Statistics in Medicine

Use of check lists in assessing the statistical content of medical studies

M J GARDNER, D MACHIN, M J CAMPBELL

Abstract

Two check lists are used routinely in the statistical assessment of manuscripts submitted to the "BMJ." One is for papers of a general nature and the other specifically for reports on clinical trials. Each check list includes questions on the design, conduct, analysis, and presentation of studies, and answers to these contribute to the overall statistical evaluation.

Only a small proportion of submitted papers are assessed statistically, and these are selected at the refereeing or editorial stage. Examination of the use of the check lists showed that most papers contained statistical failings, many of which could easily be remedied.

It is recommended that the check lists should be used by statistical referees, editorial staff, and authors and also during the design stage of studies.

Introduction

The *British Medical Journal* uses two check lists to evaluate the statistical aspects of medical studies. These check lists have been developed during statistical assessment of papers submitted to the journal[1] and have been influenced by others published previously.[2-5] One check list is intended for all studies other than clinical trials and, because of this non-specific application, is limited in detail. The second is for clinical trials and includes questions concerned with randomised or non-randomised treatment or intervention comparisons. Information on the principles behind the questions may be found, for example, in the above publications[1-5] or in the statistical guidelines of Altman *et al.*[6]

Uses of the check lists

The check lists may be used at different stages of manuscript assessment and study development.

Refereeing is difficult and time consuming,[7-10] but submitted papers clearly require subject matter referees to judge their merit within the medical specialty. Many reports, however, have some statistical content which may be outside the expertise of these particular referees and warrant separate assessment. Although the relevant considerations for this may be clear in a statistician's mind, a list of items to check and respond to serves as a useful reminder. These answers serve as the backbone for the statistician's recommendation on the paper and are supplemented usually with written comments.

Editorial staff find a check list helpful in obtaining a summary view on a paper. Because of the fixed format they can develop a familiarity which allows more rapid evaluation than from a textual report. The latter will generally be needed as well, but will be shorter than a report without the check list.

Authors receiving a copy of the completed check list from the editor can see where their paper was thought to be statistically unsatisfactory—if that is the case. Suggestions for improvements will usually be given in the report if revision is suggested. Alternatively, problems with the design or conduct of the study making the paper unsuitable for publication will be pointed out; some examples are given by Vaisrub.[11]

Planners of studies can be guided by the check lists, which indicate the need to consider relevant statistical aspects during development of protocols. Detailed advice may have to be sought from a statistician or in appropriate publications. Referral to the check lists should also improve the description of the statistical aspects of studies in submitted papers.

BMJ Ref No: _____ Date of Review: _____

Design Features			
1. Was the objective of the study sufficiently described?	Yes	Unclear	No
2. Was an appropriate study design used to achieve the objective?	Yes	Unclear	No
3. Was there a satisfactory statement given of source of subjects?	Yes	Unclear	No
4. Was there a power based assessment of adequacy of sample size?	Yes	Unclear	No
Conduct of Study			
5. Was a satisfactory response rate achieved?	Yes	Unclear	No
Analysis and Presentation			
6. Was there a statement adequately describing or referencing all statistical procedures used?		Yes	No
7. Were the statistical analyses used appropriate?	Yes	Unclear	No
8. Was the presentation of statistical material satisfactory?		Yes	No
9. Were confidence intervals given for the main results?		Yes	No
10. Was the conclusion drawn from the statistical analysis justified?	Yes	Unclear	No
Recommendation on Paper			
11. Is the paper of acceptable statistical standard for publication?		Yes	No
12. If "No" to Question 11, could it become acceptable with suitable revision?		Yes	No

Reviewer: _____

FIG 1—*Check list for statistical review of general papers for* BMJ.

MRC Environmental Epidemiology Unit (University of Southampton), Southampton General Hospital, Southampton SO9 4XY
M J GARDNER, BSC, PHD, professor of medical statistics

Community Medicine and Medical Statistics, University of Southampton Medical School, Southampton General Hospital, Southampton SO9 4XY
D MACHIN, MSC, senior lecturer in medical statistics
M J CAMPBELL, BA, PHD, lecturer in medical statistics

Correspondence to: Professor Gardner.

Outline of check lists

GENERAL CHECK LIST

Aspects covered by the general check list include design, conduct, analysis, and presentation of studies (fig 1). For each question "yes" or "no" answers are sought, but in some cases "unclear" is allowed, though its use should be minimal.

The first part of the check list relates to considerations before the start of an investigation, such as defining its main objective(s). Sometimes a choice of suitable studies to meet these is available, but some designs will be inappropriate. For example, it would not be sensible to compare elderly diseased patients with young healthy adults to determine whether a blood constituent is aetiologically important. Design considerations also include techniques for measurement and collection of data. In addition, important statistical questions relate to the source and number of subjects studied. The former will be relevant to the validity of any generalised inferences from the results. The issue of the sample size required for a study is well documented, but many studies are still too small to detect anything other than large, and often unrealistic, effects.

When the study is under way a high participation rate is needed from the recruited subjects. Those who do not participate fully are almost certain to be a biased group in some respects, with detrimental effects on the interpretation of the results. A comparison of relevant characteristics of responders and non-responders should be given.

The statistical methods used should be stated. If a technique is novel or unfamiliar then a description of its purpose and an outline of the method should be given together with a suitable reference. Aspects of presentation will also be checked, including tables and figures as well as textual content.

From the answers to the check list a summary can be made of the statistical content of a paper. Other features, which may be mentioned in the accompanying written report, contribute to the recommendation on its statistical quality.

CLINICAL TRIALS CHECK LIST

For clinical trials specific questions may be asked in addition to the items from the general check list (fig 2).

At the design stage of a clinical trial it is important to determine the diagnostic criteria for inclusion of subjects and clearly to define the treatments to be compared. Where a randomised study is appropriate, which usually is the case, a method of random allocation to treatment is mandatory and should be clearly described. Unambiguous measures of outcome must be specified for trials comparing treatments and the duration of follow up stated. There are advantages if double blind comparisons can be made, and treatment should start with a minimum delay after patient allocation. All these features should be described in the trial protocol.

In the results section the numbers and proportions of subjects treated and followed up should be stated. It is important also to describe drop outs and side effects by treatment group. In addition, treatment groups should be compared for relevant prognostic characteristics and adjustments for these made if appropriate in the analysis of outcome.

BMJ Ref No: _____ Date of Review: _____

Design Features			
1. Was the objective of the trial sufficiently described?	Yes	Unclear	No
2. Was there a satisfactory statement given of diagnostic criteria for entry to trial?	Yes	Unclear	No
3. Was there a satisfactory statement given of source of subjects?	Yes	Unclear	No
4. Were concurrent controls used (as opposed to historical controls)?	Yes	Unclear	No
5. Were the treatments well defined?	Yes	Unclear	No
6. Was random allocation to treatment used?	Yes	Unclear	No
7. Was the method of randomisation described?	Yes	Unclear	No
8. Was there an acceptable delay from allocation to commencement of treatment?	Yes	Unclear	No
9. Was the potential degree of blindness used?	Yes	Unclear	No
10. Was there a satisfactory statement of criteria for outcome measures?	Yes	Unclear	No
11. Were the outcome measures appropriate?	Yes	Unclear	No
12. Was there a power based assessment of adequacy of sample size?	Yes	Unclear	No
13. Was the duration of post-treatment follow up stated?	Yes	Unclear	No
Commencement of Trial			
14. Were the treatment and control groups comparable in relevant measures?	Yes	Unclear	No
15. Were a high proportion of the subjects followed up?	Yes	Unclear	No
16. Did a high proportion of subjects complete treatment?	Yes	Unclear	No
17. Were the drop outs described by treatment/control groups?	Yes	Unclear	No
18. Were side effects of treatment reported?	Yes	Unclear	No
Analysis and Presentation			
19. Was there a statement adequately describing or referencing all statistical procedures used?		Yes	No
20. Were the statistical analyses used appropriate?	Yes	Unclear	No
21. Were prognostic factors adequately considered?	Yes	Unclear	No
22. Was the presentation of statistical material satisfactory?		Yes	No
23. Were confidence intervals given for the main results?		Yes	No
24. Was the conclusion drawn from the statistical analysis justified?	Yes	Unclear	No
Recommendation			
25. Is the paper of acceptable statistical standard for publication?		Yes	No
26. If "No" to Question 25, could it become acceptable with suitable revision?		Yes	No

Reviewer: _____

FIG 2—*Check list for statistical review of papers on clinical trials for the* BMJ.

Experience so far

We have used the check lists on a regular basis for less than 12 months, so that only a limited amount of descriptive data are available on the main statistical problems found. We do, however, have preliminary findings based on 103 papers for which the general check list was used and 45 papers on clinical trials. Each of these papers was referred for statistical assessment because of comments by the subject matter referee or the editorial staff, and they are a small and unrepresentative sample of papers submitted to (or published in) the *BMJ*.[1] Thus the descriptive figures given below have not been subjected to any formal statistical analysis.

GENERAL CHECK LIST

For the general papers design features were the most satisfactory. Nevertheless, for 28 of the 103 papers the appropriateness of the study design was in doubt, and in 22 papers the source of subjects was not clear. In only one paper did the authors report calculating a required sample size in advance. Response rates were thought to be satisfactory in 84 of the 100 papers where the question was appropriate, but for 12 of the other 16 this information was not clearly given.

In relation to analysis about a third (34) of the papers did not describe the statistical procedures used, and in only 42 papers were the methods said to be appropriate. The main adverse comments related to lack of allowance for confounding variables, invalid use of the χ^2 test, unsuitable analysis of non-Normal data, problems of multiple comparisons, and incorrect arithmetic. Presentation was assessed as unsatisfactory for 76 of the 103 papers. The most frequent difficulties related to problems with tables, inadequate descriptions of the outcomes of hypothesis tests, lack of confidence intervals, non-Normal data, and notational ambiguities mainly associated with use of the $\pm$ sign (now banned by the *BMJ*). In only 35 papers was the conclusion drawn from the statistical analysis thought to be justified. Overall as few as 17 of the 103 papers were regarded as statistically acceptable for publication. Only six papers, however, were thought to be unsuitable for revision, though in 40 cases it was "unclear" whether revision was possible.

CLINICAL TRIALS CHECK LIST

For the 45 papers on clinical trials the design aspects were again reasonable according to the statistical assessors. The main points of exception were a lack of description of the method of randomisation in 35 papers and the absence of a power based calculation of sample size in 38. The

latter raises important ethical as well as statistical considerations,[12] which apply to the general papers also. Questions on the delay between allocation and beginning treatment and on the potential degree of blindness used were answered as "unclear" for 18 and 22 of the papers respectively.

That part of the check list concerned with statistical analysis disclosed a situation similar to that in the general papers. The method was neither described nor referenced in 25 papers and was said to be inappropriate in 19. Prognostic factors were reported to be inadequately considered in 24 papers and presentation as unsatisfactory in 41. The conclusion from the statistical analysis was said to be unjustified or in doubt in 31 of the 45 papers. For only five of the 41 papers considered unacceptable for publication, however, was suitable revision not thought possible—three of them being non-randomised studies.

Comments

These check lists have evolved over a period of time and, as shown in figures 1 and 2, differ slightly from those used initially. For example, the question on confidence intervals (question 9 in the general check list, question 23 in the clinical trials check list) is a recent addition. It has been included partly as a consequence of intended future *BMJ* policy.[13 14]

From this preliminary look at answers to the check lists improvements in reporting statistical procedures are clearly needed by some authors. Quite often the problems which have been found relate to easily rectifiable omissions of information, though sometimes there are more serious difficulties in analysis. Statistical assessment is mentioned as a possibility for any article submitted to the *BMJ* (see Instructions to authors, 4 January, p 4) and the check lists are now used routinely in this. Such a statistical evaluation is one way to prevent the publication of papers with unsatisfactory statistical content. Other approaches are, of course, possible[15]—such as the adoption of published statistical guidelines[6] or having a statistician on the editorial board. The check lists are intended for guidance on the statistical content of papers and are not presented as items to be covered at the expense of other important aspects of medical studies.[16 17]

References

1 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**:1485-8.
2 Lionel NDW, Herxheimer A. Assessing reports of therapeutic trials. *Br Med J* 1970;iii:637-40.
3 Ford BL, Tortora RD. A consulting aid to sample design. *Biometrics* 1978;**34**:299-304.
4 Sackett DL. Evaluation: requirements for clinical application. In: Warren KS, ed. *Coping with the biomedical literature: a primer for the scientist and the clinician*. New York: Praeger, 1981: 123-57.
5 Bland JM, Jones DR, Bennett S, Cook DG, Haines AP, MacFarlane AJ. Is the clinical trials evidence about new drugs statistically adequate? *Br J Clin Pharmacol* 1985;**19**:155-60.
6 Altman DG, Gore SM, Gardner MJ, Pocock SJ. Statistical guidelines for contributors to medical journals. *Br Med J* 1983;**286**:1489-93.
7 Smith R. Steaming up windows and refereeing medical papers. *Br Med J* 1982;**285**:1259-61.
8 Anonymous. Peer review at work. Dean JW, Fowler PBS. Exaggerated responsiveness to thyrotrophin releasing hormone: a risk factor in women with coronary artery disease. *Br Med J* 1985;**290**:1555-61.
9 Commentators. Peer review at work. *Br Med J* 1985;**290**:1743;**290**:1984-5;**291**:412;**291**:414;**291**: 485-6.
10 Lock S. *A difficult balance: editorial peer review in medicine*. London: Nuffield Provincial Hospitals Trust, 1985.
11 Vaisrub N. Manuscript review from a statistician's perspective. *JAMA* 1985;**253**:3145-7.
12 Altman DG. Statistics and ethics in medical research: III—how large a sample? *Br Med J* 1980;**281**:1336-8.
13 Gardner MJ, Altman DG. Confidence intervals rather than p values: estimation rather than hypothesis testing. *Br Med J* 1986;**292**:746-50.
14 Langman MJS. Towards estimation and confidence intervals. *Br Med J* 1986;**292**:716.
15 George SL. Statistics in medical journals: a survey of current policies and proposals for editors. *Med Pediatr Oncol* 1985;**13**:109-12.
16 Jones RS. Statistical assessment of papers submitted to the "British Medical Journal." *Br Med J* 1983;**286**:1971.
17 Healy MJR. Statistical guidelines for contributors to medical journals. *Br Med J* 1983;**287**:132.

(Accepted 2 December 1985)

Do anti-inflammatory drugs reduce the immunity to bee stings that beekeepers usually acquire?

Virtually all anti-inflammatory drugs inhibit the synthesis of prostanoids, but the loss of their modulating influence may have widely different effects in different subjects. Some asthmatics bronchodilate in response to aspirin[1] but others bronchoconstrict.[2] Some beekeepers find that aspirin suppresses local inflammation after a sting, but a similar drug has recently been blamed, in *Beecraft* magazine, for a transient loss of tolerance in a seasoned beekeeper—albeit to four stings. Unexpected reactions to bee stings are most often seen after multiple stings—usually when the stings have occurred in a very vascular area of the head or neck—but the role of drugs cannot be ruled out. Beta blockers potentiate anaphylaxis,[3] and those who suspect that aspirin like drugs may do the same should be encouraged to publish their evidence.—M H LESSOF, professor of medicine, London.

1 Szczeklik A, Gryglewski RJ, Nizanowska E. Asthma relieved by aspirin and other cyclo-oxygenase inhibitors. *Thorax* 1978;**33**:664-5.
2 Stevenson DD, Simon RA, Mathison DA. Aspirin-sensitive asthma: tolerance to aspirin after positive oral aspirin challenges. *J Allergy Clin Immunol* 1980;**66**:82-8.
3 Hannaway PJ, Hopper GDK. Severe anaphylaxis and drug-induced beta blockage. *N Engl J Med* 1983;**300**:1536.

Why does a woman in her 70s complain of a persistent vile taste in her mouth? Dental health and hygiene are good and she is otherwise (apart from a persistent slightly furred tongue) free of signs and symptoms. What investigations and treatment are advised?

If dental causes can be excluded the answer to this question is difficult. The furred tongue is probably not relevant. Disorders of taste have been recorded in hypothyroidism and in chronic renal failure and these should be excluded. Taste may be abnormal in neurological disease but none is evident in this patient. Taste disturbances may also occur after viral infections, including influenza and hepatitis. In the elderly dysgeusia, an abnormal taste sensation, is associated with hypogeusia, a diminished sense of taste, and with hyposmia, a diminished sense of smell. Microscopic abnormalities have been described in the structure of the taste buds, but other cases have been reported in which a central lesion seemed likely.[1] Drugs may also disturb taste and there are almost 50 references to this. Substances containing sulfhydryl groups—for example, penicillamine and captopril—are particularly likely to interfere with taste. Many cases are self limiting and there is no specific treatment. Zinc sulphate, which was in vogue ten years ago, was found no better than placebo in a controlled trial.[2] For further details Schiffner's two part review is strongly recommended.[3]—R E IRVINE, honorary consultant physician in geriatric medicine, Hastings.

1 Shafar J. Dysgeusia in the elderly. *Lancet* 1965;ii:83-4.
2 Henkin RI, Schecter PJ, Friedewald WT, Demets DL, Raff M. A double blind study of the effects of zinc sulphate on taste and smell dysfunction. *Am J Med Sci* 1976;**272**:285-99.
3 Schiffman SS. Taste and smell in disease. *N Engl J Med* 1983;**308**:1275-9, 1337-43.

To what extent do people acquire immunity to infections prevalent in the community in which they live? Would such acquired immunity diminish with age?

Such a general question can receive only a general answer and it is important to bear in mind that in some infections intrinsic non-specific immunity is important and that in others, such as measles, specific immunity is readily acquired and lifelong whereas it is limited against others—for instance, some bacterial and fungal skin infections. In many common infections individuals are infected with a series of organisms in the early years of life and gradually acquire immunity against more and more of them. In infections where the organism does not persist immunity probably declines gradually after exposure. In old age a person may be exposed less frequently and the immune system also becomes less effective so that immunity often declines and the individual becomes more susceptible to infection and disease. These phenomena may be seen in the changing clinical pattern of acute respiratory virus infections from infancy to old age. The question implies that there may be differences in the number and type of infections in different communities. This is not true for respiratory virus infections but is true for gastrointestinal infections and many parasitic diseases such as malaria. As a result visitors from the United Kingdom to developing countries are prone to get gastroenteritis and hepatitis, and those moving from one part of the developing world to another may have similar problems; those moving from areas of poor hygiene to those with good hygiene remain well. Furthermore, immunity to malaria may wane in a few years so that a student returning to an endemic area after a few years in Britain may suffer an attack of disease; nevertheless, early exposure to malaria may give lifelong resistance to severe disease such as cerebral malaria.—D A J TYRRELL, director, Common Cold Research Unit, Salisbury.