Osteoarthritis and Cartilage 20 (2012) 473–475

Osteoarthritis and Cartilage

OARSI OSTEOARTHRITIS RESEARCH SOCIETY INTERNATIONAL

Editorial

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

Keywords:
Study design
Biostatistics
Independence
Bilateral
Pseudoreplicates
Variation
Precision

SUMMARY

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[1] 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 Festing[4] and Lazic[5] 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 Hurlbert[6] 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.

[1] There is often confusion amongst researchers and in the published literature about what constitutes independent measures in an experiment or study. This is a very important issue because the assumption of independent observations is critical to the application of appropriate statistical analysis and interpretation of the data.

1063-4584/$ – see front matter © 2012 Osteoarthritis Research Society International. Published by Elsevier Ltd. All rights reserved.
doi:10.1016/j.joca.2012.02.011

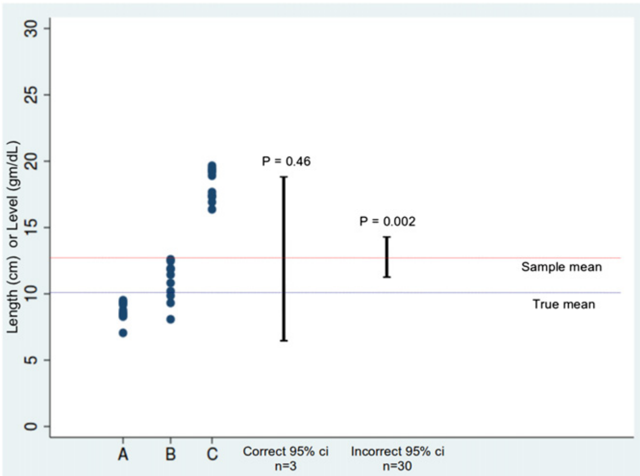

Fig. 1. A hypothetical study with three patients (A, B, and C) and 30 measurements of bone resection length (Example 1) or three repeats of a cell culture experiment on three different days (A, B and C) and 30 measurements of secreted cancer-specific protein level. Each single measurement is described with a dot. It is assumed that the unknown but true distribution of bone resection length and protein level has a Gaussian distribution with a mean value of 10 cm and gm/dL, and a standard deviation of 5 cm or gm/dL respectively. The observed mean value for the three independent observations is 12.59 cm or gm/dL and the standard deviation is 4.96 cm or gm/dL respectively. Random measurement errors are assumed to have a Gaussian distribution with mean value 0 cm or gm/dL and a standard deviation of 1 cm or gm/dL respectively, which implies that the intraclass correlation (reliability) of the measurements is 96%. The unknown mean value of bone resection length and protein level 10 cm or gm/dL and the observed sample mean value of 12.59 cm or gm/dL respectively are in the figure described with dashed lines. The 95% confidence limits and *P*-value calculated correctly using a random effects model ($n = 3$) with three analysis units and 10 repeated measurements on each analysis unit and incorrectly ($n = 30$) using a *t*-test are presented graphically as error bars while the corresponding *P*-values are presented numerically in the figure.

In the example ten measurements of the resection length have been taken on each of three randomly sampled patients. When incorrectly defining the analysis unit as the single measurement, and treating all 30 measurements as independent observations, the confidence interval and the *P*-value no longer give a good picture from these three patients of the actual statistical precision. In spite of the true resection length being exactly 10 cm these uncertainty measures support a mean length other than 10 cm. With a correct definition of the analysis unit ($n = 3$) this does not happen. False positive findings can of course emerge also with a correct definition of the analysis unit, but the risk for that particular test is then as low as the significance level.

Example 2 is also presented in Fig. 1. This describes three repeats of a cell culture experiment done on three different days; each of the dots represents the secreted cancer-specific protein level in separate replicate wells of a single cancer cell line. The expected level of this protein secreted from normal non-cancerous cells is 10 gm/dL with a standard deviation of 5 gm/dL, and the question is whether the cell secretes levels different from normal.

Consequences

The examples show what happens when "*n*" the number of independent observations (experiments) erroneously is considered to be 30 instead of three. The variability among the observations within each experiment reflects technical errors, not biological variability. The consequence of confusing the two sources of variability is a false positive result.

The examples show that the definition of the analysis unit is crucial. It is not a question of statistical orthodoxy, but a fundamental principle for rational evaluation of data. With wrong analysis unit the risk of misunderstanding data is greatly increased.

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[7]. 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[8] 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

1. Ranstam J. Sampling uncertainty in medical research. Osteoarthritis Cartilage 2009;17:1416–9.
2. Bryant D, Havey TC, Roberts R, Guyatt G. How many patients? How many limbs? Analysis of patients or limbs in the orthopaedic literature. JBJS Am 2006;88:41–5.
3. Park MS, Kim SJ, Chung CY, Cho IH, Lee SH, Lee KM. Statistical consideration for bilateral cases in orthopaedic research. JBJS Am 2010;92:1732–7.
4. Festing MFW. Principles: the need for better experimental design. Trends Pharmacol Sci 2003;24:341–5.
5. Lazic SE. The problem of pseudoreplication in neuroscientific studies: is it affecting your analysis? BMC Neurosci 2010 Jan 14;11:5.
6. Hurlbert SH. Pseudoreplication and the design of ecological field experiments. Ecol Monogr 1984;54:187–211.
7. Ranstam J. Problems in orthopedic research – dependent observations. Acta Orthop Scand 2002;73:447–50.

8. Carrière I, Bouyer J. Choosing marginal or random-effects models for longitudinal binary responses: application to self-reported disability among older persons. BMC Med Res Methodol 2002;2:15.
9. Feng Z, McLerran D, Grizzle J. A comparison of statistical methods for clustered data analysis with Gaussian error. Stat Med 1996;15:1793–806.
10. Ranstam J, Lohmander SL. Ten recommendations for Osteoarthritis and Cartilage (OAC) manuscript preparation, common for all types of studies. Osteoarthritis Cartilage 2011;19:1079–80.

J. Ranstam*
Department of Orthopedics,
Clinical Sciences Lund, Lund University,
SE-22185 Lund, Sweden

* Address correspondence and reprint requests to: J. Ranstam, Department of Orthopedics, Clinical Sciences Lund, Lund University, SE-22185 Lund, Sweden.
Tel: 46-46-171357; Fax: 46-46-177167.
E-mail address: jonas.ranstam@med.lu.se