Editorial

The importance of clear language

Most manuscripts submitted to Acta Orthopaedica, as well as to other medical scientific journals, describe treatment effects, complications, side effects, quality of life, health economy, etc. in a limited group of patients. The aim of the studies, however, is to learn something of importance for all patients, including future ones, not only for those who participated in the studies.

Statistical analysis is performed to quantify the empirical support for the findings. The evaluation is based on sample-to-population inference: an observable sample is studied in order to learn about the unobservable population it represents.

A consequence of the population not being fully observable is that the findings will always be uncertain to some degree. Statistical methods can, however, be used to quantify this uncertainty. It is often presented in terms of confidence intervals and p-values.

Unfortunately, instead of recognizing that p-values and confidence intervals describe generalization uncertainty, many authors seem to believe that these measures represent some obscure property of the observed sample itself.

For example, when two groups have different mean values, but the difference is not statistically significant, the observed difference is often described as "there was no difference". Another common phrase is that "there was no statistical difference". The first description is wrong, because it is a misrepresentation of what has actually been observed. The second characterization is ridiculous, because a statistically insignificant difference in mean values is as much "statistical" as a statistically significant one. Observed data should be described correctly: for example, "the two groups had similar mean values" or "differed in mean value, but the difference was not statistically significant".

Similar confusion is common for statistically significant findings. Statistical significance can, for example, be considered to be a token of practical significance or importance. However, uncertainty and practical significance are two fundamentally different things. Practical significance cannot be derived from a p-value.

Note that the p-value is not an effect measure; a lower p-value does not necessarily correspond to a greater difference. The same difference in revision rates between two specific knee arthroplasties is, for example, likely to get a much lower p-value when tested in the large Swedish Knee Arthroplasty Register than in a small clinical study. The reason for this is that the size of the p-value depends on sample size. It is therefore necessary to consider both the practical effects (the clinical significance of the effects) and their estimation uncertainty (related to the statistical significance of the effects). Just presenting a finding as "significant" is clearly ambiguous.

Other common language problems are related to misunderstood terminology, such as "multivariate" analyses performed with multiple linear regression (a univariate method), and four "quartiles" representing the four quarts of a sample (which are defined by three quartiles).

In addition, some statistical software packages have a bad influence on scientific writing. Attempts to improve user-friendliness are often made at the expense of specificity. For example, the description "independent groups' T-test" may be reasonable within the narrow framework of a certain software package, but it is not sufficiently specific for the statistics section of a scientific manuscript, because without knowledge about the software package used, the description can refer to any of Student's, Welch's, Satterthwaite's, Yuen's, and Pitman's t-tests, and even to Hotelling's T-test.

Vague and ambiguous descriptions and slang language may have a place in literary writing but should be avoided in science in order to facilitate correct interpretation of the findings presented.

Jonas Ranstam
Statistics consultant
jonas.ranstam@med.lu.se

DOI 10.3109/17453674.2013.842435