

Statistical Advice for Contributors to Gut

Manuscripts submitted to Gut should conform to the uniform requirements for manuscripts submitted to biomedical journals (<http://www.icmje.org>). While there may be a wide variety of styles of paper submitted, it is expected that a large majority will fall into one of three broad classifications: Experimental trial (clinical or non-clinical), observational study, or systematic review. The three styles are considered separately, but first some universal points are considered.

Guidelines for all papers (Universal points):

U1: The statistical methods are very much part of the research methodology and should be reported as thoroughly as other aspects of the methodology with a view to the ideal that the manuscript should enable a reader to replicate the work.

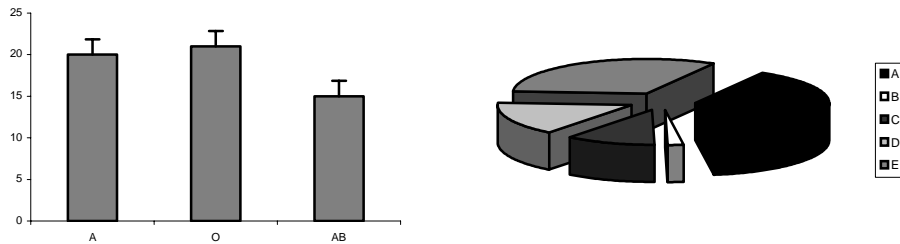
U2: Confidence intervals for effect sizes are preferable to p-values. While there may be situations where p-values are appropriate, these are in the minority. In such cases the p-values should be presented to a reasonable number of decimal places (usually three or four).

U3: In any event, merely labelling differences as significant or non-significant (including such labels as NS, <0.05, * etc) without indicating the size of the effect and 95% CI is not acceptable.

U4: It is to be hoped that such an approach will lead away from the dangers of misinterpreting statistical significance or the lack of it. Inferences drawn by comparing p-values are dubious. Interpretation of a lack of significance as evidence of no effect is usually incorrect.

U5: Figures should also adhere to the principle that confidence intervals are preferable to p-values. They should exploit their ability to display more information than a table, and as an ideal they would display all the data rather than summaries. For these reasons among others the figures described by Swinscow and Campbell as “dynamite-plunger plots” (see example below) are generally undesirable and “box and whisker” plots are preferred or, if the number of observations is small, individual data points.

U6:



Examples of undesirable graphs: (a) “dynamite-plunger plot”, (b) 3D pie chart

U7: Percentages should only be presented as a summary following the numbers that generate them. There is often little value in giving the percentage when the denominator is small, for example—less than twenty.

U8: Nearly all papers will have a concept of sample-size associated with them. The a-priori justification of this sample-size should be presented. Reporting of ethical approval does not satisfy this requirement. Note that presentation of a post-hoc calculation is likely to be of little value.

U9: If complex statistical models are being employed, then some justification of the model choice may be required, including information regarding the model fit of this and other credible models. This may not be appropriate for the paper (except perhaps in summary), but additional material may be usefully submitted.

U10: The perfect piece of research is non-existent, with all designs having to compromise between competing requirements. The paper should discuss potential weaknesses and or biases in the design, conduct, and analysis.

U11: Unless one or more of the authors are statisticians, then consider consulting a statistician or a good text. General statistical advice can be obtained from the BMJ Books publications:

Swinscow TDV, Campbell MJ, Statistics at Square One, 2002

Campbell MJ, Statistics at Square Two, 2001

Altman DG, Machin D, Bryant TN, Gardner M, Statistics with confidence, 2000

Alternative texts include Altman DG, Practical Statistics for Medical Research, 1991, Chapman and Hall/CRC Press, and Bland MJ, An Introduction to Medical Statistics, third edition, 2000, Oxford University Press.

Further guidance is also available from tutorial articles published in the BMJ, the Lancet and other journals.

U12: If the study or experiment is inconclusive (perhaps it was intended to be a pilot), then its value is in informing and enabling future research. The authors then have a greater duty to facilitate replication of the work, and inclusion in systematic reviews or meta-analyses. Consideration might be given to making the data available, e.g. by placing it on the internet.

U13: Certain words such as 'random', 'significant', and 'correlation' should be reserved for their correct technical usage.

U14: Few analyses published in the journal have been Bayesian in nature, this is a reflection of the analyses submitted and not a reflection of policy. A good Bayesian analysis is as welcome as a good frequentist analysis.

U15: Ignoring for now systematic reviews, only randomized experimental trials can have the strength of evidence to begin to claim a causal link (using phrases such as 'cause', 'effect', 'due to', 'because of' etc). Therefore any manuscript using such terminology should be written up according to the CONSORT statement.

Additional guidelines for clinical trials:

T1: These should be reported according to the latest version of the CONSORT statement (<http://www.consort-statement.org>). The purpose of this is to ensure enough detail is given, in a manner familiar to the reader, in order that critical appraisal, replication, and inclusion in systematic reviews is possible.

T2: In particular, reading the explanation and elaboration is recommended (<http://www.consort-statement.org/explanation/newene.htm>). Parts of this document are in fact a useful guide for the preparation of manuscripts not reporting experimental trials.

T3: The CONSORT checklist should be submitted with the manuscript indicating whereabouts in the paper the points are covered.

Additional guidelines for laboratory studies

Although at first sight they may seem inappropriate in general most of the CONSORT guidelines are still valid but need some interpretation. Thus for example details of participant flow may be valuable indicating how many animals failed to complete protocols since this gives an important guide to how well procedures were tolerated.

Although blinding of subjects may not be appropriate, consideration of blinding of some of the research team may be valuable to avoid bias, particularly with assessments that may have some subjectivity. The use of coded samples for assessments is good practice and should be commented on. The report should indicate how the authors attempted to avoid systematic error or confounding by the use of appropriate randomisation of treatment sequences

A useful source of general advice on the design of experiments and in particular the importance of randomization can be found in Gart, J.J. et al. (1986) *Statistical Methods in Cancer Research Volume III -the design and analysis of long-term animal experiments*, Lyon: IARC

Observational studies:

O1: These will typically require more caution in the phrasing of the conclusions and discussion, as well as a greater discussion of potential biases and weaknesses.

O2: As a minimum one might expect to see:

Title and abstract:

- details of the study design in the title or abstract

Methods:

- details of how the study groups were defined
- details of the hypothesis that inspired the study
- details of how the sample size was chosen
- details of any blinding used in assessment or analysis
- details of the statistical methods used

Results:

- baseline characteristics of the groups
- details of participant flow, early exits from the study and an account of loss through attrition, and clear accounts of the numbers involved in each analysis.
- outcomes

Discussion:

- interpretation of the results including discussion of weaknesses
- generalizability of results

O3: For studies of diagnostic accuracy, the STARD (Standards for Reporting of Diagnostic Accuracy: <http://www.consort-statement.org/stardstatement.htm>) statement may be of value. Particular note should be made that when comparing two measures of a variable mere demonstration of a correlation is inadequate and that the differences between the assays should be presented as recommended by Altman and Bland (Bland JM, Altman DG. (1986) [Statistical methods for assessing agreement between two methods of clinical measurement](#). *Lancet* **i**, 307-310) See also <http://www.mbland.sghms.ac.uk/ba.htm> where the paper is available in electronic format.

Additional Guidelines for Systematic reviews

- S1: A number of standards may be of help here including MOOSE (Meta-analysis Of Observational Studies in Epidemiology) and QUORUM (QUality Of Reporting Of Meta-analyses). Details of both of these standards are to be found at <http://www.consort-statement.org/initiatives/complements.htm>. No standard has yet been made a requirement by Gut, but consideration of the aspects raised in these documents would strengthen the majority of papers.
- S2: Books such as 'Chalmers I and Altman DG, Systematic Reviews, 1995, BMJ Publishing Group' can give further advice.
- S3: The BMJ ran a series of articles on the conduct of systematic reviews in July 2001, these are also a good source of guidance. References for these articles are
BMJ, Jul 2001; 323: 42 – 46,
BMJ, Jul 2001; 323: 157 – 162,
BMJ, Jul 2001; 323: 101 – 105
and BMJ, Jul 2001; 323: 224 – 228.

Dr A. G. Lynch and Dr C. R. Palmer
Centre for Applied Medical Statistics
Department of Public Health and Primary Care
Institute of Public Health
University of Cambridge