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 and have been influenced by others published previously. 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 or in the statistical guidelines of Altman et al.

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, 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 with 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 Vairub.

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. Referal to the check lists should also improve the description of the statistical aspects of studies in submitted papers.

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

FIG 1—Check list for statistical review of general papers for BMJ.
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 disease patients with young healthy adults to determine whether a blood constituent is aetiological 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.

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. 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 chi square 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 ± 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, 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.

References

(Accepted 2 December 1985)