Analysing clinical trial results

Introduction

When pharmaceutical companies conduct clinical trials, medical details of the patients taking part (but not their identities) are collected in a computer database together with the results of any measurements made. Statistical analyses are then conducted to formally assess the outcomes of the trial.

Analyses of clinical trial results cover three areas of interest:

- Demographic and baseline information
- Efficacy
- Safety

These areas are described further below. The type and design of the clinical trial plays an important factor in the interpretation of the statistical analyses.

Demographic and baseline information

Who took part in the trial? The effects of a medicine may differ considerably between different groups of patients. It is therefore important to know details of the all trial patients, such as:

- Age
- Sex
- Ethnic origin
- Severity of their illness

In general, the closer the match between a trial group and a population of interest, the more relevant the findings will be.

Efficacy

How well did the trial medicine work? This part of the analysis is based on pre-defined 'endpoints'. These are specific measurements related to the illness in question. Endpoints are specified in advance in the trial protocol (the document which describes in detail how the trial will be performed).

Endpoints in general can be categorised as:

- 'Hard' endpoints — those that take the form of numerical facts with intrinsic clinical importance. For example, how long the patient survived or what proportion of patients recovered from an infection.
- 'Soft' endpoints — those which are potentially influenced by the measurement process or with questionable reproducibility. For example, a quality-of-life questionnaire or the description of the patient's mood at a given moment. In order to be analysed statistically, soft endpoints have to be converted into a numerical format. This process can be controversial as it often relies on subjective data and is potentially open to inconsistencies.
- 'Surrogate' endpoints — those that are not in themselves part of the patient's experience of the illness, but may be closely related to it. For example, the results of laboratory tests.

In general, hard endpoints are preferable to soft and surrogate endpoints. Soft and surrogate endpoints need to be assessed carefully in the light of how well they represent the illness being studied.

Choosing which endpoints to use depends heavily on the nature of the illness being studied. Cancer, for instance, offers obvious hard endpoints in the form of survival, whereas an evaluation of depression must inevitably involve softer endpoints. Other illnesses, such as diabetes, are associated with well-established surrogate endpoints such as blood sugar levels.

Safety

What side effects did the medicine have? Whenever the doctor conducting a clinical trial sees a patient, he or she is asked if the patient has experienced anything untoward. Information on these 'adverse events' is collected and later analysed to give insight into possible a causal relation with the medicine studied. If such a causal relation is established, the adverse event becomes an 'adverse reaction' or side effect. Particular attention is paid to 'serious' adverse reactions — those which are life-threatening or associated with death, hospitalisation, or birth abnormalities.

Type of clinical trial

Clinical trials vary considerably in size, duration, and design. These factors play a major part in the interpretation of trial results.

The most informative clinical trial design is the 'double-blind randomised comparison', in which some patients receive the new medicine while others receive an alternative treatment. The alternative treatment, sometimes called the 'control', may be either:

- A **placebo** — an inactive 'dummy' treatment
- An **active comparator** — generally a well-established treatment for the illness being studied.

Participants are allocated to each study group by chance. The

trial is set up so that while the study is going on, neither the doctor nor the patient knows who is receiving which treatment. A trial set up like this is said to be 'double-blinded'. Double-blinding reduces the potential for bias in the results.

In such trials, the results are presented in terms of the difference between the group receiving the new medicine and the group receiving the control treatment:

- Where the comparison is against a placebo, this difference is a measure of the real effect of the new medicine.
- Where the comparison is with an active comparator, the difference gives insight into how the new medicine compares with current medical practice.

In both cases, two aspects of the difference are likely to be reported:

- **Size***:* This is often reported as the actual difference recorded in a particular trial together with a '95% confidence interval'. This is the range within which we can be 95% sure that the true difference would lie for the population. Although you may detect a statistical significance, it may not be clinically relevant. Generally speaking, the larger this difference, the more likely it is to be clinically relevant (to increase survival by a year is of more clinical relevance than to increase it by a day).
- **Statistical significance**: Because some individuals respond better than others to treatment, there is always a risk that the difference between groups seen in a clinical trial may have arisen by chance. For example, if all the inherently good responders were randomised to one group, and the bad responders to the other. Statisticians can calculate how likely it is for this scenario to have occurred in a particular clinical trial

and they express their result as a 'p-value'.

A p-value of 0.05 means that there is a 5% or 1 in 20 chance that the difference happened by chance. It is conventionally taken as the threshold for accepting results as 'statistically significant'. It is important to realise that the word 'significant' used in this sense says nothing about the medical importance of the results — it merely offers reassurance that the result is unlikely to be accidental. For example, a one-meter increase in a six-minute walk distance might, in a large enough trial, be shown to be statistically significant (i.e. unlikely to have arisen by chance) but it would never be regarded by a heart-failure patient or his doctor as being of any clinical value.

A second important group of clinical trials, often conducted to investigate long-term safety, takes the form of 'open-label' trials. In these there is no control group — everyone is treated with the new medicine, and their experience is recorded. No differences between groups can arise (either accidentally or through genuine therapeutic effects), and hence there is no place for significance testing. Balanced against these shortcomings, open-label trials often include large numbers of patients (up to several thousand) studied for long periods of time (several years in some cases). These trials therefore make it easier to detect rare side effects and those that take a long time to develop.

The results of such trials are presented as straightforward tables listing different adverse events and how frequently they were seen.