

# Statistics in clinical trials: Key concepts

## Introduction

Statistical methods provide formal accounting for sources of variability in patients' responses to treatment. The use of statistics allows the clinical researcher to form reasonable and accurate inferences from collected information, and sound decisions in the presence of uncertainty. Statistics are key to preventing errors and biases in medical research. This article covers some key concepts of statistics and their applications to clinical trials.

## Hypothesis testing

A hypothesis is an assumption, or set of assumptions, that either a) asserts something on a provisional basis with a view to guiding scientific investigation; or b) confirms something as highly probable in light of established facts.

For our purposes here, we are interested in the hypothesis that asserts something – for example that a new treatment for a disease is better than the existing standard of care treatment. If the new treatment is called 'B', and the standard of care treatment is called 'A' then the hypothesis states that 'B' is better than 'A'.

You might presume that scientists would set about proving this hypothesis, but that is not the case. Instead this objective is approached indirectly. Rather than trying to prove the B hypothesis, scientific method assumes that in fact A is true – that there is no difference between the standard of care and the new treatment. This is known as the 'Null' hypothesis. The scientists then try to disprove A. This is also known as

proving the null hypothesis false. If they can do this – prove that hypothesis A is false, and that that the standard of care isn't better than the new treatment – it follows that B is true, and that the new treatment is better than the standard treatment.

## **Why is this done?**

There is no simple answer for that, this is the widely accepted method that has evolved in modern science, but it may help to use a legal analogy. The null hypothesis covers our current situation or knowledge (so, in a courtroom analogy, that 'the accused is innocent'), which we need to trust unless we have sufficient evidence otherwise. But if we sought to prove the 'alternative hypothesis' (as it is known, opposite the 'null hypothesis') then in effect 'the accused is guilty'.

Another, perhaps easier way of getting at this point is to quote Albert Einstein:

*"No amount of experimentation can ever prove me right; a single experiment can prove me wrong."*

This seems to suggest that trying to prove the null hypothesis false or wrong is a more rigorous, and achievable objective than trying to prove the alternative hypothesis is right. Please note that this does NOT properly explain why science adopts this approach, but perhaps it can help us here to comprehend and accept a tricky concept more easily.

## **Type I and Type II errors**

If you look at the table below you can see what the difference between Type I errors (false positives) and Type II errors (false negatives) is.

	<b>Null hypothesis is true</b>	<b>Null hypothesis is false</b>
Reject the null hypothesis	Type I error 'False Positive'	Correct outcome 'True Positive'
Fail to reject the null hypothesis	Correct outcome 'True negative'	Type II error 'False negative'

This is still very confusing, so in order to express this in simpler terms here is a very stark example:

- Type I errors could kill a patient – imagine a study that incorrectly found that the standard of care wasn't better than the new treatment, and consequently gave new treatments to people with catastrophic results. Committing Type I errors will incorrectly detect an effect that isn't present.
- Type II errors mean that potentially valuable research goes to waste. Perhaps this research could have been really useful, but as no further study takes place, no harm is done to patients. Committing Type II errors will fail to detect an effect that is present.

It is clear, then, that Type I errors are more serious than Type II errors when it comes to patients.

## **Significance level**

Significance level is the probability of committing a Type I error. This is affected by the size of the sample, and by the 'statistical power' of the test.

## **Statistical power**

The 'power' of a statistical test is the probability that it will correctly lead to the rejection of a null hypothesis – or in other words, the ability of the test to detect an effect, if that effect actually exists. Another way of describing this

is to say that the 'power' of a test is the probability of NOT making a Type II error.

## P-values

P-values, or 'probability' values, weigh the strength of the evidence on a scale between 0 and 1. A small p-value (typically less than 0.05, or 5%) indicates that there is strong evidence *against* the null hypothesis, which might lead you to reject the null hypothesis, while a large p-value (greater than 0.05) indicates the opposite.

## Correlation versus causation

When analysing the results from a trial, it is important to remember that correlation is not the same thing as causation. Correlation is when two variables are linked in some way; however this does not mean that one will cause the other (there is an association between both variables). An example of this involves hormone replacement therapy (HRT) and coronary heart disease (CHD), where women taking HRT were at less risk from CHD. This, however, was not due to the actual HRT process, but rather due to the fact that the group of people receiving HRT tended to belong to a higher socio-economic group, with better-than-average diets and exercise regimes.

Causation can be observed when a factor causes an outcome. A causal factor is often a partial cause of an outcome. To differentiate between correlation and causation it is important to record as much information as possible about the participants in trials. It is also necessary to apply carefully the scientific methodology in clinical trials design and to assess the possible bias in the trial.

# Data manipulation

Data manipulation is the practice of selectively reporting data incorrectly or creating false results. An example of this would be when data that disagree with the expected result are intentionally discarded to increase the proportion of results that would confirm the stated hypothesis. When a researcher removes the outliers (a result that is very much bigger or smaller than the next nearest result) from the results, it is important to verify that those are truly outliers and not just results that differ from the expected or wanted results. Another example of data manipulation would be when a data collector randomly generates a whole set of data out of a single patient measurement collected.

# Data transformation

Data transformation is the application of a mathematical formula to some data gained through a trial. This is often used to make the presentation of data clearer or easier to understand. For example if measuring fuel efficiency for cars, it is natural to measure efficiency in the form of 'kilometres per litre'. However, if you were assessing how much additional fuel would be required to increase the distance travelled, it would be expressed as 'litres per kilometre'. Applying an incorrect formula to obtain the new presentation of this data in this case, would affect the overall results of the trial.

# Data merging

Data merging is the act of combining data from multiple studies in order to gain a better understanding of the situation. One of the most common forms of this is meta-analysis where the results from several published trials are put together to be aggregated and compared. It is important whilst performing a meta-analysis to carefully check that the

trial methodologies are the same or comparable. Any differences on design need to be taken into account, so that there are no underlying different variables (confounding variables). An example of incorrect data merging might be aggregating data from several trials with different species of mice as an animal test.

## Attachments

- Presentation: Statistics in Clinical Trials: Key Concepts

Size: 381,778 bytes, Format: .pptx

A presentation describing the key concepts of statistics in clinical trials, which can be adapted for own use.

A2-4.33.3-V1.1