Statistics for clinical trials and audit

Ian Kestin

Abstract

This article covers the application of statistics to clinical trials and audit, including the basic types of study design, bias, power analysis, guides to good clinical practice, the presentation of results and applications in quality assurance.

Keywords Bias; crossover; randomization

Royal College of Anaesthetists CPD matrix: 3J01, 3J03

It is only relatively recently that the importance of scientific evidence has been recognized in clinical medicine, and most of this evidence has been obtained from clinical trials. We need to be sceptical when reading the journals because there are commercial, institutional, methodological and personal influences on the 'impartial' use of science in medicine; one-third of all major original clinical research turns out to be wrong.[1]

Study designs

Bias is any factor that may alter the results and lead to false conclusions; over 30 different types of bias have been described. Some common types of bias in clinical trials are listed in Box 1.

Bias in medical research may occur at all stages and may be entirely unrelated to the conduct of the researchers. Statistical techniques have been used to demonstrate the effects of bias in clinical trials (Box 2).

Case studies describe the outcomes of an intervention in one or more individual patients. They lack any control patients for comparison or methods to avoid bias.

Retrospective studies are observations of patients who have completed their treatment, and the data obtained after the events, for example, from written records. A common type of retrospective study is the case–control study in which patients who have the disease or condition of interest are compared with control patients who do not. These control patients are selected to match the patients as closely as possible in all respects except the disease in question. This selection inevitably has a risk of introducing hidden bias, the effect of which cannot be assessed. Missing data is another common problem of retrospective studies.

Prospective studies are those in which the patients are selected in advance and then studied in a structured format according to the study protocol.

Ian Kestin *FRCA is Consultant Anaesthetist at the Western Infirmary, Glasgow, UK. Conflicts of interest: none declared.*

Common types of bias in clinical trials

- **Selection bias** occurs when the patients are selected in a manner that introduces systematic differences between the groups. This can occur in many ways, e.g. poor methods of randomization. It may be accidental or a deliberate manipulation of the study by the investigators
- **Measurement bias** can arise if the measurements made on the patients have systematic errors that affect some groups more than others. This can occur if equipment is not calibrated uniformly, and is especially likely if different observers are making subjective assessments of the patients. Observers inevitably make different assessments of the same observation
- **Publication bias** is a significant cause for bias in medical knowledge when not all data are submitted for publication*
- **Commercial bias** occurs because studies sponsored by pharmaceutical companies generally show more favourable results for drug therapy than independent studies,[2] and meta-analyses also seem biased by the commercial interests of the authors[3]
- **Attrition bias** may lead to errors in interpretation if patients who have entered the study from analysis are excluded, as the rate or causes of drop-out from the study may not be equal for all groups. For example, when comparing a medical and a surgical treatment, if those who died as a direct or indirect result of surgery are excluded from analysis, a bias towards the surgical treatment is introduced. In general, all patients should be analysed in the original groups to which they were allocated (called **intention to treat**)

*See example in *Anaesthesia and Intensive Care Medicine* 2015; ***16:*** 200–07.

Box 1

A randomized controlled trial is an experiment in which the eligible patients are randomly allocated to receive one of the treatments. Usually one or more groups receive the drug of interest and one group, the control group, is used for comparison. Depending on the purpose of the study, the control group may receive an inert substance, a placebo or a standard treatment for the disease studied. In some studies the patients receive all the

Effects of bias on medical research

There have been four meta-analyses comparing studies with adequate techniques to ensure that the investigators were unaware of the treatment allocation (i.e. adequate 'blinding') with studies of the same topic in which the investigators probably could have discovered the treatment of the patients. All reported that there was an obvious exaggeration of the benefits of treatment in studies in which masking was inadequate. The effects were similar but less marked if the trials had inadequate randomization of the subjects or if the patients were aware of the treatment allocation[4]

Box 2

treatments in sequence, and thus serve as their own controls. These are called **crossover studies**, and confounding factors will be equal across all treatments. Not all trials can be done using a crossover design, and the limitations are given in Box 3.

The purpose of **randomization** is to distribute the confounding factors that may affect the response equally across all treatments. Some of these factors may be known or obvious, for example age, gender and smoking, but, more importantly, there will nearly always be unknown factors (e.g. genetic) that may affect outcome. Recruiting an adequate number of patients and randomly allocating them to the different treatments is the only method of minimizing the effect of confounding variables. The method of randomization is important and a recognized method, such as random number tables, should be used by someone unconnected with the conduct of the study. Allocation by days of the week, hospital number or birthday is not random, and may introduce bias into the characteristics of the groups. It is often easy for the investigators to 'adjust' the randomization if they are doing it themselves.

There are several types of randomization. **Simple randomization** allocates the patients to a one of the treatment groups entirely by chance. If the patients are initially subdivided according to baseline characteristics, for example age or gender, and then these subgroups are allocated randomly to one of the treatments this is called **stratified randomization**. Stratified randomization will reduce the risk that the groups are unbalanced at the end of the study, and is used if there are important baseline characteristics known to affect the outcome of treatment. The disadvantage is that it may be difficult to recruit sufficient patients to all the categories, so delaying the study. **Minimization** is a technique that is particularly useful if the study has a small number of patients. The first patient is allocated randomly, and then the second and subsequent patients are allocated using a weighted randomization. The weighting is adjusted in each patient to increase the chance that the patient is allocated to a treatment group that would minimize the differences in baseline characteristics already present between the groups. The principle of randomization is maintained while minimizing the chance of unequal groups at the end of the study.

Randomization cannot ensure that the confounding variables are equally distributed across the groups as it is still possible by chance for the groups to end up unequal, for example all the males being allocated to one group. It is common to use statistical tests to check whether the groups are similar in ages, weights, etc. after the study has been completed. These tests will detect only major differences, and the unknown confounding variables remain exactly that — unknown. If the two groups are found to differ in important characteristics after the study has been completed, all is not lost. The results can still be analysed using statistical techniques that compensate for differences in **baseline characteristics**, such as analysis of covariance or multiple regression analysis.

Blinding or masking means that the investigator and/or patient are unaware of their treatment group; if both are unaware of the treatment, this is called a **double-blind trial**. Masking is important, as the response to treatment is often considerably altered by expectation, either by the patient or by the investigator. If the patients knew they were receiving the placebo, they would not expect to improve, whereas in practice there can often be a considerable response to placebo. A randomized double-blind controlled trial is the gold standard for obtaining medical evidence. Sometimes it is not possible, for example in studies comparing a surgical with a medical treatment, and these are known as open studies; however they should still be randomized.

There are a number of guides to good practice in the conduct of research (Box 4). A poorly conducted study is unethical (see *Anaesthesia and Intensive Care Medicine* 2012; **13:** 7–10).

New drugs undergo a series of clinical studies in order to be given a product licence (Box 5).

Limitations of crossover trials

- **Period effects:** there should not be any significant change with time in the condition of the disease during the study period. If the disease significantly worsened or improved between the first and second treatments, the two treatments would not be studied under similar conditions. For example, transient diseases such as the common cold cannot be studied using a crossover design. The order of the two treatments under investigation is usually varied between the patients to avoid bias from period effects
- **Treatment–period interactions:** one of the treatments may work differently if given in one of the study periods. The treatment may work more effectively earlier or later in the disease process, or its effects may be modified by the other treatment
- **Carryover effects:** there must be adequate time for the effects of the first treatment to disappear before starting the second treatment, otherwise the true effects of the second treatment are not being measured

The data can be tested statistically for the presence of each of these effects after the study has been completed, but these tests will detect only major effects. It is better to ensure that a crossover trial is the appropriate design and is well conducted

Box 3

Power analysis

An essential part of the design of any clinical trial is a power analysis, which is a statistical technique to estimate the number of patients required to reduce the risk of a type II error (see *Anaesthesia and Intensive Care Medicine* 2015; **16**: 200–207) to an acceptable value. The actual calculations in a power analysis depend on the type of data, that is, categorical, ordinal or continuously variable data. The probability of a type II error is denoted as β, and should be 20% or less, and the power of the study is defined as (1−β). For example, if β has been chosen to be 10%, the power of the study is 90%; with this number of patients, the study has a 90% probability of demonstrating a treatment difference if such a difference exists.

The main determinants of the power of a study are:

- the magnitude of the difference between the treatment and control groups and the variability of the data; this information is often unknown and is usually the purpose for doing the study, but estimates can be obtained from pilot studies, previously published work, or chosen by the