example, appearance, taste, administration); (c) control of allocation schedule (for example, the location of the schedule during the trial, when the code was broken for the analysis, and circumstances under which the code could be broken for individual cases); and (d) some statement on the perceived success (or failure) of the double blinding efforts. Ideally this last guideline would include tests of efficacy, but such tests may have problems with validity. Clearly, information on allocation concealment⁵ should also be provided. Hopefully these guidelines will provoke further dialogue on requirements for implementing and reporting double blinding.³

An intention to treat analysis uniquely preserves the qualities of bias reduction of randomisation. Improper handling of exclusions after randomisation introduces bias.¹⁰⁻¹⁹ Unfortunately, many trials in our study provided inadequate information on exclusions, which corroborates previous research.^{20,21,24} Authors must provide complete information on exclusions after randomisation and whether the primary analysis used the intention to treat principle.³ If authors present other analyses—for example, per protocol¹⁷—details should be provided in the report. While not unbiased, those other analyses may prove useful and illuminating.¹⁷

Investigators sufficiently diligent to produce a trial without exclusions would also seemingly be more likely to produce a trial with proper allocation and analysis. That may have happened in those trials in which authors explicitly reported no exclusions but not in those with no apparent exclusions. We suggest that some investigators who do not report exclusions may in fact have made exclusions during the trial but ignored them in the report. That supports earlier findings.⁴ Indeed, two trials with documented exclusions have published reports that indicate no exclusions.²⁵

Bias due to exclusions may exist in some of those trials that report no apparent exclusions, while many comparatively better trials may come from those in which authors reported exclusions. Some readers, however, may view trials without exclusions as less biased. Thus, some of the more biased trials may be mistaken as unbiased and many of the less biased trials as biased. Until authors comprehensively report ex-

Table 1—Two indicators of trial quality by reported information on exclusion after randomisation. Values are percentages (numbers) of trials

Indicator	No reported exclusions			At least one exclusion (n=52)
	Explicitly stated (n=11)	No apparent exclusions (n=38)	Total (n=49)	
Reported intention to treat analysis	9 (1)	0	2 (1)	12 (6)
Adequate allocation concealment	45 (5)	16 (6)	22 (11)	27 (14)

Key messages

- Only about half of the published randomised controlled trials that could have been double blinded actually were
- Poor reporting of methods used for double blinding raises concerns about how effective blinding was in many of the studies
- Some investigators may have excluded some patients from the analysis but not reported these exclusions in their published paper
- Given the apparent failure to report exclusions, some of the more biased trials may be mistaken as unbiased and vice versa
- Investigators must pay more attention to executing and reporting their approaches to blinding and the handling of exclusions

clusions after randomisation, editors, readers, and those conducting systematic reviews should all be wary of this disquieting paradox.

Funding: Division of Sexually Transmitted Diseases Prevention, Centers for Disease Control.

Conflict of interest: None.

- 1 Chalmers TC, Levin H, Sacks HS, Reitman D, Berrier J, Nagalingam R. Meta-analysis of clinical trials as a scientific discipline. I. Control of bias and comparison with large co-operative trials. *Stat Med* 1987;6:315-25.
- 2 Chalmers I, Hetherington J, Elbourne D, Keirse MJNC, Enkin M. Materials and methods used in synthesizing evidence to evaluate the effects of care during pregnancy and childbirth. In: Chalmers I, Enkin M, Keirse MJN, eds. *Effective care in pregnancy and childbirth*. Vol 1. *Pregnancy*. Oxford: Oxford University Press, 1989:39-65.
- 3 The Standards of Reporting Trials Group. A proposal for structured reporting of randomized controlled trials. *JAMA* 1994;272:1926-31.

- 4 Schulz KF, Chalmers I, Hayes RJ, Altman DG. Empirical evidence of bias: dimensions of methodologic quality associated with estimates of treatment effects in controlled trials. *JAMA* 1995;273:408-12.
- 5 Schulz KF, Chalmers I, Grimes DA, Altman DG. Assessing the quality of randomization from reports of controlled trials published in obstetrics and gynecology journals. *JAMA* 1994;272:125-8.
- 6 Altman DG, Dori CJ. Randomisation and baseline comparisons in clinical trials. *Lancet* 1990;335:149-53.
- 7 Williams DH, Davis CE. Reporting of assignment methods in clinical trials. *Control Clin Trials* 1994;15:294-8.
- 8 Sonis J, Joines J. The quality of clinical trials published in the Journal of Family Practice. *J Fam Pract* 1994;39:225-35.
- 9 Dean AG, Dean JA, Coulombier D, Burton AH, Brendel KA, Smith DG, et al. *Epi Info, version 6: a word processing, database, and statistics program for epidemiology on microcomputers*. Atlanta, Georgia: Centers for Disease Control and Prevention, 1994.
- 10 Pocock SJ. *Clinical trials: a practical approach*. Chichester: Wiley, 1983.
- 11 Altman DG. *Practical statistics for medical research*. London: Chapman and Hall, 1991.
- 12 Chalmers TC, Smith H Jr, Blackburn B, Silverman B, Schroeder B, Reitman D, et al. A method for assessing the quality of a randomized control trial. *Control Clin Trials* 1981;2:31-49.
- 13 Pocock SJ. Statistical aspects of clinical trial design. *Statistician* 1982;31:1-18.
- 14 De Jonge H. Deficiencies in clinical reports for registration of drugs. *Stat Med* 1983;2:155-66.
- 15 May GS, DeMets DL, Friedman LM, Furberg C, Passamani E. The randomized clinical trial: bias in analysis. *Circulation* 1981;4:942-51.
- 16 Sackett DL, Gent M. Controversy in counting and contributing events in clinical trials. *N Engl J Med* 1979;301:1410-2.
- 17 Lewis JA, Machin D. Intention to treat—who should use ITT? *Br J Cancer* 1993;68:647-50.
- 18 Bulpitt CJ. *Randomised controlled clinical trials*. The Hague: Martinus Nijhoff, 1983.
- 19 Mosteller F, Gilbert JP, McPeck B. Reporting standards and research strategies for controlled trials: agenda for the editor. *Control Clin Trials* 1980;1:37-58.
- 20 Gxtzsche PC. Methodology and overt and hidden bias in reports of 196 double-blind trials of nonsteroidal anti-inflammatory drugs in rheumatoid arthritis. *Control Clin Trials* 1989;10:31-56.
- 21 DerSimonian R, Charette LJ, McPeck B, Mosteller F. Reporting on methods in clinical trials. *N Engl J Med* 1982;306:1332-7.
- 22 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.
- 23 Arnett RM, Jones JS, Horgoe EO III. Effectiveness of 1% lidocaine dorsal penile nerve block in infant circumcision. *Am J Obstet Gynecol* 1990;163:1074-80.
- 24 Meinert CL, Tonascia S, Higgins K. Content of reports on clinical trials: a critical review. *Control Clin Trials* 1984;5:328-47.
- 25 Gxtzsche PC. Multiple publication of reports of drug trials. *Eur J Clin Pharmacol* 1989;36:429-32.

(Accepted 4 January 1996)

Endothelium and inelastic arteries: an early marker of vascular dysfunction in non-insulin dependent diabetes

Jonathan Goodfellow, Mark W Ramsey, Lisa A Luddington, Christopher J H Jones, Phillip A Coates, Frank Dunstan, Malcolm J Lewis, David R Owens, Andrew H Henderson

Cardiovascular Sciences Research Group, University of Wales College of Medicine, Cardiff CF4 4XN
Jonathan Goodfellow, research fellow
Mark W Ramsey, lecturer
Lisa A Luddington, research nurse
Christopher J H Jones, lecturer
Malcolm J Lewis, professor
Andrew H Henderson, professor

Diabetes Research Unit, University of Wales College of Medicine
Phillip A Coates, senior registrar (medicine)
David R Owens, senior lecturer

Department of Medical Statistics and Computing, University of Wales College of Medicine
Frank Dunstan, lecturer

Correspondence to: Dr Goodfellow.

BMJ 1996;312:744-5

Atheroma is the main cause of mortality and morbidity in patients with non-insulin dependent diabetes mellitus. Endothelial dysfunction underlies atheroma formation. Flow mediated vasodilatation is a measure of endothelial function as well as a determinant of arterial distensibility,¹ measurements of which have not previously been reported in patients with non-insulin dependent diabetes.

Subjects, methods, and results

We studied 12 symptom free, non-smoking, healthy patients (six men) with well controlled non-insulin dependent diabetes diagnosed 1-7 (mean 3.8) years previously (mean age 50 (SD 9) years, body mass index 26.9 (4.8) kg/m², serum cholesterol concentration 5.8 (0.5) mmol/l). Patients with known causes of endothelial dysfunction were excluded. All patients had insulin resistance with basal hyperinsulinaemia,² no biochemical evidence of renal impairment, and no

microalbuminuria. A group of 12 normal non-smoking subjects was matched for sex, age, body mass index, blood pressure, and serum cholesterol concentration and served as a control group.

Artery diameter (D) and distensibility were measured by high resolution ultrasonic vessel wall tracking³; blood pressure (P) by finger photoplethysmography; blood flow by ultrasound as the product of mean velocity corrected for Doppler angle and internal brachial artery diameter. Distensibility ($\Delta D/D \cdot \Delta P$) was derived from systolic distension (systolic minus diastolic intravascular diameter, ΔD) and simultaneous pulse pressure (ΔP).

Fasting subjects were studied in a temperature controlled room (21-23°C) after 15 minutes of supine rest. Measurements were made at baseline, during reactive hyperaemia 60 seconds after deflation of a wrist cuff inflated to suprasystolic pressure for five minutes (flow related, endothelium dependent vasodilatation) three minutes after 10 µg (submaximal dose) sublingual glyceryl trinitrate and three minutes after 400 µg (supramaximal dose) sublingual glyceryl trinitrate (endothelium independent vasodilatation). Haemodynamic variables returned to baseline 15 minutes after each intervention.

Groups were compared with unpaired *t* tests, apart from the glyceryl trinitrate distensibility data, which failed a test for normality and were analysed with the Mann-Whitney test. Paired *t* tests were used within groups to assess changes from baseline. *P* < 0.05 was regarded as significant.

Baseline characteristics and resting brachial artery measurements were similar between diabetic and normal subjects (heart rate 61 (SD 4) *v* 56 (8) per minute; blood pressure 127 (12)/69 (13) *v* 124 (15)/63