Statistical Reporting Guidelines – European Journal of Applied Physiology

Introduction:
The aim of these guidelines is to provide prospective authors with succinct indications of the expectations of statistical quality that the European Journal of Applied Physiology considers appropriate. They are written assuming a familiarity with common statistical terminology and methods, and are advisory rather than mandatory. If authors are in any doubt, they should consult a bio-statistician at the planning stages of their study. Further consultation during the analytical and/or interpretational stages of the study may also be prudent.

General Policy Statement:
The European Journal of Applied Physiology’s policy on statistical reporting is to only accept manuscripts for publication which display experimental design, statistical data analysis, and interpretation thereof which are both sound and appropriate to the hypothesis being posed. In particular, the design and methods should be described with sufficient detail and accuracy to enable the reader to repeat the study without ambiguity. Authors are encouraged to provide supporting citations for methods using readily available pertinent original articles and/or reviews.

Experimental Design:
Good experimental design features are a requirement of all manuscripts considered for publication. The overriding concern is that the design employed is such as to permit the aims of the study to be fully and efficiently achievable. Specifically, the following represents a listing of several of the more commonly employed good design features; many of these features are intended to enhance statistical power and through the control of extraneous variation.

- Avoidance of confounding (assuring that treatment effects are separable)
- Balance, counterbalance and crossover (negating temporal or learning effects)
- Blocking (the use of smaller more homogeneous blocks of experimental units)
- Use of covariate(s) (such as adjustment for differing initial or baseline values)
- Randomisation of treatment allocation (often, but not necessarily restricted)
- Single- or double-blinding (assuring independence of effects)
- Inclusion of controls or placebos (for normative or comparative purposes)
- Matched controls (for improved precision of comparisons)

Sample Size (n) Considerations:
Sample sizes must always be clearly stated. They should be justifiable, and large enough to provide reliable results and inferences. In hierarchical designs, effective sample sizes at each level should be clearly evident.

Case studies (n = 1) will be considered on their merits.
Intervention studies with small sample size (n less than about 15) should be treated with appropriate caution. For example, in such situations, sensitivity may be compromised, inferences to a wider population may not be feasible, and nonparametric methods may be more appropriate. Other small sample studies will be assessed on their merits.
Larger sample sizes should be subject to an investigation of their cost:benefit aspects, though it is recognised that in the study of rare events where incidence is low, large samples are frequently a necessity.
Significance Level ($\alpha$), P-value and Power:
The minimum P-value to achieve significance (the chosen $\alpha$) should be stated and justified before any data are collected. The commonly adopted $\alpha = 0.05$ is not necessarily always appropriate. In particular, it is important to recognise that when several tests are being performed (not just in multiple comparisons), the likelihood of rejecting at least one true null hypothesis (a Type 1 error) is increased. Consideration should therefore be given to downward adjustment of the chosen $\alpha$ level(s) *a priori*. In addition, after analysis, it is informative to readers if authors report the exact P-value for any important results.

Statistical power (probability of rejection of a false null hypothesis, or of not committing a Type 2 error) is affected by the chosen $\alpha$ level, sample size, inherent variability, and the real, but often unknown, value of the tested parameter under the alternate hypothesis (often referred to as the effect size). All of these should be considered. It is advisable to carry out at least an approximate power analysis for any investigation, where feasible. While it is not always necessary to report this in the manuscript itself, it is useful supportive information to include in the submission.

Selection and Validity of Analytical Methods:
In most instances several analytical methods may be available for use, and it is not normally the case that one is correct and all others are wrong. One may be more appropriate, or more informative than another, and the choice of which to use lies with the author. In some cases it may be helpful to indicate the reason(s) for selection of one method in preference to another.

For analysis of data, the *European Journal of Applied Physiology* has no specific preferred statistical software of choice. Authors are free to select whichever they prefer, but such selection should always be stated and referenced. Most parametric analyses are based on a variety of assumptions about the distribution of the data (or of the residuals after modelling). Some techniques are sensitive to departures from these assumptions, while others are robust. Authors should note the relevant assumptions for the method(s) they employ, and confirm their validity. If there is a degree of departure, this should be reported also and appropriate remedial action taken. This may necessitate a transformation of the data, application of an adjustment or correction factor, or a change in method. Failure to account for any such departures usually invalidates the method and any subsequent inferences.

Disclosure of Variability, Uncertainty and Error:
Authors are urged to accurately disclose measures of variability, uncertainty, measurement error etc. For variability of individual values, the standard deviation should be used; whereas for variability of mean values, the standard error should be used. Uncertainty should normally be expressed using either confidence intervals or limits of agreement. The disclosure of measurement error, and the concomitant number of significant digits used when reporting measured values, is an important adjunct to interpretation of the results.

References:
Readers should consult appropriate texts and references for further information and more specific detail. The following list may be of assistance.

Altman DG, Moher D, Schulz KF. Peer review of statistics in medical research. Reporting power calculations is important. BMJ. 2002; 325(7362): 491; author reply 491.