Repeated measurements, bilateral observations and pseudoreplicates, why does it matter?1

S U M M A R Y

A common requirement of statistical methods, critical to the interpretation of the data, is that the analyzed observations are independent. This is not always the case in experiments and clinical studies, a mistake which can be expected to lead to erroneous study results. The phenomenon is explained, its consequences described, and suggestions to avoid the problems presented.

Correlated observations

Repeated measurements on the same subject, bilateral observations, and laboratory replicates of specimens from the same donor are often more alike than observations on different subjects. This relationship is known as intraclass correlation. In contrast to Pearson’s product–moment correlation coefficient, which measures the linear dependence between two variables, the intraclass correlation coefficient is not affected by the order of the observations within each class (here subject).

That observations are independent is a fundamental assumption on which most statistical methods rely. The assumption is often neglected, both in clinical research and laboratory science. For example, in clinical research patients contribute independent observations but analyses are often performed on knees, hips, ankles, shoulders and elbows, and in vitro experiments are often done on cartilage pieces from two or three patients but analyzed as if they represented a larger number of patients. The definition of the analysis unit is a central issue because it strongly influences the results and the interpretation of the findings in a study or experiment.

Several papers have been written about the independence assumption and the frequent violations of it. For example, Bryant et al.2 and Park et al.3 focus on how often the independence assumption is violated in clinical research, while Festing4 and Lazic5 concentrate on dependence problems in laboratory experiments.

Lazic underlines that the terms experiment and replicate often are used ambiguously in laboratory science. Cell culture experiments can be repeated three times and be reported as three independent replicate experiments, but the word experiment can also refer to the entire procedure. The word replicate is often used to describe technical replicates, repeated measurements on the same analysis unit, but can also be used to describe biological replicates, independent analysis units. More than 20 years ago Hurlbert6 recognized the confusion between correlated and independent observations in ecologic field research and coined the word pseudo–replication to describe multiple observations on the same analysis unit. This term is now used also in other scientific fields.

The effect of analyzing correlated observations with statistical methods requiring independence is that both the variability and the number of observations (or degrees of freedom) is incorrect. This is problematic as these two properties, variation and number, determine the statistical precision. Confidence intervals and P-values, calculated using correlated observations and assuming that these are independent, may not give a fair representation of the sampling uncertainty they purport to measure. Statistical significance can be greatly exaggerated.

Examples

Two examples of this phenomenon are presented in Fig. 1. The first example, Example 1, describes a hypothetical example of a study aiming to assess whether the mean length of bone resection, in patients treated with bone resection for bone tumors, differs from 10 cm. Let us say that it is a common opinion, which we wish to challenge, that the resection length is 10 cm, and that both a longer and a shorter mean resection length would be clinically interesting.
A correct statistical analysis of correlated data can be performed in different ways. One way is to fit a random or mixed effects model. This is a method often included in statistical software packages, and the calculations are technically fairly simple to perform\(^1\). In contrast to a conventional (fixed effects) statistical models, which includes only fixed (e.g., between-subject) effects, a random or mixed effects model includes random (e.g., subject-specific) effects or a mixture of fixed and random effects. A random effects model was used in the examples to calculate the correct statistical precision. This model included all 30 observations, but structured into three independent clusters with 10 correlated observations in each. The analysis was similar to performing a one-sample t-test on the three mean values from each cluster.

Other ways can be to fit a marginal model\(^5\) or to use bootstrapping techniques\(^9\). While mixed-effects models include estimation of subject-specific effects, marginal models are based on estimating population-averaged effects. Bootstrapping is a general resampling technique, which uses a number of resamples of the observed dataset to estimate effects, each sample being obtained by random sampling with replacement. However, all methods for dealing with correlated data rest on more complex theories, and require greater statistical proficiency, than methods traditionally used in clinical and laboratory research. Severe analysis mistakes can easily be done.

In addition, not all study designs yield meaningful results. Between-subject and subject-specific observations can be combined in ways that are impossible to analyze correctly, however advanced the statistical methodology is. It is important to plan studies and experiments carefully, and often useful to consult a statistician already in the planning stage of the study. As R.A. Fisher stated in 1938: “To consult the statistician after an experiment is finished is often merely to ask him to conduct a post mortem examination. He can perhaps say what the experiment died of”.

**Recommendations**

When writing a research report it is important to describe the design of the experiment or study, the data collection and the statistical analysis in sufficient detail. The analysis unit should be clearly defined with respect to its independence, and the number of independent and repeated observations included in summary statistics and analyses should be clearly presented. Results from statistical analyses of correlated observations using traditional methods should not be considered reliable.

More information on this subject can be found in a previous editorial\(^10\) and in the *Osteoarthritis and Cartilage* guide for authors available on the journal website.

**References**

