EDITORIAL

Show the data, don't conceal them

GB Drummond[1] and SL Vowler[2]

[1]*Department of Anaesthesia and Pain Medicine, University of Edinburgh, Edinburgh, UK, and*
[2]*Medical Statistician, Cancer Research UK, Cambridge, UK*

Correspondence
GB Drummond, Department of Anaesthesia and Pain Medicine, Royal Infirmary, 51 Little France Crescent, Edinburgh EH16 4SA, UK. E-mail: g.b.drummond@ed.ac.uk

----

Gordon Drummond is Senior Statistics Editor for *The Journal of Physiology*. Sarah Vowler is a medical statistician with Cancer Research UK.

Data presentation and statistical analysis in scientific writing are agreed to be in need of improvement, despite the profusion of advice and instruction. Recent evidence supports the need for better planning and analysis of animal experiments. This series of short articles aims to provide advice in small easily digested pieces, on a variety of topics, both basic and more specialized, that are relevant to readers of the journal. The present article encourages authors to present data clearly, preferably as a dot plot, so that the distribution of the values can be recognized. The use of different measures of distribution of a population, and different measures of precision of an estimate is contrasted.

This short editorial is the first of a planned series to guide the handling of data from laboratory experiments. We hope that authors (and editors) may find this helpful in writing and assessing future journal articles.

It's clear that data handling in science needs to be improved, if only from the number of attempts that have been made to rectify the problem. Books, articles and websites give copious words of advice, too numerous to cite. Many are indeed helpful, but others are often too dense and discouraging for the non-specialist. Often the advice is 'too general in nature, too limited in scope, and too specialized in vocabulary to be useful to most authors and editors' (Lang and Secic, 2010). On one occasion the American Physiological Society gave specific advice (Curran-Everett and Benos, 2004). When the guidelines were followed up, the results were found to be mixed, with some praise but also criticism and controversy, with little overall effect on the quality of publication (Curran-Everett and Benos, 2007). In general, encouraged by the slow but substantial change that has occurred in medical journals, there seems to be a mood to improve, although changes in basic science journals have been slow, if they have occurred at all. Comparison of basic science reports with clinical studies is not flattering (Watters and Goodman, 1999).

We are well aware that advice alone has been ineffective. We could easily list a comprehensive set of guidelines, which would probably suffer a fate similar to those that have gone before. In contrast we hope that by keeping the articles in this series short and focused, they will offer advice that is digestible and palatable. They may even attract readers who find such topics difficult and unappealing when they are presented in larger portions. We shall try to avoid technical terms and complicated maths.

Much advice starts with a suggestion that one should consult a statistician. Considering the time and cost of some experiments, this would often be prudent. Admittedly statisticians are a rare breed and can be hard to find. Often when the data are given to a statistician there is a problem with them that might have been avoided if the consultation had occurred at the start. Fisher once said: 'To consult a 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' (Fisher, 1938). In practice, it's likely that basic scientists do not know that they should seek advice beforehand, as they may not be sure what they will find, or they consider that what they will find will be easily described and analysed with a simple toolbox of tests. On occasion, physiological studies resemble a random walk through a series of 'what ifs?' and 'how can we prove this mechanism?' stages that may leave a statistician searching for a single testable hypothesis. When basic scientists do have data to analyse, they may think the exciting part of the study

is done. They look around for an easy way to deal with their data, and often ask someone with possibly more experience but probably no more training what to do. The advice is often 'well, this is what I did . . .', almost designed to sustain the status quo and perhaps perpetuate error.

A recent article repeated the often heard advice: 'The choice of how to express the data is very important and should not be made solely on the basis of habit or convention. Always inspect the data in its raw form' (Lew, 2007).

Data presentation to reveal the distribution of data

> Display data in their raw form.
>
> A dot plot is a good start.
>
> 'Dynamite plunger plots' conceal data.
>
> Check the pattern of distribution of the values.
>
> Take care when samples are small.
>
> Choose appropriate measures of precision or dispersion.

In this editorial, we argue that we should use methods of data presentation that allow inspection, not concealment, of the nature of the distribution of the data. This is a first important step in good statistical practice. The next step is entirely dependent on the distribution of the data, so it is important that this is verified rather than just assumed.

In many journals, including *The Journal of Physiology*, the current convention used to illustrate a set of data is to use a solid vertical bar, with a T at the top to indicate the precision of the estimate: this is often called the 'dynamite plunger plot'. Most often, the bar indicates a mean value and the plunger the standard error of the mean (SEM). In most circumstances this convention is unnecessary, and can conceal important features of the data.

Is there a better way? To some extent this depends on what the figure intends to show. However, in many cases this concept of 'what do we wish to show?' may not even have been considered: the author's aim is often just to present 'a result' without first considering the underlying implication or characteristics of this 'result'. Time spent inspecting the raw data allows a visual impression, which can indicate their nature.

Figure 1 shows two sets of data, each with 50 observations, randomly sampled, with very similar mean values, but drawn from two populations with differing distributions. The dot plots of the raw data clearly demonstrate that the distribution of the points is different between the two sets. The first set has a Gaussian (or normal) distribution, so the values are symmetrically distributed. The second set is right skewed (lognormal), with a limited number of smaller values and a few large values. This type of distribution of values is quite common in biology, for example plasma concentrations of immune or inflammatory mediators. Imagine presenting the plunger plots only: who would know that the values were skewed – and that the common statistical tests would be inappropriate?

The dot plot shows each observation we have sampled. On the other hand, the plunger plot obscures both the number of values and their distribution. The entire column of the plunger plot attracts attention. The width, colour, or
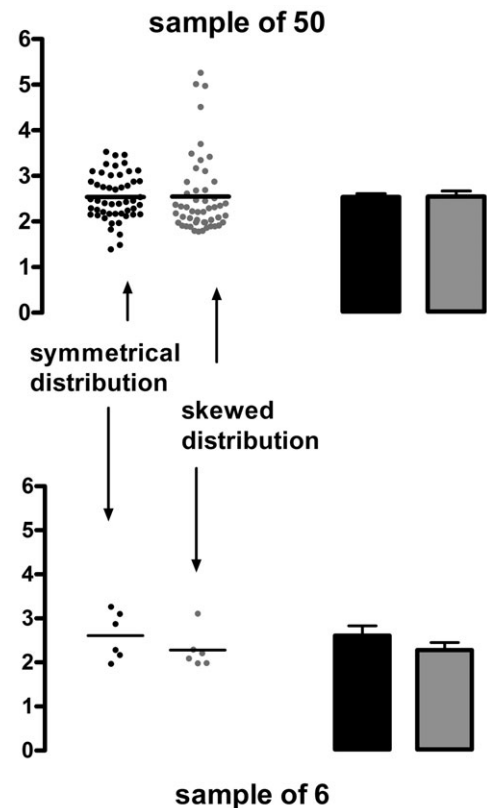

Figure 1

Comparison of a dot plot with a plunger plot. A. two datasets, with the same mean value, displayed as either dot plots or plunger plots. The bars are mean values: error bars indicate SEM. B. subsamples of 6 data points from the above datasets. The line indicates the mean of these values.

shading of the column can distract the eye, whereas in fact the only feature that's relevant is the height of the column itself. When a bar plot is used, any error bars that are shown are often not properly defined, and often only display the upper error, assuming, often incorrectly, the lower error to be symmetric. If measures are paired or repeated several times, then it's preferable to be able to show these relationships. This is possible with a dot plot by joining up the dots with lines indicating relationship. This cannot be done with plunger plots.

Dynamite plunger plots are never an appropriate way to plot the data. If dot plots are inappropriate, because the sample size is too large and the plot becomes too cluttered, then a better way to plot the data is to have a point or line representing the mean, rather than the commonly used bar. Errors bars illustrating the 95% confidence interval (CI) should also be included. The figure should be appropriately annotated to indicate that the mean and 95% CI are being shown.

We also show a subsample of six observations. This is a more frequent size of sample in lab studies (Watters and Goodman, 1999). The precision with which we can estimate the exact features of the entire population is reduced. Notice first that the asymmetry of the skewed data results in a different mean value and that, once more, these variations are
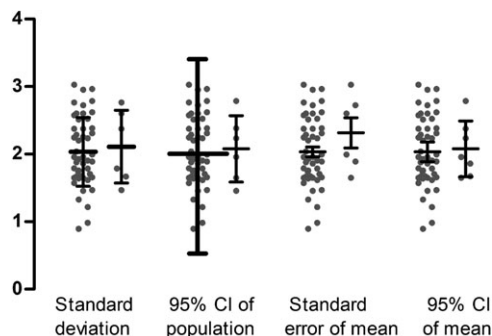
Figure 2

Different types of error bar. In each pair of examples, 50 and 6 samples have been taken. The central line represents the mean value, and the error bars show the values stated. Note how the precision of the mean value (i.e. the SEM and the CI of the mean) is affected by the small sample, whereas the indices of dispersion (SD and CI of population) are less affected.

not evident with the plunger plots. Also notice that the skew of the values is less evident. However, if the mathematical basis of the data was known to give the data a propensity for this effect, then this could still be reasonably suspected.

Once we have had the chance to 'see' the data with the dot plot, we will appreciate more clearly how we may wish to demonstrate other features of the data. What do we wish to present from our sample, in addition to the actual values? If we want to summarise the entire population and in particular indicate the extent of variation of these values, then the appropriate values to calculate are the mean and the standard deviation (SD). When data are skewed, the mean is not always the best choice of summary. An alternative and perhaps preferable index of scatter would be the 95% confidence limits. However, 95% confidence limits should only be used when a substantial sample size is available.

The values we sampled are thus used to estimate the characteristics of the entire population. The larger the sample, the more likely it will be representative and the better our estimate of the entire population will be. Note that this is only true if the selection of samples has been random. If there has been bias in the selection process, then the data may not become more representative as more observations are added.

Using such estimates of the population, we could determine if another subsequent value that we obtained was abnormal (because it fell outside an expected range that we had defined). In some cases this may be why the original sampling of the population was done. To define reference ranges a very large sample of normal values is needed. Then the 95% CI is chosen as the reference range. This implies, by definition, that those 5% of 'normal values' that lie outside the range would be considered abnormal. We can then say that a new value drawn from this population would be likely, 95% of the time, to have a value within these confidence limits. These features of the entire population are shown on the left in Figure 2.

However, in experimental biology data are usually obtained to compare one group with another. To summarise the characteristic feature of each group, the mean is calcu-

lated. To assess the precision with which we believe we know this value, we calculate the standard error of this mean. Again, a better index of this precision would be the 95% confidence values for the mean, which indicate that we'd be 95% likely to get another estimate of the mean within this range, using another set of data samples from the same population. Note how subsamples of the data give similar estimates of the entire population, but reduce the precision of the estimate of the mean (i.e. the SEM is greater, because it is inversely proportional to the number of observations).

To summarise, although the dynamite plunger is a frequent plot in many subject fields, it has been progressively discarded in others, because it emphasises the wrong features. Data can be better presented and compared using alternative methods. This is important where inferences have to be drawn about differences between samples, such as the effects of treatment (a topic we will address later). We argue, as many others have done before, that the simplest method may often be the best: just plot the actual values.

A big advantage of the plots we illustrate is that they allow easy application of an under-used, but familiar and comforting method of analysis: the 'eyeball' test. It may be crude, but it's robust, and if used in obvious cases, can become the 'barn door' test: all that's needed.

Later editorials will deal with using samples for statistical comparison, and other frequently used statistical tests.

Acknowledgements

Conflicts of interest

None.

References

Curran-Everett D, Benos DJ (2004). Guidelines for reporting statistics in journals published by the American Physiological Society. Am J Physiol Cell Physiol 287: C234–C245.

Curran-Everett D, Benos DJ (2007). Guidelines for reporting statistics in journals published by the American Physiological Society: the sequel. Adv Physiol Educ 31: 295–298.

Fisher RA (1938). Presidential address to the First Indian Statistical Congress. Sankhya 4: 14–17.

Lang TA, Secic M (2010). Preface. *How to Report Statistics in Medicine. Annotated Guidelines for Authors, Editors, and Reviewers*, American College of Physicians: Philadelphia, pp. xv–xvi.

Lew M (2007). Good statistical practice in pharmacology. Problem 1. Br J Pharmacol 152: 295–298.

Watters MPR, Goodman NW (1999). Comparison of basic methods in clinical studies and *in vitro* tissue and cell culture studies reported in three anaesthesia journals. Br J Anaesth 82: 295–298.