# The foundations of evidence-based medicine

Ian Mason, Phd, Occupational health and medical author

In this short article we will highlight the following:

- Placebo vs active treatment
- Double-blind vs single-blind trials
- Randomization vs non-randomized
- Multi-centre vs single-centre trials
- Large vs small study
- Power and sample size calculation
- Prospective vs retrospective study
- Primary vs secondary endpoint
- Meta-analysis
- Statistical significance and the p-value

**Ian Mason looks at the history of clinical trials and some key features used in current studies.**

Can you recall your first hospital admission? Mine was in the late 1950s, to have my tonsils removed. I was six years old and terrified. However it was some comfort to be amongst a sizeable gang of children on the same ward. You see tonsillectomy was deeply fashionable in the 50s – a surgical panacea for upper respiratory tract problems. One in four children in the UK had their tonsils removed – some quarter of a million procedures per year[1]. In the decades that followed accumulating evidence showed removing tonsils was of questionable clinical value[2]. Consequently the number of children undergoing tonsillectomy in the UK today is around one tenth of that in the 1950's[3]. Cases are now carefully selected with the technique used guided by well conducted clinical trials[4].

The story of mass tonsillectomy is a good example of how once popular medical interventions were found to be archaic, useless or downright dangerous; 'Interventions' such as blood-letting, giving pregnant women enemas during labour, or prescribing bed rest to speed recovery after a heart attack – classic examples of antiquated medical practice that was based on clinical anecdote, misguided expert opinion or woefully out of date teaching in medical schools.

## Evidence era

Today is undeniably the era of evidence-based medicine. Well-designed clinical trials are used to give objective information about the safety and efficacy of healthcare interventions – be they drugs, prosthetic implants or other medical innovations. Carefully conducted clinical trials not only offer the fastest and safest way to determine if an intervention works, but they are absolutely necessary

Mölnlycke®

for regulatory approval and for acceptance by clinicians.

Evidence-based medicine took off in the early 1990s but its roots lay much earlier. Arguably one of the earliest controlled clinical trials took place in 1747 when Dr James Lind decided to find out if oranges and lemons could prevent the debilitating (and potentially fatal) illness called scurvy amongst sailors on long sea voyages. The basic principles of his experimental model stand today. Under controlled conditions, he compared several nutritional interventions in sick patients well matched for the seriousness of their scurvy. 'The most visible and good effects were perceived from the use of oranges and lemons', he wrote[5]. It took a half-century for his findings to be put into practice and for the British Navy to make lime juice compulsory for naval sailors (which is why the British are widely known as 'Limeys'). This simple innovation immeasurably strengthened the British Navy.

Further building blocks of the modern clinical trial were added slowly. In the mid-nineteenth century the use of a **placebo** or 'dummy remedy' against which an **active treatment** could be compared, was added. Two further important developments took place in the 1940s. The first was an early double blind controlled trial (to investigate the use of patulin against the common cold)[6].

The important difference between a **double- and single-blind trial** is that in a single blind study, subjects do not know which treatment they are taking until the end of the study. In a double blind study, neither investigators nor subjects know who received which treatment until the end of the study. Incidentally the patulin study showed no benefit for the active treatment – it was nevertheless a useful finding. The second development took place in 1946 and was the first use of randomisation in a major clinical trial (to evaluate streptomycin in tuberculosis patients[7]).

In order to minimize the risk of bias in a comparative study, patients who agree to take part can be randomly allocated to different treatment groups. In the case of a study comparing two interventions, the allocation process is equivalent to the flip of a coin to decide which intervention a patient is given. Randomisation also balances any other factors that could influence the treatment effect such as age, gender or weight. A **randomised study** is always considered to be more robust than a **non-randomised study**.

## Other features of modern clinical trials

Modern randomised controlled clinical trials may employ several other important features. Many studies today are multi-centre (carried out at more than one site) whereas a **single-centre trial** takes place at one site. Although a **multi-centre study** may sound more 'credible' this is not necessarily the case. In a single centre study, it is easier to consistently control variables that may influence outcome. Moreover, multi-centre studies are sometimes performed in order to recruit a more diverse patient population, or because it is the only viable means of recruiting a sufficient number of subjects to reliably evaluate a small treatment effect. A **large study** is not necessarily more credible than a **smaller one**; a good statistician will have specified in advance the subject numbers needed to provide a reliable result in either case by so called power and sample size calculation.

A **prospective study** is designed in advance to answer one or more research questions, while a **retrospective study** reviews existing information, such as patient registry data or electronic health records. Therefore a prospective study follows patients into the future, a retrospective looks at data from their past. Of course a study could be both – if a researcher reviews data from a retrospective study group and then follows that patient cohort into the future.

I have already mentioned the use of a placebo as a control comparator to the intervention being investigated. Other comparators include the best existing (or 'gold standard' ) treatment. In the case of medical device studies, placebos are often called sham devices. An appropriate control adds credibility to a study.

It is not always possible to design a blinded study, for example if medical devices or wound dressings are being compared and the subject and experimenter can clearly see these. In such cases a cross over study may be used. In this type of study the subject receives both treatments in a random order.

## End points

Some trials are described having **primary and secondary endpoints**[8]. The primary endpoint of a clinical trial is the endpoint for which the trial is statistically powered. Secondary endpoints are additional endpoints, preferably also pre-specified, for which the trial may not be powered, but which may provide useful information or clues about avenues for further research.

Another type of study that has become more widely used over recent years is **meta-analysis**[9]. This involves the use of statistical methods to combine results from published studies that address the same question. The results are collated into a systematic review that combines the results to improve the precision of estimates for treatment effect, and assess whether treatment effects are similar in similar situations. Cochrane Reviews are examples of systematic reviews that are generally recognised as the highest standard in evidence based health care. Of course solid statistics and

robust trial results are only part of the story. Increasingly the importance has been recognised of the need to combine critical appraisal of the evidence with the patient's values and preferences through shared decision making[10].

## Statistical significance and Clinical relevance

The interpretation of 'statistical significance' is often misinterpreted as 'clinically important'. **Statistical significance** quantifies the probability of a study's results being due to chance. Clinical significance or relevance, on the other hand, refers to the magnitude of the actual treatment effect. For a successful study you would like to see a difference between the intervention and control group that is clinically relevant, something that is likely to impact medical practice, and statistical significant so that the result is not only due to chance. To measure the statistical significance a p-value is used. The **p-value** will give an indication whether the result is due to chance or to a real treatment effect. Commonly it is used to say that if the p-value is less than 0.05 then statistical significance has been achieved (see Table 1) and that the result is due to chance is less than 5%[11].

Finally, some trials are undertaken for commercial purposes. If a company intends to make claims that compare one of its products to another product, then it is essential to base those claims on the results of a study that directly compares those products. It is generally considered unacceptable to make such claims unless such evidence from a comparative study is available.

This short article can only skim the surface of clinical trial design – whole textbooks have been written on the subject. If your interest has been piqued, a fascinating website that celebrates the work of Dr James Lind (author of The Scurvy Study) has been set up by the Royal College of Physicians of Edinburgh. It can be found at www.jameslindlibrary.org. In a series of excellent and freely available articles the site illustrates the evolution of methods to assess the effects of treatments – from setting out the principles of testing treatments a thousand years ago, to systematic analyses of the experiences of tens of thousands of patients today.


**References:**

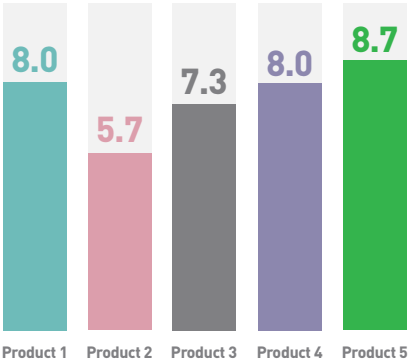
1. Illingworth RS Proceedings of the Royal Society of Medicine Vol 54 May 1961:393-399
2. Dwyer-Hemmings L. 'A Wicked Operation'? Tonsillectomy in Twentieth-Century Britain. Med Hist. 2018 Apr; 62(2): 217–241.
3. https://onlinelibrary.wiley.com/doi/pdf/10.1111/coa.13707
4. Akural EI et al Post-tonsillectomy pain: a prospective, randomised and double-blinded study to compare an ultrasonically activated scalpel technique with the blunt dissection technique. Anaesthesia 2001 Nov;56(11):1045-50.
5. Lind J. 'A Treatise of the Scurvy.' Edinburgh 1753
6. Bhatt A. Evolution of Clinical Research. Perspect Clin Res. 2010 Jan-Mar; 1(1): 6–10.
7. MRC Streptomycin in TB Trials Committee. BMJ 1948;2:769-83
8. Vetter TR. Defining the Primary Outcomes and Justifying Secondary Outcomes of a Study: Usually, the Fewer, the Better. Anesth Analg. 2017 Aug;125(2):678-681.
9. Haidich AB. Meta-analysis in medical research. Hippokratia. 2010 Dec; 14(Suppl 1): 29–37.
10. Djulbegovic B. Progress in evidence-based medicine: a quarter century on. Lancet. 2017 Jul 22;390(10092):415-423.
11. Common pitfalls in statistical analysis: Clinical versus statistical significance. Priya Ranganathan, C. S. Pramesh, and Marc Buyse. Perspect Clin Res. 2015 Jul-Sep; 6(3): 169–170.


Table 1. P-value and the related significance levels

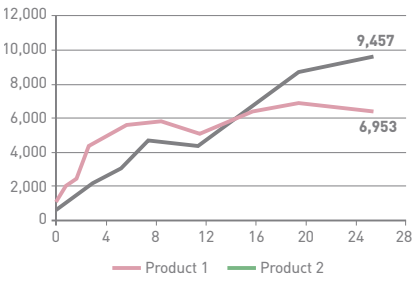
| Significance level | Specification |
|---|---|
| p > 0.05 | not significant |
| p ≤ 0.05 (5%) | significant |
| p ≤ 0.01 (1%) | very significant |
| p ≤ 0.001 (0.1%) | highly significant |

There are many ways to display the results from a comparative study. A graph should contain enough information so it does not need additional text to be understood, and you should select the graph that is best suited for the comparison you have made.

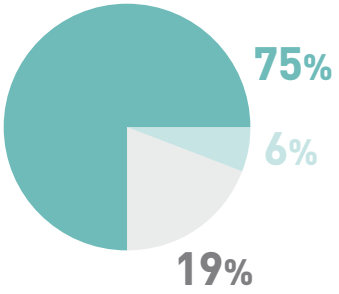
If relevant, also add the number of subjects included as the "n=", as well as the p-value for each comparison you are making.




Bar chart




Line chart




Pie chart

Find out more at www.molnlycke.se



Mölnlycke®