Available online at www.sciencedirect.com

SCIENCE DIRECT®

Contemporary Clinical Trials 26 (2005) 331–337

Contemporary Clinical Trials

www.elsevier.com/locate/conclintrial

The effect of scientific misconduct on the results of clinical trials: A Delphi survey

Sanaa Al-Marzouki*, Ian Roberts, Tom Marshall, Stephen Evans

Department of Epidemiology and Population Health, London School of Hygiene & Tropical Medicine, 49-51 Bedford Square, London, WC1B 3DP, United Kingdom

Received 27 August 2004; accepted 14 January 2005

Abstract

Objectives: To discover what types of scientific misconduct are most likely to influence the results of a clinical trial.
Design: Delphi survey of expert opinion with three rounds of consultation.
Setting: Non-industry clinical trial "community".
Participants: Experts identified from invitees to a previous MRC consultation on clinical trials. 32 out of the 40 experts approached agreed to participate.
Results: We identified thirteen forms of scientific misconduct for which there was majority agreement (>50%) that they would be likely or very likely to distort the results and majority agreement (>50%) that they would be likely or very likely to occur. Of these, the over-interpretation of 'significant' findings in small trials, selective reporting and inappropriate subgroup analyses were the main themes.
Conclusions: According to this expert group, the most important forms of scientific misconduct in clinical trials are selective reporting and the opportunistic use of the play of chance. Data fabrication and falsification were not rated highly because it was considered that these were unlikely to occur. Registration and publication of detailed clinical trial protocols could make an important contribution to preventing scientific misconduct.

Keywords: Scientific misconduct; Clinical trial; Delphi survey

* Corresponding author.
E-mail address: sanaa.al-marzouki@lshtm.ac.uk (S. Al-Marzouki).

1551-7144/$ - see front matter © 2005 Elsevier Inc. All rights reserved.
doi:10.1016/j.cct.2005.01.011

1. Background

Scientific misconduct has been defined as behaviour by a researcher, whether intentional or not, that falls short of good ethical and scientific standards [1], and in particular can arise in the context of clinical trials. However, because the results from clinical trials are used to decide whether or not treatments are effective, decisions that may influence treatment choices for large numbers of patients, the prevention and detection of scientific misconduct in clinical trials is particularly important. Although any form of scientific misconduct can discredit the findings of a clinical trial, misconduct that distorts the estimate of the treatment effect or its precision is of special importance since it may lead to patients being given useless or harmful treatments or to patients being denied effective treatments. Nevertheless, there is currently little information about what types of scientific misconduct are most likely to distort the results of, or conclusions from, clinical trials.

This study used the Delphi methodology [2] among experts in clinical trials to provide an insight into what types of scientific misconduct are most likely to influence the results of a trial. The Delphi technique is a consensus method used to determine the extent of agreement on an issue. A panel of experts is asked to take part in a series of rounds to identify, clarify, refine, and finally to reach agreement on a particular issue. Because the panel do not meet, individuals can express their opinion without being influenced by others. In the Delphi method, anonymity of response enhances objectivity, the use of feedback through multiple iterations allows for a complete and thorough consideration and response, and the use of statistical analysis of the group response quantifies the strength of agreement and the pattern of agreement.

2. Methods

A group of 40 experts in clinical trials was assembled from the list of people invited to respond to the UK Medical Research Council (MRC) Clinical Trials for Tomorrow consultation [3]. Each expert was sent a letter explaining the aims and methods of the study and invited to take part in a Delphi survey with three rounds. Panel members were selected on the basis of their knowledge of the subject area and their willingness to be involved in research as is recommended when using the Delphi approach [4].

In the first round, each participating expert was asked to list, briefly and concisely, four suggestions about how scientific misconduct can arise in the design, conduct, analysis and reporting of a clinical trial. These suggestions were then collated and any duplicates were removed from the list in preparation for the second round.

In the second round, the list of collated suggestions was sent to each participant, whether or not they had responded to the first round. Participants were asked to rate each form of scientific misconduct on two dimensions: (1) the likelihood that it would occur in a clinical trial and (2) the likelihood that it would distort the results (i.e. have an effect on the magnitude of the treatment effect or its precision). Participants rated each suggestion on a five point scale from “very unlikely” to “very likely”. A score of one indicated that the form of misconduct would be very unlikely to occur or would be very unlikely to distort the results. A score of five indicated that that form of misconduct would be very likely to occur or would be very likely to distort the results.

For round three, a list was prepared of all the forms of misconduct, showing the frequency distributions of the scores on both dimensions. Each participant’s response in the second round was indicated under the appropriate number on the frequency distribution. Each participant was offered the

Table 1
Types of misconduct for which majority agreement was reached on the criterion of likely or very likely to distort the result, with percentages at this level of agreement and the percentage breakdown of respondents' views on the likelihood of occurrence

Types of misconduct	Percentage indicating likely or very likely to distort results	Likelihood to occur (%)				
		Very unlikely				Very likely
		1	2	3	4	5
Design						
Failure to use random allocation	92	12	68	16	0	4
Failure to specify in the protocol the main outcome measure	88	8	48	28	16	0
Inadequate allocation concealment	84	0	24	48	20	8
Different follow-up schedules in arms	80	8	40	52	0	0
Use of a cross-over where carry-over is expected	79	8	46	46	0	0
Intentional use of non-optimum comparison treatment	76	0	40	44	16	0
Precision of measurement is avoided in an equivalence trial	74	0	30	55	15	0
Inadequate blinding of outcome assessment	72	0	12	72	12	4
Inappropriate timing of measurement of treatment effects	60	4	20	68	8	0
In an equivalence trial, choice of an inappropriate outcome measure	56	0	28	56	16	0
Conduct						
Tampering with treatment packs so as to un-blind allocation	95	17	75	4	4	0
Selective withdrawals on basis of knowledge of allocation	92	8	52	28	12	0
Data falsification	92	64	32	4	0	0
Data fabrication	92	72	24	4	0	0
Treatment recognition in blinded trials	64	4	36	36	24	0
Post-hoc changes in protocol	52	0	20	56	20	4
Analysis						
Altering analysis methods until finding a significant result	100	4	28	60	8	0
Use of battery of methods of comparison to get the right answer	100	0	24	64	12	0
Altering results in knowledge of allocation	100	76	16	8	0	0
Excluding patients or results to exaggerate effects or remove adverse events	99	17	46	21	16	0
Use of primary outcome measure that was not pre-specified	96	12	48	28	12	0
Selecting covariates to bias treatment effect in a particular direction	96	16	40	32	12	0
Selective exclusion of "protocol violation outliers"	88	0	32	44	24	0

(continued on next page)

Table 1 (*continued*)

Types of misconduct	Percentage indicating likely or very likely to distort results	Likelihood to occur (%)				
		Very unlikely				Very likely
		1	2	3	4	5
Analysis						
Inappropriate subgroup analyses	88	0	8	28	48	16
Claiming equivalence by dint of failure to demonstrate a difference	88	0	8	42	38	12
Rely on biased comparisons as the primary analysis	87	0	57	30	13	0
Missing data ignored when informative	84	0	20	36	32	12
Using a different primary endpoint from that specified in the protocol	84	16	48	20	16	0
Post-hoc analysis not admitted	83	0	4	37	42	17
Trial stopped for marketing and not scientific reasons	83	0	32	45	14	9
Reducing data in a biased fashion	77	9	43	24	19	4
Incorrectly imputing values for missing data	76	4	36	44	12	4
Subgroup analyses done without interaction tests	75	0	0	25	50	25
Failure to account for 'clustering' issues (multi-level)	72	0	12	44	32	12
Fail to comply with a pre-specified analysis plan	68	0	32	48	16	4
Deviation from intention to treat analysis	68	0	8	60	24	8
Ignore data on side-effects	64	8	40	32	4	16
Fail to specify a reasonable analysis plan in advance	56	0	12	52	20	16
Use of inappropriate statistical methods	56	0	32	48	16	4
Analysis conducted by the sponsor of the trial	54	0	4	42	33	21
Inappropriate analysis for example comparison of survival time by *t*-test	52	4	32	56	8	0
Reporting						
Failure to report unfavourable results	100	0	8	56	20	16
Selective reporting of positive results or omission of adverse events data	96	0	8	32	24	36
Selective reporting based on *p*-values	92	0	0	20	64	16
Report of subgroup without reference to wide study	92	0	48	28	24	0
Pos hoc analyses reported as a main conclusion	92	0	32	44	24	0
Negative or detrimental studies not published	88	0	8	24	28	40
Over-interpretation of 'significant' findings in small trials	87	0	0	17	50	33
Putting undue stress on results from subgroup analysis	84	0	4	28	48	20
Selective reporting of (i) subgroups (ii) outcomes (iii) time points	80	0	4	32	40	24

Table 1 (*continued*)

Types of misconduct	Percentage indicating likely or very likely to distort results	Likelihood to occur (%)				
		Very unlikely				Very likely
		1	2	3	4	5
Reporting						
Report of single variable where multiple variables assessed and not reported	68	0	20	52	20	8
Failure to report results or long delay in reporting	68	0	16	24	24	36
Clinically important effect sizes may be declared to suit results	63	0	12	63	17	8
Poor use of figures which mislead/distort results	60	0	28	56	12	4
Unjustified extrapolation	58	0	17	46	33	4
Selective reporting of outcomes in the abstract	56	0	0	24	44	32
Conclusion drawn that cannot be linked with evidence provided in report	56	4	16	44	20	16
Reporting under control of sponsor	56	0	20	64	8	8
Claim an analysis is by "intention-to-treat" when it is not	52	4	24	48	12	12
Giving incomplete information about analyses with non significant results	52	0	4	40	32	24

opportunity to change his or her response in the light of the group's opinion by ticking a new value for the score or if they did not wish to change their opinion to tick the same number as before.

For the analyses, majority agreement was considered to have been achieved if more than half of the expert group gave the same score. Forms of misconduct for which there was majority agreement that it would be likely (score 4) or very likely (score 5) to distort the results of a clinical trial (these two scores being combined for this purpose) were listed with the distribution of opinions on the likelihood that this form of misconduct would actually occur.

3. Results

Of the 40 experts invited to take part, 32 agreed to participate in the study, of whom 26 (81%), 27 (84%), and 25 (78%) completed rounds one, two and three, respectively. The 26 respondents in round one generated a list of 84 suggestions for the design stage of clinical trials, 93 suggestions for the conduct stage, 88 suggestions for the analysis stage and 85 suggestions for the report stage. Editing and combining similar items reduced the list to 35 suggestions (design), 30 suggestions (conduct), 36 suggestions (analysis) and 42 suggestions (reporting).

At the end of the third round, there was majority agreement that 60 forms of scientific misconduct were likely or very likely to distort the results of a clinical trial (Table 1). The types of scientific misconduct for which there was majority agreement that they would be likely or very likely to distort the results and majority agreement that they would be likely or very likely to occur are shown in Table 2. Of

Table 2
Types of misconduct for which there was majority agreement (>50%) that they would be likely or very likely to distort the results, and that they would be likely or very likely to occur

Types of misconduct	Indicating likely or very likely to occur (%)
Over-interpretation of 'significant' findings in small trials	83
Selective reporting based on *p*-values	80
Selective reporting of outcomes in the abstract	76
Subgroup analyses done without interaction tests	75
Negative or detrimental studies not published	68
Putting undue stress on results from subgroup analysis	68
Inappropriate subgroup analyses	64
Selective reporting of (i) subgroups (ii) outcomes (iii) time points	64
Selective reporting of positive results or omission of adverse events data	60
Failure to report results or long delay in reporting	60
Post-hoc analysis not admitted	59
Giving incomplete information about analyses with non significant results	56
Analysis conducted by the sponsor of the trial	54

the 13 types of misconduct shown in Table 2 the most likely to occur was over-interpretation of 'significant' findings in small trials, while selective reporting and inappropriate subgroup analyses were the main themes, these being given as likely to occur by more than three quarters of the respondents.

4. Discussion

This study used an expert consensus approach to determine what experts in clinical trials believe are the most important forms of scientific misconduct in clinical trials. We had specified a-priori that the criterion for important in this context would be forms of misconduct believed to occur commonly and to distort the trial results. The results fall into two main categories: selective reporting of trial results and inappropriate subgroup analyses.

The main strength of the Delphi technique is that it optimises input from respondents and minimises the bias that can be encountered in face to face group interaction. In this case, each expert offered their opinions freely and without any peer pressure from others in the expert group. The expert panel was chosen because of their knowledge and experience in the conduct of clinical trials. There are no recommendations regarding the most appropriate panel size for the Delphi technique with typical panel sizes varying between 10 and several hundred members, nor are there any recommendations concerning the sampling techniques [5]. The Delphi technique is qualitative approach and although we believe it was an appropriate method for eliciting the opinions of the particular group of experts chosen, the extent to which our results can be generalised is open to question.

A limitation of this study was that some of the suggestions elicited in the first round were vague or ambiguous. As a result, it was difficult to accurately exclude duplicates and so the list that was used in the second and third Delphi rounds was somewhat repetitive. On the other hand, the consistent high ranking of selective reporting and inappropriate subgroup analyses does suggest that we have accurately identified the most important issues.

Although there has been considerable attention in the scientific literature on the problems of data fabrication and data falsification these were absent from our list of the most important forms of misconduct because there was majority agreement that these problems were very unlikely to occur. Our results suggest that selective reporting and the opportunistic use of the play of chance (inappropriate subgroup analyses) are more important considerations in ensuring that patients receive only effective treatments. Indeed, the two problems can be closely related. Multiple post-hoc subgroup analysis with selective reporting might easily result in authors making exaggerated subgroup claims about treatment effectiveness [6].

A publicly accessible inventory of trial protocols that include a clear description of the statistical analysis plan is a potential solution to the problems of selective reporting and subgroup analyses. Such an initiative is already underway and was given further impetus earlier this year when the UK NHS joined the worldwide effort to register clinical trials at inception [7]. This could be combined with rigorous and thorough statistical review in the peer review process of clinical trials to ensure that the subgroup analyses undertaken and reported were those specified in the protocol. Future research will need to assess the extent to which this initiative has been successful.

Acknowledgements

The authors thank the members of the expert panel for their support with this study: Kamran Abbasi, Douglas Altman, David Braunholtz, Phil Edwards, Diana Elbourne, Lelia Duley, Emma Hall, Julian Higgins, John Imeson, Chris Jennison, David Jones, Mike Kenward, Betty Kirkwood, John Lewis, Gordon Murray, Chris Palmer, Tim Peters Ruth Pickering, Peter Rothwell, Peter Sandercock, Stephen Senn, Haleema Shakur, Anne Truesdale, Paula Williamson.

References

[1] Nimmo WS. Joint consensus conference on misconduct in biomedical research. Proc R Coll Phys Edinb 2000; 30(Supplement 7).
[2] Linstone H, Turoff M. The Delphi Method: Techniques and Applications. Addison-Wesley; 1975.
[3] M.R.C. Clinical Trials for Tomorrow. London: Medical Research Council; 2003.
[4] Erlandson DA, Harris EL, Skipper BL, Allen SD. Doing Naturalistic Inquiry. A Guide to Methods. London: Whurr Publishers; 1993.
[5] Reid NG. The Delphi techniques: its contribution to the evaluation of professional practice. In: Ellis R, editor. Professional Competence and Quality Assurance in the Caring Professions. Beckenham, Kent: Croome-Helm, 1988.
[6] Pocock SJ, Assmann SE, Enos LE, Kasten LE. Subgroup analysis, covariate adjustment and baseline comparisons in clinical trials reporting: current practice and problems. Stat Med 2002;21:2917–30.
[7] Staessen JA, Bianchi G. Registration of trials and protocols. Lancet 2003;362:1009–10.