

Statistical Advice and Checklist

- [A Word from the RNOH Statistical Consultants](#)
- [Study Design Checklist](#)
- [Descriptive Statistics](#)
- [Quantifying the Uncertainty in your Sample Estimates](#)
- [Choosing the Appropriate Statistical Analysis](#)
- [Sample Size](#)
- [Recommended Resources](#)

A word from the RNOH Statistical Consultants

In our capacity as statistical consultants to various members of staff engaging in research at the RNOH, we have observed a recurrence in the type of advice sought. To this end we have produced a statistical checklist, which researchers could consult as a starting point.

The following set of self-ask questions should allow investigators to conduct their research correctly and with greater confidence. Should any problems remain unanswered then advice should be sought from a statistician. We aim to highlight the standard problems a researcher will encounter at the design stage by linking the following considerations with specific examples from a published study about a randomized controlled trial (RCT) looking at prevention of falls in the elderly.

We also present some general statistical advice on descriptive statistics, quantifying the uncertainty in your sample estimates, selecting the appropriate statistical analysis and sample size, along with some recommended resources for more in-depth reading.

Study Design Checklist

Is the investigator planning a study or has data already been collected? Ideally the advice of a statistician should be sought right at the beginning of any research undertaken. Contributions from the statistician given towards the design of the study and in the drafting of the proposal could avoid problems at a later stage.

Planning: What question is the investigator trying to answer? A clear aim should be stated as motivation for undertaking the research. The design and conduct of your study should be undertaken with a view to primarily answering your main aim or aims.

In falls in the elderly example, the whole trial has been conducted with the clear main aim of ascertaining whether a structured assessment of elderly people who have had falls could decrease the rate of further falls. Other outcome measures considered included death, functional status and use of health care.

In determining the best ways to answers your aims, a number of important considerations should naturally follow:

- **Population:** What is the population of interest? (i.e. to what sort of patients/subjects are the findings of this study to be extrapolated?). How will patients be sampled for the study? Are they representative of the intended population?

In the example, the population of interest are elderly people. Clear inclusion and exclusion criteria are given, for example, all patients aged 65 years and above who attended A&E with a primary diagnosis of a fall were potentially eligible. Patients with strong cognitive impairment were excluded; could this affect the applicability of the findings?

- **Intervention or Exposure:** If undertaking a trial, is the intervention well defined? Is the intervention confounded with other factors? If undertaking an observational study, is exposure status readily determined?

In the example, patients were randomly assigned to one of two groups: The intervention group underwent an assessment with referral to the relevant services indicated, and those assigned to the control group received usual care only. Demographics and other potential confounders were similar in the two groups because of the random assignment.

- **Outcome:** What outcome measures will you be collecting? Will these outcomes help you to answer your primary research question? Can your outcomes be measured accurately/unambiguously? Will measurement be blind?

Follow-up was done by postal questionnaire. Information about subsequent falls was requested. No mention is made of whether assessors were unaware as to which group patients were from (blinding).

Descriptive Statistics

Once your study has been conducted and your data has been collected descriptive statistics should be used to summarize your data in a meaningful way.

When you have Categorical data: The best way to summarize categorical data is to present the number and proportion/percent of individuals or items falling into each category of the variable.

When you have Numerical data: Numerical data should be summarized with a measure of location (to describe the typical data value) and spread (to describe how spread out the data observations are). The most appropriate measure of location and spread will depend on the distribution of the data. If your numerical data is normally/symmetrically distributed the **mean** and **standard deviation (SD)** should be used as respective measures of location and spread. The mean is often referred to as the average value and is found by adding all observations and dividing this by the number of observations in the data set. The SD can be thought of as an average of the deviations of all the observations from the mean of the observations. If your numerical data is skewed the **median** and **Inter-quartile range (IQR)** should be preferred. The median is the middle value in the data set when the data values are arranged in increasing order of magnitude. The IQR contains the central 50% of the ordered observations.

Quantifying the Uncertainty in your Sample Estimates

Generally we are interested in information about populations and hope to use estimates from our sample data (e.g. means/differences in means, or proportions/difference in proportions) to make inferences about the population. The fact that there are many possible samples however means that there are many possible estimates that could be recorded. You will get different estimates from different samples. How good is your estimate?

Along with your sample estimate you also need a measure of uncertainty. The best way to quantify the variability/precision of your estimate is with a 95% Confidence Interval (CI).

The 95% CI gives the range of values within which the true population parameter would lie on 95% of sampling occasions

Choosing the Appropriate Statistical Analysis

In many cases we are interested in carrying out a formal statistical hypothesis test to assess whether there is a statistical difference between a number of groups. There are 4 basic elements behind statistical hypothesis testing:

1. Specify null and alternative hypotheses under study
2. Collect relevant data and choose the appropriate test
3. Calculate *test statistic*
4. Obtain p-value

The p-value is the probability of the observed data *given that the null hypothesis is true* (i.e. that there is no difference between the groups). It is the p-value which measures the strength of evidence *against* the null hypothesis, allowing you to either reject or not reject your null hypothesis.

Very large p-value (p = 0.7)	<p>=> data could occur often when H_0 is true</p> <p>=> can't reject null H_0</p> <p>=> NOT evidence that H_0 is true</p>
Very small p-value (p = 0.001)	<p>=> appears implausible since these data would rarely arise by chance when H_0 is true</p> <p>=> reject H_0, in favour of H_A</p>

It is common to use a cut-off of $p=0.05$ to determine whether something is 'significant' or not.

Guidance for Analysis with Numerical Data: When you have numerical data, if your data is normally/symmetrically distributed (best determined by a histogram of your data) then use the appropriate parametric analysis; otherwise, use the non-parametric equivalent.

Appropriate Hypothesis Test	Parametric Test	Non-Parametric Test
---	------------------------	----------------------------

To compare two paired groups on the same variable , for example, <i>a comparison of the number of falls suffered in one group of patients before and after occupational therapy assessment.</i>	Paired t-test	Wilcoxon signed-rank test
To compare two independent/unpaired groups on the same variable , for example, <i>a comparison of the number of falls suffered between the assessment group and the control group.</i>	Independent t-test	Mann-Whitney U test
To compare three or more independent/unpaired groups on the same variable , for example, <i>a comparison of the number of falls suffered between the control group, the assessment group and another group that received another form of intervention.</i>	One way analysis of variance (ANOVA)	Kruskal Wallis test

When analyzing the strength of the relationship between two numerical variables **correlation analysis** can be employed. Pearson's correlation coefficient quantifies the strength of the linear association between two numerical variables and goes from -1 to 1; 1 is a perfect positive linear association (as one variable increases so does the other), 0 suggests no association and -1 is a perfect negative association (as one variable increases the other decreases) e.g. You might look at the relationship *between number of falls suffered and age in a certain group of patients.*

In many cases it may be preferable to conduct **linear regression analysis** which quantifies the linear relationship between one numerical variable and another with an equation, the regression line. This can be a more useful method of analysis when exploring the relationship between two numerical variables when one variable depends on the other.

For more details on linear regression see Petrie and Sabin (2000).

Guidance for Analysis with Categorical Data: *As an example - a comparison of the proportion of patients suffering subsequent falls in the assessment group compared to the proportion in the control group.* If a comparison of proportions between two independent samples is being undertaken then use a **Chi-squared test** if numbers are large or **Fisher's exact test** when numbers are small. If a comparison between three or more independent samples is being undertaken then use a Chi-squared test. If numbers are small, then some categories may need to be combined if it makes sense to do so.

Guidance for Analysis with Other Types of Data: Survival data, i.e. when the time for a certain event to occur is of interest. In this situation, a special type of analysis called survival analysis is required, *for example, a group of patients that received assessment, how long did each subject take to suffer a subsequent fall? If no subsequent fall occurred in some subjects, then how long were each of these subjects followed up for?* Survival analysis centers on describing the pattern of survival which can be achieved using **Kaplan-Meier** methods and survival curves. The **log rank test** can be used to compare the survival times of two or more groups.

Sample Size

One of the most commonly asked questions when planning a statistical study is how many observations should be made? Other things being equal, the greater the sample size the more precise your estimates will be. Depending on what type of study is being undertaken, an estimate of sample size can be reached by setting a few assumptions and then referring to special sample size formulae or tables.

The first question to ask yourself is; is your study a descriptive study or a comparative one?

Descriptive study

1. What is being estimated? e.g. a mean or proportion? What does the researcher expect the estimate to be (approximately)?
2. How wide would the confidence interval (CI) be for this estimate? (e.g. if $n=10, 25, 100, 400$)
3. What width of CI can the investigator tolerate?

Sample Size Formula for a Descriptive Study

Comparative studies

What (2) groups are being compared? What sort of outcome?

a) Binary Outcome

1. Difference in percentages. The difference between the two groups that would be deemed to be of clinical importance needs to be specified, for example: *The proportion of patients who suffered further falls in the intervention group compared to the proportion of patients who suffered further falls in the control group. How many subjects need to be recruited to detect a 30% decrease in falls in the intervention group compared to the controls?*

2. Assign power, significance level. The power of a study is defined as the probability of correctly detecting the difference between two treatments as significant; this is usually set at a high value such as 80% or 90%. The significance level is referred to as the Type 1 error, i.e. the probability of incorrectly rejecting the null hypothesis. The Type 1 error is fixed as part of the study, usually at 5%, for example: *The power was set at 90% and the significance level at $p < 0.05$. Then, to detect a 30% reduction in the rate of falls between the two groups a sample size of 352 would be required.*

b) Numerical Outcome

1. Specify a difference in means that would be deemed to be clinically important.
2. Estimate of standard deviation in each group, perhaps based on previous studies or on a pilot study.
3. Assign power, significance level (see point two under binary study)
4. 'n' required.

Sample Size Formula for a Comparative Study

Whenever a sample size is calculated, it has to be remembered that the value will probably be an underestimate due to non-response, patient withdrawals, etc. The calculated sample size should thus be appropriately inflated to allow for this (inflate by anticipated dropout rate).

Recommended Resources

If you would like to do better research, a highly recommended text is:

Petrie and Sabin (2000) *Medical Statistics at a Glance*, Blackwell Science. Topics cover basic statistics in detail and give an indication of the underlying concepts of, and methods used for, more extensive analyses. Each topic is displayed on a two- or three-page spread and includes examples in medicine or dentistry. There is an accompanying website (www.medstatsaag.com).

An excellent paper from one of the above authors that addresses statistical concepts and methods in an entirely orthopaedic setting is:

Petrie A, Statistics in Orthopaedic papers, *The Journal of Bone and Joint Surgery (Br)*, 2006; Vol 88B; No.9;1121-1136.

Other reading material is also suggested at the bottom of these [online medical statistics](#) pages adapted by Ed Juszcak, based on an idea by Mike Bradburn and Sharon Love on the Oxford Radcliffe NHS Trust website

References

- Bland M, 1995, *An introduction to Medical Statistics*, second edition, Oxford University Press.
- Campbell MJ & Machin D, 1993, *Medical Statistics: A commonsense approach*, second edition. Wiley.
- Close J, Ellis M, Hooper R, et al. Prevention of falls in the elderly trial (PROFET): A randomised controlled trial. *Lancet* 1 999;353 :93-7.
- Kirkwood BR & Sterne JAC, 2003, *Medical Statistics*, second edition, Blackwell Science
- Petrie and Sabin, 2000, *Medical Statistics at a Glance*, Blackwell Science.