British Journal of Pharmacology (2007) 152, 295–298
 0007–1188/07 $30.00
www.brjpharmacol.org

REVIEW

Good statistical practice in pharmacology Problem 1

M Lew

Department of Pharmacology, University of Melbourne, Parkville, Victoria, Australia

Background and purpose: This paper is intended to assist pharmacologists in making the most of statistical analysis and in avoiding common errors that can lead to false conclusions.
Approach: A scenario is presented where a pathway inhibitor increased blood pressure responses to an agonist by about one third. The fictional experimenter concludes that the inhibitor enhanced the responses to the agonist, but has not applied any statistical analysis. Questions are asked of the reader, and a discussion of the author's answers is presented.
Results: The agonist responses have unequal standard errors, as often seen in data like these concentration-response curves with responses expressed as change from baseline. The uneven variability (heteroscedasticity) violates an assumption of conventional parametric statistical analyses, but can be corrected by data transformation. Expressing the data as absolute blood pressure and then transforming it to log blood pressure eliminated the heteroscedasticity, but made evident an effect of the inhibitor on baseline blood pressure.
Conclusions and implications: Statistical analysis is a sensible precaution against mistakes, but cannot protect against all erroneous conclusions. In this scenario, the inhibitor reduced the blood pressure and increased responses to the agonist. However, it is likely that the latter effect was a consequence of the former and thus no conclusion can be safely drawn about any direct interaction between the agonist and the pathway inhibitor from this experiment. Where results are awkward to interpret because of confounding factors such as an altered baseline, statistical analysis may not be very useful in supporting a safe conclusion.
British Journal of Pharmacology (2007) **152**, 295–298; doi:10.1038/sj.bjp.0707370; published online 9 July 2007

Keywords: data interpretation; statistical analysis; data transformation; research design; initial conditions; Monte Carlo method; concentration-response curve

Problem

The following data were obtained from an experiment intended to determine the effect of pretreatment with an inhibitor of an intracellular signalling pathway (test) on the blood pressure responses of anaesthetized rats to an α-adrenoceptor agonist (agonist). Each rat was randomly allocated to either control or treatment groups ($n=8$ rats in each group), and the bolus doses of the agonist were applied in an ascending sequence after the effect of the previous dose had worn off (Figure 1).

Without any statistical analysis, the experimenter concluded that the pathway inhibitor enhanced the responses to the agonist.

(1) Is the conclusion reasonable given the data? Should a statistical test be applied to the data before deciding whether the intervention was effective?
(2) Would any re-expression of the data be necessary before analysis of the data?
(3) What conclusion do the data support?

Correspondence: Dr M Lew, Department of Pharmacology, University of Melbourne, Parkville, Victoria 3010, Australia.
E-mail: michaell@unimelb.edu.au
Received 4 October 2006; revised 20 December 2006; accepted 30 May 2007; published online 9 July 2007

Analysis of problem 1: A case of expressive obscurity

1. Is the conclusion reasonable given the data? Should a statistical test be applied to the data before deciding whether the intervention was effective?

The conclusion looks quite solid, and you do not need always to apply statistical analyses before making a conclusion. However, effects have to be quite large before they can be reliably identified by even an experienced experimenter – particularly if the experimenter has an interest in the result! Statistical analysis is a powerful tool for avoiding mistaken conclusions, and given that the small cost involved for analysis of most common experimental designs – just a little thought and effort – there is scant justification for leaving that tool in the toolbox.

Statistical analysis of the data shown in this problem would indicate that the effect has only a very low probability of being the result of chance alone, but the obvious conclusion that follows – that the intervention enhanced the responses to the agonist – is very misleading, as we will soon see. Statistical analyses generally provide guidance about the probabilities of very specific hypotheses, and it

turns out that hypotheses regarding the responses as percent of the baseline are not very interesting in this experiment. The results of a statistical test are not the only thing to consider when making scientific inferences.

2. Would any re-expression of the data be necessary before analysis of the data?

Yes. There is an uneven distribution of the variability in the data, with little variability in small responses and lots in the large responses. (Note: the condition where variances are unevenly distributed is often called 'heteroscedasticity', but we should avoid words with more than seven syllables!) The unevenness is not particularly pronounced and might commonly be ignored, but doing so can lead to biases in the results of ordinary least-squares regression and statistical analyses like analysis of variance – the analytical approaches most likely to be applied to this type of data – because those analyses are based on the assumption (among others) of equal distribution of variability across the data.

The particular pattern of heterogeneity of variance visible in these data is quite common in pharmacological studies and so we will explore it a bit further. Some of the heterogeneity is a consequence of the way that the data are expressed: expression of the measurements as percent of the baseline inevitably results in the baseline values having no variability. To fix that part of the problem, we only have to return the data to the raw form of blood pressure. However, for these data, the variation is still unevenly distributed in that form because there is a strong positive correlation between the means and their variation – a common pattern of unevenness that occurs when measurements are bounded at the low end, when data have a natural tendency to scale exponentially, or whenever the underlying population is right-skewed. Many variables commonly measured in pharmacology have such properties. Concentrations are always bounded because they cannot be less than zero, and it has been shown empirically that the concentrations of many blood constituents have right-skewed distributions (Flynn *et al.*, 1974). The distributions of concentration–response curve EC_{50}s (which are concentrations) are right-skewed. Cultured cells grow exponentially, at least to a degree, and counted values are theoretically distributed according to the right-skewed Poisson distribution.

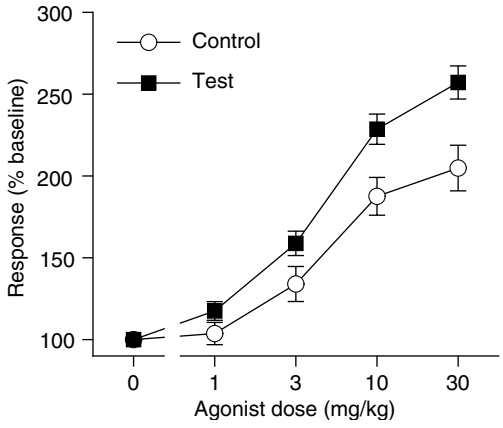

Figure 1 Blood pressure responses to an α-adrenoceptor agonist in untreated rats (control) or in rats pretreated with an intracellular pathway inhibitor (test). Each rat was randomly assigned to either treatment or control ($n = 8$ per group) and the agonist doses were applied in ascending sequence.

The log-normal distribution is a common example of a right-skewed distribution, and so it is not surprising that a logarithmic transformation of such data can even out the distribution of variance where the variance is correlated with the mean. Other transformations such as taking the square root or reciprocal of the values will also fix the unevenness of variance in some situations, but the log transform is a good one to try first. For the present data, log transformation of the blood pressure values almost eliminates the unevenness of variance and completely eliminates the correlation between mean and variance (Figure 2).

Statistical analysis of log-transformed data does require that the hypotheses tested relate to the logarithmic values, so one needs to be sure that the log transform does not make the outcome difficult to interpret. Conveniently, in most cases, a hypothesis relating to the magnitude of a property is logically equivalent to a hypothesis relating to the logarithm of the magnitude of the property, so that issue usually does not matter in practice. It certainly is not an issue in this experiment, so we will apply a log transform (Figure 3).

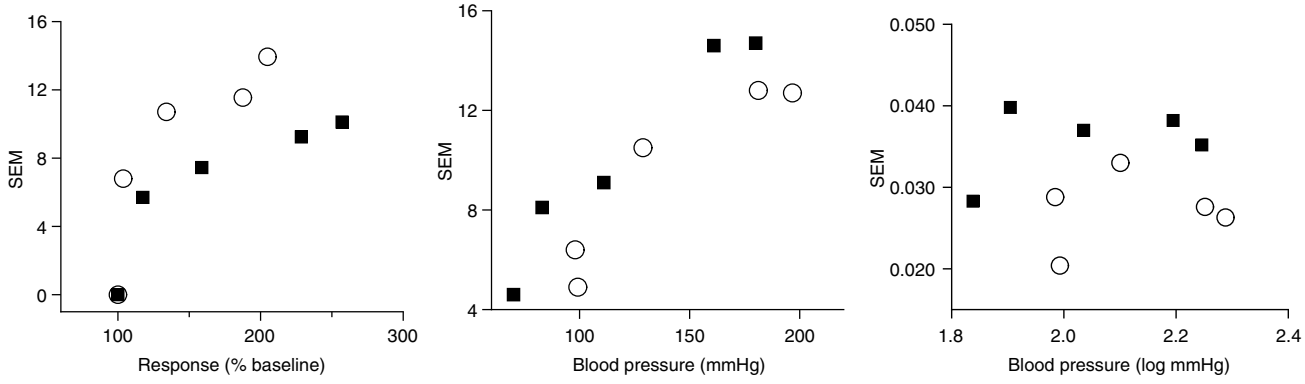

Figure 2 Variability of standard error of the mean (s.e.m.) with responses expressed as percentage of the baseline (left panel), blood pressure (centre panel) and log blood pressure (right panel). The logarithmic transformation removes the tight correlation between the size of the response and the variability in the response measurement.

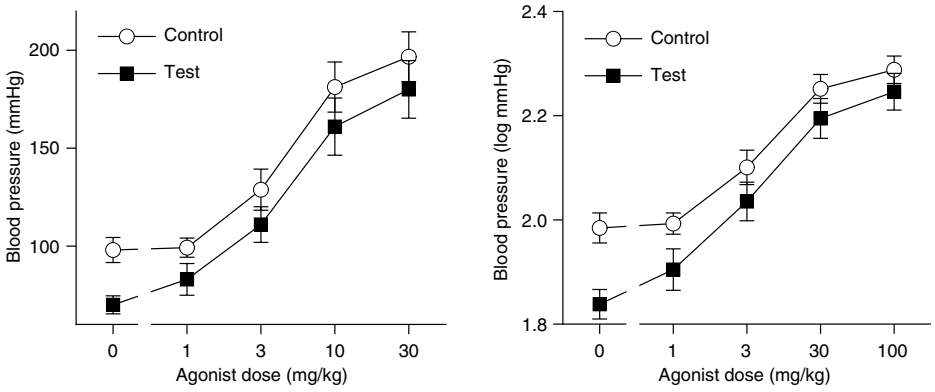

Figure 3 Data from Figure 1 re-expressed as raw blood pressure (left) and as log blood pressure (right). Note that these illustrations of the data allow the baseline blood pressure to be displayed. Compare the patterns of effect with that in Figure 1.

It is important to note that while the log transform is effective in eliminating the heterogeneity of variance in these data, other patterns of uneven variation may not be so easily identified and corrected (see Keppel and Wickens, 2004, section 7.4, for an extensive discussion of heterogeneity of variance). The best approach in those cases might be to use statistical analyses that do not assume equal variances in the first place. For comparison of two means, one might use the Mann–Whitney *U*-test or Welch's version of the *t*-test that is modified specifically for unequal variances, although it is arguable that an exact permutation test would be the best approach (Ludbrook, 2000). For comparisons between more than two means, the Kruskal–Wallis analysis of variance on ranks might be used, or a more powerful exact permutation procedure. Those tests will not be dealt with any further here, because they would not be the best approach for the data in the current problem. However, the topic of non-parametric tests will be expanded in another paper in this series.

3. What conclusion do the data support?

Based on the percent baseline data (Figure 1), one might reasonably conclude that the intervention has increased the responses to the agonist, but based on the raw blood pressure data or the log blood pressure data (Figure 3), one would probably conclude that the intervention has decreased the baseline blood pressure and not substantially altered the responses to the agonist. Which interpretation is correct? Both … or perhaps neither.

It is possible that there was a failure of randomization in the experiment, or an uneven randomization, such that the rats in the control and test groups had different blood pressures before treatment. In that case, there would be no point in considering effects of the intervention any further. However, let us assume that the baseline difference was an effect of the intervention. Even so, if we *define* a response as having units of percent of baseline, then there really was an increase in the responses as a consequence of the treatment. The change in baseline blood pressure does not negate the conclusion that responses were enhanced, rather it provides the likely mechanism for their enhancement. While such a conclusion is sound, it is dangerous. The experimenter and readers might easily be misled into thinking that there was a direct interaction between the signalling pathway inhibited and the effectiveness of the agonist. A convincing demonstration of such an interaction would require determination of what the intervention did to the agonist responses *independently* of any effect on baseline pressure. Using a percent baseline scale for responses would be a particularly poor choice because that scale is 'fragile', in that it can be confounded by any change in the baseline.

The experimenter may not have expected the intervention to change the baseline blood pressure – in fact, it would be uncharitable to suggest that anyone would choose to express the data as percent of baseline if an alteration in the baseline was expected – but the results suggest quite strongly that it did: analysis of the baseline blood pressures values with Student's *t*-test gives $P = 0.003$. However, it has to be noted that test was unplanned and consequently carries an inflated risk of a false-positive outcome. Unplanned comparisons are selectively applied to test for (unexpected) differences that are apparent in the data. Some of the time the unexpected differences will be random rather than real treatment effects, but if they are apparent to the eye then they are likely to be 'declared' significant by statistical analysis.

It is impossible to accurately correct for that inflation of the false-positive outcome rate without making an assumption about the (unknown) true effect rate, but consideration of the scientific rationale for an effect can help. If it is difficult to explain how the intervention could have decreased blood pressure, then we might suspect that the statistical outcome is a type I error. On the other hand, if it is easy to explain how the intervention could have lowered blood pressure, then we might be justified in accepting that it really did so. Unlikely results need to be supported by strong evidence; predictable results need less.

We do not know how likely the blood pressure lowering effect of inhibiting the particular pathway was, but $P = 0.003$ is low enough to absorb quite a strong correction for being unplanned without becoming unconvincing. Nonetheless, it is important that the results of an unplanned comparison be treated differently from the results of a planned comparison, at least by being identified as such in any publication. Ideally, the result should be confirmed with a new experiment before publication. Some statisticians go so far as recommending that the new experiment be conducted by a different laboratory. Good and Hardin (2003) suggest that

'no reputable scientist would ever report results before successfully reproducing the experimental findings twice, once in the original laboratory and once in that of a colleague.' They probably had in mind a more profound result than what we are discussing here, but there is a sharp contrast between their caution and the more cavalier approach to 'significance' used by most of us.

So where are we? We have a substantial effect of the pathway inhibitor on the percent baseline blood pressure responses to the agonist but no effect on the raw blood pressure responses, and we have a large unpredicted effect of the intervention on the baseline blood pressure. We are not really in a position to draw any clear conclusion about the biology, and even the most thorough statistical analysis of the available data will not really change our situation. Looked at in that way, the experiment is not much more than a preliminary study suggesting a new hypothesis that the intervention lowers blood pressure. Whether the experimenter should continue with the original hypothesis that the intervention alters the effectiveness of the agonist would depend mostly on what rationale there was for that hypothesis in the first place. The experiments would need to be altered and expanded substantially to provide convincing evidence of an effect of the intervention on the agonist responses that is independent of the blood pressure lowering effect. A range of doses of the signalling pathway inhibitor should be investigated and controls for the effect of altered baseline blood pressure would need to be devised.

Conclusions and recommendations

(1) Uneven distribution of variance is commonly encountered in pharmacological data and is generally associated with skewed data distributions. Those conditions violate assumptions built into conventional parametric statistical analyses. When the means and the variances are correlated, you should consider using a logarithmic transformation of the data to reduce or eliminate the problem.

(2) 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 before any normalization that can alter the nature as well as the extent of apparent effects, and think about how normalization might obscure or exacerbate any confounding effects of uncontrolled influences.

(3) Statistical analysis is a powerful tool for supporting decisions about the meaning of experimental results. However, where results are awkward to interpret because of confounding factors, statistical analysis is rarely very helpful. We have to be prepared to treat each set of experiments as preliminary and use the results as a guide to forming specific hypotheses and to designing better experiments.

References

Flynn FV, Piper KAJ, Garcia-Webb P, McPherson K, Healy MJR (1974). The frequency distributions of commonly determined blood constituents in healthy blood donors. *Clinical Chimica Acta* **52**: 163–171.

Keppel G, Wickens TD (2004). *Design and Analysis, a Researcher's Handbook*, 4th edn. Pearson Prentice Hall: New Jersey.

Good PI, Hardin JW (2003). *Common Errors in Statistics (and How to Avoid Them)*. Wiley Interscience: New Jersey.

Ludbrook J (2000). Computer-intensive statistical procedures. *Crit Rev Biochem Mol Biol* **35**: 339–358.