Research Review
EDUCATIONAL SERIES

Clinical Trials – An Overview

About the Reviewers

Associate Professor Chris Frampton

Chris is an Associate Professor in the Department of Biostatistics at the University of Otago, Christchurch. He has 25 years of experience in statistical analysis and design. He has a particular interest in the statistics associated with drug development studies, with an emphasis on phase III RCTs.

Professor Shaun Holt

Shaun holds Pharmacy and Medicine degrees, has been the Principal Investigator in over 50 clinical trials and has over 80 publications in the medical literature. He is Ex-Medical Director of Clinical Trials in the Wellington Asthma Research Group and is an Advisor to the Asthma and Respiratory Foundation of New Zealand.

About Research Review

Research Review is an independent medical publishing organisation producing electronic publications in a wide variety of specialist areas.

Research Review publications are intended for New Zealand medical professionals.

Medical science owes much to the development of the clinical research method known as the randomised clinical trial (RCT), which plays a key role in modern clinical research and is considered to provide the best level of evidence in evidence-based medicine.[1] According to this classification, results of an RCT are more definitive than any other type of clinical research information.[1]

This paper is intended for people who need to understand clinical trial terminology and is directed in particular towards doctors, nurses and pharmacists reading clinical trial reports. The coverage of material is meant to assist with understanding what clinical trials involve – it covers the majority of terms that are likely to be encountered and is a broad summary of the process of clinical trials. A variety of medical, statistical, epidemiological, ethical and data management terms are included; the style of explanations and definitions is aimed at being pragmatic rather than all-encompassing. It is hoped that this paper will assist readers who understand little or nothing of terms relating to clinical trials to appreciate their essential meanings. Full and complete explanations of all terms included would entail a large and unwieldy tome, which is not the intention of this paper. References are provided for readers to follow-up the complete details of all the terms used.

Design of RCTs[2-6]

A clinical trial is defined as a prospective scientific experiment usually conducted to assess the safety and effectiveness of an intervention in groups of subjects. Interventions may be diagnostic, preventative, or therapeutic in nature and may include drugs, biologics, medical devices, or methods of screening. Interventions may also include procedures that aim to improve the quality of life or to better understand how the intervention works in the study participants.

In a **randomised clinical trial (RCT)**, most commonly each patient is assigned at random either to receive the new drug, the standard treatment for that disease, or a placebo treatment (a non-functional substitute (e.g., a sugar pill). After randomisation, the groups of subjects are followed up in exactly the same way, and the only differences between the care they receive, for example, in terms of procedures, tests, outpatient visits, follow-up calls etc. should be those intrinsic to the treatments being compared. The advantage of randomisation is that it helps to ensure that the groups in the trial have similar characteristics, making it easier to compare outcomes between groups. Randomisation also minimises **bias** or **systematic errors**. Bias can influence a clinical trial by introducing systematic errors associated with the design, conduct, analysis, and reporting of the results. The most common types of bias in clinical trials involve subject selection and outcome measurement:

- If the researcher knows which treatment a patient is given, it could affect how s/he collects information on the outcome during the trial
- The researcher might select patients in a certain way that could favour the new treatment, resulting in a selection bias
- Excluding subjects from statistical analysis because of noncompliance or missing data could bias an estimate of the true benefit of a treatment
- Clinical trial designs should seek to reduce these systematic errors.

Confounding is another potential problem that an effective design can mitigate. Confounding makes it difficult to isolate the specific effects of the intervention from those due to some additional factor therefore it may not be possible to determine whether a new intervention is truly effective or non-effective. Confounding occurs when the additional factor influences the treatment outcome and as a consequence of treatment group assignment differs between the treatment groups. For example, perhaps by chance random allocation to two groups in a trial of two treatments for hypertension meant that the mean age for group A was higher. At the end of the study this group showed a smaller reduction in blood pressure compared to group B. The comparison between the two groups is, therefore, confounded by age – we cannot be

sure whether the difference in blood pressure reduction is a consequence of the effects of the treatment or the age difference.Confounding can hide an existing treatment difference and also create an apparent difference when there is not one in reality.

Phases of clinical trials[2-6]

Clinical trials must follow certain procedures, to satisfy regulation requirements for development of a new drug in humans. The US Food and Drug Administration (FDA) first described the four `phases' of clinical trials (see Table 1); this terminology is now widely accepted throughout the pharmaceutical industry. Under this system, a new drug or intervention begins testing in phase I trials and then proceeds to phase II and III trials in a sequential manner that ends with the intervention being established as the new standard or in its licensing. After licensing, a phase IV trial may be undertaken to explore the long-term morbidity and effects that would be too uncommon to be detected in previous studies.

Table 1. Phases of clinical trials[2-6]

	Objective	Typical No. of patients
Phase I	♦ First investigation of a new drug in humans (often called `first in man' studies) ♦ To investigate the pharmacokinetics and the pharmacological effects of a drug, including dose-response and side effects	10 to 30, usually healthy volunteers
Phase II	♦ Provides preliminary efficacy and safety data	Fewer than 100
Phase III	♦ To compare new treatment to the standard therapy or a control or placebo (if no standard treatment exists) ♦ Phase IIIb studies investigate new indications for already licensed drugs	Hundreds or thousands
Phase IV	♦ Long-term surveillance of patients to identify morbidity and late effects (post-marketing study)	Many thousands

Different kinds of phase I studies:[2-6]

Phase I studies are frequently undertaken with normal healthy males and occasionally with patients, e.g., oncology drugs.

SAD

In **Single Ascending Dose studies**, small groups of subjects receive a single dose of the drug while they are observed and tested for a period of time. If tolerated, and the pharmacokinetic data is broadly in line with predicted safe values, the next group of subjects receives a higher dose. This is continued until pre-calculated pharmacokinetic safety levels are reached, or until the administered dose is associated with unacceptable toxicity. The **maximally tolerated dose (MTD)** is usually the dose below the one that produces unacceptable toxicity. The MTD is also defined as the dose that has an acceptable number of side effects and is therefore used in further studies.

MAD

Multiple Ascending Dose studies follow the SAD studies both temporally and in process, as these allow determination of MTDs with repeat dosing. MAD studies assess the pharmacokinetics and pharmacodynamics of multiple doses of the drug: patients receive multiple low doses of the drug, while samples (of blood and other fluids) are collected at various time points and analysed to understand how the drug is processed within the body. The dose is subsequently escalated for further groups, up to a predetermined level.

Food effect

An investigation into any differences in absorption of the drug by the body, caused by eating before the drug is given. These studies are often run as a **crossover study**, with volunteers being given two identical doses of the drug on different occasions; one while fasted, and one after being fed.

Outcome measures

Another measure of toxicity in phase I trials involves finding the **dose-limiting toxicity (DLT)**. In healthy volunteers a DLT occurs when a serious adverse event involving any reaction related to the trial drug requires treatment and the person has to stop taking the new drug.

Other trial endpoints that are also measured may include the monitoring of drug uptake, metabolism and excretion, body temperature, blood pressure, drug plasma concentration and other biological and physiological markers. Many variables have to be measured, to collect sufficient data to determine whether the drug is safe enough and worth investigating further.

Phase II studies[2-6]

The goals of phase II studies are:

(i) to learn more about safety and side effects

(ii) to provide data allowing selection of optimal doses for subsequent trials

(iii) know within a short period of time whether the drug is likely to be effective.

Phase II trials also serve as **pilot** (or **feasibility**) **studies**, assessing whether a phase III trial is likely to be successful.

Phase II studies can be divided into phase IIA and phase IIB.

- Phase IIA assesses dosing requirements (how much drug should be given).
- Phase IIB studies efficacy (how well the drug works at the prescribed dose(s)).

Trial design

Phase II designs may include a control arm (standard treatment or placebo) and the new treatment arm could be one or more doses.

Single-arm study

The simplest design has only one arm – all subjects receive the same intervention.

Single-arm two-stage study

The intervention is first tested on a small number of subjects – if a certain number respond without indication of toxicity, the trial continues and a second group of subjects is recruited, otherwise the trial stops: this is known as the **stopping rule**.

Randomised phase II study with control arm

This involves two trial groups – the new intervention and a control (standard treatment or placebo). The control arm is often used when it is not well known how subjects respond generally. Outcomes from both arms are used to design the corresponding arms in a phase III trial.

Randomised phase II study with several intervention arms

- Investigates two or more new treatments or multiple doses of the same treatment at the same time
- Each arm is designed as a single-arm study, and subjects are randomised to the different groups

- Treatment response rates determine which treatment should be taken further (i.e. those with response rates exceeding the expected response on standard treatments)
- Can include a control arm using standard treatment or placebo.

Randomised phase II study with several intervention arms: two-stage design

An extension of the single-arm two-stage design.

- At Stage 1, some subjects are randomised to each new treatment
- Efficacy outcomes determine which treatments can proceed to Stage 2.

Phase III studies[2-6]

The central question of a phase III study is whether the intervention works

- Often conducted at several medical centres to see if people treated at different places have similar experiences
- Central question: does the new treatment work and warrant a change in practice?
- Often have a short follow-up period for evaluation (i.e. short *vs* the time the intervention might be used in clinical practice)
- Phase III studies allow an extensive evaluation of any side effects
- Common efficacy endpoints include:
  - mortality
  - occurrence of the disease of interest
  - disease progression
  - cure or relief of chronic symptoms
  - change in lifestyle or behaviour.

Phase III studies will often continue accumulating outcome data while waiting for regulatory approval, allowing patients to continue to receive possibly lifesaving drugs until the drug can be obtained by purchase. Other reasons for continuing the trial could result in post-approval "label expansion" on evidence that the drug works for additional types of patients/diseases beyond the original use for which the drug was approved for marketing, it may obtain additional safety data, or it may support marketing claims for the drug. Studies in this phase are sometimes categorised as `phase IIIB studies'. Any reported adverse events relating to the drug may require it to have stronger side effect warnings, more limited conditions for use, or even force it to be withdrawn from the market.

Double-blind means that neither the doctor nor the trial participant knows whether the participant is receiving the experimental treatment.

In **single-blind trials**, doctors know what treatments their patients are getting. Only the trial participants do not know which group they are in. Double-blind trials are considered better because they prevent doctors from acting on preconceived notions they may have about whether or not the drug works.

Open trials

In **unblinded trials**, or **open trials**, both doctors and participants know what treatments are being given. Open trials, like single-blind trials, are considered to be more prone to error than double-blind procedures.

Equivalence or **noninferiority trials** are designed to show that two interventions have a similar efficacy, but usually one of the interventions is safer, more cost-effective, or easier to administer and therefore this intervention may have an advantage even if efficacy is comparable.

Clinical equivalence trials are based on clinical outcomes such as death, stroke, heart attack, or hospitalisation.

Bioequivalence drug trials compare pharmacokinetic (PK) measurements from two forms of the same drug (involving a new or different formulation). For example, a medicine traditionally used as a capsule might be formulated as a liquid and the bioequivalence of this formulation change would need to be tested. The actual compound or a marker is used to determine that a similar amount of drug is taken into the body (similar PK). Major advantages of using a pharmacokinetic approach are that the outcomes are clearly defined (PK measurements) and these outcomes have a lower variability than some outcomes associated with clinical equivalence trials (e.g. where outcomes such an improvement in depression are difficult to measure objectively and reproducibly). This may mean that similar efficacy can be inferred from comparable (bioequivalent) PK parameters.

Common trial designs:[2-6]

Most trials have **parallel groups**: independent groups of subjects, where each subject receives only one treatment.

In a **crossover trial**, each participant gets both treatments being tested. Some participants are assigned at random to receive drug A, and later, drug B. Others receive B, then A. Each subject serves as his/her own comparison: this is a common design for bioequivalence trials.

To produce valid results, there should be no **residual** (**carryover**) **effect** from the first treatment before the second treatment is assessed, and vice-versa. To satisfy this requirement:

- A sufficient **washout period** is required – the time between the two trial treatments when neither is given – the length of time depends on the aetiology of the disorder being examined and the pharmacological properties of the trial treatments.
- The extent of the disorder should return to baseline levels by the end of the washout period, i.e. there is no cure after the first treatment.
- In some crossover studies, there is a **sequence effect**, in which treatment order matters: people given A then B respond differently to those given B then A. The statistical analysis has to allow for this.

Split-person design: when two interventions are administered at the same time, (i) in dentistry, when comparing the effects of two types of fissure sealants on future caries risk, one sealant method could be randomly applied to the left side of the mouth and the other sealant to the right side (called a **split-mouth design**) (ii) in medicine, a new topical cream for psoriasis could be randomly applied to one arm and outcomes compared with a standard cream applied to the other arm.

Factorial trials

When patients are being treated with a combination of drugs, a new drug may be evaluated by testing it in combination with other drugs rather than by itself.

A simple factorial design would have one group testing therapy A, another testing therapy B, a third group testing A and B combined, and a control group testing neither A nor B.

Factorial designs are considered an efficient way to test medicines in combination, but their results are not always easy to interpret, particularly if the medicines interact.

Orphan drug trials

Orphan drug trials test drugs designed to treat rare diseases – defined by the US FDA as affecting fewer than 200,000 Americans; the EU defines rare disease as one that affects less than 5 in 10,000. Some are rare genetic diseases that occur when missing or defective enzymes prevent essential biochemical reactions from happening. Because affected individuals are so few, an orphan drug may be tested only on a small number of participants, who generally are so sick that if the drug works, their improved health is obvious.

Phase IV studies[2-6]

Results from phase III trials may need further validation for ongoing regulatory approval. The trials may not have tested for interactions with other drugs, tested the effects in certain populations such as pregnant women, or did not include enough people to detect rare side effects. After the drug enters the market and many thousands of people start taking it, these rare side effects and drug interactions can appear. These phase IV studies are also known as **Post-marketing Surveillance Trials**.

During the safety surveillance (pharmacovigilance) period, harmful effects may result in a drug being no longer sold, or restricted to certain uses: recent examples involve cerivastatin (Baycol, Lipobay), troglitazone (Rezulin) and rofecoxib (Vioxx).

Good Clinical Practice[7]

Clinical trial conduct is highly influenced by a well-established document called **Good Clinical Practice** (**GCP**), a set of recommendations intended to standardise clinical trial conduct. It defines roles and responsibilities for trial staff, and protects the rights, safety and well-being of trial subjects. The International Conference on Harmonisation (ICH) provides the international standard, based on the Declaration of Helsinki, although other organisations have developed their own similar guidelines. The guideline provides a unified standard for the EU, Japan and the USA, which assists the mutual acceptance of clinical data by regulatory authorities in these jurisdictions.

The 13 core principles of ICH GCP guidelines for clinical trials are:

1. Clinical trials should be conducted in accordance with the ethical principles of the Declaration of Helsinki, and consistent with Good Clinical Practice and the appropriate regulatory requirement(s).
2. A trial should only be conducted if the potential risks and inconveniences are outweighed by the expected benefit for the trial subject and society.
3. The rights, safety and well-being of trial subjects are the most important considerations and should prevail over the interests of science and society.
4. Non-clinical and clinical information about a new intervention (especially an investigational medicinal product) should be used to justify the proposed trial.
5. A clinical trial should be scientifically sound, and described in a clear and sufficiently detailed protocol.
6. A proposed trial and its protocol must have approval from an independent ethics committee. Researchers should follow the protocol when conducting the trial.
7. Trial subjects should be the responsibility of a qualified clinician (or dentist), who will make decisions about the medical care.
8. All researchers involved in conducting a trial should be qualified by education, training and experience relevant to their tasks.
9. All human subjects should give informed consent before they participate in a trial.
10. Clinical trial information should be recorded, handled and stored in a way that allows its accurate reporting, interpretation and verification.
11. Data should be kept confidential and protected, particularly when it identifies a particular subject. The regulations that govern privacy and confidentiality should be followed, where required.
12. Investigational medicinal products should be manufactured, handled and stored in accordance with Good Manufacturing Practice and used as specified in the trial protocol.
13. Systems for assuring the quality of the trial conduct and data should be in place.

The principles of GCP may be applied to **any** clinical research investigation that may impact upon the safety and well-being of human subjects.

Sample size[2-6]

Calculating sample size, allowing for adequate levels of significance and power, is an important part of trial planning.

Sample size estimation needs:

(i) an estimate of the expected outcome in the control group of the trial (those subjects receiving standard care or placebo)

(ii) an estimate of the likely effect on the outcome, assuming the new treatment is beneficial (e.g., reduces mortality by 20%).

Calculating sample size for an RCT in which a specific outcome is being assessed requires consideration of the statistical power of the study effect. A study is said to have adequate power if it can reliably detect a true difference in outcome between the standard or control arm and the intervention arm, if a clinically important difference actually exists. The power of a study increases with the more events and more participants that are included, or when its measurement of outcomes is more precise.

Statistical power is commonly set at 80%. This definition accepts a likelihood of one in five (i.e., 20%, β error) of not showing a statistically significant difference between two treatments when a clinically important effect really exists. Large trials therefore occasionally set the power at 90% to reduce to 10% the possibility of a so-called "false-negative" result. Apart from the false-negative result a RCT may also falsely conclude that one treatment is significantly superior to another, 'false-positive'. The probability of this occurring can also be controlled as part of determining the statistical power when designing the study. This probability is routinely set at 5% and is known as the α error.

Types of outcomes[2-6]

Outcome measures involve `counting people' or `taking measurements on people'. Frequently, `counting people' includes assessing the `time to the event' of the outcome of interest, e.g., the time to death.

For trial outcomes, the unit of interest is usually a person. So, the outcome measure will involve either **counting** how many people have a particular characteristic (i.e., put them into mutually exclusive groups, such as `dead' or `alive'), or **taking measurements** on them (see box).

Examples of outcome measures when the unit of interest is a person
Counting people (binary or categorical data)
Dead or alive
Admitted to hospital (yes or no)
Suffered a first heart attack (yes or no)
Recovered from disease (yes or no)
Severity of disease (mild, moderate, severe)
Ability to perform household duties (none, a little, some, moderate, high)
Taking measurements on people (continuous data)
Blood pressure
Body weight
Cholesterol level
Size of tumour
White blood cell count
Number of days in hospital
Number of units of alcohol intake per week

Counting people asks how many have the health outcome of interest, and may involve calculating the **percentage** or **proportion**, e.g.:

- The number of vaccinated individuals who develop flu is divided by the total number of people in the group. This proportion (or percentage) is the **risk**, i.e. the risk of developing flu if vaccinated. The same calculation can be made for the unvaccinated group.

Taking measurements on people: these outcomes vary between individuals, e.g., New York Heart Association class, systolic blood pressure, cholesterol levels. We can summarise these data with two parameters: the `average' level and the `spread' of a variable.

The `average level' is calculated as the location that summarises where the middle (central tendency) of the distribution lies. The three measures commonly used to describe the central tendency are:

- Mean: the sum of all the values of observations is divided by the total number of observations
- Median: the median is the middle value in a data set which divides a **distribution** exactly in half so that 50% of its **scores** are higher than it and 50% are lower. The major advantage with the median is that it is not overly influenced by **extreme observations.** It is often used in describing the typical income of a group of individuals, or residential real estate values. The median can be more appropriate than the mean for skewed distributions. When the distribution is symmetrical, the median equals the mean
- Mode: the most frequently occurring value, i.e., the most typical value. There may be more than one mode if two values are equally frequent.

For statistical analysis, the mean is most commonly reported. The mean is only a reliable measure of location if the data set it relates to is symmetrically distributed. If distribution is skewed (more data lie on either side of the mean), the mean is not useful, since it is greatly influenced by extreme observations.

The spread of a variable is summarised by:

- Standard deviation: indicates the average distance of all observations from the mean. The standard deviation has an important role in statistical analysis
- Range: the difference between the highest and lowest values is called the full range of values
- Range between percentiles: percentiles are the value below which a given percentage of the data observations occur. A common range used is the interquartile range, which is the range between the 25th and 75th percentile. Using this overcomes the problem of extreme data values lying away from the mean or median.

The standard deviation is more commonly used in statistical inference. The median and range are often used to describe the central location and spread of survival data.

Time-to-event outcome measures differ between trials, applying to any specified event occurring after a certain amount of time – e.g., time from entry into a trial until the occurrence or recurrence of a disorder (such as an asthma exacerbation, or time until hospital discharge). Commonly used time-to-event endpoints are described in the box below:

Endpoint	An event is defined as follows. All other subjects are censored	Comments
Overall survival	Death from any cause	Easily defined May mask the effects of an intervention if it only affects a specific disease
Disease-free survival	First recurrence of the disease Death from any cause	Useful when patients are thought to be free from disease after treatment, so patients have a good prognosis Needs date of recurrence
Event-free survival	First recurrence of the disease First recurrence of other specified diseases Death from any cause	Similar to disease-free survival
Progression-free survival	First sign of disease progression Death from any cause	Useful for advanced disease, where patients have not been `cured' after treatment, and are expected to get worse in the near future Needs date of progression
Disease (or cause)-specific survival	Death from the disease of interest	Useful when examining interventions that are not expected to have an effect on any disease apart from the one of interest Needs accurate recording and confirmation of cause of death Assumes treatment is not associated with deaths from other causes
Time-to-treatment failure	First sign of disease progression Death from any cause Stopped treatment	Similar to progression-free survival
Recurrence: there was no clinical evidence of the disease shortly after treatment, but the disease returned later on. Progression (or relapse): the patient still had the disease after treatment, but it got worse later. Disease and event-free survival may be used interchangeably, so it is useful to be clear about the precise definition.		

When the outcome measure is based on two or more event types and a subject could have both events (e.g., disease occurrence followed by death), the analysis will usually consider only the date of the first event. This is because the patient may be managed differently afterwards: the trial treatment changes or stops, non-trial therapies are given, or patients may receive the treatment from the other trial arm. When this occurs, it is difficult dealing with subsequent events, and how to attribute differences in the endpoint to the trial treatments.

Take home messages

Professor Shaun Holt: Tests of treatments must be designed properly in order to determine whether they are safe and effective. Even if well designed research does not find the preferred outcome, this is still useful as it can close avenues of research that are unlikely to lead to the safe or effective treatments and focus future research towards areas that are more likely to achieve this.

Good research design leads to a fair test of a treatment and it factors in the main reasons why we can be fooled into thinking a treatment may work, when in fact it does not. Such reasons include

1. natural history – many conditions will get better on their own and this is summarised by the cynical view on treating the common cold, that with good treatment it will resolve in seven days....and left on its own it would take around a week.
2. placebo effect – we are learning more and more about how optimism, wishful thinking and confirmation bias can (usually) positively effect our health. The placebo effect is remarkable – up to 40% improvements can be seen in some medical conditions in the placebo group of a study and in some conditions, such as depression, up to 90% of study participants may have a positive placebo response.
3. regression to the mean – many conditions improve and worsen over time, in a roughly cyclical fashion, but treatment is usually sought when the condition is at its worst, with the likely scenario being that improvements would have occurred anyway. People do not tend to wake up one day and say that their chronic back pain is the best it has been for years and then decide to see a chiropractor!

Fair tests of treatments and good clinical trial design aim to obtain reliable information for clinicians to use by, amongst other things, comparing like with like, by reducing biases, by accounting for the effects of chance and by assessing all relevant information.

The randomised controlled trial is the gold standard of medical research and one of the reasons that it is so effective at determining the truth about a treatment's effectiveness is that, as similar patients are randomly allocated to the different treatment groups, although the above factors discussed such as natural history and placebo effect will still be present, they will likely be present in each group of patients to approximately the same extent. This, in effect, will eliminate these factors and therefore any differences between the two groups that are found will likely be a result of the treatments and not these potentially confounding factors.

Study size is crucial and put simply, the larger the study, the more chance there is that the study results are an accurate reflection of the truth. If you spin a coin 100 times there will be very close to a 50:50 split in terms of the numbers of heads and tails, but if you were to spin the coin only 10 times, you could by chance get eight, nine or even 10 of one type.

Statistical and clinical significance are terms that are often confused, particularly by the media, and are best illustrated with an example. In a large study of 10,000 patients taking a new drug for high blood pressure, the main finding may be that the drug reduces systolic blood pressure by 1 mmHg, with a p value of less than 0.05. This would be statistically significant, but not clinically significant, as such a small reduction in blood pressure is not likely to affect health outcomes.

Systematic reviews – even well-designed studies – can occasionally give an incorrect answer. In addition, we may have other studies that were not large enough, or had other problems, but still provide useful data.

Therefore the highest form of medical research evidence is a systematic review which looks at all relevant, important data, and places more importance on that which is most accurate. A good systematic review sets out in advance which study information will be collected and how the data will be analysed, thereby minimising the chance of bias by "cherry-picking", i.e., only including positive or negative studies. One of the best examples of the power of systematic reviews can be seen in the logo of the Cochrane collaboration.

The logo shows the situation with respect to the use of corticosteroids in prematurely born babies in that, before the systematic review, their usefulness or otherwise was simply not known. Small studies with different findings are represented on the logo and represent the confused clinical picture at the time. However, the systematic review determined with certainty that the treatment was effective, and based on this finding, the practice was introduced into mainstream practice resulting in thousands of babies not dying.

THE COCHRANE COLLABORATION®

Whole textbooks are devoted to the subject of good clinical study design, but the aspects described above are some of the most important.

REFERENCES

1. Sackett DL, et al. Evidence-based medicine: How to practice and teach EBM. 2nd ed. Edinburgh, United Kingdom: Churchill Livingstone, 2000:173-7.
2. Peace KE, Chen D-G. Clinical Trial Methodology (Chapman & Hall/CRC Biostatistics Series). 2010.
3. Stanley K. Statistical primer for cardiovascular research. *Circulation* 2007;115:1164-9.
4. Wang D, Bakhai A. Clinical trials: a practical guide to design, analysis, and reporting. ReMedica Publishing, 2006.
5. Hackshaw A. A concise guide to clinical trials. Chichester: Wiley-Blackwell, 2009.
6. Flather M, Aston H, Stables R. Handbook of clinical trials. ReMedica Publishing, 2001.
7. The International Conference on Harmonisation of Technical Requirements for Registration of Pharmaceuticals for Human Use (ICH). www.ich.org

Roche

This publication has been created with an educational grant from Roche Products (NZ) Limited.
The content is entirely independent and based on the authors' opinions.