Know when your numbers are significant

Experimental biologists, their reviewers and their publishers must grasp basic statistics, urges David L. Vaux, or sloppy science will continue to grow.

The incidence of papers in cell and molecular biology that have basic statistical mistakes is alarming. I see figures with error bars that do not say what they describe, and error bars and P values for single, "representative" experiments. So, as an increasingly weary reviewer of many a biology publication, I'm going to spell out again[1] the basics that every experimental biologist should know.

Simply put, statistics and error bars should be used only for independent data, and not for identical replicates within a single experiment. Because science represents the knowledge gained from repeated observations or experiments, these have to be performed more than once—or must use multiple independent samples—for us to have confidence that the results are not just a fluke, a coincidence or a mistake. To show only the result of a single experiment, even if it is a representative one, and then misuse statistics to justify that decision, erodes the integrity of the scientific literature.

It is eight years since *Nature* adopted a policy of insisting that papers containing figures with error bars describe what the error bars represent[2]. Nevertheless, it is still common to find papers in most biology journals—*Nature* included—that contain this and other basic statistical errors. In my opinion, the fact that these scientifically sloppy papers continue to be published means that the authors, reviewers and editors cannot comprehend the statistics, that they have not read the paper carefully, or both.

Why does this happen? Most cell and molecular biologists are taught some statistics during their high school or undergraduate years, but the principles seem to be forgotten somewhere between graduation and starting in the lab. Often, the type of statistics they learnt is not relevant to the kinds of experiment they are now doing And, once in the lab, people generally just do what everyone else does, without always understanding why.

Even if experimental biologists do not need to use statistical evidence for their own experiments, they should have an understanding of the basics so that they can interpret others' work critically. They don't all need to understand complex statistics, or to hire professional statisticians, but there would be fewer sloppy papers if every author, reviewer and editor understood statistical concepts such as standard deviation, standard error of the mean (s.e.m.), sampling error and the difference between replicate and independent data (see *Statistics Glossary*).

Back to Basics

In the life sciences there are typically two types of publication: those that use large data sets and rely mostly or wholly on statistical evidence (for example, epidemiology, psychology, clinical trials and genome-wide association studies), and those that do not—such as much cell and molecular biology, biochemistry and classical genetics.

For papers with large data sets that rely purely on statistical evidence, recommendations exist for computing sample size, reporting on outlying results and other issues[3,4]. But these guidelines do not serve authors of the other category of papers. Cell and molecular biologists have the luxury of being able to probe their experimental systems in multiple, independent ways and can therefore often get by with *N*s

"Experimental biologists should know what sort of sampling errors are to be expected".

of three, without the need for sophisticated statistics.

The first figure in a typical paper in cell or molecular biology, for example, might show the difference in phenotype between three wild-type and three gene-deleted mice. The second figure might compare the levels of proteins in cells derived from the mice, looking at both the deleted protein and one of its substrates, or the effects of treating wild-type cells with an inhibitor of the protein encoded by the deleted gene. If the evidence from these experiments is consistent, and gives support to a coherent model, it would be unnecessary to analyze 30 mice of each type, or to repeat the Western blots of protein levels 30 independent times. Watson and Crick's paper on the structure of DNA[5] does not contain statistics, graphs with error bars or large *N*s.

Understanding the rudiments of statistics would stop experimental biologists from calculating a *P* value and a s.e.m. from triplicates from one representative experiment, and might stop the reviewers and editors from letting these pass unquestioned. If the results from one representative experiment are shown, then *N* = 1 and statistics do not apply. Besides, it is always better to include a full data set, rather than withholding results that are not representative. When *N* is only 2 or 3, it would be more transparent to just plot the independent data points, and let the readers interpret the data for themselves, rather than showing possibly misleading *P* values or error bars and drawing statistical inferences.

If the data in an experiment are equivocal, or the effect size is small, it is much better to come up with an extra,

[1]Cumming, G., Fidler, F. and Vaux, D. L. J. Cell Biol. 177, 7–11 (2007).
[2]Vaux, D. L. Nature 428, 799 (2004).

[3]Landis, S. C. *et al.* Nature 490, 187–191 (2012).
[4]Nakagawa, S. and Cuthill, I. C. Biol. Rev. Camb. Philos. Soc. 82, 591–605 (2007).

[5]Watson, J. D. and Crick, F. H. Nature 171, 737–738 (1953).

Statistics Glossary

Some common statistical concepts and their uses in analyzing experimental results.

Term	Meaning	Common uses
Standard deviation (s.d.)	The typical difference between each value and the mean value.	Describing how broadly the sample values are distributed. $s.d. = \sqrt{(\Sigma (x - mean)^2/(N - 1))}$
Standard error of the mean (s.e.m.)	An estimate of how variable the means will be if the experiment is repeated multiple times.	Inferring where the population mean is likely to lie, or whether sets of samples are likely to come from the same population. $s.e.m. = s.d./\sqrt{N}$
Confidence interval (CI; 95%)	With 95% confidence, the population mean will lie in this interval.	To infer where the population mean lies, and to compare two populations. $CI = mean \pm s.e.m. \times t_{(N-1)}$
Independent data	Values from separate experiments of the same type that are not linked.	Testing hypotheses about the population.
Replicate data	Values from experiments where everything is linked as much as possible.	Serves as an internal check on performance of an experiment.
Sampling error	Variation caused by sampling part of a population rather than measuring the whole population.	Can reveal bias in the data (if it is too small) or problems with conduct of the experiment (if it is too big). In binomial distributions (such as live and dead cell counts) the expected s.d. is $\sqrt{(N \times p \times (1 - p))}$; in Poisson distributions (for example, cells per field) the expected s.d. is $\sqrt{mean}$.

N, number of independent samples; t, the t-statistic; p, probability.

mechanistically different, experiment to test the hypothesis, than to repeat the same experiment until P is less than 0.05.

If statistics are shown, it should be for a good reason. Descriptive statistics, such as range or standard deviations, are only necessary when there are too many data points to visualize easily. Inferential statistics (an s.e.m., confidence interval or P value) should be shown only if they make it easier to interpret the results, and they should not detract from other key considerations such as the magnitude of the effects or their biological significance.

Figure legends should state the number of independent data points and, for experiments in which replicates were performed, only the mean of the replicates should be shown as a single independent data point. For replicates, no statistics should be shown, because they give only an indication of the fidelity with which the replicates were created: they might indicate how good the pipetting was, but they have no bearing on the hypothesis being tested[6].

All experimental biologists and all those who review their papers should know what sort of sampling errors are to be expected in common experiments, such as determining the percentages of live and dead cells or counting the number of colonies on a plate or cells in a microscope field. Otherwise, they will not be able to judge their own data critically, or anyone else's.

Repeat after me

How can the understanding and use of elementary statistics be improved? Young researchers need to be taught the practicalities of using statistics at the point at which they obtain the results of their very first experiments.

To encourage established researchers to use statistics properly, journals should publish guidelines for authors, reviewers and editors on the use and presentation of data and statistics that are relevant to the fields they cover. All journals should follow the lead of the *Journal of Cell Biology*[7] and make a final check of all figures in accepted papers before publication. They should refuse to publish papers that contain fundamental errors, and readily publish corrections for published papers that fall short. This requires engaging reviewers who are statistically literate and editors who can verify the process. Numerical data should be made available either as part of the paper or as linked, computer-interpretable files so that readers can perform or confirm statistical analyses themselves.

When William Strunk Jr, a professor of English, was faced with a flood of errors in spelling, grammar and English usage, he wrote a short, practical guide that became *The Elements of Style* (also known as Strunk and White)[8]. Perhaps experimental biologists need a similar booklet on statistics.

David L. Vaux is professor of cell biology at the Walter and Eliza Hall Institute of Medical Research and at the University of Melbourne, Parkville, Victoria 3052, Australia. e-mail: vaux@wehi.edu.au

This article was first published in Nature 492, December 2012; *re-published with permission.*

[6]Vaux, D. L., Fidler, F. and Cumming, G. EMBO Rep. 13, 291–296 (2012).

[7]Rossner, M. The Scientist 20, 24–25 (2006).
[8]Strunk, W. Jr and White, E. B. The Elements of Style (5th edn) (Allyn and Bacon, 2009).