

Discrepancies in sample size calculations and data analyses reported in randomised trials: comparison of publications with protocols

An-Wen Chan, clinical fellow,¹ Asbjørn Hróbjartsson, senior researcher,² Karsten J Jørgensen, researcher,² Peter C Gøtzsche, director,² Douglas G Altman, director³

¹Mayo Clinic, Rochester, USA

²Nordic Cochrane Centre, Copenhagen, Denmark

³Centre for Statistics in Medicine, University of Oxford, Oxford

Correspondence to: A-W Chan
chan.anwen@mayo.edu

Cite this as: *BMJ* 2008;337:a2299
doi:10.1136/bmj.a2299

ABSTRACT

Objective To evaluate how often sample size calculations and methods of statistical analysis are pre-specified or changed in randomised trials.

Design Retrospective cohort study.

Data source Protocols and journal publications of published randomised parallel group trials initially approved in 1994-5 by the scientific-ethics committees for Copenhagen and Frederiksberg, Denmark (n=70).

Main outcome measure Proportion of protocols and publications that did not provide key information about sample size calculations and statistical methods; proportion of trials with discrepancies between information presented in the protocol and the publication.

Results Only 11/62 trials described existing sample size calculations fully and consistently in both the protocol and the publication. The method of handling protocol deviations was described in 37 protocols and 43 publications. The method of handling missing data was described in 16 protocols and 49 publications. 39/49 protocols and 42/43 publications reported the statistical test used to analyse primary outcome measures.

Unacknowledged discrepancies between protocols and publications were found for sample size calculations (18/34 trials), methods of handling protocol deviations (19/43) and missing data (39/49), primary outcome analyses (25/42), subgroup analyses (25/25), and adjusted analyses (23/28). Interim analyses were described in 13 protocols but mentioned in only five corresponding publications.

Conclusion When reported in publications, sample size calculations and statistical methods were often explicitly discrepant with the protocol or not pre-specified. Such amendments were rarely acknowledged in the trial publication. The reliability of trial reports cannot be assessed without having access to the full protocols.

INTRODUCTION

Sample size calculations and data analyses have an important impact on the planning, interpretation, and conclusions of randomised trials. Statistical analyses often involve several subjective decisions about which data to include and which tests to use, producing

potentially different results and conclusions depending on the decisions taken.¹⁻⁷ Methods of analysis that are chosen or altered after preliminary examination of the data can introduce bias if a subset of favourable results is then reported in the publication.

The study protocol plays a key role in reducing such bias by documenting a pre-specified blueprint for conducting and analysing a trial. Explicit descriptions of methods before a trial starts help to identify and deter unacknowledged, potentially biased changes made after reviewing the study results.⁸⁻¹⁰ To evaluate the completeness and consistency of reporting, we reviewed a comprehensive cohort of randomised trials and compared the sample size calculations and data analysis methods described in the protocols with those reported in the publications.

METHODS

We included all published parallel group randomised trials approved by the scientific-ethics committees for Copenhagen and Frederiksberg, Denmark, from 1 January 1994 to 31 December 1995. We defined a randomised trial as a research study that randomly allocated human participants to healthcare interventions. We manually reviewed the committees' files in duplicate to identify eligible studies as part of a separate study of outcome reporting.⁹ We confirmed journal articles for each trial by surveying investigators and searching PubMed, Embase, and the Cochrane Controlled Trials Register.⁹

For each trial, we reviewed the protocol, statistical appendices, amendments, and publications that reported the primary outcome measures. Two reviewers used electronic forms to independently extract data relating to the design, sample size calculation, and statistical analyses in duplicate; we resolved disagreements by discussion. No reviewer extracted data from both the protocol and publication for the same trial.

The pre-specified primary outcomes of our study were the proportion of trial protocols and publications that did not provide key information (described below) about sample size calculations and statistical methods

and the proportion of trials with discrepancies between the information presented in the protocol and the publication. We considered a 10% or greater difference between calculated sample sizes in the protocol and publication to be a discrepancy, as well as any qualitative or quantitative difference in other information we examined.

From publications, we noted the achieved sample size and any statements about early stopping or whether analyses were pre-specified or amended. From both protocols and publications, we categorised the design framework of the trial as superiority, non-inferiority, or equivalence (box 1). We also recorded the fundamental components of sample size calculations (table 1), the basis on which the minimum clinically important effect size (δ) and estimated event rates were derived (for example, previous literature, pilot data), and any reference to re-estimation of sample size on the basis of an interim analysis. When multiple sample size calculations were reported, we focused on the one for the primary outcome measure.

We also documented how deviations from the treatment protocol were handled in terms of both the stated type of analysis (such as intention to treat) and any additional text description (box 1). For the primary outcome measure of the trial (box 1), we recorded the method of handling missing data and the type of statistical test used. We recorded the factors and outcome measures used in any subgroup analyses; the covariates and outcome measures used in any adjusted analyses; and the number of statistical comparisons described or reported between randomised groups, excluding baseline comparisons.

Box 1 Definitions of collected data

Design framework

Superiority trial—Explicitly described as a study designed to show a difference in effects between interventions or not explicitly described as an equivalence or non-inferiority trial

Non-inferiority trial—Explicitly described as a study designed to show that one intervention is not worse than another, or a non-inferiority margin is specified, or a one sided confidence interval is presented

Equivalence trial—Explicitly described as a study designed to show that one intervention is neither inferior nor superior to another or an equivalence margin is specified

Handling of protocol deviations

Intention to treat analysis—All participants with available data are analysed in the original group to which they were randomly assigned (as randomised), regardless of adherence to the protocol. No data are excluded for reasons other than loss to follow-up

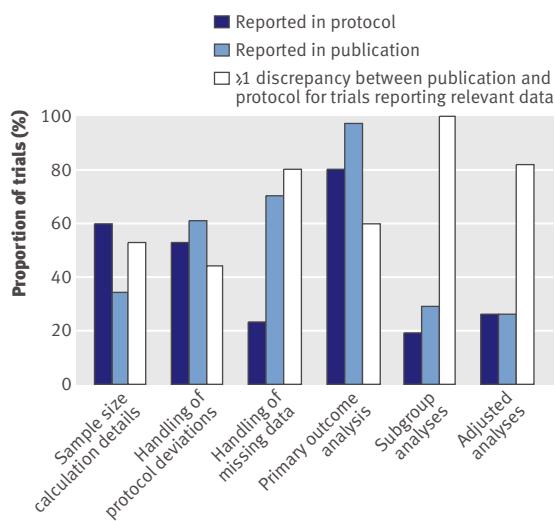
Per protocol analysis—Participants with available data are analysed as randomised provided they meet some defined level of adherence to the protocol

As treated analysis—Participants are analysed in the group corresponding to the actual intervention received (ignoring original randomisation)

Primary outcome

Main outcome(s) of interest, in the following hierarchical order:

1. Explicitly defined as primary or main
2. Outcome used in the power calculation
3. Main outcome stated in the trial objectives



Reporting of sample size calculations and data analyses in publications compared with protocols

Finally, we recorded the use of interim analyses and data monitoring boards.

RESULTS

We identified 70 parallel group randomised trials that received ethics approval in 1994-5 and were subsequently published (table 2).⁹ The median publication year was 1999 (range 1995-2003). Fifty two of the trials evaluated drugs, and 56 were funded in full or in part by industry. Most trials involved two study arms ($n=47$), multiple centres (46), and some form of blinding (49). The median achieved sample size per study arm was 66 (10th-90th centile range 13-324). The most common specialty fields were endocrinology ($n=11$), anaesthesiology (5), cardiology (5), infectious diseases (5), and oncology (5).

Sixty nine trials were designed and reported as superiority trials. One trial was stated to be an equivalence trial in the protocol but reported as a superiority trial in the publication; no explanation was given for the change.

Sample size calculation

Overall, only 11 trials fully and consistently reported all of the requisite components of the sample size calculation in both the protocol and the publication.

Completeness of reporting—An a priori sample size calculation was reported for 62 trials; 28 were described only in the protocol and 34 in both the protocol and the publication. Thirty seven protocols and 21 publications reported all of the components of the sample size calculation (figure). Individual components were reported in 74-100% of protocols and 48-75% of publications (table 1). Nine protocols provided only the calculated sample size without any further details about the calculation. Among trials that reported an estimated minimum clinically important effect size (δ), 20/53 protocols and 10/33 publications stated the basis on which the figure was derived.

Table 1 | Reporting of sample size calculations in trial protocols and publications

Component of sample size calculation	No of trials reporting each component (n=62)*		
	Protocol	Publication	Both
Name of outcome measure	51	31	28
Minimum clinically important effect size (delta)	53	33	32
Estimated event rate in each study arm†	20/27	12/16	9/14
Standard deviation for delta‡	18/24	9/14	7/12
Alpha (type 1 error rate)	50	33	31
Power	51	34	32
Calculated sample size	62	30	30
All components reported	37	21	18

*Among trials reporting at least one component of sample size calculation in trial protocol.

†For trials reporting sample size calculations using binary outcome measures.

‡For trials reporting sample size calculations using continuous outcome measures.

Comparison of calculated and actual sample sizes—Sixty two trials provided a calculated sample size in the protocol. Of these, 30 subsequently recruited a sample size within 10% of the calculated figure from the protocol; 22 trials randomised at least 10% fewer participants than planned as a result of early stopping (n=3), poor recruitment (2), and unspecified reasons (17); and 10 trials randomised at least 10% more participants than planned as a result of lower than anticipated average age (1), a higher than expected recruitment rate (1), and unspecified reasons (8). A calculated sample size was as likely to be reported accurately in the publication if there was a discrepancy with the actual sample size compared with no discrepancy (11/32 v 14/30).

Discrepancies between publications and protocols—Both the publications and the protocols for 34 trials described a sample size calculation. Overall, we noted discrepancies in at least one component of the published sample size calculation when compared with the protocol for 18 trials (figure). Publications for eight trials reported components that had not been pre-specified in the protocol, and 16 had explicit discrepancies between information contained in the publication and protocol (table 3, box 2). None of the publications mentioned any amendments to the original sample size calculation.

Protocol deviations

The specific method of handling protocol deviations in the primary statistical analysis (as defined in box 1) was named or described in 37 protocols and 43 publications (figure). Overall, the primary method described for handling protocol deviations in the publication differed from that described in the protocol for 19/43 trials; table 4 provides details. None of these discrepancies was acknowledged in the journal publication.

Thirty protocols and 33 publications used the term “intention to treat” analysis and applied a variety of definitions (table 5). Few of these protocols (n=7) and publications (3) made it explicit whether study participants were analysed in the group to which they were originally randomised. Most protocols (22) and publications (18) incorrectly excluded participants

from the intention to treat analysis for reasons other than loss to follow-up (table 5).

Missing data

The method of handling missing data was described in only 16 protocols and 49 publications (figure). Methods reported in publications differed from the protocol for 39/49 trials. Published methods were often not pre-specified in the protocol (38/49). For one trial, the protocol stipulated that missing data would be counted as failures, whereas in the publication they were excluded from the analysis.

Primary outcome analysis and overall number of tests

Fifty four trials designated at least one outcome measure as primary in the protocol (n=49) or publication (43). The statistical method for analysing the primary outcome measure was described in 39 protocols and 42 publications. Overall, 25 publications that described the statistical test for primary outcome measures differed from the protocol (figure, box 2).

The median number of between group statistical tests defined in 44 protocols was 30 (10th-90th centile range 8-218); the other 26 protocols contained insufficient statistical detail. Publications for all 70 trials reported a median of 22 (8-71) tests. Half of the protocols (n=36) and publications (34) did not define whether hypothesis testing was one or two sided. Interestingly, we found one neurology trial that used two sided P values in one publication (all P values >0.1) and a one sided P value in another (P=0.028).

Table 2 | Characteristics of published parallel group randomised trials

Characteristic	No of trials (n=70)
Intervention:	
Drug	52
Surgery/procedure	9
Counselling/lifestyle	7
Equipment	2
No of study groups:	
2	47
3	15
>3	8
Blinding:	
Blinded in some way	49
None	13
Unclear	8
Study centres:	
Single	24
Multiple	46
Funding:	
Industry only	45
Industry and non-industry	11
Non-industry only	10
None	3
Unclear	1

Subgroup analysis

Overall, 25 trials described subgroup analyses in the protocol (n=13) or publication (20). All had discrepancies between the two documents (figure, box 2). Twelve of the trials with protocol specified analyses reported only some (n=7) or none (5) in the publication. Nineteen of the trials with published subgroup analyses reported at least one that was not pre-specified in the protocol. Protocols for 12 of these trials specified no subgroup analyses, whereas seven specified some but not all of the published analyses. Only seven publications explicitly stated whether the analyses were defined *a priori*; four of these trials claimed that the subgroup analyses were pre-specified even though they did not appear in the protocol.

Adjusted analysis

Overall, 28 trials described adjusted analyses in the protocol (n=18) or publication (18). Of these, 23 had discrepancies between the two documents (figure, box 2). Twelve of the trials with protocol specified covariates reported no adjustment (n=10) or omitted at least one pre-specified covariate (2) from the published analysis. Twelve of the trials with published adjusted analyses used covariates that were not pre-specified in the protocol. Ten of these trials did not mention any adjusted analysis in the protocol, whereas two trials added new covariates to those specified in the protocol. Publications for only one trial explicitly stated whether the covariates were defined *a priori*.

Interim analyses and data monitoring boards

Interim analyses were described in 13 protocols, but reported in only five corresponding publications. An additional two trials reported interim analyses in the publications, despite the protocol explicitly stating that there would be none. A data monitoring board was described in 12 protocols but in only five of the corresponding publications.

DISCUSSION

We identified a high frequency of unacknowledged discrepancies and poor reporting of sample size calculations and data analysis methods in an unselected cohort of randomised trials. To our knowledge, this is the largest review of sample size calculations and statistical methods described in trial publications compared with protocols. We reviewed key methodological information that can introduce bias if misrepresented or altered retrospectively. Our broad sample of protocols is a key strength, as unrestricted access to such documents is often very difficult to obtain.¹¹ Previous comparisons have been limited to case reports,⁶ small samples,^{12 13} specific specialty fields,¹⁴ and specific journals.¹⁵ Other reviews of reports submitted to drug licensing agencies did not have access to protocols.^{4 16 17}

One limitation is that our cohort may not reflect recent protocols and publications, as this type of review can be done only several years after protocol submission to allow time for publication. Whether the

Box 2 Anonymised examples of unacknowledged discrepancies in sample size calculations and statistical analyses reported in publications compared with protocols

Sample size calculation

Changed delta (1)

- Outcome: disease progression or death rate
- Protocol: delta 10%; event rates unspecified
- Publication: delta 6%; event rates 16% and 10%

Changed delta (2)

- Outcome: mean number of active joints
- Protocol: delta 2.5 joints
- Publication: delta 5 joints

Changed standard deviation

- Outcome: mean symptom score
- Protocol: 1.4
- Publication: 0.49

Changed power

- Outcome: survival without disease progression
- Protocol: 90%
- Publication: 80%

Changed sample size estimate

- Outcome: thromboembolic complication rate
- Protocol: 2200
- Publication: 1500

Statistical analyses

Changed primary outcome analysis

- Outcome: global disease assessment
- Protocol: χ^2 test
- Publication: analysis of covariance

New subgroups added to publication

- Outcome: time to progression or death
- Protocol: baseline disease severity
- Publication: duration of previous treatment*, type of previous treatment*, blood count*, disease severity

Omitted covariates for adjusted analysis in publication

- Outcome: neurological score at six months
- Protocol: baseline neurological score, pupil reaction, age, CT scan classification, shock, haemorrhage
- Publication: no adjusted analysis reported

*Described explicitly as pre-specified despite not appearing in the protocol

widespread adoption of CONSORT and other reporting guidelines for publications has improved the quality of protocols or reduced the prevalence of unacknowledged amendments in recent years is unclear.¹⁸ However, our results are consistent with more recent examples of discrepancies.^{3 12 13} Furthermore, we previously found that the prevalence of publication restrictions stated in industry initiated trial protocols did not change between 1995 and 2004.¹⁹

We also acknowledge that detailed statistical analysis plans may not always be included in the application for scientific or ethical review as a result of varying

Table 3 | Discrepancies in sample size calculations reported in trial publications compared with protocols

Component of sample size calculation	No of trials with discrepancy		
	Total	Not pre-specified*	Different from protocol description
Outcome measure (n=31)†	7	3	4
Estimated delta (n=33)†	12	6	6: 3 larger in protocol; 3 larger in article
Estimated event rates (n=16)‡	3	3	0
Estimated standard deviation (n=14)§	5	2	3: 2 larger in protocol; 1 larger in article
Alpha (n=33)†	2	2	0
Power (n=34)†	9	2	7: 5 larger in protocol; 2 larger in article
Calculated sample size (n=30)†	8	0	8¶: 7 larger in protocol; 1 larger in article
Any component (n=34)**	18	8	16

*Reported in publication but not mentioned in protocol.

†Among trials reporting component in publication.

‡Among trials reporting event rates for binary outcome measures in publication.

§Among trials reporting standard deviations for continuous outcome measures in publication.

¶Greater than 10% difference in calculated sample size.

**Among trials reporting any component in publication.

standards, even though this information has a role in evaluating the validity of a study. However, this does not explain the frequent discrepancies we found between explicit descriptions in protocols and publications (box 2).

We found that sample size calculations and important aspects of statistical analysis methods were often incompletely described in protocols and publications. When reported in the publication, they were discrepant with the protocol for 44-100% of trials. The 70 trials in this study were part of a larger review that found unacknowledged changes to primary outcome measures in more than half of 102 trials,⁹ so we are not surprised to find frequent discrepancies in other aspects of study conduct.

Because the choice of parameters for sample size calculations and statistical analyses is based on somewhat subjective judgments, specifying them *a priori* in order to avoid selective reporting or revisions on the

basis of the data is important. This includes defining each component of the sample size calculation²⁰; the analysis plan for primary outcome measures⁸; the primary population for analysis²¹; and subgroup,^{22,23} adjusted, and interim analyses.²⁴ Selective reporting of a favourable subset of analyses done can lead to inaccurate interpretation of results, similar to selective publication or selective outcome reporting.^{25,26}

Even when analysis plans are reported, concerns may exist that published plans were constructed after exploration of the data. Accurate reporting of sample size calculations and data analysis methods is important not only for the sake of transparency but also because the choice of methods and the reasons for choosing them can directly influence the interpretation and conclusions of study results.¹⁻²⁷ Public access to full protocols is thus needed to reliably appraise trial publications. Several journals have recognised this principle and require submission of protocols with manuscripts.²⁸⁻³⁰

We identified two types of discrepancies: lack of pre-specification, where published information about the sample size calculation or data analysis methods did not appear in the protocol; and unacknowledged amendments, where published information differed from the protocol. Whereas the first type of discrepancy could be explained by varying standards in protocol content, the second represents an explicit change from the time of ethical approval to analysis and publication of the trial. Retrospective amendments to methods of statistical analysis can sometimes be justifiable before unblinding of the trial, but no good reason exists to misrepresent the pre-trial sample size calculation or misleadingly describe the methods as pre-specified when the protocol suggests otherwise. Substantive amendments should be submitted to ethics committees and explicitly acknowledged in the publication so that readers can judge the potential for bias.⁸

Sample size calculations

Although good reasons exist for reporting sample size calculations in publications,^{20,31} previous reviews of publications have found poor reporting of sample size

Table 4 | Discrepancies in primary method of handling protocol deviations, as reported in publications compared with protocols

Discrepancy and No	Primary method(s) described	
	Protocol	Publication
Not pre-specified in protocol*—11:		
5	None	ITT
4	None	Per protocol
1	None	ITT and per protocol
1	None	ITT and as treated
Added new method in publication†—3:		
2	Per protocol	ITT and per protocol
1	ITT	ITT and on treatment analysis
Omitted protocol-specified method‡—2:		
	ITT and per protocol	Per protocol
Changed method†—3		
1	Per protocol	ITT
1	As treated	ITT
1	ITT	Per protocol
Total discrepancies*—19		

ITT=intention to treat.

*Among 43 trials that described methods of handling protocol deviations in publication. †Among 32 trials that described methods of handling protocol deviations in both publication and protocol.

calculations and their underlying assumptions.^{14 32-36} These reviews involved specific specialty fields or journals, and only one had access to trial protocols.¹⁴

Inadequate reporting and unacknowledged changes in sample size calculations can introduce bias and lead to misinterpretation of trial results. For example, favourable results could be highlighted by changing the primary outcome measure stated to have been used to calculate sample size,⁹ or by modifying the pre-specified minimum clinically important effect size (delta), which helps to determine whether one study intervention can be declared superior or equivalent to another.^{6 31} In addition, publications of studies with poor recruitment might report a modified sample size calculation to give the appearance that a smaller sample size had originally been planned. Although methods for valid changes to sample size parameters mid-trial are emerging,³⁷ no such method was mentioned in any study in our sample.

Data analyses

Although results for primary outcome measures are key determinants of an intervention's benefits or harms profile, we found that the analysis method for primary outcome measures was often undefined in the protocol or altered in the publication. Such unacknowledged changes compound previously identified discrepancies in primary outcome measures.^{9 10}

The recommended primary analysis population for superiority trials is defined by the intention to treat principle, whereby data from all trial participants are analysed according to their randomly allocated study arm, irrespective of the degree of compliance or crossover to other interventions.³⁸ Missing data from losses to follow-up can be handled with various statistical methods that can each produce different results.³⁹ Previous reviews of high impact journals have found that a high proportion of publications provide

insufficient details about methods of handling protocol deviations or missing data.^{40 41} In addition, a comparison of publications with protocols of cancer trials revealed that intention to treat analysis was done more often than was reported in the publications.¹⁴

We found omissions from the trial protocol as well as retrospective choices of analysis populations in almost half of the publications in our cohort. This creates the potential for preferential reporting of per protocol analyses over less favourable intention to treat analyses.⁴⁵ The definition of the term intention to treat was also highly variable and often inaccurately applied in our cohort, as has been found in other reviews.⁴⁰⁻⁴² To avoid ambiguity, the study arm in which participants are analysed and the criteria for excluding participants from analyses should be explicitly defined in the protocol and publication.⁸

Likewise, decisions to examine particular subgroups or to include particular covariates in adjusted analyses may be influenced by extensive exploratory analyses of the data.²² Subgroup analyses in publications are often over-interpreted, poorly reported, and lacking in clarity about whether they were pre-specified.^{13 27 43-45} Although exploratory analyses should be identified as such, we found that assertions about analyses being pre-specified were often misleading when compared with protocols. In addition, most interim analyses described in protocols were not mentioned in the publication, which deprives readers of important information about the decisions to stop or continue a trial as planned.²⁴

Conclusions

Our findings support the need to improve the content of trial protocols and encourage transparent reporting of amendments in publications through research training. In collaboration with journal editors, trialists, methodologists, and ethicists, we have launched the SPIRIT (standard protocol items for randomised trials) initiative to establish evidence based recommendations for the key content of trial protocols.⁴⁶ Public availability of trial protocols and submissions to regulatory agencies is also necessary to ensure transparent reporting of study methods.^{16 17 47 48} Prospective trial registration is an effective means of ensuring public access to protocol information, although a limited amount of methodological information is currently recorded on registries.⁴⁹⁻⁵¹

To improve the reliability of published results, investigators should document the sample size calculations and full analysis plans before the trial is started and should then analyse the results with fidelity to the study protocol or describe major amendments in the publication.⁸ As the guardians of clinical research before study inception, scientific and ethical review committees can help to ensure that statistical analysis plans are well documented in protocols. Only with fully transparent reporting of trial methods and public access to protocols can the results be properly appraised, interpreted, and applied to care of patients.

Table 5 | Definitions of "intention to treat" analysis used in protocols and publications

	Protocols (n=30)	Publications (n=33)
Group assignment for analysis		
According to original randomisation:	5	20
Explicitly described	5	3
Assumed on basis of participant flow diagram	0	8
Assumed on basis of same participant numbers at baseline and in results	0	9
According to intervention actually received*	2	0
No description	23	13
Exclusion of participants from analysis		
Exclude losses to follow-up only	4	13
Exclude data from participants owing to†:	22	18
Inadequate adherence to treatment	19	12
Other protocol violations	6	5
Adverse events or lack of efficacy	10	1
Other reasons	6	1
Unclear reasons for exclusion	4	2

*Regardless of original randomised group assignment.

†Some trials had multiple reasons for exclusion.

WHAT IS ALREADY KNOWN ON THIS TOPIC

The results and conclusions of randomised trials are influenced by the choice of statistical analysis methods and individual components of sample size calculations

If these methodological choices are defined or altered after examination of the data, the potential for biased reporting of favourable results is substantial

WHAT THIS STUDY ADDS

Trial protocols and publications are often missing important methodological information about sample size calculations and statistical analysis methods

When described, methodological information in journal publications is often discrepant with information in trial protocols

Contributors: All authors contributed to the study design, data collection, data interpretation, and drafting of the manuscript. A-WC also did the statistical analyses and is the guarantor.

Funding sources: None.

Competing interests: A-WC chairs and AH, DGA, and PCG are members of the steering committee for the SPIRIT (standard protocol items for randomised trials) initiative, a project that aims to define key protocol content for randomised trials.

Ethical approval: The scientific-ethics committees for Copenhagen and Frederiksberg granted unrestricted access to trial protocols. The research did not undergo the usual ethical review process, as patients' data were not reviewed.

Provenance and peer review: Not commissioned; externally peer reviewed.

- 1 Psaty BM, Kronmal RA. Reporting mortality findings in trials of rofecoxib for Alzheimer disease or cognitive impairment: a case study based on documents from rofecoxib litigation. *JAMA* 2008;299:1813-7.
- 2 Porta N, Bonet C, Cobo E. Discordance between reported intention-to-treat and per protocol analyses. *J Clin Epidemiol* 2007;60:663-9.
- 3 Curfman GD, Morrissey S, Drazen JM. Expression of concern reaffirmed. *N Engl J Med* 2006;354:1193.
- 4 Melander H, Ahlgqvist-Rastad J, Meijer G, Beermann B. Evidence based medicine—selective reporting from studies sponsored by pharmaceutical industry: review of studies in new drug applications. *BMJ* 2003;326:1171-3.
- 5 Schiffner R, Schiffner-Rohe J, Gerstenhauer M, Hofstädter F, Landthaler M, Stoltz W. Differences in efficacy between intention-to-treat and per-protocol analyses for patients with psoriasis vulgaris and atopic dermatitis: clinical and pharmacoeconomic implications. *Br J Dermatol* 2001;144:1154-60.
- 6 Murray GD. Research governance must focus on research training. *BMJ* 2001;322:1461-2.
- 7 Hrachovec JB, Mora M. Reporting of 6-month vs 12-month data in a clinical trial of celecoxib. *JAMA* 2001;286:2398-400.
- 8 Altman DG, Schulz KF, Moher D, Egger M, Davidoff F, Elbourne D, et al. The revised CONSORT statement for reporting randomized trials: explanation and elaboration. *Ann Intern Med* 2001;134:663-94.
- 9 Chan AW, Hróbjartsson A, Haahr MT, Gøtzsche PC, Altman DG. Empirical evidence for selective reporting of outcomes in randomized trials: comparison of protocols to published articles. *JAMA* 2004;291:2457-65.
- 10 Chan AW, Krleža-Jeric K, Schmid I, Altman DG. Outcome reporting bias in randomized trials funded by the Canadian Institutes of Health Research. *CMAJ* 2004;171:735-40.
- 11 Chan AW, Upshur R, Singh JA, Ghersi D, Chapuis F, Altman DG. Research protocols: waiving confidentiality for the greater good. *BMJ* 2006;332:1086-9.
- 12 Hahn S, Williamson PR, Hutton JL. Investigation of within-study selective reporting in clinical research: follow-up of applications submitted to a local research ethics committee. *J Eval Clin Pract* 2002;8:353-9.
- 13 Hernandez AV, Steyerberg EW, Taylor GS, Marmarou A, Habbema JD, Maas AI. Subgroup analysis and covariate adjustment in randomized clinical trials of traumatic brain injury: a systematic review. *Neurosurgery* 2005;57:1244-53.
- 14 Soares HP, Daniels S, Kumar A, Clarke M, Scott C, Swann S, et al. Bad reporting does not mean bad methods for randomised trials: observational study of randomised controlled trials performed by the Radiation Therapy Oncology Group. *BMJ* 2004;328:22-4.

- 15 Al-Marzouki S, Roberts I, Evans S, Marshall T. Selective reporting in clinical trials: analysis of trial protocols accepted by the Lancet. *Lancet* 2008;372:201.
- 16 Hemminki E. Study of information submitted by drug companies to licensing authorities. *BMJ* 1980;280:833-6.
- 17 Turner EH, Matthews AM, Linardatos E, Tell RA, Rosenthal R. Selective publication of antidepressant trials and its influence on apparent efficacy. *N Engl J Med* 2008;358:252-60.
- 18 Moher D, Schulz KF, Altman DG. The CONSORT statement: revised recommendations for improving the quality of reports of parallel-group randomised trials. *Lancet* 2001;357:1191-4.
- 19 Gøtzsche PC, Hróbjartsson A, Johansen HK, Haahr MT, Altman DG, Chan AW. Constraints on publication rights in industry-initiated clinical trials. *JAMA* 2006;295:1645-6.
- 20 Schulz KF, Grimes DA. Sample size calculations in randomised trials: mandatory and mystical. *Lancet* 2005;365:1348-53.
- 21 Fergusson D, Aaron SD, Guyatt G, Hebert P. Post-randomisation exclusions: the intention to treat principle and excluding patients from analysis. *BMJ* 2002;325:652-4.
- 22 Rothwell PM. Treating individuals 2. Subgroup analysis in randomised controlled trials: importance, indications, and interpretation. *Lancet* 2005;365:176-86.
- 23 Brookes ST, Whitley E, Peters TJ, Mulheran PA, Egger M, Davey SG. Subgroup analyses in randomised controlled trials: quantifying the risks of false-positives and false-negatives. *Health Technol Assess* 2001;5:1-56.
- 24 Montori VM, Devereaux PJ, Adhikari NK, Burns KE, Egger CH, Briel M, et al. Randomized trials stopped early for benefit: a systematic review. *JAMA* 2005;294:2203-9.
- 25 Song F, Eastwood AJ, Gilbody S, Duley L, Sutton AJ. Publication and related biases. *Health Technol Assess* 2000;4(10):1-115.
- 26 Dwan K, Altman DG, Arnaiz JA, Bloom J, Chan AW, Cronin E, et al. Systematic review of the empirical evidence of study publication bias and outcome reporting bias. *PLoS ONE* 2008;3:e3081.
- 27 Lagakos SW. The challenge of subgroup analyses—reporting without distorting. *N Engl J Med* 2006;354:1667-9.
- 28 Jones G, Abbasi K. Trial protocols at the BMJ. *BMJ* 2004;329:1360.
- 29 McNamee D, James A, Kleinert S. Protocol review at the Lancet. *Lancet* 2008;372:189-90.
- 30 PLoS Med. Guidelines for authors. <http://journals.plos.org/plosmedicine/guidelines.php#supporting>.
- 31 Altman DG, Moher D, Schulz KF. Peer review of statistics in medical research: reporting power calculations is important. *BMJ* 2002;325:491.
- 32 Weaver CS, Leonardi-Bee J, Bath-Hextall FJ, Bath PMW. Sample size calculations in acute stroke trials: a systematic review of their reporting, characteristics, and relationship with outcome. *Stroke* 2004;35:1216-24.
- 33 Vickers AJ. Underpowering in randomized trials reporting a sample size calculation. *J Clin Epidemiol* 2003;56:717-20.
- 34 Hebert R, Wright S, Dittus R, Elasy T. Prominent medical journals often provide insufficient information to assess the validity of studies with negative results. *J Negat Results Biomed* 2002;1:1.
- 35 Freedman KB, Back S, Bernstein J. Sample size and statistical power of randomised, controlled trials in orthopaedics. *J Bone Joint Surg Br* 2001;83:397-402.
- 36 Moher D, Dulberg CS, Wells GA. Statistical power, sample size, and their reporting in randomized controlled trials. *JAMA* 1994;272:122-4.
- 37 Schafer H, Timmesfeld N, Muller HH. An overview of statistical approaches for adaptive designs and design modifications. *Biom J* 2006;48:507-20.
- 38 Lachin JM. Statistical considerations in the intent-to-treat principle. *Control Clin Trials* 2000;21:167-89.
- 39 Molenberghs G, Thijs H, Jansen I, Beunckens C, Kenward MG, Maitlinckrodt C, et al. Analyzing incomplete longitudinal clinical trial data. *Biostatistics* 2004;5:445-64.
- 40 Hollis S, Campbell F. What is meant by intention to treat analysis? Survey of published randomised controlled trials. *BMJ* 1999;319:670-4.
- 41 Baron G, Boutron I, Giraudeau B, Ravaud P. Violation of the intent-to-treat principle and rate of missing data in superiority trials assessing structural outcomes in rheumatic diseases. *Arthritis Rheum* 2005;52:1858-65.
- 42 Gravel J, Opatrný L, Shapiro S. The intention-to-treat approach in randomized controlled trials: are authors saying what they do and doing what they say? *Clin Trials* 2007;4:350-6.
- 43 Moreira ED Jr, Stein Z, Susser E. Reporting on methods of subgroup analysis in clinical trials: a survey of four scientific journals. *Braz J Med Biol Res* 2001;34:1441-6.
- 44 Parker AB, Naylor CD. Subgroups, treatment effects, and baseline risks: some lessons from major cardiovascular trials. *Am Heart J* 2000;139:952-61.

- 45 Assmann SF, Pocock SJ, Enos LE, Kasten LE. Subgroup analysis and other (mis)uses of baseline data in clinical trials. *Lancet* 2000;355:1064-9.
- 46 Chan AW, Tetzlaff J, Altman DG, Gøtzsche PC, Hróbjartsson A, Krleža-Jeric K, et al. The SPIRIT initiative: defining standard protocol items for randomised trials. *German J Evid Quality Health Care (suppl)* 2008;102:S27.
- 47 MacLean CH, Morton SC, Ofman JJ, Roth EA, Shekelle PG. How useful are unpublished data from the Food and Drug Administration in meta-analysis? *J Clin Epidemiol* 2003;56:44-51.
- 48 Turner EH. A taxpayer-funded clinical trials registry and results database. *PLoS Med* 2004;1:e60.
- 49 Laine C, Horton R, DeAngelis CD, Drazen JM, Frizelle FA, Godlee F, et al. Clinical trial registration—looking back and moving ahead. *N Engl J Med* 2007;356:2734-6.
- 50 Sim I, Chan AW, Gulmezoglu AM, Evans T, Pang T. Clinical trial registration: transparency is the watchword. *Lancet* 2006;367:1631-3.
- 51 Krleža-Jeric K, Chan AW, Dickersin K, Sim I, Grimshaw J, Gluud C. Principles for international registration of protocol information and results from human trials of health related interventions: Ottawa statement (part 1). *BMJ* 2005;330:956-8.

Accepted: 10 September 2008