An Exploratory Study of Statistical Assessment of Papers Published in the *British Medical Journal*

Martin J. Gardner, PhD, Jane Bond, MSc

Statistical assessment of papers submitted to the *British Medical Journal* has increased to some 300 papers annually. The assessment produces a recommendation to the editor on each paper from a statistical viewpoint together with a completed checklist that indicates the quality of certain important features. This exploratory study was aimed at monitoring the process. It reports a comparison of checklist answers on 45 papers as originally submitted with those on the papers as subsequently published. Of the 45 papers, only 5 (11%) were considered statistically acceptable at submission, but this increased to 38 (84%) after publication. Revisions had not been made adequately in 4 of the 7 unsatisfactory published papers, and the 3 others were thought to be of dubious validity. A major omission from at least 28 papers was information on sample size calculations. It is concluded that statistical assessment is beneficial but that further efforts by authors and assessors could make it even more effective.

(*JAMA*. 1990;263:1355-1357)

THE FIRST important analysis of the quality of statistics to appear in medical journals was published in the *Journal of the American Medical Association*.[1] This article, based on a survey of publications in 10 different journals, suggests that, from a statistical viewpoint, 41 (28%) of 149 analytic studies were acceptable, 7 (5%) should have been rejected, and 101 (68%) should have been revised before publication. In the same article, Schor and Karten report the outcome of an experimental program of statistical review of manuscripts submitted to 1 journal. Of 514 such manuscripts, 161 were published during the period of the study, with 119 (74%) being deemed after publication to be statistically acceptable. The authors also report a further study of random samples of papers published before and after the inception of this program that showed an increase from 35% to 70% in statistical acceptability. Overall, this is an indication that much improvement was needed and that much of it was possible. Since this report, a large number of studies on the quality of statistics in medical journals have been carried out.[2-10] Altman[11] provides a useful overview of the findings from some of these surveys.

From the MRC Environmental Epidemiology Unit (Dr Gardner) and the Department of Statistics (Ms Bond), University of Southampton, and Southampton General Hospital (Dr Gardner), Southampton, England.

Presented at The First International Congress on Peer Review in Biomedical Publication, Chicago, Ill, May 10-12, 1989.

Reprint requests to MRC Environmental Epidemiology Unit, Southampton General Hospital, Southampton SO9 4XY, England (Dr Gardner).

The *British Medical Journal* has for a long time, and progressively through the tenure of its current editor,[12] recognized the importance of statistical aspects of medical research for a valid basis to scientific development.[13-15] The increased statistical involvement is shown in the Figure, indicating that now some six submitted manuscripts are assessed statistically per week. This process includes submitted papers for which subject-matter reviewers or the editorial board believe that statistical assessment is required. In addition, over the last 2 years a statistician has been attending *British Medical Journal* "hanging committee" meetings 2 weeks per month.

The statistical review covers all aspects of the study and typically takes about 2 hours. It contains a written report that gives a conclusion about the paper's acceptability for publication and a completed "checklist" on major statistical aspects (Tables 1 and 2). Two separate checklists are being used: one for the general type of submitted papers and the other specifically for papers that report clinical trials.[16]

The limited objective of this study was to gather preliminary information on whether submitted papers that subsequently are published in the *British Medical Journal* are improved by the current approach to statistical assessment.

METHODS

Because the study was intended to be exploratory and to serve as a baseline for future, more detailed investigation, only a restricted period was covered. In total, 25 general papers published from January to March 1988, and 20 papers that reported clinical trials published from January to June 1988 are included. These 45 papers are all those that were published during the periods that had been assessed statistically on submission—averaging about 3 published papers per week.

The originally submitted manuscripts together with the statistical reviewers' comments and checklists were obtained from the archives. Review of the published articles was carried out by one of us (J.B.), who had not been involved in the statistical assessment procedure previously. The study findings given herein are a comparison of the before and after checklist responses.

RESULTS

Submitted Papers

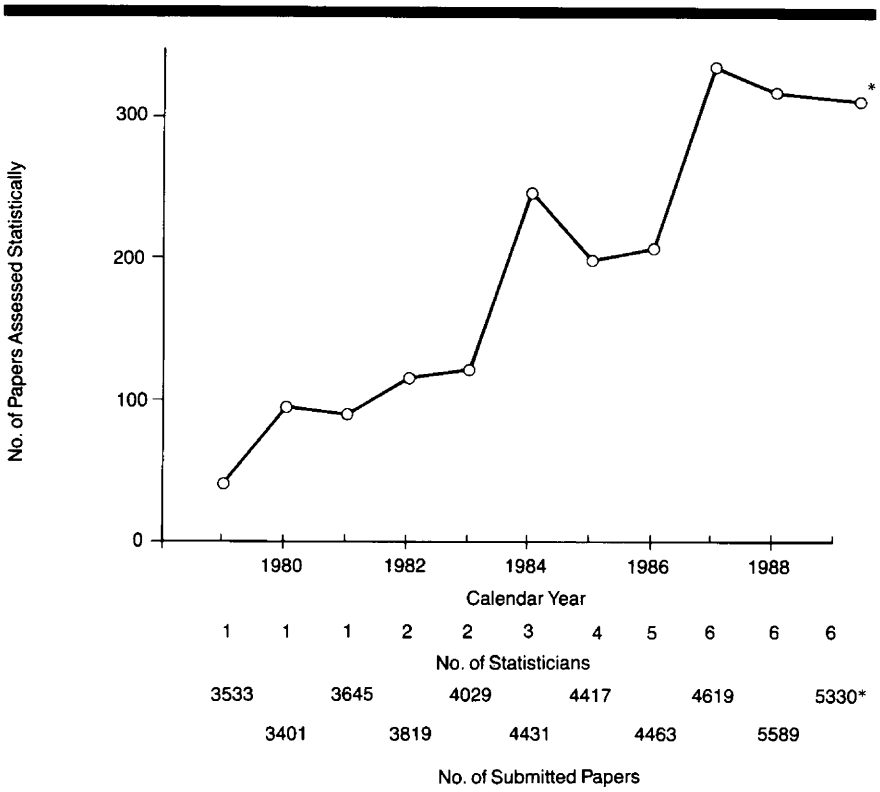
Table 1 shows the findings on the checklist for the 25 general papers. It can be seen that aspects of design and conduct were well covered, the major exception being the lack of reporting of a prestudy calculation of sample size requirements. Other unsatisfactory aspects included the lack of confidence intervals for the main results, although these had been requested from authors previously.[17]

Table 2 shows the findings on the checklist for the 20 papers on clinical trials. Among features of design and conduct, those that were least satisfactory were the lack of mention of the method of randomization and the lack of a prestudy sample size calculation. Aspects of analysis and presentation were thought to be of a lower standard overall, including the omission of confidence intervals.

Only 5 of 45 submitted papers were considered suitable for publication from a statistical viewpoint, though for most others it was thought possible to make them suitable with appropriate revision.

Published Papers

In comparing submitted and published papers, attention has been given



Number of papers assessed statistically, number of statisticians involved, and number of papers summitted to the *British Medical Journal*, by year. Asterisk indicates that the numbers were estimated, November 20, 1989.

Table 1.—Responses to Checklists on Submitted Manuscripts and Published Articles for 25 'General' Papers

	Response					
	Submitted Manuscripts			Published Articles†		
	Yes	Unclear	No*	Yes	Unclear	No
Design features						
1. Was the objective of the study sufficiently described?	25	0	0	...	...	...
2. Was an appropriate study design used to achieve the objective?	17	8	0	5	3	0
3. Was there a satisfactory statement given of source of subjects?	19	2	4	5	0	1
4. Was a prestudy calculation of required sample size reported?	0	0	15 (10)	1	0	14
Conduct of study						
5. Was a satisfactory response rate achieved?	18	5	1 (1)	6	0	0
Analysis and presentation						
6. Was there a statement adequately describing or referencing all statistical procedures used?	15	0	10	6	0	4
7. Were the statistical analyses used appropriate?	7	10	8	13	4	1
8. Was the presentation of statistical material satisfactory?	7	0	18	13	0	5
9. Were confidence intervals given for the main results?	3	0	21 (1)	11	0	10
10. Was the conclusion drawn from the statistical analysis justified?	11	13	1	10	3	1
Recommendation on paper						
11. Is the paper of acceptable statistical standard for publication?	3	0	22	18	0	4
12. If "No" to question 11, could it become acceptable with suitable revision?	20	1	1 (3)	0	1	1
Total	**145**	**39**	**101 (15)**	**88**	**11**	**41**

*Numbers in parentheses are the number of manuscripts for which no response was included.
†Only for responses of "unclear" or "no" on submitted manuscripts.

particularly to questions for which either "unclear" or "no" answers were given on the checklists for the submitted manuscripts.

Table 1 shows the answers given by our reviewer (J.B.) for these questions on the general papers. Some improvements are seen throughout and, overall, 88 (63%) of 140 originally "unclear" or "no" answers became "yes" after revision and publication. Two aspects, sample size determination and confidence interval reporting, nevertheless, still were poor.

Table 2 shows the answers given by our reviewer (J.B.) for the published articles on clinical trials. Again, improvements are seen in general, and, overall, 98 (60%) of 164 originally "unclear" or "no" answers became "yes" after revision and publication. There were, however, two exceptions: the lack of description of sample size determination and method of randomization.

In total, 7 of 45 published articles were considered not to be of an acceptable statistical standard.

COMMENT

The impression gained from this exploratory study of the current process of statistical assessment for the *British Medical Journal* is mainly, although not totally, encouraging. A fuller analysis, including an evaluation of the statisticians' written reports, is available.[18] The study itself has been limited in size and scope, and ideally requires extension and repetition.

The worst feature shown in Tables 1 and 2 is the continued lack of reporting of sample size calculations. This was particularly surprising for clinical trials, although it agrees with previous findings.[9] Other poor features were lack of description of the method of randomization and lack of confidence intervals for the main results. The former is seeking information such as whether random numbers were generated by tables or a computer algorithm, whether sealed envelopes were used and who opened them, or whether the randomization was simple or balanced. The omission of confidence intervals was corrected fully in papers on clinical trials but not in general papers.

Three published studies were regarded to be of dubious validity from a statistical viewpoint, each containing problems pointed out in the original assessments. In one it was unclear whether the control subjects (school friends of the case subjects) in a case-control study were appropriate, in another the response rate was very low (57%), and in the third only a partially randomized design was used. These

Table 2.—Responses to Checklists on Submitted Manuscripts and Published Articles for 20 'Clinical Trials' Papers

	Response					
	Submitted Manuscripts			Published Articles†		
	Yes	Unclear	No*	Yes	Unclear	No
Design features						
1. Was the objective of the trial sufficiently described?	20	0	0	...	...	...
2. Was there a satisfactory statement given of diagnostic criteria for entry to trial?	19	0	1	0	0	1
3. Was there a satisfactory statement given of source of subjects?	16	1	3	3	0	1
4. Were concurrent controls used (as opposed to historical controls)?	20	0	0	...	...	...
5. Were the treatments well defined?	19	1	0	0	1	0
6. Was random allocation to treatment used?	17	0	3	0	0	3
7. Was the method of randomization described?	7	0	10 (3)	2	0	8
8. Was there an acceptably short delay from allocation to commencement of treatment?	9	6	0 (5)	4	2	0
9. Was the potential degree of blindness used?	11	8	0 (1)	4	4	0
10. Was there a satisfactory statement of criteria for outcome measures?	16	1	1 (2)	1	0	1
11. Were the outcome measures appropriate?	19	1	0	1	0	0
12. Was a prestudy calculation of required sample size reported?	5	1	14	0	1	14
13. Was the duration of posttreatment follow-up stated?	19	1	0	1	0	0
Conduct of trial						
14. Were the treatment and control groups comparable in relevant measures?	14	4	0 (2)	1	3	0
15. Were a high proportion of the subjects followed up?	16	1	3	1	1	2
16. Did a high proportion of subjects complete treatment?	18	1	1	0	1	1
17. Were the dropouts described by treatment/control groups?	12	0	4 (4)	4	0	0
18. Were side effects of treatment reported?	17	0	2 (1)	0	0	2
Analysis and presentation						
19. Was there a statement adequately describing or referencing all statistical procedures used?	7	0	13	13	0	0
20. Were the statistical analyses used appropriate?	5	3	12	12	1	2
21. Were prognostic factors adequately considered?	10	3	4 (3)	2	1	4
22. Was the presentation of statistical material satisfactory?	5	0	15	12	0	3
23. Were confidence intervals given for the main results?	4	0	15 (1)	15	0	0
24. Was the conclusion drawn from the statistical analysis justified?	8	11	1	7	5	0
Recommendation on paper						
25. Is the paper of acceptable statistical standard for publication?	2	0	18	15	0	3
26. If "no" to question 25, could it become acceptable with suitable revision?	17	0	1 (2)	0	0	1
Total	**332**	**43**	**121 (24)**	**98**	**20**	**46**

*Numbers in parentheses are the number of manuscripts for which no response was included.
†Only for earlier responses of "unclear" or "no" on submitted manuscripts.

are instances where statistical review clashed with editorial decision to publish, presumably on other aspects of the papers and topics; the area needs further discussion.

In four published articles the suggested revisions had not been carried out adequately. The main concerns related to the methods of analysis, including the use of a nonparametric rather than parametric test, not analyzing a clinical trial on an "intention-to-treat" basis, and inappropriate analysis of repeated measurements.

The following points emerge from this study. Possibly more papers should be rereviewed by a statistician after revision. Constraints are time and resources: how to allocate best the available statistical manpower. In some instances, authors are referred specifically to local statistical help, and if this is obtained, rereview possibly should be less relevant. Action on unsatisfactory checklist items should be clarified and made mandatory. Finally, not all authors adhere to sound statistical principles during preparation of their studies and manuscripts. More emphasis should be placed on the use of published guidelines.[19,20]

References

1. Schor S, Karten I. Statistical evaluation of medical journal manuscripts. *JAMA*. 1966;195:145-150.
2. Lionel NDW, Herxheimer A. Assessing reports of therapeutic trials. *Br Med J*. 1970;3:637-640.
3. Gore SM, Jones IG, Rytter EC. Misuse of statistical methods: critical assessment of articles in *BMJ* from January to March 1976. *Br Med J*. 1977;1:85-87.
4. White SJ. Statistical errors in papers in the *British Journal of Psychiatry*. *Br J Psychiatry*. 1979;135:336-342.
5. DerSimonian R, Charette LJ, McPeek B, Mosteller F. Reporting on methods in clinical trials. *N Engl J Med*. 1982;306:1332-1337.
6. Emerson JD, Colditz GA. Use of statistical analysis in the *New England Journal of Medicine*. *N Engl J Med*. 1983;309:709-713.
7. Gardner MJ, Altman DG, Jones DR, Machin D. Is the statistical assessment of papers submitted to the *British Medical Journal* effective? *Br Med J*. 1983;286:1485-1488.
8. Avram MJ, Shanks CA, Dykes MHM, Ronai AK, Stiers WM. Statistical methods in anesthesia articles: an evaluation of two American journals during two six-month periods. *Anesth Analg*. 1985;64:607-611.
9. Pocock SJ, Hughes MD, Lee RJ. Statistical problems in the reporting of clinical trials: a survey of three medical journals. *N Engl J Med*. 1987;317:426-432.
10. Murray GD. The task of a statistical referee. *Br J Surg*. 1988;75:664-667.
11. Altman DG. Statistics in medical journals. *Stat Med*. 1982;1:59-71.
12. Lock S. *A Difficult Balance: Editorial Peer Review in Medicine*. London, England: The Nuffield Provincial Hospitals Trust; 1985.
13. Swinscow TDV. *Statistics at Square One*. 8th ed. London, England: British Medical Association; 1983.
14. Gore SM, Altman DG. *Statistics in Practice*. London, England: British Medical Association; 1982.
15. Gardner MJ, Altman DG, eds. *Statistics With Confidence*. London, England: British Medical Journal; 1989.
16. Gardner MJ, Machin D, Campbell MJ. Use of check lists in assessing the statistical content of medical studies. In: Gardner MJ, Altman DG, eds. *Statistics With Confidence*. London, England: British Medical Journal; 1989:101-108.
17. Langman MJS. Towards estimation and confidence intervals. *Br Med J*. 1986;292:716.
18. Bond J. *An Aspect of Statistical Refereeing in the British Medical Journal*. Southampton, England: University of Southampton; 1988. Thesis.
19. Altman DG, Gore SM, Gardner MJ, Pocock SJ. Statistical guidelines to contributors to medical journals. In: Gardner MJ, Altman DG, eds. *Statistics With Confidence*. London, England: British Medical Journal; 1989:83-100.
20. Bailar JC, Mosteller F. Guidelines for statistical reporting in articles for medical journals. *Ann Intern Med*. 1988;108:266-273.