Title:
Ophthalmic statistics note 9: parametric vs non-parametric methods for data analysis

Corresponding author:
Simon S Skene
Comprehensive Clinical Trials Unit, University College London, Gower Street, London, WC1E 6BT, UK
Email: s.skene@ucl.ac.uk
Tel: + 44 (0)20 3108 3941

Co-authors:
Simon S Skene
Comprehensive Clinical Trials Unit, University College London, London, UK

Catey Bunce
NIHR Biomedical Research Centre at Moorfields Eye Hospital NHS Foundation Trust and UCL Institute of Ophthalmology, London, UK

Nick Freemantle
Department of Primary Care and Population Health, University College London, London, UK

Caroline J Doré
Comprehensive Clinical Trials Unit, University College London, London, UK

On behalf of the Ophthalmic Statistics Group

Keywords: non-parametric tests, Mann-Whitney test, Wilcoxon signed rank test, t-test, hypothesis testing

Word count: 1542
Abstract
This is the ninth statistics note produced by the Ophthalmic Statistics Group (OSG) which is designed to be a simple guide to ophthalmic researchers on a statistical issue with an applied ophthalmic example. The OSG is a collaborative group of statisticians who have come together with a desire to raise the statistical standards of ophthalmic research by increasing statistical awareness of common issues.

INTRODUCTION/SCENARIO
Distributions of measured data are often well modelled by known probability distributions, which provide a useful description of their underlying properties such as location (average), spread (variation) and shape. Statisticians use probability distributions to interpret and attribute meaning, draw conclusions and answer research questions using the measurements or data that researchers gather during their studies. Different types of data follow different probability distributions and these distributions are characterised by certain features called parameters. Even the most statistically averse of researchers is likely to have heard of the Normal distribution, which is often used to approximate the distribution of continuous or measurement data such as intraocular pressure, central retinal thickness and degree of proptosis. The Normal distribution follows a ‘bell shaped curve’ (albeit with a rim stretching to ± infinity) the shape of which is specified by the mean and standard deviation, with different values of each, giving rise to different bell shaped curves, see Figure 1.

Other distributions such as the Binomial and Poisson probability distributions are less commonly reported in ophthalmic research and are characterised by different parameters. The Binomial distribution is used for dichotomous data and is characterised by the probability of success, i.e. the number of ‘successes’ out of a total number of observed events e.g. the proportion of graft transplants which fail within 6 months of transplantation. The Poisson distribution is used for counts data and is characterised by the mean number of events e.g. endophthalmitis rates.

The assumption that the observed data follow such probability distributions allows a statistician to apply appropriate statistical tests which are known as parametric tests. The Normal distribution is a powerful tool provided the data plausibly arise from that distribution, or can be made to reasonably approximate this following a suitable transformation such as by taking natural logarithms to reduce
Figure 1. Examples of the Normal $N(\mu, \sigma)$ distributions parameterised by different values of mean ($\mu$) and standard deviation ($\sigma$)

asymmetry. The Normal distribution also serves as an approximating distribution to the Poisson or Binomial distribution under certain circumstances, or can be used for large samples to approximate the distribution of the sample mean via the Central Limit Theorem. Tests based on the Normal distribution are therefore extremely useful and form the basis of many analyses, the usual tests being z-tests or t-tests which rely on approximate normality or normality respectively. ¹ If we can assume a normal distribution, then we expect 95% of values to lie within 1.96 standard deviations of the mean.

Parametric tests make assumptions about the distribution of the data and sometimes it may be impossible to assess these assumptions, perhaps because the sample size is small, or because that data do not follow any of the more common probability distributions. Alternatively, we may be interested in making inferences about medians rather than means or about ordinal or ranked data. In such circumstances, statisticians may adopt an alternative class of statistical tests which are
known as non-parametric, or distribution free methods. These methods work by ranking the data in numerical order and analysing these ranks rather than the actual measurements observed. Two of the most well-known non-parametric methods are the Mann-Whitney test (or U-test) and the Wilcoxon matched pairs signed-rank test, which are suitable for data from two unpaired samples or two paired samples respectively.\(^2\)\(^3\)

The Wilcoxon matched pairs signed rank test calculates the differences between each matched pair in the two samples and replaces their absolute values with their ordered ranks (1, 2, 3 etc), ignoring zeros. Under the null hypothesis of no difference between samples, the sum of the positive and negative ranks should be similar. The test statistic is usually taken to be the smaller of the two sums, and exact p-values can be found using statistical software or by comparison with statistical tables.

The Mann Whitney U-test effectively considers all pairs of observations from two independent samples and calculates the number of pairs for which an observation in one sample is preceded by an observation from the other. Again, the U-statistic can be calculated from the summed ranks within each sample, found by ordering the pooled observations.

Such tests depend only on the rank ordering of the observed values and not on any assumptions about their underlying distributions, so that there are no associated parameters to be estimated, and in that sense such methods are considered non-parametric or distribution-free. These are easily implemented in standard statistical software packages such as R, Stata, SAS or SPSS for example.

**Scenario 1**

A colleague has conducted an exploratory randomised controlled clinical trial evaluating a novel treatment for ocular trauma in 40 patients, 20 of whom received standard care and 20 of whom received the novel treatment. The primary outcome measure is visual acuity in the treated eye 6 months after surgery, measured using ETDRS charts at a starting distance of 4 metres. In the analysis of the trial, a decision has to be made between using parametric and non-parametric methods, and she asks me for advice. A histogram of visual acuity is highly asymmetric i.e. the distribution is skewed, so that these data appear to violate the assumption of approximate normality. I decide therefore to propose the Mann-Whitney test, and a p-value of 0.76 leads to the conclusion that there is little evidence of any difference between the medians in the two groups. My colleague asks to see an estimate of the treatment effect. Whilst the t-test would have provided me with an estimate of the mean difference with a confidence interval, no such result is directly
forthcoming from the Mann-Whitney test, although it is possible to calculate the difference in medians and a 95% confidence interval for the difference.4,5

Discussion
Non-parametric tests can be useful, but careful thought should be given on a case-by-case basis as to whether they are the most appropriate method of analysis. Where an assumption of normality is tenable parametric tests will be more powerful, offering greater opportunity to detect differences where they exist, and have the advantage that they provide useful information about the size of treatment effect and confidence intervals directly. The relative loss of power (the probability of finding statistically significant results where differences exist) when adopting non-parametric methods, even where these are appropriate, is well known. It is not unusual to inflate sample sizes by 10% or more to accommodate possible non-parametric analyses resulting in poorer efficiency (i.e. a larger sample size is required to identify a given difference between treatment groups).

It is important to note that, in contrast to what is stated in certain statistical text books, non-parametric tests are not a solution to the problem of small sample sizes.6,7 In fact for comparative samples with less than 4 per group, the Mann Whitney test cannot produce a significant p-value (<0.05) whatever the values of observations in the samples. For small samples where normality is difficult to assess, it may be reasonable to assume a Normal distribution, or to use a transformation, based on the distribution of data from previous larger studies. Non-parametric methods are themselves not free of assumptions, for example the Mann Whitney test assumes that samples arise from the same underlying distribution and differ only in location.

Scenario 2
My colleague brings me results of her comparison of the visual acuity data captured in an exploratory randomised controlled clinical trial. She has used a non-parametric test to compare visual acuity 6 months post surgery. She states that she has used this test because the data were highly skewed. She shows me histograms of the data from each group and looking at this, I realise that she herself has violated an assumption used by the non-parametric test. The distributions differ not only by central location but also in the spread of the data. Examination of the data in both groups suggests that patients either respond greatly or not at all – average change is misleading in this instance.
SUMMARY
With highly skewed or otherwise awkward data the median may be more robust than the mean as a measure of central tendency and is used with non-parametric methods of analysis. However, it should be noted that this approach separates the p-value from the effect size, since the Mann Whitney test for example tells us only whether there is a shift in location between samples, and is by design divorced from the actual estimation of effect. Confidence intervals for the difference in medians should be presented to give an indication of the size of the effect. An alternative to non-parametric methods is given by bootstrapping or re-sampling, but such methods should not be considered without reference to a statistician. Where assumptions of normality are plausible, possibly following a transformation, parametric methods are preferable providing extra power and allowing adjustment for other factors such as differences between treatment groups at baseline in the case of clinical trials.

LESSONS LEARNT
- If the data appear to follow a probability distribution, use the appropriate parametric test. This will maximise power and interpretation.
- If data are highly skewed, see whether a simple transformation will achieve normality but don’t forget to back transform when presenting results.
- If data appear not to follow any common distribution and there are no reports of these data elsewhere to allow you to assume normality, you should consider a non-parametric test.
- If in doubt, analyse your data both ways and see whether the conclusions agree.

Collaborators

Contributors
SSS drafted the paper, CB reviewed and revised the paper, NF and CJD conducted an internal peer review and provided additional comment.

Funding
This research received no specific grant from any funding agency in the public, commercial or not-for-profit sectors.

Acknowledgements
The post of CB is partly funded by the NIHR Biomedical Research Centre based at Moorfields Eye Hospital NHS Foundation Trust and UCL Institute of Ophthalmology. The views expressed in this article are those of the authors and not necessarily those of the NHS, the NIHR or the Department of Health.

Competing Interests
None.
REFERENCES


