

Statistical analysis

Descriptive statistics and *P* values

Clinical trials often involve comparing an intervention against a control according to their effect on some outcome. Descriptive statistics are used to summarise the observed treatment effect as a point estimate. For continuous variables (for example, blood pressure) the difference between the mean values in the intervention and control arms could be used as the point estimate for the treatment effect. For binary variables (for example, response versus no response), the absolute difference, or the relative difference, between the proportion experiencing a response in the intervention and control arms could be used.

Although a point estimate may numerically favour one or other treatment arms, it does not guarantee that a true treatment effect actually exists. There are three other possible explanations for such a result: bias, error and chance. Observed differences may be interpreted as evidence of a true treatment effect, provided each of these alternative explanations can be ruled out.

Bias: Study design features such as **randomisation** and **treatment concealment** (blinding) are used to control the potential of bias to explain observed effects. The application of sound analysis approaches is another important method to control bias. An important example is the **intention-to-treat** analysis principle whereby all volunteers who are randomised to a trial treatment contribute to the analysis (that is, no data are discarded), and they are counted in the treatment arm they were randomised to, irrespective of the treatment they actually received.

Error: The implementation of high-quality data management and clinical practices is used to rule out data errors as a plausible explanation for observed effects.

Chance: Statistical inference principles are used to quantify whether the observed effect could be attributed to chance alone. The probability of chance alone generating a result at least as extreme as that observed when, in reality, no treatment effect exists is quantified by a statistic known as the ***P* value**. *P* values are calculated by statisticians using probability

theory, and the types of calculations depend on the types of outcomes (continuous, binary, etc).

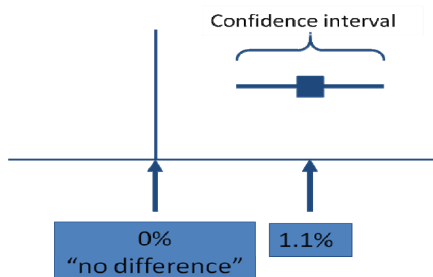
We conclude that an observed result is due to a real treatment effect if the *P* value is small (for example, less than 0.05), assuming we can also rule out bias and error as plausible explanations. If we cannot rule out chance as a plausible explanation, then the study is inconclusive.

Confidence intervals

Point estimates and *P* values do not provide all the information required to fully interpret the results of a clinical trial. These statistics do not indicate how large (or small) the real difference could be, nor what range of treatment differences are consistent with the data from the trial. These questions can be addressed by calculating a confidence interval. Confidence intervals come in various strengths; the most common is a 95% confidence interval. An intuitive, but not quite technically correct, interpretation of a confidence interval is that it contains the range of plausible values for the true treatment effect where the 95% reflects the level of plausibility. A more technically correct interpretation is that if the current trial design were repeated many times under identical conditions, 95% of these trials would yield confidence intervals that include the true treatment effect.

An example of a confidence interval is presented below. Note that the interval shown in the figure does not contain 0% (no difference) as a plausible value for the true treatment effect. This implies that chance alone is not a likely explanation for the observed result. We would consequently expect the associated *P* value to be small and the result to be considered statistically significant. A statistically significant treatment effect does not necessarily imply that the results are clinically meaningful — the point estimate and confidence interval estimated for the treatment effect may lie in a region that is considered clinically insignificant. Confidence intervals can therefore be much more useful than *P* values to help evaluate the clinical significance of results.

Statistical analysis



95% Confidence interval for difference in probability of experiencing a response

Planning an analysis

There are three basic components to a complete statistical analysis comparing intervention and control treatments in a clinical trial. The first is the calculation of a point estimate for the treatment effect. This is the best single estimate available of the true treatment effect. The second component is the *P* value, which quantifies the probability that chance alone could produce a treatment effect at least as extreme as that observed under the null hypothesis that in reality there is no treatment effect. The third component is a confidence interval, to identify the range of plausible values for the true treatment effect.

Finalising a **statistical analysis plan** before starting an analysis and the generation of preliminary results is highly recommended. This approach ensures that the analysis techniques are chosen on the basis of their methodological appropriateness rather than whether they generate results that support preconceptions of what the conclusion of the trial should be.

Contact the Outreach team (trials@ctc.usyd.edu.au) or through the website (www.clinicaltrialsoutreach.org.au) for further advice on this topic.