Primary Care
RESPIRATORY JOURNAL
www.thepcrj.org
Journal of the Primary Care Respiratory Society UK and the International Primary Care Respiratory Group

GUIDE FOR AUTHORS

Statistical Considerations in Preparing Manuscripts for the *PCRJ*

> "*... good statistical analysis requires commonsense and judgement, as well as a repertoire of formal techniques, so that there is an art in statistics as well as in medicine.*"
> (Altman, Gore, Gardner, & Pocock, 1983; p1489)

The art in statistics extends to the accurate reporting of studies – objectives, methods, results, interpretations, and all. The *PCRJ* agrees with and follows the recommendations of the International Committee of Medical Journal Editors (ICMJE) in the preparation of manuscripts for submission, and these same recommendations are also followed for the statistical reporting of studies.

> "*Describe statistical methods with enough detail to enable a knowledgeable reader with access to the original data to verify the reported results. When possible, quantify findings and present them with appropriate indicators of measurement error or uncertainty (such as confidence intervals). Avoid relying solely on statistical hypothesis testing, such as P values, which fail to convey important information about effect size. References for the design of the study and statistical methods should be to standard works when possible (with pages stated). Define statistical terms, abbreviations, and most symbols. Specify the computer software used.*" (ICMJE, 2009)

The following guidelines are intended to augment the above general principles:

1. If a statistician was consulted and/or is among the author list, mention and identify this person(s). This will help the editors to determine if further statistical refereeing is needed. Also make sure you indicate the extent of this involvement when discussing 'Author contributions'.

2. Describe the study design in sufficient detail to enable someone to replicate the study if necessary and also to allow the reader to judge how valid the findings from the study are. Different study types – like observational and diagnostic studies, clinical trials, and others – have existing guidelines and checklists such as STROBE, STARD and CONSORT (all accessible from http://www.equator-network.org/resource-centre/library-of-health-research-reporting/). Make use of these. It should be explained how the elements of study design that could impact on the statistical analysis – like randomisation, primary sampling units, sample size and power, known confounders, repeated measurements, attrition and missing data – are treated. Some common mistakes include testing for differences in baseline characteristics of randomised groups, post-analysis power calculations to establish sufficiency of effective sample sizes, and treating cluster samples as simple random samples. These should be avoided.

3. Specify how the sample size was calculated, with sufficient details as to the assumptions used and the methods. A study may use samples of different sizes in different analyses intentionally or as a consequence of missing data. Sometimes the study may not achieve the intended sample size. These deviations must be reported.

4. Identify precisely the statistical methods used, and, except for the more common methods, reference these citing appropriate sources. Mention the assumptions made. When these are not mentioned the reader assumes that the assumptions required for a test are all met, which might be belied later in the paper. For example, tests where normality assumption is critical may be carried out on samples clearly in departure from normality as revealed by their sample statistics. How much detail should be given about a statistical method depends on how common they are and how familiar the reader will be with it. While the details of t test may not be expected, less common (relative to this journal and its readership) methods like correction for regression to the mean – for example, in repeated lung function measurements – should be detailed. The reason for the choice of the method used in the paper should be given, especially if alternative methods are used. The choice of the significance level (usually $\alpha = 0.05$) and the power (usually $\beta = 0.80$ or 0.90) should be reported. Where multiple comparisons are made, the method(s) chosen for accounting for it and the altered α level should be stated. Any alteration to data like dropping of outliers and transformations should be clearly stated. In testing inferences, state whether one- or two-sided tests were used; two-sided tests are the norm so there will need to be a clear justification for using one-sided tests.

5. If a statistical programme was used for analysis, clearly identify it. Common programmes like SPSS, SAS, and STATA require only the acronym and the version. If a less common or specific programme was used, mention the author(s) and/or the source also. Some authors require you to reference their work; if so, please respect this wish.

6. Describe the data so that the reader can validate the analysis. Descriptive statistics should take into account the scale of measurement and the distribution; for example, continuous measures can be the mean and standard deviation for a normally distributed variable, and the median and inter-quartile range for one deviating from normality. A common mistake is not recognising that when presenting the variability rather than the precision in the measurement one should use the mean and the standard deviation, not the standard error or the confidence interval. Categorical variables can be described using proportions and percentages. With small samples, reporting the actual numbers is more appropriate than just giving the percentage in each category.

7. Maintain a balance between the text, tables and figures when reporting the results of the statistical analyses. Try to avoid repeating the results in the tables and figures in the text.

8. The following points relate to some more specific issues:

8.1 Numbers in text

8.1.1 Spell out numbers one through nine and use numerals for 10+. This is true for ordinal numbers too.

8.1.2 Exceptions:

8.1.2.1 Use numerals for units of measurement, times and dates (e.g. 5 mL not five mL).

8.1.2.2 Spell out numbers when beginning a sentence.

8.1.2.3 When reporting similar quantities numerals are used even for those below 10 (e.g. 92% White, 5% South Asians and 3% other minority ethnic groups).

8.1.2.4 When it is necessary to differentiate between two consecutive numbers belonging to different classes (e.g. thirty 5-mg tablets).

8.1.2.5 Large numbers in general expressions are always in words.

8.1.3 Use spacing between numbers and units (e.g. 70 mL) except for per cent sign (<70%).

8.1.4 Report decimals using a zero.

8.1.5 Do not use hyphen for presenting ranges, instead use "to" or "through". The measurement unit need to be given only at the end of the range (300 to 400 mL) except in the case of percentages (50% to 70%). Duplicate numerals should not be left out (300 to 340 days not 300 to 40 days).

8.1.6 Use "/" for proportions (e.g. 3/5) and rates (62/1000) and ":" for ratios (1: 1.2).

8.1.7 Common fractions should be spelt out (e.g. half the sample).

8.2 Do not indicate standard deviation or standard error using "±" sign but using SD or SE; e.g. 180 (SD 45), 180 (SE 4.5).

8.3 Indicate the confidence level when reporting confidence interval; e.g. 180 (95%CI 171.2 to 188.8). In tables the '95%CI' is placed in the header; e.g. 180 (171.2 to 188.8). A common mistake associated with the use of confidence intervals is to infer non-significance in differences in means of two groups when the confidence intervals overlap. Even with considerable overlap, the difference in mean might be significant when tested using, for example, the t test.

8.4 P-values are reported using 'P' not 'p'. Report the precise value (e.g. P = 0.456 or P =0.012 not NS or P <0.05) up to three digits then report as P= <0.001 (not 0.0000 as in the output of some statistical software. In genetic epidemiological studies, it might be necessary to show precisely how high the significance is, in which case use scientific notation (e.g. P = 3.008 X 10^{12}).

8.5 Results presenting t, χ^2, and F should specify the degrees of freedom within parentheses. E.g. t (67) = 2.082, P = 0.041; χ^2 (2) = 12.879, P <0.001; F (1, 190) = 3.140, P = 0.078.

8.6. When statistical tests have different forms (e.g. paired and unpaired t test), report which form of the test was used.

8.7. In reporting regression, in addition to the coefficients, their precision and significance, measures of model fit like R^2 should be given. Use β for unstandardised coefficients and std β for standardised coefficients.

Dr Gopalakrishnan Netuveli, Statistical Editor, *PCRJ*
October 2011

Information about the *Primary Care Respiratory Journal* is available on the World Wide Web at the following address: **http://www.thepcrj.org**

The Primary Care Respiratory Society UK, formerly known as the General Practice Airways Group, is a registered charity (Charity No: 1098117) and a company limited by guarantee registered in England (Company No: 4298947)
VAT Registration Number: 866 1543 09
Registered offices: 2 Wellington Place, Leeds, LS1 4AP
Address for correspondence: PCRS-UK, Smithy House, Waterbeck, Lockerbie, DG11 3EY
Tel: +44 (0)121 629 7741 Fax: +44 (0)121 336 1914 Email: info@pcrs-uk.org Website: http://www.pcrs-uk.org