Project Identification: 2021-1-CZ01-KA220-HED-000023177

Investor/Program/Project type: European Union - Erasmus+ Key Action 2: Cooperation for innovation and the exchange of good practices; Strategic Partnerships in the field of education, training and youth

Curriculum Development of Human Clinical Trials for the Next Generation of PhD Students and Early Career Researchers in the Medical, Science, Pharmacy and Health Professions

CHAPTER 1

CLINICAL TRIAL DESIGN

Authors: Frances Shiely
University College Cork, Cork, Ireland

Reviewers: Marta del Alamo, Keiko Ueda, Christine Kubiak
European Clinical Research Infrastructure Network (ECRIN)

Date first created: 12/09/2022

Last revision: 07/05/2024

Content

Co-funded by the Erasmus+ Programme of the European Union

Time required to complete this chapter

Core content:	2h 00m
Additional/advanced content (yellow boxes):	3h 00m
Activities/practical exercises (blue framed boxes):	30m
Total time:	**5h 30m**

Co-funded by the Erasmus+ Programme of the European Union

1 Introduction to the chapter

Randomised trials are conducted to provide evidence to support better and more informed decisions about medicine and other healthcare initiatives. Trials support these decisions through the data they collect which is guided by the research question and appropriate trial design to answer this research question. Trials are designed to minimise bias in so far as possible. Typical methods to reduce bias for example are the use of a control group, randomisation etc. The purpose of the randomised trial is assessment of efficacy, safety, or risk benefit ratio and the goal may be superiority, non-inferiority, or equivalence. We typically classify trials into four phases: Phase 1, 2, 3 and 4. These may include trials of therapeutic agents, prophylactic agents, diagnostic agents, surgical procedures or health service strategies. Who is involved in the design of the trial is usually at the discretion of the principal investigator, but we would like to see a full trial design team, including, statisticians, methodologists, patients and the public representatives, research nurses, trial managers, data scientists and quality and regulatory affairs representatives. The Randomised Controlled Trial design is crucial to ensure the trial is effectively designed to collect the trial outcomes and answer the research question with high quality data. The first chapter in this lecture series of 12 is focused on trial design. In the CONSCIOUS project[1], we introduced you to trial design. We will briefly revise some of that here, and then move on to more advanced trial design methods. In total we will talk about the following: Review of superiority, non-inferiority and equivalence trials; Review of phases of trials; Cluster randomised trials; Adaptive trial design; Explanatory versus pragmatic trials; Looking at SPIRIT, PICO, Equipoise, the Research question; Eligibility criteria and outcome data; Sample size, randomisation; and blinding.

2 Review of Superiority, Non-Inferiority and Equivalence Trials

Table 1: Objectives of superiority, non-inferiority and equivalence trials

Name	Objective
Superiority trial	To determine a clinically relevant difference between two interventions
Equivalence trial	To determine whether a (new) intervention is neither worse nor better than another (established) intervention
Non-inferiority trial	To determine whether a (new) intervention is not inferior to another (established) intervention

2.1 Superiority Trials

In the past, the typical trial question was, "is the new intervention better, or superior to, no treatment or standard treatment"? However, as medicine, care and treatments improve, many patients are now on a standard regimen, we ask "do alternative treatments that may be equal to, or at least no worse than, existing treatments with regard to the primary outcome convey other important advantages in terms of safety, adherence, patient convenience, and or/cost? These trials are referred to as non-inferiority trials. For example, first line multiple sclerosis treatments were all given by subcutaneous injections. Patients did this at home themselves. About 5 years ago, the first oral therapy was introduced. These were trialled as non-inferiority trials. They were trying to establish if they were no worse than existing treatments, but of

[1] CONSCIOUS: Chapter 2, Why clinical trials are the gold standard?, http://conscious.novaims.unl.pt/my/

course the oral regimen is preferable for patients. It is widely accepted that important differences exist between superiority and non-inferiority trials in terms of their design, analysis and interpretation. This is reflected in regulatory agency guidelines, CONSORT statements on the reporting of trials and review articles.

Superiority trials are used to demonstrate the study treatment is superior to the control. So, when the aim of the study is to show that an experimental (E) treatment is superior to a control (C) treatment, the RCT is called a superiority trial and the associated statistical test is a superiority test. Often a non-significant superiority test is wrongly interpreted as proof of no difference between the two treatments. This is really important to get right. This is not the case. Proving that two treatments are equal in performance is impossible with statistical tools; at most, one can show that they are equivalent. But this is a different type of study which we will discuss below in the Equivalence Section.

If the purpose of the superiority trial is to detect a difference between two drugs, the goal of establishing a new drug is statistically superior to the active control and/or placebo, is easier than the goal of establishing the new drug is clinically superior to the active control and/or placebo. The null hypothesis (hypothesis written in the negative) is that treatment X is NOT more effective than treatment Y for a given condition.

The key point to remember, is that the question in a superiority trial is different to the question in a non-inferiority trial and this affects the design and conduct of the trial.

Look at the Figure 1. It shows the superiority margin. The upper confidence interval (black line) shows that the line is fully to the right, showing that the new treatment is superior to the control. The one below it is also superior but the error bars are closer to the blue line, the edge of when it's determined that the new treatment is better than the old. The third line shows that the confidence interval crosses the blue line of superiority therefore the new treatment is not better. NB. Just because a new treatment is not better than the control treatment, it does not mean that the new treatment is worse, or inferior.

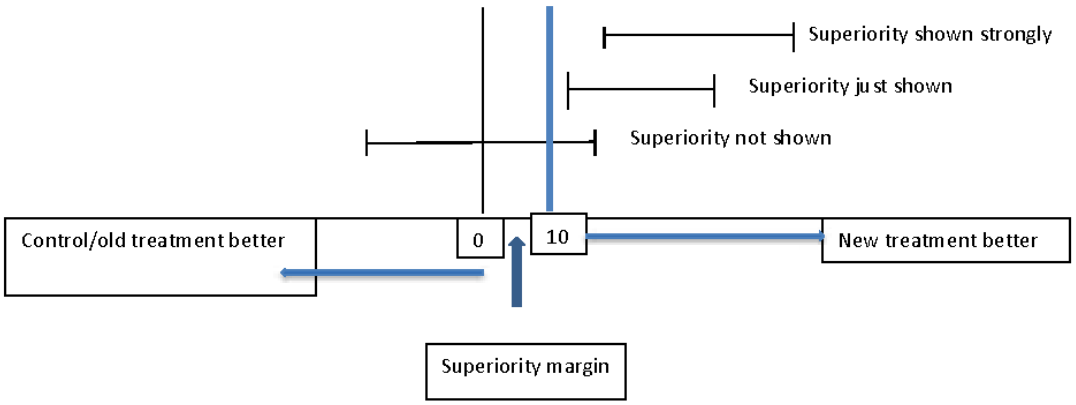

Figure 1: Superiority margin

2.2 Equivalence Trials

In equivalency trials, the objective is to test whether a new intervention is equivalent to an established one. This is different from superiority and non-inferiority. This is also the trial design

typically used to get market authorisation for generic drugs. The null hypothesis for an equivalence trial is treatment X is either worse or better than treatment Y for a given condition by greater than Δ (Δ is the equivalency margin - the amount of difference that is acceptable by the investigators to state that the two treatments are the same).

In equivalency studies, there are design aspects that need to be considered. Firstly, the control or standard treatment must have been shown conclusively to be effective, i.e., truly better than placebo or no therapy. The circumstances under which the active control was found to be useful (populations, dosage etc.) ought to be reasonably close to those of the planned trial. These requirements also mean that the trials that demonstrated efficacy of the standard should be recent and properly designed, conducted, analysed and reported.

If you look at Figure 2 and follow the x-axis, you will see that the equivalence margin goes from -8 to +8. These parameters are decided, based on existing evidence from either observational studies or other trials. The 0 is the line of no association. Take the "inconclusive" that crosses the line -8. The length of this horizontal line indicates the error around the true value. As this line is on both sides of the -8 limit which has been set, then our result is inconclusive, and we cannot say that the treatment is equivalent. The same can be said for the "inconclusive" line that crosses at +8. However, if our line as evident in the "equivalent" line is within the -8 and + 8 boundary, then we can say that our study drug is equivalent to the one it was measured against. If the line is totally to the left of the graph, then we can say our study drug is superior to the one it is being tested against. If the line is totally to the right, as in the "inferior" line show, then our study drug is inferior to the one it is tested against.

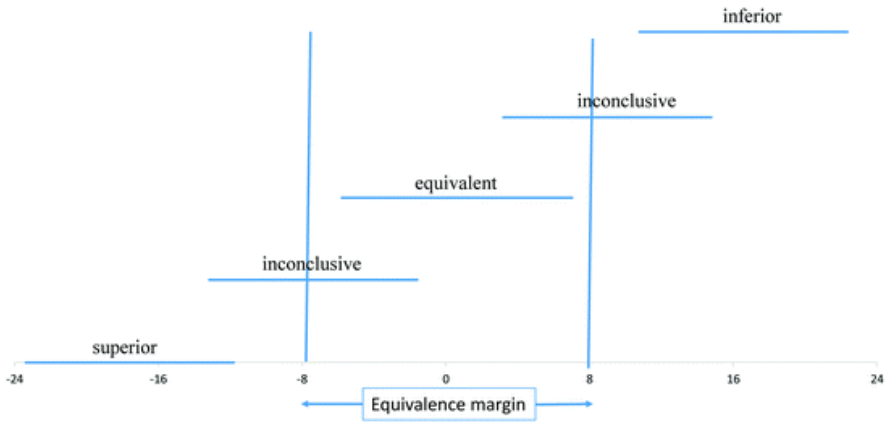

Figure 2: Equivalence margin

2.2.1 Issue with showing equivalence trials are identical

It cannot be shown statistically that two therapies are identical, as an infinite sample size would be required. Therefore, if the intervention falls sufficiently close to the standard, as defined by reasonable boundaries, the intervention is claimed to be "the same" as the control in an equivalence trial. Selecting the margin of indifference is a challenge. In an equivalence trial, it is necessary to determine a “zone of clinical equivalence” prior to the trial onset.

2.3 Non-inferiority Trials

Non-inferiority trials test whether a new experimental treatment is not unacceptably less efficacious than an active control treatment already in use. With continuous improvements in

Co-funded by the Erasmus+ Programme of the European Union

health technologies, standard care, and clinical outcomes, the incremental benefits of newly developed treatments may be only marginal over existing treatments. Sometimes assigning patients to a placebo is unethical. In such circumstances, there has been increasing emphasis on the use of non-inferiority trial designs. Non-inferiority trials are more complex to design, conduct, and interpret than typical superiority trials.

2.3.1 What is an Active Control?

The term "active control trial" refers to clinical trials in which the control treatment employed is an active one. There are several reasons for using active controls in clinical trials.

In trials involving serious outcomes such as mortality, it is unethical to use a placebo when active treatments are available. Clinical equipoise, referring to the state of true uncertainty about the relative benefits of alternative treatments under the "null" hypothesis to be tested, is an ethically necessary condition in all clinical research. Active controls are sometimes used to demonstrate the efficacy of a drug that may have large placebo effects. Active controls are also used to determine how experimental treatments compare to alternative treatments. Active control trials aim to demonstrate that treatments of interest have either superior effects or similar effects to the control.

The research question in a non-inferiority trial is whether the experimental therapy is not inferior to the active control (whereas the experimental therapy in an equivalence trial should not be inferior to, nor superior to, the active control). Thus, a non-inferiority trial is one-sided, whereas an equivalence trial is two-sided. (For non-inferiority, we want experimental therapy to be better than the active control). Many of the same issues that are critical for designing an equivalence trial also are critical for designing a non-inferiority trial, namely, appropriate selection of an active control and appropriate selection of the “zone of clinical non-inferiority” (margin).

2.3.2 Why Non-inferiority Trials?

Non-inferiority trials aim to show that the new drug is no worse than standard treatment. Equivalence trials aim to show the new treatment is no better and no worse. Non-inferiority trials may also be used when the new treatment may offer important advantages over currently available standard treatments, in terms of improved safety, convenience, better compliance, or cost. In addition, clinical trials are increasingly required to demonstrate benefits in clinical endpoints rather than surrogate endpoints, even though the incremental benefits from new treatments is diminishing, which is also an important factor in determining sample size. Such practical considerations are also driving a trend towards designing clinical trials that aim to demonstrate experimental treatments have similar effects to active controls of a proven efficacy rather than a superior effect. However, testing for non-inferiority makes trial design and interpretation of results less straightforward than typical superiority trials.

2.3.3 Interpreting Non-inferiority Trials

Interpreting non-inferiority trials can be difficult to get your head around as we are mixing our terminology with superior while we are not actually conducting a superiority trial. Try to come up with some means by which you can distinguish these terms. Figure 3 is very useful as it shows the zone of inferiority (shaded green). To interpret it, you also need to be able to interpret a 95% confidence interval.

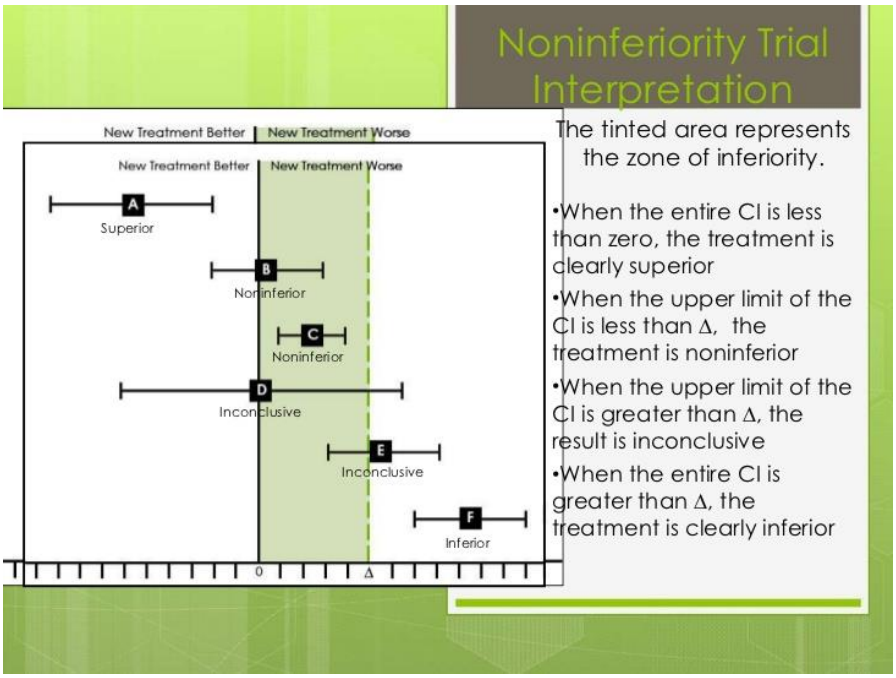

Figure 3: Non-inferiority trial interpretation

Practical exercise
Read the following article by Dunn et al. 2018 from Trials [Dunn DT, Copas AJ, Brocklehurst P. Superiority and non-inferiority: two sides of the same coin? Trials. 2018 Dec;19(1):1-5].[2] Trials Journal has free access to all articles. Within the text, take the CAP-IT study and decide if you think it is a superiority study or an inferiority study. Justify your response.

3 Trial Phases

Look at Figure 4 which details the phases and timelines of a new drug for example.

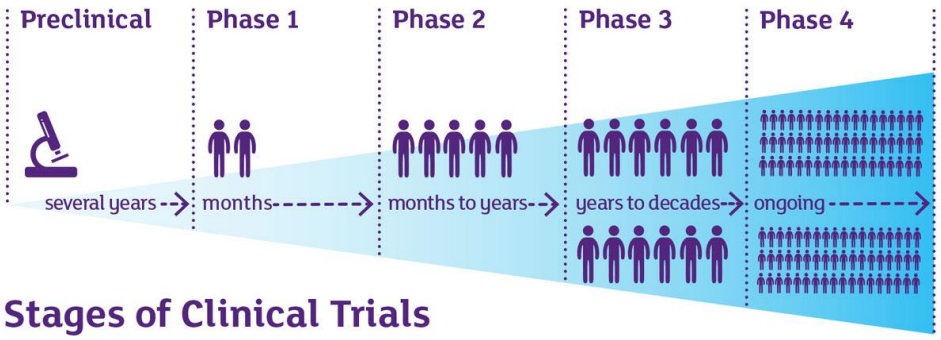

Figure 4: Phases of clinical trials

[2] https://trialsjournal.biomedcentral.com/articles/10.1186/s13063-018-2885-z

3.1 Pre-clinical phase

This is the so-called move from bench to bedside, i.e., deciding if a new drug or therapy is suitable for clinical trials. Trials are only initiated if pre-clinical findings suggest that the new drug or treatment is likely to be safe and will work in people.

3.2 Phase I Clinical Trials

The hallmark of phase I clinical trials is the involvement of human participants at this stage. Though useful pre-clinical information is obtained from animal models or cells, it is essential to get early data from human participants. People who partake in phase I studies are usually healthy volunteers. In some cases however, people who have already tried and failed to improve on existing therapies may be permitted to partake. This is usually in circumstances where a person's illness is so advanced that they will most likely not survive.

The principle objectives in Phase I drug trials are to:

- make sure that the new medicine presents no major safety issues,
- determine an acceptable single drug dosage (how much can be given without causing serious side-effects),
- clarify that it can reach the targeted body area, and remain there long enough to deliver its benefits,
- gain preliminary evidence that it could offer therapeutic value or prevent the disease or condition.

One of the first steps in evaluating drugs is to estimate how large a dose can be given before unacceptable toxicity is experienced by patients. This is usually referred to as the maximally tolerated dose. In estimating the maximally tolerated dose, the investigator usually starts with a very low dose and escalates the dose until a prespecified level of toxicity is obtained. Typically, a small number of participants, usually three, are entered sequentially at a particular dose. IF not specified level of toxicity is observed, the next predefined higher dose level is used. If unacceptable toxicity is observed in any of the three participants, additional participants, usually three, are treated at the same dose. If no further toxicity is seen, the dose is escalated to the next higher dose. If additional unacceptable toxicity is observed in any of the three participants, additional participants, usually three, are treated at the same dose. If no further toxicity is seen, the dose is escalated to the next higher dose. If additional unacceptable toxicity is observed, then the dose escalation is terminated and that dose, or perhaps the previous dose, is declared to be the maximally tolerated dose. This particular design assumes that the maximally tolerated dose occurs when approximately one-third of the participants experience unacceptable toxicity. Variations of this design exist, but most are similar.

Phase I trials are not randomised. They may or may not have a control group. Usually Phase I trials are a series of cases in which the participants are given incremental doses of the drug while they are monitored carefully by the investigators. Phase I studies attempt to estimate tolerability and characterise pharmacokinetics (movement of drugs within the body) and pharmacodynamics (effects of drugs). Once the safety of the drug or vaccine is established, researchers then progress to stage two. Typically phase I studies might require a total of around 10-20 patients but this could be up to 80-100 patients.

3.3 Phase II Clinical Trials

If phase I is successful, approval will be sought for a trial involving a larger group of people. Phase II trials will usually (but not always) include patients who have the condition the potential medicine is targeting. They are fairly small scale investigations into the efficacy and safety of a drug, and require close monitoring of each patient. Phase II trials can sometimes be set up as a screening process to select out those relatively few drugs of genuine potential from the larger number of drugs which are inactive or over-toxic, so that chosen drugs may proceed to phase III trials. Phase II trials typically have 100-200 patients.

The objectives of Phase II trials are to establish:

- effectiveness in treating the condition,
- effectiveness in preventing the condition (if the volunteer does not already have it),
- appropriate dosing levels.

Once a dose or range of doses is determined, the next goal is to evaluate whether the drug has a biological activity or effect. The comparison may consist of a concurrent control group, historical control, or pre-treatment status versus post-treatment status. Because of uncertainty with regard to dose-response, phase II studies may also employ several doses, with perhaps four or five intervention arms. Participants in phase II studies are usually carefully selected, with narrow inclusion criteria. Generally phase II studies are performed to make a decision as to whether to further develop a new drug or device. As such the purpose is to refine an estimate of the probability of success in phase III. Success depends on a variety of factors, including estimated beneficial and adverse effects, feasibility, and event rates of the target population. Because phase II trials by definition don't have adequate power to define the effect on major clinical outcomes, the estimate of treatment effect and harm may depend on multiple inputs, including effects on biomarkers, on more common but less definitive clinical outcomes, and on more minor safety signals.

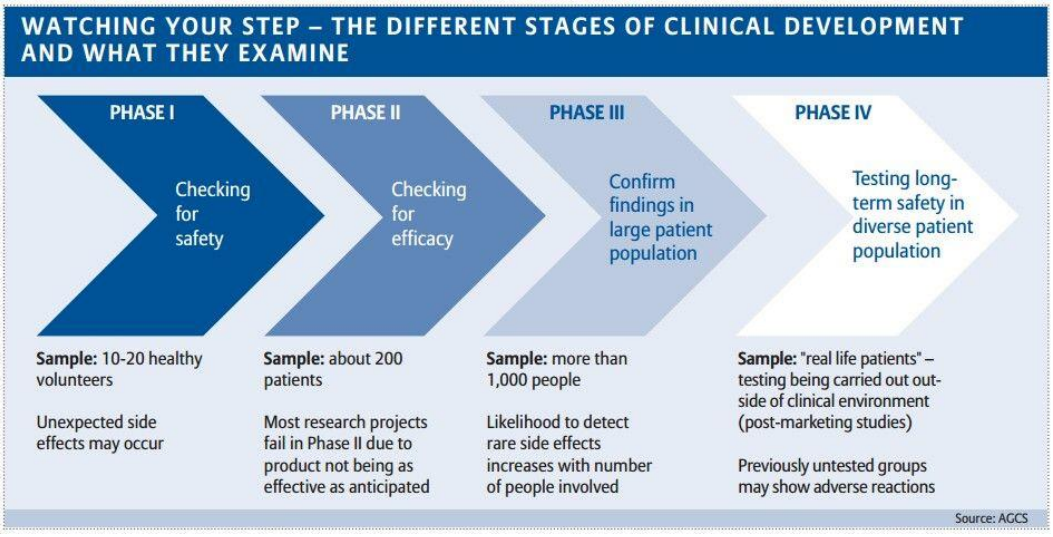

Figure 5: Number of participants and trial phases

3.4 Phase III Trials

Phase III trials aim to find out:

- which treatment works better for a specific disease/illness
- more about the side effects
- how the treatment affects people's quality of life

They may compare standard treatment with:

- a completely new treatment
- different doses of the same treatment
- having the same treatment more, or less, often
- a new way of giving a standard treatment

Phase III trials usually involve many more patients than phase I or II. This is because differences in success rates may be small. So, the trial needs many patients to be able to show the difference. Sometimes phase III trials involve thousands of people in many different hospitals and even different countries. Most phase III trials are randomised. This means the people taking part are put into treatment groups at random.

When we conduct a Phase III clinical trial, that trial occurs in various stages as detailed in the table below.

Table 2: Stages of a Phase III Trial

Stage	Activities
1. Initial Concept	Scientific questions, endpoints/outcomes, sample size
2. Protocol design	Operational manual, invited meetings to design the trial
3. Recruitment and conduct of the trial	Enrolment, monitoring, DSMB (data safety monitoring board), interim analysis (if appropriate)
4. Follow-up	Events (safety, efficacy)
5. Close-out	Prepare a 'clean and locked' database
6. Analyse and report the results	Data analysis, presentations and manuscripts for publication, final report to the regulatory agencies if it is a regulated trial
7. Long-term follow-up	Link findings with initial trial data

1. As we consider our clinical trial, we have an initial concept. We also know what scientific question we would like to answer. What endpoints/outcomes would we evaluate to determine whether or not our scientific question has been answered correctly? This is really important as the outcomes define the sample size. We then consult with our biostatistical colleagues and determine the sample size and the power that we would need in order to adequately test our hypothesis, our scientific question for the trial.

2. In the second stage, we write a very detailed protocol, essentially a recipe for how investigators are going to conduct themselves with respect to their interface with the participant at every site that is participating in the trial. We distribute this detailed recipe in the form of an operations manual to every investigator and their research team. This could be in multiple languages if it is an international multi-centre trial. We hold investigator meetings so that all

individuals who are conducting the trial are doing it the same way so
that we can have a degree of assurance that there's no variation in how the protocol is being implemented site to site.

3. The longest phase of a Phase 3 trial is often the period of recruitment and conduct of the trial. It is during this phase we enrol participants who fulfil the enrolment criteria for the trial. We monitor activities during the trial. One form of monitoring is to determine whether or not investigators are actually deviating from the protocol. And we make sure to remind them of the steps that they need to follow when they conduct the protocol. We also consider the fact that as data are accumulating, there could be some evidence which is starting to accumulate suggesting that one treatment may be better than another. As investigators, we don't know this because you will remember that we have to have equipoise in our study as we approach a patient and ask them for informed consent. We have to be able to truly face them and say we don't know whether one treatment is better than another. But we do have an obligation to have an independent party who is charged with that responsibility of monitoring any emerging evidence. That is the data safety monitoring board (DSMB) and they periodically look at the evidence in a clinical trial. This is referred to as interim (in the middle) analyses.

4. The follow-up phase of the trial is where we monitor the patients until we have acquired a sufficient number of events. Efficacy events are usually the target for us in an event driven trial. These again are our outcomes, specifically the primary outcome which determines the sample size. We hypothesised what the treatment effect might be, we hypothesised what the event rate might be in the control group, and we make a determination as to how many events we would need if that hypothesis were true. We continue to catalogue safety events during the follow-up phase as well.

5. We then enter the close-out phase of the trial once we have acquired the appropriate number of participants. During the close-out phase of the trial, our major activities are to achieve a clean and locked database. This is easily done if the database is correctly completed and maintained throughout the trial. In the situation where investigators fill out a case report form, sometimes their entries are a little bit unclear and a query is sent from the data management team to the investigator asking for clarification of what they meant. The database is kept open until we have clarified all these open issues. Then we arrive at a point where we say no further changes will be made to the database and we consider the database to be locked.

6. The next phase of the trial is where we analyse and report the results. Data analysis is obviously very important during this phase of the trial. Our focus is to do that analysis of the data so that we can present our findings in the form of oral presentations and written manuscripts. Phase III regulatory studies (typically medical device or drug trials) also require a final report to be sent to regulatory agencies to determine whether or not the drug can be approved.

7. More and more Phase III clinical trials also have a long-term follow-up phase. Sometimes this is actually continued in a double blind fashion. Once that long term follow-up phase is complete, investigators link the findings of the long-term follow-up phase with the initial trial data.

3.5 Phase IV Trials

Phase IV trials are usually conducted after marketing authorisation is granted and the medicine is in general use. Phase IV studies are also known as post-authorisation safety studies (PASS) and may be voluntary or imposed by the regulatory authorities. The possibility also exists of requesting the marketing authorisation holder to conduct post-authorisation efficacy studies (PAESs) in order to complement efficacy data that are available at the time of the initial authorisation. Phase IV studies collect additional information about side-effects and safety, long-term risks and benefits, and/or how well the medicine works when used widely.

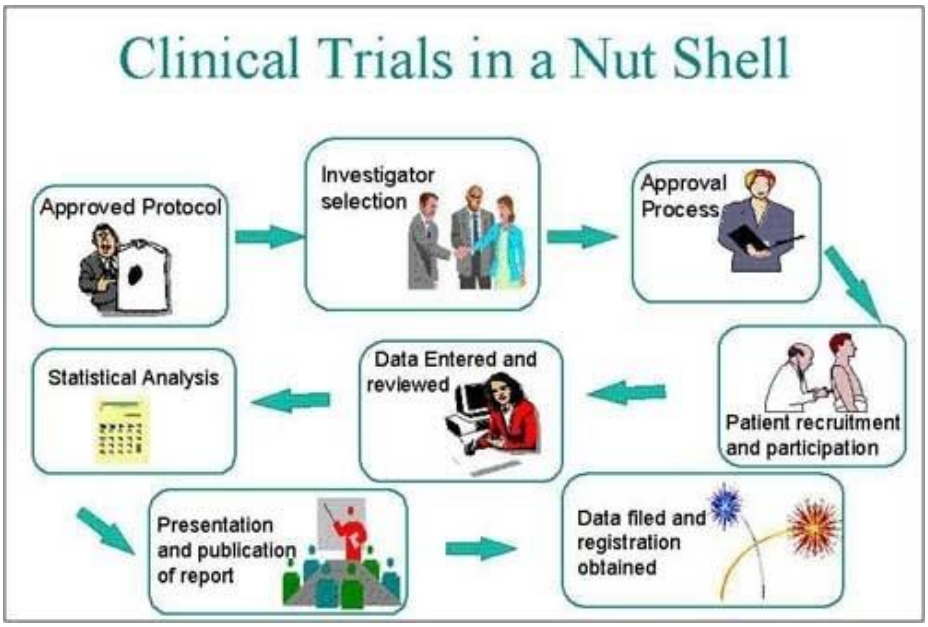

Figure 6: Clinical Trials in a Nut Shell

4 Cluster Randomised Trials

A cluster randomised trial (CRT), also known as a group randomised trial or community randomised trial, is a randomised controlled trial in which groups of individuals (clusters e.g. school, clinic, village) are randomised. Cluster randomised controlled trials are so called because they unit of randomisation is a cluster, rather than individual. Think back to when we discussed randomisation in detail in CONSCIOUS[3]. We randomised individual people. In a CRT we randomise groups of individuals. For example, if we were doing a cluster RCT in GP practices, we would randomised by the GP practice. So everybody in one GP if randomised to the intervention, would all receive the intervention. If a GP practice was randomised to the control group, each individual enrolled in the study would be in the control arm. Sometimes we wish to use an intervention at certain times, then time would form the cluster unit of allocation.

[3] CONSCIOUS: Chapter 2, Why clinical trials are the gold standard?, http://conscious.novaims.unl.pt/my/

Listening
Listen to this three minute clip of Prof. Sandra Eldridge talking about CRTs ("Sandra Eldridge - Question 10 What is a cluster-randomised trial and when is it useful?"[4]). She is one of the leaders in this area and maintains the website that I mention below too.

4.1 Why do a CRT

CRTs can be harder to design, require more participants than individually randomised trials, and are often more prone to biases. However, reasons why we might choose to conduct a CRT rather than an individually-randomised trial include that the intervention is implemented at the cluster level, there are practical and/or ethical difficulties in randomising at individual level (although randomising clusters to avoid consent is not acceptable), to avoid issues of contamination, or to estimate indirect effects, for example in vaccination trials.

Below are various examples of trials that have been conducted for these reasons. Read them.

a) The intervention is implemented at the cluster level or it is logistically easier or more ethical to administer to groups of individuals.

- School based education programme to reduce salt intake in children and their families (School-EduSalt)[5]
- Effects of water quality, sanitation, hand-washing, and nutritional interventions on diarrhoea and child growth in rural Bangladesh: a cluster randomised controlled trial[6]

b) To avoid issues of contamination, e.g. for infectious diseases, or to estimate indirect effects, for example in vaccination trials.

- Village-integrated eye worker trial (VIEW): rationale and design of a cluster-randomised trial to prevent corneal ulcers in resource-limited settings[7]

4.2 Types of CRT Design

The standard design, the parallel two-arm design, randomises clusters to one of two arms or conditions, e.g. treatment or control, and measures outcomes in individuals in both arms on the same follow-up measurement schedule. Depending on the research question, the same individuals may be followed up over time (a cohort design) or different individuals may be sampled at different time points (a cross-sectional design). When the parallel CRT design is augmented by the addition of baseline measures before randomisation, this is referred to as the parallel cluster randomised trial with before and after observations (CRT-BA).

[4] https://youtu.be/kO70HO90Xow
[5] https://www.bmj.com/content/350/bmj.h770
[6] https://www.thelancet.com/journals/langlo/article/PIIS2214-109X(17)30490-4/fulltext
[7] https://bmjopen.bmj.com/content/bmjopen/8/8/e021556.full.pdf

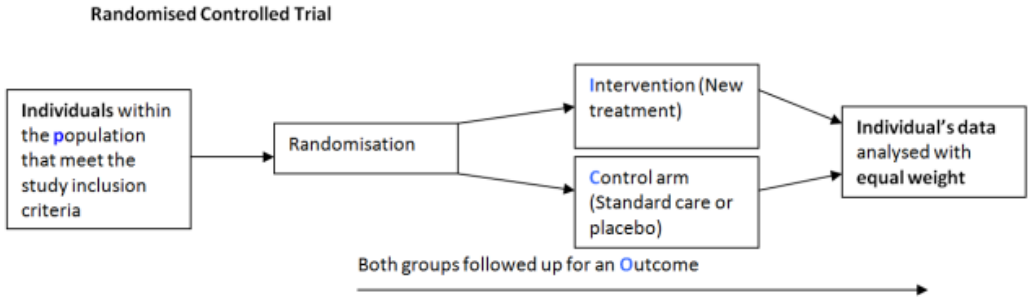

Figure 7: Parallel two-arm CRT

Further reading on CRTs can be accessed at the following link.[8]

Alternatives to the parallel two-arm design include the crossover design and the stepped wedge design, which is a modification of the crossover design. In the crossover design, clusters are randomised to a treatment sequence, and in the stepped wedge design all clusters start in the control condition and eventually end in the treatment condition, where the treatment start time is randomly allocated.

If you are interested in study design and want to read more about stepped wedge CRTs see the following resource.[9] This website is maintained by a group of individuals who specialise in CRTs and Stepped Wedge CRTs (SW-CRTs).

4.3 Advantages of CRTs

1. Sometimes it is the only correct methodology, e.g., if implementing a new school curriculum. A randomised trial evaluating the curriculum in this instance would to randomise some schools to adopt the new curriculum while others would be allocated to continue with the existing curriculum.

2. Cluster randomisation avoids, or reduces, the risk of contamination between the intervention group and the control group. Contamination occurs when some people exposed to the intervention transfer knowledge to those in the control group. This will dilute any intervention effects and make it more difficult to show any difference in effect between the two groups.

4.4 Disadvantages of CRTs

1. CRTs require a larger sample size than trial with individually randomised participants. Typically between 50 percent and 200 percent more participants are required because the standard statistical methods used for analysis and power calculation assume that outcomes for individuals within a trial have no relationship with the outcomes of others within the trial. This is tricky in a CRT because take for example an intervention in a school setting, children of similar characteristics are often selected into a class and they

[8] https://clusterrandomisedtrials.qmul.ac.uk/what-is-a-parallel-crt/
[9] https://clusterrandomisedtrials.qmul.ac.uk/what-is-a-sw-crt/

are taught by the same teacher in the same environment. This correlation between individuals within a cluster is known as the intracluster correlation coefficient (ICC). When we do a power calculation to determine how large our trial needs to be, we take the ICC into account.

2. Design bias is an issue, particularly selection bias after randomisation. For example, CRTs are prone to biased recruitment. This is because many CRTs first recruit the cluster, then randomise and finally recruit the participants. This approach leads to bias. Given the allocation is usually known to those recruiting the participants, it can be selective, and this introduces post randomisation selection bias. To avoid this, the potential participants within a cluster must ideally be identified in advance of randomisation.

3. Post-randomisation exclusion is another issue in CRTs. This relates to lack of blinding by the person applying the inclusion and exclusion criteria.

4. Dilution bias is also an issue. This occurs because participants are allocated without consent and then consent is required from the individuals after. Depending on the intervention, there can be a high refusal rate. If 30% of the intervention group refuse to be part of the intervention group, then the intervention effects will be diluted by 30%. In order to improve the quality of cluster trials participants should be identified before the cluster randomisation which can avoid both recruitment bias and dilution effects.

Thanks to Prof. David Torgerson for some of the contents of this lecture from his book 'Designing Randomised Trials in Health, Education and the Social Sciences: An Introduction' (Torgerson, D. (2008) Springer) Ch. 9. CRTs.

5 Adaptive Trial Design; Explanatory and Pragmatic Trials

An adaptive design is defined as a design that allows modifications to the trial and/or statistical procedures of the trial after its initiation without undermining its validity and integrity. The purpose is to make clinical trials more flexible, efficient and fast. Due to the level of flexibility involved, these trial designs are also termed as “flexible designs.” Flexibility here does not mean that the trial can be modified any time at will. The modification and adaptations have to be pre-planned and should be based on data collected from the study itself. In other words, they cannot be redesigned at random. From the beginning, a plan is made to create an opportunity for modification as the trial progresses if it is necessary. Do NOT mistake this design with a platform trial design whereby multiple treatments can be added to a trial as it progresses. Figure 8 shows the difference between a fixed design and an adaptive design.

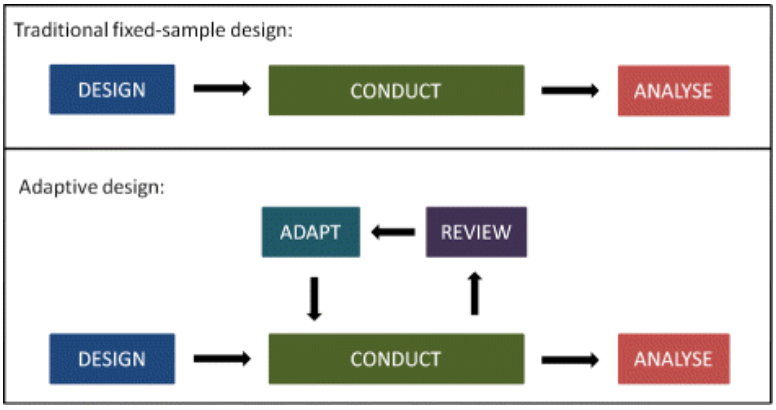

Figure 8: Adaptive design versus the traditional fixed design

5.1 Advantage of Adaptive Trial Designs

The central advantage of adaptive design is the ability to include prospectively planned opportunities for modifying study design elements and hypotheses based upon interim data analyses. Such modifications must be prospectively planned in the protocol and any interim analyses need to control for statistical bias. Adaptive designs can lead to improved efficiency (either fewer participants on average to achieve the same level of statistical power to detect a true treatment effect, or higher power for the same number of participants) and trial attractiveness to enrolled participants (e.g. by closing ineffective arms earlier and allocating more participants to treatments that have shown more promise from data accrued so far).

5.2 Disadvantage of Adaptive Trial Design

It sounds so positive, but the first law of thermodynamics in lay man's terms is, we don't get something for nothing. So what are the considered disadvantages. For a thorough read on this, please read Wason, Brocklehurst and Yap's paper in BMC Medicine. They go through some good study examples of adaptive designs.

Wason J, Brocklehurst P, Yap C. When to keep it simple–adaptive designs are not always useful. BMC medicine. 2019 Dec;17(1):1-70.[10]

Read the following freely accessible article on why you should consider an adaptive trial design and how to run and report them.

Pallmann P, Bedding AW, Choodari-Oskooei B, Dimairo M, Flight L, Hampson LV, Holmes J, Mander AP, Odondi LO, Sydes MR, Villar SS. Adaptive designs in clinical trials: why use them, and how to run and report them. BMC medicine. 2018 Dec;16(1):1-5.[11]

[10] https://bmcmedicine.biomedcentral.com/articles/10.1186/s12916-019-1391-9
[11] https://bmcmedicine.biomedcentral.com/articles/10.1186/s12916-018-1017-7

5.3 RCTs defined by intervention, exposure, no. of participants etc.

There are many different types of RCTs. Jadad (1998) gives a comprehensive list and I have modified this list to include newer types of designs.[12]

Table 3: Types of Randomised Controlled Trials (RCTs)

RCTs according to the aspects of the interventions they evaluate
Explanatory and pragmatic trials Efficacy and effectiveness trials Phase I, II, II and IV trials
RCTs according to how the participants are exposed to the interventions
Parallel trials Crossover trials Trials with factorial design Adaptive trials Cluster trials
RCTs according to the number of participants
From n-of-1 to meg-trials Fixed size Sequential trials
RCTs according to whether the investigators and participants know which intervention is being assessed
Open trials Single blind trials Double blind trials Triple and quadruple-blind trials
RCTs according to whether the preferences of non-randomised individuals are participants are taken into account
Zelen's design Comprehensive cohort design Wennberg's design

5.4 Explanatory versus Pragmatic Trials

Explanatory trials address whether or not an intervention works. Typically, these trials are designed in such a way that the results are likely to yield a 'clean' evaluation of the interventions. So the situation is controlled very tightly by setting strict inclusion criteria for example, or by using homogeneous study groups. In general, explanatory trials are optimized to demonstrate the efficacy of an intervention in a highly selected patient group; however, findings from these studies may not be generalizable to the larger clinical problem.

On the other hand, pragmatic trials are designed not only to determine whether the intervention works, but also to describe all the consequences of its use, good and bad, under circumstances mimicking clinical practice. In other words, these are trials conducted in the real-world setting.

[12] Jadad AR. Randomised controlled trials: a user's guide. Health Technology Assessment. 1998;2(13), 214.

Pragmatic trials attempt to understand the real-world benefit of an intervention by incorporating design elements that allow for greater generalizability and clinical applicability of study results.

Listening
Listen to the following short video of Professor Sandra Eldridge "S. Eldridge - Q. 2 How are pragmatic trials different from ... ?".[13] Sandra is retired now but was a Professor of Biostatistics and Director of the UKCRC Registered Pragmatic Clinical Trials Unit at Barts and The London School of Medicine and Dentistry.

As our goal is that trial results are applicable and relevant, it is our desire to conduct trials that are more pragmatic than explanatory. The reality though is that some elements of our design are more pragmatic and some are more explanatory so our trials are on a pragmatic-explanatory continuum. PRECIS-2, (Pragmatic Explanatory Continuum Indicator Summary 2) is a tool to help trialists design clinical trials and consider where they would like their trial to be on this pragmatic/explanatory continuum. The tool was published with a toolkit with information on how to use it when designing a trial. The PRECIS 2 website has all of this information available for free at https://www.precis-2.org/.[14] Figure 9 shows the PRECIS 2 tool.

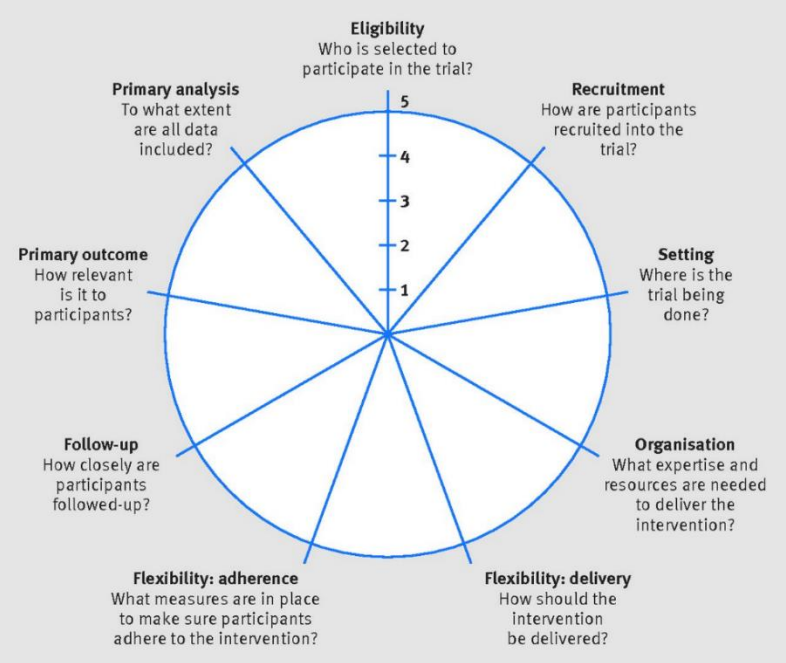

Figure 9: The PRECIS 2 tool

When you use the PRECIS 2 tool, if your trial is focused on being highly pragmatic, then it should look like a beautiful spider's web. Figure 10 below shows an almost perfectly designed pragmatic trial.

[13] https://youtu.be/NfJQ9rX1acs
[14] https://www.precis-2.org/

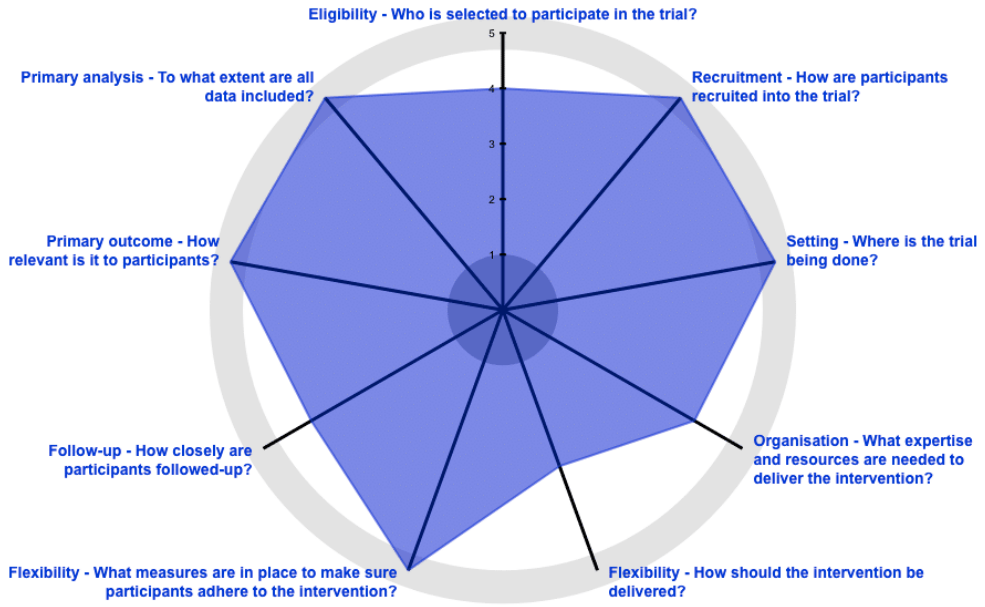

Figure 10: Completed PRECIS 2 wheel

Read the following article by Professor Merrick Zwarrenstein, one of the original authors of PRECIS 2. Here he elucidates on the history of pragmatic and explanatory attitudes to trials.[15]

6 The SPIRIT 2013 Statement

The SPIRIT (Standard Protocol Items: Recommendations for Interventional Trials) 2013 Statement provides evidence-based recommendations for the minimum content of a clinical trial protocol. SPIRIT is widely endorsed as an international standard for trial protocols.

6.1 Background

The trial protocol provides guidance to individuals conducting the study, serves as the basis for trial registration, and facilitates study appraisal by participants and external reviewers, including research ethics committees/institutional review boards, funders, regulators, journal editors, and systematic reviewers. However, there is accumulating evidence that many protocols do not fully address important study elements. Incomplete protocol content can impair understanding and implementation of the trial; reduce efficiency of protocol review; and lead to burdensome protocol amendments.

In 2007, an international group of stakeholders (the SPIRIT Group) launched the SPIRIT initiative to help improve the completeness and quality of trial protocols. The evidence-based SPIRIT recommendations were developed using systematic, transparent methodology and

[15] Zwarenstein M. 'Pragmatic' and 'explanatory' attitudes to randomised trials. Journal of the Royal Society of Medicine. 2017;110(5):208-218.

broad consultation with 115 experts representing diverse stakeholders involved in the design, funding, conduct, review, and publication of trial protocols.

The final products consist of the SPIRIT 2013 Statement and the accompanying Explanation & Elaboration paper, which were published in January 2013. SPIRIT 2013 builds on other applicable international guidance by citing empirical evidence and providing additional recommendations. It adheres to the ethical principles mandated by the 2008 Declaration of Helsinki, and encompasses the protocol items recommended by the International Conference on Harmonisation Good Clinical Practice E6 guidance. SPIRIT 2013 also supports trial registration requirements from the World Health Organization and the International Committee of Medical Journal Editors.

6.2 The SPIRIT Checklist

This consists of a 33 item checklist. Table 4 (taken from the SPIRIT website)[16] gives the title with an example of what is expected for each response.

Reading
The SPIRIT Checklist consists of a 33-item checklist. Go through the document which gives the title with an example of what is expected for each response.[17]

6.3 The main paper

Full details of the development of SPIRIT and a detailed explanation of the SPIRIT checklist can be found in the following published paper: Chan A-W, Tetzlaff JM, Altman DG, Laupacis A, Gøtzsche PC, Krleža-Jerić K, Hróbjartsson A, Mann H, Dickersin K, Berlin J, Doré C, Parulekar W, Summerskill W, Groves T, Schulz K, Sox H, Rockhold FW, Rennie D, Moher D. SPIRIT 2013 Statement: Defining standard protocol items for clinical trials. Ann Intern Med. 2013;158:200-207.[18]

A website is also dedicated to SPIRIT with many online resources for filling in the checklist, including some excellent examples. A website is also dedicated to SPIRIT with many online resources for filling in the checklist, including some excellent examples.[19]

7 PICO, Equipoise and the Research Question

7.1 The Research Question

Formulating a research question is essential before beginning any research. The planning of a clinical trial depends on this question however, the specific question to be answered by the trial is often not well stated. Most clinical trials are, and should be, designed to answer a single question. This can vary but it is useful to think of trials in these terms. A researchable question is an uncertainty about a problem that can be challenged, examined, and analysed to provide useful information. A successful research project depends upon how well an investigator

[16] https://www.spirit-statement.org/
[17] https://www.spirit-statement.org/wp-content/uploads/2013/08/SPIRIT-Checklist-download-8Jan13.doc
[18] https://doi.org/10.7326/0003-4819-158-3-201302050-00583
[19] https://www.spirit-statement.org/

formulates the research question based on the problems faced in day-to-day research activities and clinical practice. A well-formulated research question needs extreme specificity and preciseness which guides the implementation of the project, keeping in mind the identification of variables and population of interest.

Listening
Listen here to Fiona Godlee, editor of the BMJ as she details why the BMJ rejects paper ("Research to Publication - making medical research better"[20]). Notice her reference to the research questions.

7.2 Clinical Equipoise

Clinical equipoise is a state of disagreement or uncertainty in the informed, expert medical community about the relative clinical merits of the intervention arms in a trial. In other words, the experts cannot agree on the best treatment, for example, so they decide to test it in a RCT. The principle of equipoise reconciles two potentially conflicting ethical imperatives: to ensure that research involving human participants generates scientifically sound and clinically relevant information while demonstrating proper respect and concern for the rights and interests of study participants. The ethics of clinical research requires equipoise. Should the investigator discover that one treatment is of superior therapeutic merit, he or she is ethically obliged to offer that treatment.

Two features of medical research pose special challenges for the goal of ensuring respect and concern for the rights and interests of participants. First, to generate reliable information, research often involves design features that alter the way participants are treated. For example, randomisation and blinding are commonly used to reduce selection bias and treatment bias. Controlling how interventions are allocated and what researchers and participants know about who is receiving which interventions helps to more clearly distinguish the effects of the intervention from confounding effects. But randomisation severs the link between what a participant receives and the recommendation of a treating clinician with an ethical duty to provide the best possible care for the individual person. Second, medical research involves exposing people to interventions whose risks and potential therapeutic, prophylactic, or diagnostic merits may be unknown, unclear, or the subject of disagreement within the medical community.

The principle of equipoise states that if there is uncertainty or conflicting expert opinion about the relative therapeutic, prophylactic, or diagnostic merits of a set of interventions, then it is permissible to allocate a participant to receive an intervention from this set, so long as there is not consensus that an alternative intervention would better advance that participant's interests.

Listening
Watch this short video on equipoise "4. Equipoise".[21]

[20] https://youtu.be/A8EINvlnktw
[21] https://youtu.be/E8dnHMqk1-0

However, there is another argument in relation to equipoise in clinical trials. The requirement which entails that the investigator have no "treatment preference" throughout the course of the trial, presents nearly insurmountable obstacles to the ethical commencement or completion of a controlled trial and may also contribute to the termination of trials because of the failure to enroll enough patients.

Equipoise has its origin in medical ethics and the law. The principles of medical ethics were set forth in the Nuremberg Code[22] and subsequently enunciated as principles having to do with the need for beneficence, competence and justice.

Beneficence - A principle in medical ethics that asserts that the options available in treating or caring for human beings is limited to that set which is justifiable on the basis of beneficence. Translated to research settings involving human beings, the principle means that only those acts, procedures and treatments that meet this test are justifiable.

Competence - A principle in medical ethics that asserts the care and treatment performed or offered in a research setting must be offered and administered in a competent fashion, consistent with accepted standards of care.

Justice - The maintenance or administration of what is just and fair, especially by impartial methods and procedures. This is an elusive consent. Often that which is seen as being just from one perspective is seen as unjust from a different perspective. The issue of justice in research involving human beings arises in regard to selection on both an individual and societal level. The Belmont Report states, "Justice is relevant to the selection of participants of research at two levels: the social and the individual. Individual justice in the selection of participants would require that researchers exhibit fairness: thus they should not offer potentially beneficial research only to some patients who are in the favour or select only 'undesirable' persons for risky research. Social justice requires that distinction be drawn between classes of participants that ought, and ought not, to participate in any particular kind of research, based on the ability of members of that class to bear burden and on the appropriateness of placing further burdens on already burdened persons. Thus it can be considered a matter of social justice that there is an order of preference in the selection of classes of participants, e.g., adults before children, and that some participants e.g., institutionalised or mentally infirm, may be involved as research participants if at all only on certain conditions.

7.3 PICOT(S)

PICO, PICOT or PICOTS (all pronounced PICO) is used to help us describe our research in a concise manner. Table 4 summarises what each represents.

Table 4: PICOT(S) description

PICOTS		
P	Population	Patient or the problem to be addressed
I	Intervention	Exposure to be considered-treatments/tests
C	Control	Control or comparison intervention treatment/placebo/standard of care
O	Outcome	Outcome of Interest

[22] CONSCIOUS: Chapter 1, Clinical Research/Trials; http://conscious.novaims.unl.pt/my/

T	Time	Timing of the outcome assessment or when follow-up will occur or for how long the intervention will occur
S	Setting	Setting where the study takes place

Practical Exercise: applying PICOT(S)

A 2-year-old boy presents in an outpatient clinic with fever and severe pain in his right ear. He has a history of recurrent ear infections, and his mother expresses a concern that he has been on the antibiotic amoxicillin for the past few weeks. She is worried about the consequences of the long-term antibiotic use. She is also concerned about the outcome associated with recurrent ear infections. She wants to know if the prescribed amoxicillin is effective, or it can be substituted with another antibiotic because of its side effects such as frequent diarrhoea. Several questions arise from this case which can be broadly classified into background and foreground questions. The general questions about a clinical problem or a disease are called “Background Questions". These questions generally ask what, when, how, and where about the disease, disorder, or treatment for instance, “What is otitis media?” or “How does amoxicillin work?” etc. These types of questions can be answered by going through review articles or text books.

The patient-oriented questions involving interpretation of a therapy or disease and consideration of risk vs. benefit for a patient or a group of patient are called “Foreground Questions.” These types of complex clinical questions are best answered by primary or pre-assessed studies in the literature. These questions mostly compare the two, either two drugs or treatments or two diagnostic methods, etc.

The PICO (population, intervention, control, and outcomes) format is considered a widely known strategy for framing a “foreground” research question. Breaking the question into four, five or six components will facilitate the identification of relevant information.

Further information on PICOT(S) can be found here.[23]

Answer to activity

Population: addressing a specific population, its important characteristics and demographic information. From the above case, you can identify paediatric population with otitis media, the age range, sex, presenting complaint, and history.

Intervention: the intervention can be a treatment, procedure, diagnostic test, and risk or prognostic factors. In this case, the intervention will be your plan to treat the patient which can be a new therapy, a diagnostic test, prognostic factor, or a procedure. For example, based on your observation in clinic, cefuroxime is another better treatment option as compared to amoxicillin in treating otitis media but you are not sure about its efficacy in paediatric population with otitis media.

Comparator or control: when a new therapy is compared with the existing one.

Outcome: the effect of the intervention. For example, its effectiveness in controlling pain. Therefore, the outcome in the above case can be the relief of pain, the resolution of infection, or decreasing the risk of developing resistance. A good primary outcome should be easily quantifiable, specific, valid, reproducible, and appropriate to your research question.

[23] https://www.cochranelibrary.com/about/pico-search

8 Eligibility Criteria and Outcome Data

All clinical trials require a precise definition of which participants are eligible for inclusion. The early stages of protocol development may proceed with only a rough outline of the intended type of patient, but before the trial gets underway, this must be transformed into a detailed specification.

The main objective is to ensure that participants in the trial may be identified as representative of some future class of participants to whom the trial's findings may be applied. In addition, one wishes to focus on the type of patient considered most likely to benefit from the new treatment under investigation. However, one does not wish to be so restrictive about participant entry that the trial remains small and its findings lack generality. A study should define the study population in advance, stating unambiguous inclusion (eligibility) criteria. The impact that these criteria will have on study design, ability to generalise, and participant recruitment must be taken into account.

There are three considerations when considering inclusion and exclusion criteria

1. The source of participant recruitment
2. The disease state under investigation
3. The specific criteria for exclusion of participants

8.1 The Source of the Participants

The issue of representatives needs to be considered carefully. For instance, in the study of depressive illness if one recruits hospital in-patients one ends up with an atypical group. Likewise, in a study on excessive alcohol consumption, if it were to take place in the Emergency Department of a hospital, this setting is likely to bias the study because people who consume alcohol excessively, can often end up in the emergency room. So when we are conducting a full-scale trial then we must try and aim for a group of patients that truly represent the disease under investigation, even if this restricts the extent of patient evaluation, so that other clinicians can relate the trial’s conclusions to future patients in their clinical practice.

8.1.1 Specific Criteria for Exclusion of Participants

All trials will have exclusion criteria as well as inclusion criteria. These are supplements to the main definition of the disease. However, one should avoid making requirements too stringent since one might then have difficulty finding enough participants, causing those participants in the trial to be an unduly select group. For instance, it is common practice to exclude elderly participants, say over 65. Such patients may be less responsive to therapy, more affected by side-effect or more difficult to evaluate properly. However, if the best of the therapies under investigation are liable to be used on future elderly participants, then it may be wise to include them in the trial. In general, one needs to strike a balance between including all participants who may potentially benefit from trial therapy and aiming for a more select group of participants who are most suited to the trial's purpose.

8.1.2 Inclusivity in Trials

Be aware also of restrictive inclusion criteria. For instance, in countries where English is the first language, it is common to have an inclusion criterion of “must speak English”. This will

naturally exclude those whose first language is not English which is usually minority groups. Other common exclusion criteria are pregnancy, vulnerable populations, participants with co-morbidities. We must ask ourselves then if we are really designing pragmatic trials and trials that represent the communities we are targeting.

For more information on inclusivity in trials go to trialforge.org.[24]

8.2 Outcomes

A crucial issue is how to measure whether or not something works using a good measure of outcome. The choice of outcome measure will clearly depend upon the research question. An ideal outcome measure will be sensitive to important effects, reliable, in that it will return the same findings when participants are re-measured in the same circumstances, and valid, i.e., the outcome instrument will give us an accurate assessment of the actual outcome we wish to measure. This is important because many outcomes that are measured are not 'true' outcome measures. For example, quality of life measurement do not truly measure a person's quality of life - they only give a general indication. Objective measures of outcomes such as death may not give us the complete picture either.

Read the following: Clarke M. Standardising outcomes for clinical trials and systematic reviews. Trials. 2007 Dec;8(1):1-3.[25] Though an old article now, it sets the scene for what was to come and the current programme of research in outcomes.

8.2.1 Selecting the Outcome for the Trial

Outcomes should be meaningful for those people who may make decisions about the intervention in a trial. The outcomes can be clinical measures; health related quality of life; disease specific measures; and resource use. Trials usually have one primary outcome, on which sample size is calculated; and secondary outcomes. The choice of outcome measure is one of the most important decisions you will take in your trial. The question you need to ask yourself when choosing outcome measures is whose decision am I trying to support with my trial? Those people are the people who drive your choice of outcome. Of course you also need to keep your research question and objective in mind.

So how do we choose the primary outcome, especially when it is so important to our trial. The first thing to do is look at the literature. What primary outcome measures have other people used when conducting trials in the same area? The reason that it's a great idea to use outcome measures that other have used is because it means you can compare your outcomes with their outcomes, i.e., the trials can be compared.

An organisation called COMET does exactly that. It specialises in choosing relevant outcomes for clinical trials. COMET stands for **C**ore **O**utcome **M**easures in **E**ffectiveness **T**rials. Learn more at the Comet Initiative website.[26]

[24] https://www.trialforge.org/trial-forge-centre/include/
[25] https://trialsjournal.biomedcentral.com/articles/10.1186/1745-6215-8-39
[26] https://www.comet-initiative.org/

If you don't think that the literature supports a choice of outcome measure for the decision you are trying to support, then a great alternative would be to ask the people that your study will affect. Talk to patients. Talk to the doctors. Talk to policymakers, if it's a policy level decision. And ask them, what would be relevant primary outcome for my trial to support your decision, and then measure that. Remember, you and not trying to convince yourself. You are trying to convince somebody else - a doctor, a patient, a policymaker.

8.2.2 Secondary Outcome Measures

Trials generally collect several other outcomes, and these are called secondary outcomes. These are not as important as the primary, but provide extra data from your trial. And here's a top tip with regard to secondary outcomes. People often collect far too many. There's a great study that looks at eight cancer trials, and they asked a great question. How much of the data that were collected in those trials actually made it out into the public domain so that other people could learn from the data that were collected, where it can actually increase the evidence base behind decisions. So they counted the items that were collected in those trials, and then looked to publications that came from those trials to see how much data made it into the public domain and how much did not make it into the public domain? Well, they calculated a median, and they found that that median, unused data, was 82%. Think of all the work that everybody involved in that trial did to collect the data? This data was not used so it is research waste. It is a waste of money, time and effort. We also know that collecting huge amounts of information in trials can be off putting for patients, so potentially, collecting all this outcome data could affect your recruitment to your trial. Additionally, we are adding to participants' burden and risk as the participants are the ones that take the risks when participating in trials. So with regard to secondary outcomes, choose them very wisely, and collect only those that will really support the scientific question.

8.2.3 Timing of Outcome Assessment

Once it has been decided what to measure, it is important to consider when outcomes should be measured. Typically, the period of follow-up should be sufficient to ensure that the wider range of effects, other than those that are evident immediately after the intervention is introduced, are observable. If quality of life measures are to be included, it is also important to consider whether a baseline assessment should be included (required if the researcher wishes to evaluate any changes in a participant's well-being during or after treatment). The timing of each outcome measurement should be made explicit in the protocol of the trial.

8.2.4 Surrogate Outcomes/Endpoints

We tend to use the word endpoint and outcome interchangeably in clinical trials but they are not the same. Keep this in mind when you are reading any literature. A surrogate endpoint is a physical measurement of a specific outcome which is considered to be a valid predictor (or representative) of the real outcome or final result. In simple words, a surrogate endpoint is like a measurable indicator that can help us know what the real result is. One example of this is cholesterol levels and the risk of having a heart attack. Generally, a high level of cholesterol in the bloodstream indicates a greater risk of having a heart attack. Thus, measuring higher cholesterol levels (which is the surrogate endpoint) allows us to predict that the patient has a greater risk of suffering from heart attacks (which is the real outcome). Another example is blood pressure and the incidence rate of heart attacks and strokes. Evidence shows that having high blood pressure increases the risk of suffering from strokes or from heart attacks. In this case, blood pressure is the surrogate endpoint as a higher blood pressure is a predictor

of a greater risk of strokes and heart attacks. In another example,
having an irregular heart beat (the surrogate endpoint) suggests that there is a greater risk of suffering from a sudden cardiac death (the real outcome). In other words, surrogate endpoints are useful since they can, in some situations, be representatives of the final clinical outcomes.

8.2.5 Benefits of Surrogate Endpoints

Surrogate endpoints are very appealing to use in research for several reasons. Firstly, measuring real outcomes (unlike measuring surrogate outcomes) is often very time-consuming. For instance, it could take researchers several years to collect enough information to determine whether a new cholesterol lowering drug significantly reduces the incidence rate of heart attacks in the group being tested. It would be easier, instead, to measure cholesterol levels in patients a few weeks after starting the clinical trial and to judge the efficacy of the drug based on the degree of reduction in the cholesterol levels. Secondly, using surrogate outcomes reduces costs for manufacturers since it reduces the duration of the clinical trial which means the new intervention could be introduced to the market faster. In addition to this, surrogate endpoints could, on some occasions, be useful since they could give more balanced findings than using real outcomes. For instance, when testing a new painkiller, measuring the related surrogate endpoints (before and after administering the new painkiller) would probably give a less subjective finding than performing pain assessment on the patient. This means that, on some occasions, it would be better to consider and assess both the surrogate and real outcomes.

Despite these benefits, it is important to remember that using surrogate endpoints might be problematic on some occasions since they are not always good representatives of the real clinical outcome. In a general sense, a surrogate outcome would only be a good representative of the real outcome if the surrogate endpoint itself is the sole or major contributor to the progression of the disease (or disorder) towards the real endpoint.

8.2.6 Composite Outcomes

A composite outcome consists of two or more component outcomes. Patients who have experienced any one of the events specified by the components are considered to have experienced the composite outcome. The main advantages supporting the use of a composite outcome are that it increases statistical efficiency because of higher event rates, which reduces sample size requirement, costs, and time; it helps investigators avoid an arbitrary choice between several important outcomes that refer to the same disease process; and it is a means of assessing the effectiveness of a patient reported outcome that addresses more than one aspect of the patient's health status.

Unfortunately, composite outcomes can be misleading. This is especially true when treatment effects vary across components with very different clinical importance. For example, suppose a drug leads to a large reduction in a composite outcome of "death or chest pain." This finding could mean that the drug resulted in fewer deaths and less chest pain. But it is also possible that the composite was driven entirely by a reduction in chest pain with no change, or even an increase, in death.

Studies show that treatment effects often vary, and typically, the effect is smallest for the most important component and biggest for the less important components. Unless authors clearly present data for all components and take care in how they discuss composite findings, it is easy for readers to assume mistakenly that the treatment effect applies to all components.

8.2.7 Qualitative Outcomes

Generally trials use quantitative measures to assess outcomes. However, it is also possible to use qualitative outcomes. Qualitative research may be conducted concurrently with pilot or full RCTs to understand the feasibility and acceptability of the interventions being tested, or to improve trial conduct.

For more information on qualitative research within trials the QUESTS group at the University of Galway specialise in quests.ie.[27]

A great read is also this article by Cooper C, O'Cathain A, Hind D, Adamson J, Lawton J, Baird W. Conducting qualitative research within Clinical Trials Units: avoiding potential pitfalls. Contemporary Clinical Trials. 2014;38(2):338-343.[28]

9 Sample Size, Randomisation

9.1 Sample Size

Sample size determination is an essential step in planning a clinical study. It is critical to understand that different study designs need different methods of sample size estimation. Although there is a vast literature discussing sample size estimation, incorrect or improper formulas continue to be applied. The best advice when it comes to needing to know how many participants is appropriate for your trial to answer your research question is - consult a trial statistician. If you are a statistician reading this, you will know that you want this engagement as early as possible in the trial design process. A very common error researchers and trialists make is to consult the statistician too late in the process. In brief though, what information is necessary to conduct a sample size calculation. Many software packages are available for this but you need to know:

1. The research question or research hypothesis.
2. The significance level of the test, i.e., the alpha level, typically 0.05.
3. The standard deviation (can be unknown but may be able to use one from a pilot study or prior study).
4. The power of the test. Usually you will see 80% power (a one in five chance of not rejecting the null hypothesis) at a minimum but 90% power is better.
5. The effect size (either by specifying the minimum clinically important difference (MCID) or a realistic difference based on prior evidence and information).
6. The intracluster correlation (if it is a cluster randomised trial).

Trials are all about teams. Consult a statistician but know the above parameters so you can discuss them with the statistician.

9.2 Randomisation

The main difference between other study design methods and the randomised controlled trial is that in the latter two or more groups are formed by random allocation. Randomisation is the best approach to dealing with and controlling for selection bias and temporal changes. Other

[27] https://quests.ie/
[28] https://doi.org/10.1016/j.cct.2014.06.002

terms such as blinding, or the use of placebos, may be associated with some types of randomised trial but their use is neither a necessary nor a sufficient condition for a study to be identified as a randomised trial.

Read the following half page article by Professor Doug Altman. He is recently deceased but has left behind a legacy. Some of you will have seen his books and some of you might know him for his Altman and Bland plots. In any event, he gives a nice summary of randomisation back in 1991 in the BMJ. Don't be disappointed that it is old. Some methods are still relevant today and sometimes the original version is superior.

Altman DG. Randomisation. British Medical Journal. 1991 Jun 6;302(6791):1481[29]

9.2.1 Randomisation Defined

- Random allocation means that all participants have a defined probability, or an equal chance, of assignment to a particular intervention.
- Allocation is not determined by the investigator, clinicians, or participants.
- Allocation is not predictable based on a pattern.

The method used to assign interventions to trial participants is a crucial aspect of clinical trial design. Random assignment is the preferred method. It has been successfully used regularly in trials for more than 50 years. Randomisation has three major advantages.

First, when properly implemented, it eliminates selection bias, balancing both known and unknown prognostic factors, in the assignment of treatments. Without randomisation, treatment comparisons may be prejudiced, whether consciously or not, by selection of participants of a particular kind to receive a particular treatment.

Second, random assignment permits the use of probability theory to express the likelihood that any difference in outcome between intervention groups merely reflects chance.

Third, random allocation, in some situations, facilitates blinding the identity of treatments to the investigators, participants, and evaluators, possibly by use of a placebo, which reduces bias after assignment of treatments. Of these three advantages, reducing selection bias at trial entry is usually the most important.

9.2.2 Sequence Generation and Allocation Concealment

Successful randomisation in practice depends on two interrelated aspects - adequate generation of an unpredictable allocation sequence and concealment of that sequence until assignment occurs. A key issue is whether the schedule is known or predictable by the people involved in allocating participants to the comparison groups. The treatment allocation system should thus be set up so that the person enrolling participants does not know in advance which treatment the next person will get, a process termed allocation concealment. Proper allocation concealment shields knowledge of forthcoming assignments, whereas proper random sequences prevent correct anticipation of future assignments based on knowledge of past assignments.

[29] https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1670173/

9.2.3 What purpose is served by random allocation?

- Covariates are distributed equally across the groups at baseline (Not always (especially if N is small)).
- Affects both measured and unmeasured variables
- The risk of imbalance remains even after properly executed randomisation
- The first table presented in most RCTs will provide a comparison of treatment and comparison groups (sometimes with p-values though there is debate about this in the literature. North American journals allow p-values to be presented with it's associated statistical test, but European journals do not. Some researchers in Europe are changing their views and say that presenting the p-values may be a useful way of identifying subversion bias.
- If randomisation has been performed correctly, chance is the only explanation for any observed difference between groups, in which case statistical tests are considered superfluous

9.2.4 Allocation Sequence

The first part of any randomisation schedule is the allocation sequence. This is a list of intervention groups, randomly ordered, used to assign sequentially enrolled participants to intervention groups. Also termed the "assignment schedule", "randomization schedule", or "randomization list" (CONSORT Statement). There are many ways to do this such as simple randomisation, block randomisation or stratified randomisation. There are other methods also but again, we recommend you work with a statistician.

9.2.5 Simple Randomisation

The easiest method is simple randomization. If you assign participants into two groups A and B, you assign participants to each group purely randomly for every assignment. Even though this is the most basic way, if the total number of samples is small, sample numbers are likely to be assigned unequally. For this reason, we recommend you to use this method when the total number of participants is more than 100.

9.2.6 Block Randomisation

We can create a block to assign sample numbers equally to each group and assign the block.

If we specify two in one block (the so-called block size is two), we can make two possible sequences of AB and BA. When we randomize them, the same sample numbers can be assigned to each group. If the block size is four, we can make six possible sequences; these are AABB, ABAB, ABBA, BAAB, BABA, BBAA, and we randomize them.

However, there is a disadvantage in that the executer can predict the next assignment. We can easily know the fact that B comes after A if the block size is two and if the block size is four; we can predict what every 4th sample is. This is discordant with the principle of randomization. To solve this problem, the allocator must hide the block size from the executer and use randomly mixed block sizes. For example, the block size can be two, four, and six.

9.2.7 Stratified Randomization

Randomization is important because it is almost the only way to assign all the other variables equally except for the factor (A and B) in which we are interested. However, some very important confounding variables can often be assigned unequally to the two groups. This

possibility increases when the number of samples is smaller, and we can stratify the variables and assign the two groups equally in this case.

For example, if the smoking status is very important, what will you do? First, we have two methods of randomization that we learned previously. There are two randomly assigned separate sequences for smokers and non-smokers. Smokers are assigned to the smoker's sequences, and non-smokers are assigned to the non-smoker's sequences. Therefore, both smokers and non-smokers groups will be placed equally with the same numbers.

So we can use 'simple randomization with/without stratification' or 'block randomization with/without stratification.' However, if there are multiple stratified variables, it is difficult to place samples in both groups equally with the same numbers. Usually two or fewer stratified variables are recommended.

9.2.8 Allocation Concealment

As vital as generating the allocation sequence is, concealing it is as important. In fact some trialists say that if they were asked to choose between the two, concealing the allocation is the most important. Without adequate allocation concealment, even random, unpredictable assignment sequences can be undermined. Knowledge of the next allocation in the sequence can disrupt the unbiased allocation of patients. Perhaps a doctor has a very sick patient, the he/she feels will really benefit from the new drug on the trial, if he/she knows the next allocation, he/she would be able to direct the patient to the intervention arm. However, this will of course bias the results. Avoidance of such bias depends on the prevention of foreknowledge of treatment assignment.[30]

If those making the decision about patient eligibility are aware of the arm of the study to which the patient will be allocated - if randomisation is unconcealed - they may systematically enrol sicker - or less sick - patients to either treatment or control groups. This will defeat the purpose of randomisation and the study will yield a biased result. Inadequate allocation concealment leads to exaggerated estimates of treatment effect, on average, but with scope for bias in either direction.

Example

RCT of open vs laparoscopic appendectomy

- Trial ran smoothly during the day.
- At night, however, the attending surgeon's presence was required for the laparoscopic procedure but not the open one; and the limited operating room availability made the longer laparoscopic procedure an annoyance.
- Reluctant to call in a consultant, and particularly reluctant with specific senior colleagues, the residents sometimes adopted a practical solution.
- When an eligible patient appeared, the residents checked the attending staff and the line up for the operating room and, depending on the personality of the attending surgeon and the length of the line up, held the translucent envelopes containing orders up to the light.
- As soon as they found one that dictated an open procedure, they opened that envelope. The first eligible patient in the morning would then be allocated to a laparoscopic appendectomy group according to the passed-over envelope.

[30] https://pubmed.ncbi.nlm.nih.gov/11867132/

- If patients who presented at night were sicker than those who presented during the day, the residents' behaviour would bias the results against the open procedure.

10 Blinding

Blinding, also referred to as masking, refers to the concealment of group allocation from one or more individuals involved in the research study. In practice, that means that if you're taking part in a trial, you will not know what treatment arm you have been allocated to. Often, your doctor or healthcare professional will not know either. There are open label, single-blind, double blind and triple blind studies. It is not to be confused with allocation concealment.

Definition from the CONSORT Statement "The practice of keeping the trial participants, care providers, those collecting data, and sometimes even those analyzing data unaware of which intervention is being administered to which participant. Blinding is intended to prevent bias on the part of study personnel. The most common application is "double-blinding", in which participants, caregivers and those assessing outcome are blinded to intervention assignment. The term "masking" may be used instead of "blinding".

10.1 What Does Blinding Prevent?

- Performance bias - different response to treatment
- Ascertainment bias - when the results or conclusion of a trial are systematically distorted by knowledge of which intervention each participant is receiving
- Observer bias - avoids situations where the observer 'sees' an imaginary benefit for those participants treated with the observer's preferred treatment

However, blinding is not always possible. In an ideal world every study would be triple blind - participants, clinicians and researchers would all be blind to the treatment that the participant has been allocated to. The world isn't ideal though, and lots of the trials that are going on involve complex interventions (i.e. not something as simple as a tablet that you can easily duplicate the look and feel of to ensure allocation remains concealed). Some trials are only able to run if they are single blinded, or completely unblinded - surgical trials for example. Innovative trial designs and techniques are often incorporated in an effort to overcome potential bias in these situations. For example, treating patients according to a strict protocol to reduce the risk of differential behaviours by patients and healthcare providers. An attempt to blind participants and personnel does not ensure successful blinding in practice. Blinding can be compromised for most interventions. For many blinded drug trials, the side effects of the drugs allow the possible detection of which intervention is being received for some participants, unless the study compares two rather similar interventions, e.g. drugs with similar side effects, or uses an active placebo.[31]

[31] Boutron I, Estellat C, Guittet L, Dechartres A, Sackett DL, Hróbjartsson A, Ravaud P. Methods of blinding in reports of randomized controlled trials assessing pharmacologic treatments: a systematic review. PLoS Medicine. 2006;3(10): e425.

Example
In a study by Howard, J et al. Clin Pharmacol Ther. 1982;32(5):543-53, the AMIS study (Aspirin Myocardial Infarction Study), which is well known amongst cardiologists, the study examined aspirin versus placebo and the outcome was survival for 3-4 years after myocardial infarction. 95 / 285 (33%) deliberately tested the capsule. 67% of testers guessed the treatment correctly compared to 47% of non-testers.

10.2 What's the Difference Between Allocation Concealment and Blinding?

Concealment of allocation

- Procedure to protect randomisation process before the participant enters the trial
    - Failed concealment from the investigator or clinician
    - Failed concealment from the patient
- Concealment of allocation is ALWAYS feasible
- If not done, results in selection bias. The randomisation benefits are lost, and treatment assignment is no longer truly random

Blinding

- Masking of the treatments after randomisation (once trial begins)
    - Failed masking of patients, investigators, outcome assessors, etc.
- Blinding is NOT always feasible
- If not done, can result in patients biasing their responses because of their knowledge of treatment; can also lead to biased outcome assessment because investigators have knowledge of treatment

11 Conclusion

This concludes the chapter on study design. We have covered many new concepts in detail. Take the opportunity to read further on these topics by exploring the literature in reputable journals. This chapter is just the beginning of your clinical trial journey.

Co-funded by the
Erasmus+ Programme
of the European Union