

How robust are your data?

New rules for the presentation of statistics.

Thanks to advanced imaging technologies and better integration with molecular and systems approaches, cell biology is undergoing something of a renaissance as a quantitative science. Robust conclusions from quantitative data require a measure of their variability. Cell biology experiments are often intricate and measure complex processes. Consequently the number of independent repeats of a measurement can be limited for practical reasons, yet the variability of the measurements can be rather high. Cell biologists have developed good intuition to guide their analysis of such constrained datasets. Biological complexity and the reliance on intuition can cause culture shock to physical scientists crossing over into cell biology (a kind of extension of the celebrated 'two cultures' concept of C. P. Snow).

With the arrival of quantitative information and '-omic' datasets, statistical analysis becomes a necessity to complement instinct. The problem is that statistical tools are built on basic assumptions such as the independence of replicate measurements and the normality of data distribution. Usually, sizeable datasets are prerequisite for statistical analysis. Alas, these can be as hard come by as a biostatistician (n is typically well below 5). The result is that all too often statistics (frequently undefined 'error bars') are applied to data where they are simply not warranted.

There are no easy solutions to rectify the prevalence of poor statistics in cell biology studies. However, an obvious recommendation is to consult a statistician when planning quantitative experiments. Consider whether n represents independent experiments (you may actually be publishing a measure of the quality of your pipette!) and whether it is large enough for the test applied. Avoid showing statistics when they are not justified; instead, show 'typical' data or, better still, all the measurements. Importantly, displaying unwarranted statistics attributes a misleading level of significance to the data. Always describe and justify any statistical analysis applied. We have updated our guidelines to reflect these recommendations (www.nature.com/ncb/pdf/gta.pdf). One key rule: if the number of independent repeats is less than the fingers of one hand, show the actual measurements rather than error bars. If you wish to present error bars, include the actual measurements alongside them.

Finally, please remember that you are interrogating a complex system — be careful not to discard 'outlier' data points on a whim, as they may well be as relevant as clustered measurements. One is naturally inclined to ignore data that do not match the hypothesis tested, but biology is rarely as black and white as we would like. Do not make 'hypothesis driven' research become 'hypothesis forced'!

Attribution and accountability

Author contribution statements are now mandatory and author responsibilities have been clarified.

Nature Cell Biology has encouraged author declarations for a decade and now almost 90% of papers carry them. The declaration is important as it adds transparency and accountability, as well as assigning credit. In an effort to standardize the information, we have now made it mandatory.

While authors can structure them as they see fit, we require that every author is listed.

Although the corresponding author(s) of a paper is implicitly responsible for the accuracy and integrity of the data presented, we have refined our guidelines on author responsibility: in a collaboration, a senior researcher must take responsibility for the contribution of each group. This includes verifying that the data and conclusions accurately reflect the source data, that data analysis and image manipulations have been made in accordance with our guidelines (www.nature.com/ncb/authors/index.html), that the original data have been archived and that materials, data and algorithms have been deposited in the appropriate databases and will be distributed to interested parties. Corresponding authors are responsible for ensuring that all co-authors agree with the content and author list of a manuscript and for informing them of issues that arise before and after publication; they must certify that their manuscript fulfils the policies outlined at www.nature.com/authors/editorial_policies/authorship.html as part of the online submission process.

Reproducible methods

Nature Cell Biology will publish online methods in more detail.

A central tenet of academic research is that the data are exposed to public scrutiny. Data that survive debate and experimental validation form the basis of new research, which nurtures the tree of knowledge (the shedding of the odd branch in a storm of controversy is a healthy part of the process). The dissemination of new research is commonly facilitated by publication in peer reviewed journals, ensuring that data pass a set of standards set by the research community and applied by referees and editors.

An essential part of the process is that scientific papers are sufficiently detailed to allow for assessment of the data and for independent reproduction of experiments (we have commented previously on the decline of reproduction of data; Editorial *Nature Cell Biol.* 8, 541; June 2006). In an effort to improve the information provided by the papers we publish, we have previously called for a ban on 'data not shown' (Editorial *Nature Cell Biol.* 8, 541; June 2006) and it is our policy to display uncropped data (Editorial *Nature Cell Biol.* 8, 203; March 2006), as well as the sequences of nucleotide probes and antigens (Editorial *Nature Cell Biol.* 9, 481; May 2007). We have also set clear policies for the sharing of research materials (Editorial *Nature Cell Biol.* 8, 425; May 2006).

However, a criticism rightly levied at journals with an intentionally terse format is that excessively tight word limits are not compatible with methods sufficiently detailed for reproducibility. Most read papers online, and this format allows for the cost-effective display of limitless information. As a result, we have joined the other *Nature* titles to present our 'Methods' sections online only. Although we have relaxed our format guidelines for this section, we suggest that methods are limited to around 1,600 words.

Notably, the 'Methods' section will remain integral to the main online paper and it will be copy edited. Although references in the 'Methods' will only be included in the online edition of the manuscript, they will be taken into account for impact factor calculations.

All the policies discussed above are listed in our online 'Guide to authors' (<http://www.nature.com/ncb/authors/index.html>)