

## The art of reporting numerical data

Jonathan A Cook, Dongquan Bi, Jonas Ramstan

It is unfortunate that in English and a number of other languages we use the same term “Statistics” to refer to both numerical data (e.g. national statistics) and also the science of collecting and analysing data (e.g. a degree in Statistics). The former can be very mundane and routine, whereas the latter, which utilises the former to understand variability and carry out inference, enlightening and often sobering. The reporting of numerical data should be informed by statistical principles (the science of statistics). One area where this can be counterintuitive is the level of numerical precision to report data (both of individual values and summaries like means and standard deviations). A review of three recent BJS articles identified over 1000 statistics across the articles which either summarised data from the respective study, or reported an output from an analysis of study data. Often in manuscripts submitted to BJS, values are reported to an excessive level of precision.

Simplicity, neatness of presentation, a desire to report fully a numerical calculation, or a false understanding of the value of the data collected, can all lead to this reporting precision to an excessive level. For example it is tempting, but somewhat misleading to report 13 out of 39 as (33.33%). The use of 2 decimal points (4 significant figures) gives the false impression of greater precision than really occurred. In this example, each observation accounted for over 3% of the final percentage. There is a quirk of numerical data that percentages will often not add up to exactly 100% where there are 3 or more categories. However, as long as the number of observations within each category are accurately reported with the percentage in the group this apparent discrepancy is but a small price to pay for greater clarity and a more honest representation of the data. A simple rule of thumb is that if there are less than 100 observations in a sample reporting percentages to fractions of a percent is not helpful. Arguably even for larger samples it is rarely necessary except perhaps for reporting the extremes (e.g. 0.1% versus 0%). Many people will find tables filled with redundant decimal points more difficult to read and understand, lost in a sea of unnecessary figures. The number of people who are able to differentiate between tenths of a percent let alone one hundred of a percent are in our view rather miniscule. Equivalent arguments can be made for reporting proportions where 2 decimal points should be the standard level of precision.

Continuous data is perhaps the hardest to fairly report; to do so requires understanding of the quantity measured, how well it can be measured, how it is or might be used, and how well the original data was recorded. While an operation time reported as 134.682341 minutes can be recorded and in principle could occur it is of no greater value (and certainly less clarity) than recording 135 minutes. You would be a brave person to believe an operation time is accurately recorded to this level of precision (microseconds) in practice. For the poor reader of papers, this is even more so the case when it is recognised that the reported definition of procedure times in studies (it is often not stated), and that operating practices vary so much between surgeons, surgical teams and institutions. My 135 minutes could easily be your 140 minutes. Reporting mean operation times beyond minutes is of little clinical value irrespective of how large the study sample size is.

Another area where excessive decimal points are commonly used is in reporting p-values as if more zeros creates stronger evidence. This probably reflects both some misunderstanding of what a p-value is, which is an indirect measure of evidence against a hypothesis, and not a measure of the strength, nor the magnitude<sup>1</sup> of statistical disagreement (only of the unlikelihood of compatibility assuming the hypothesis). If the conventional statistical approach is being used with a-prior statistical significance level specified (say, the typical 2-sided 5% level), then reporting to sufficient level to see how the p-value related to that marker of statistical significance is what is important. As a

consequence reporting a p-value as “ $p < 0.05$ ” or “NS” does not provide sufficient detail as it is unclear how close or not the p-value was to the cut-off point. However, once the value is some distance from the cut-off marker then further precision is of little consequence. P-values above 0.1 can happily be reported to 1 decimal point (e.g. 0.2) and if 0.05 is the significance level of interest, then the differences between p-values  $< 0.001$  are of little merit. Where large numbers of hypothesis are tested (e.g. analyses of genetic data), the significance level of interest should be reduced<sup>2</sup> but the same principle can be applied to the corresponding lower significance level. Here, as with reporting other outputs of statistical tests and models such as mean difference or an odds ratio with confidence intervals, the level of precision that is appropriate should be determined by the question we are seeking to answer and how the findings might be applied, as well as the level of precision the data can bear.

Helpful guidance on reporting numbers is available elsewhere<sup>3</sup> for a range of statistical metrics beyond those covered here. It is difficult to be too prescriptive as surgery and science are wonderfully complex. Nevertheless, it would benefit our research, and the readers of it, if we were more circumspect in reporting our data, and a bit more humble in our data presentation.

## References

1. Cook JA, Ranstam J. Statistical methods that provide an effect size are to be preferred. *BJS* 2016; 103:1365. DOI: 10.1002/bjs.10237
2. Cook JA, Ranstam J. Spurious findings. *BJS* 2017; 104:97. DOI: 10.1002/bjs.10239.
3. Cole TJ. Too many digits: the presentation of numerical data. *Arch Dis Child* 2015; published online 15 April.