Chapter 4
Alternative Interventional Study Designs

Stephen P. Glasser

A man who does not habitually wonder is but a pair of spectacles behind which there is no eye

(Thomas Carlyle) [1]

Abstract There are many variations to the classical randomized controlled trial. These variations are utilized when, for a variety of reasons, the classical randomized controlled trial would be impossible, inappropriate, or impractical. Some of the variations are described in this chapter and include: equivalence and non-inferiority trials; crossover trials; N of 1 trials, case-crossover trials, and externally controlled trials. Large simple trials, and prospective randomized, open-label, blinded endpoint trials are discussed in another chapter.

Keywords Equivalence/noninferiority testing • Superiority testing • PROBE design • Factorial design • Assay sensitivity • Consistency assumption • N of 1 trial • Crossover design • Case-crossover design • Adaptive design • Registry randomized control trial • Null hypothesis

There are a number of variations of the 'classical' RCT design. For instance, many view the classical RCT as having an exposure group compared to a placebo control group, using a parallel design, and a 1:1 randomization scheme. However, in a given RCT, there may be several exposure groups (e.g. utilizing different doses of the drug under study), and the comparator group may be an active control rather than a

S.P. Glasser, M.D. (✉)
Division of Preventive Medicine, University of Alabama at Birmingham,
1717 11th Ave S MT638, Birmingham, AL 35205, USA
e-mail: sglasser@uabmc.edu

S.P. Glasser (ed.), *Essentials of Clinical Research*, DOI 10.1007/978-3-319-05470-4_4,

placebo control; and, some studies may have both. By an active control, it is meant that the control group receives an already approved intervention. For example, a new anti-hypertensive drug could be compared to placebo or could be compared to a drug already approved by the FDA and used in the community (frequently, in this case, the manufacturer of the investigational drug will compare their drug to the currently most frequently prescribed drug for the indication of interest). The decisions regarding the use of a comparator are based upon a number of considerations and discussed more fully under the topic entitled equivalence testing. Also, the randomization sequence may not be 1:1, particularly if (for several reasons, ethical issues may be one example) one wanted to reduce the number of subjects exposed to placebo. Also, rather than parallel groups, there may be a titration schema built into the design. On occasion, the study design could incorporate a placebo withdrawal period in which at the end of the double blind comparison, the intervention group is subsequently placed on placebo (this can be done single-blind or double-blind). In this latter case, retesting 1 or 2 weeks later occurs with comparison to the original placebo group. Other common variants to the classical RCT are discussed in more detail below.

Traditional Versus Equivalence/Non-inferiority Testing (See Tables 3.6 in Chap. 3 and 4.1 in This Chapter)

As discussed in Chap. 3, most clinical trials have been designed to assess if there is a difference in the efficacy to two (or more) alternative treatment approaches (with placebo ideally being the comparator treatment). Consider the fact that for evidence of efficacy there are two distinct approaches: to demonstrate a difference-showing superiority of the investigational drug to control (placebo, active, lower dose) which then demonstrates the drug effect; or, to show equivalence or non-inferiority to an active control (i.e. the investigational drug is of equal efficacy or not worse than an active control). That is, one can attempt to demonstrate that there is similarity to a known effective therapy (active control) and attributing the efficacy of the active control drug to the investigational drug, thereby demonstrating a drug effect (i.e. equivalence). Since nothing is perfectly equivalent, equivalence means within a margin predetermined by the investigator (termed the equivalence margin). Non-inferiority trials on the other hand aim to demonstrate that the investigational drug is not worse than the control, but once again by a defined amount (i.e. not worse by a given amount – the non-inferiority margin), the margin (M or δ) being that amount no larger than the effect the active control would be expected to have in the study. As will be discussed later, this margin is not easy to determine and requires clinical judgment; and, this represents one of the limitations of these kinds of trials [2]. These aforementioned approaches are presented diagrammatically in Fig. 4.1a–c.

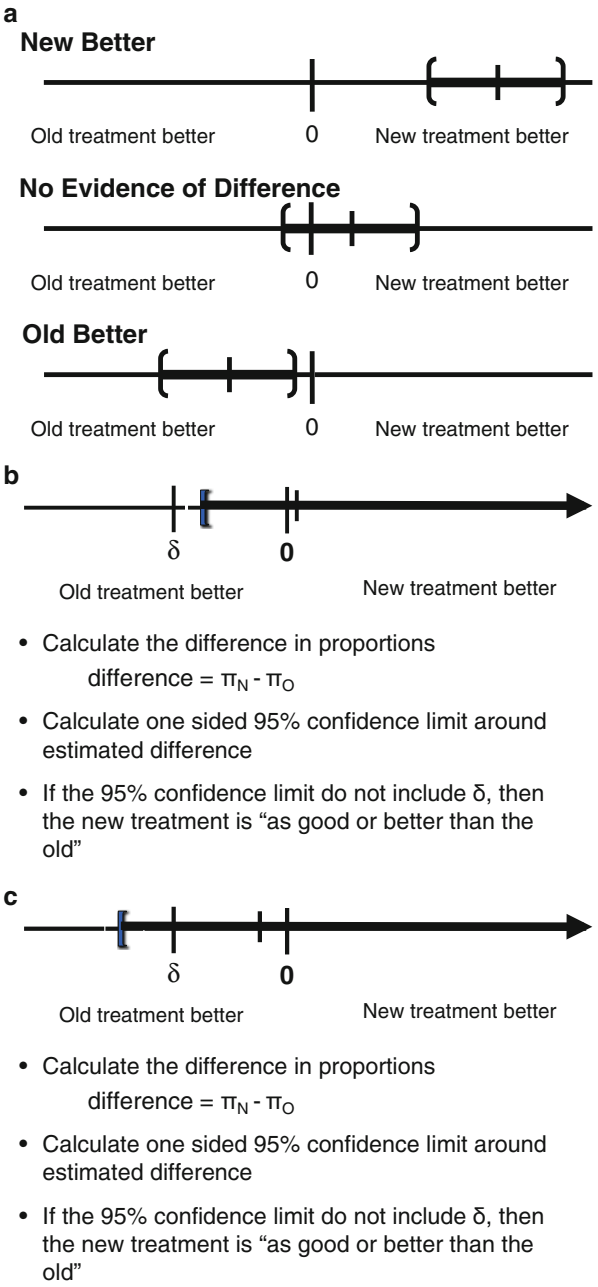

Fig. 4.1 (**a**) Outcomes for traditional (superiority) testing. (**b**) Outcomes for equivalence testing. Since the lower confidence bound is not beyond theta, the null has not been rejected. (**c**) Outcome in equivalence testing. Since the lower confidence bound is beyond theta, the null is rejected and therefore the two treatments are "equivalent"

As also discussed in Chap. 3, there are a number of reasons for the increased interest in equivalence and non-inferiority trials including the ethical issues associated with placebo controls. In general, for studies of efficacy, placebo-controls are preferable to active controls, due to the placebo's ability to distinguish an effective treatment from a less effective treatment. The ethical issues surrounding the use of a placebo-control aside, there are other issues that have led to the increasing interest and use of equivalence and non-inferiority studies. For example, clinical trials are increasingly being required to show benefits on clinical endpoints rather than on surrogate endpoints at the same time that the incremental benefit of new treatments is getting smaller. This has led to the need for larger, longer, and more costly trials; and, this has resulted in the need to design trials that are less expensive. Additional issues are raised by the use of equivalence/non-inferiority trials, such as assay sensitivity, the aforementioned limitations of defining the margins, and the constancy assumption.

Assay Sensitivity

Assay sensitivity is a property of a clinical trial defined as the ability of the trial to distinguish effective from ineffective treatments [3]. That is, assay sensitivity is the ability of a specific clinical trial to demonstrate a treatment difference if such a difference truly exists [3]. Assay sensitivity depends on the effect size one needs to detect. One, therefore, needs to know the effect of the control drug in order to determine the trials assay sensitivity. There is then an inherent, usually unstated, assumption in an equivalence/non-inferiority trial, namely that the active control is similarly effective in the particular study one is performing (i.e., that one's trial has assay sensitivity), compared to a prior study that utilized a placebo comparator. However, this aforementioned assumption is not necessarily true for all effective drugs, is not directly testable in the data collected (if there is no placebo group to serve as an internal standard); and this, in essence, causes an active control equivalence study to have elements of a historically controlled study [4].

A trial that demonstrates superiority has inherently demonstrated assay sensitivity; but, a trial that finds the treatments to be similar, cannot distinguish (based upon the data alone) between a true finding, and a poorly executed trial that just failed to show a difference. Thus, an equivalence/non-inferiority trial must rely on the assumption of assay sensitivity, based upon quality control procedures and the reputation of the investigator. The International Conference on Harmonization (ICH) guidelines (see Chap. 6) list a number of factors that can reduce assay sensitivity, and includes: poor compliance, poor diagnostic criteria, excessive measurement variability, and biased endpoint assessment [5]. Thus, assay sensitivity can be more directly ascertained in an active control trial only if there is an 'internal standard,' a control vs. placebo comparison as well as the control vs. test drug comparison (e.g. a three-arm study).

Advantages of the Equivalence/Non-inferiority Approach

As discussed above, the application of equivalence testing permits a definitive statement that the new treatment is 'as good' (if the null hypothesis is rejected), and depending upon the circumstances, this statement may meet the needs of the manufacturer, who may only want to make the statement that the new treatment is as good as the established treatment, with the implication that the new treatment is preferred because it may require less frequent dosing, or be associated with fewer side effects, less invasiveness etc. On the other hand, the advantage of superiority testing is that one can definitively state if one treatment is better (or worse) than the other, with the downside that if there is not evidence of a difference, you cannot state that the treatments are the same (recall, that the null hypothesis is never 'accepted' – it is simply a case where it cannot be rejected, i.e. 'there is not sufficient evidence in these data to establish if a difference exists').

Disadvantages or Limitations of Equivalence/Non-inferiority Studies

The disadvantages of equivalence/non-inferiority testing include: (1) that the choice of the margin chosen to define whether two treatments are equivalent is difficult; (2) that it requires clinical judgment and should have clinical relevance (variables that are difficult to measure); (3) the assumption that the control would have been superior to placebo (assumed assay sensitivity) had a placebo had been employed (constancy assumption- that is, one expects the same benefit in the equivalence/non-inferiority trial as occurred in a prior placebo controlled trial); and, (4) having to determine the margin such that it is not greater than the smallest effect size (that of the active drug vs. placebo) in prior placebo controlled trials [6]. In addition, there is some argument as to whether the analytic approach in equivalence/non-inferiority trials should be ITT or Per Protocol (Compliers Only) [7]. While ITT is recognized as valid for superiority trials, the inclusion of data from patients not completing the study in equivalence/non-inferiority trials, could bias the results towards the treatments being the same, which could then result in an inferior treatment appearing to be non-inferior or equivalent. On the other hand, using the compliers only (per protocol) analysis may bias the results in either direction. Most experts in the field argue that the Per Protocol (some like to say non ITT analysis implying that it is as close to ITT analysis as possible) analysis is preferred for equivalence/non-inferiority trials but some still argue for the ITT approach [7]. Also, blinding does not protect against bias as much in equivalence/non-inferiority trials as it does with superiority trials-since the investigator, knowing that the trial is assessing equality may subconsciously assign similar ratings to the treatment responses of all patients.

Table 4.1 Approaches to hypothesis testing in clinical trials

RCT Hypothesis testing		
Hypothesis	Superiority	Equivalence/noninferiority
Null	New = Old	New < Old ± margin
Alternative	New > Old	New = Old
Null rejected	New is different than Old	New is at least as effective as Old
Failure to reject the null	Did not show that New is different that Old	Did not show that New is as effective as Old

The Null Hypothesis in Equivalence/Non-inferiority Trials (Table 4.1)

> It is a beautiful thing, the destruction of words…Take 'good' for instance, if you have a word like 'good' than is there need for the word "bad"? 'Ungood' will do just as well [8]

Recall that with traditional (superiority) hypothesis testing, the null hypothesis states that 'there is no difference between treatment groups' (i.e. New = Established, or placebo). Rejecting the null, then allows one to definitively state if one treatment is better than another (i.e. New > or < Established). The disadvantage is if at the conclusion of an RCT there is not evidence of a difference, one cannot state that the treatments are the same, or as good as one to the other.

Equivalence/non-inferiority testing in essence 'flips' the traditional null and alternative hypotheses. Using this approach, the null hypothesis is that the new treatment is worse than the established treatment (i.e. New < Old); that is, rather than assuming that there is no difference, the null hypothesis in equivalence/non-inferiority trials is that a difference exists and the new treatment is inferior. Some distinguish between equivalence and noninferiority, since strictly speaking equivalence means that the treatment effect is between the + and – margins and is therefore 2-sided, while noninferiority implies that the new treatment is "no worse than the old treatment and therefore is 1-sided. However, many in the field and an extension of the CONSORT Statement [9] suggest that two-sided confidence intervals are appropriate for most noninferiority trials, so the need for separating the two approaches is questionable.

Just as in traditional testing, the two actions available resulting from statistical testing is: (1) reject the null hypothesis, or (2) failure to reject the null hypothesis. However, with noninferiority/equivalence testing, rejecting the null hypothesis is making the statement that the new treatment is not worse than established treatment, implying the alternative, that is, that the new treatment is as good as (i.e. New $\geqq$ Established). Hence, this approach allows a definitive conclusion that the new treatment is at least as good, or is not inferior to the established.

As mentioned before, a caveat is the definition of 'as good as,' which is defined as being in the 'neighborhood' or having a difference that is so small as to be considered clinically unimportant (generally, event rates within ±2 % – this is known as the equivalence or non-inferiority margin usually indicted by the symbol δ). The need for this 'neighborhood' that is considered 'as good as' exposes the first shortcoming of equivalence/non-inferiority testing – having to make a statement that "I reject the null hypothesis that the new treatment is worse than the established, and

accept the alternative hypothesis that it is as good *and by that I mean that it is within at least X % of the established*" (the wording in italics are rarely included in the conclusions of a manuscript). A second caveat of equivalence/non-inferiority testing is that no definitive statement can be made that there is evidence that the new treatment is better or worse. Just as in traditional testing, one never accepts the null hypothesis – one only fails to reject it. Hence if the null is not rejected, all one can really say is that there is no evidence in these data that the new treatment is as good as or better than the old treatment. In equivalence trials, the conventional significance testing has little relevance, since failure to detect a difference does not imply equivalence. Rather, results should be reported with point estimates and confidence limits with the equivalence margin kept in mind.

In summary**,** the design of equivalence trials should mirror that of earlier successful trials of the active comparator as closely as possible [10] and, analysis strategies should not center on intention-to-treat (since ITT tends to reduce the difference between the intervention and control, it biases towards equivalence). Jones et al. also discuss why equivalence trials generally need to be larger than their placebo controlled counterparts, and why the standard of conduct needs to be especially high in terms of withdrawals, losses, and protocol deviations.

A potential concern has been raised over the rapid growth of noninferiority trials. For example, If novel therapy "A" is non-inferior to existing therapy "B" which itself was brought to market based upon non-inferiority data compared to therapy "C", the non-inferiority margin becomes more difficult to ascertain. Some potential ways one can overcome this is by comparing A to C directly but this may not be feasible if B has supplanted C in clinical practice. Alternatively, the margin for comparing A to B can be set to narrow limits, but this will increase the sample size.

One might ask; which is the 'correct' approach, superiority or equivalence testing? There is simply no general answer to this question; rather, the answer depends on the major goal of the study. But, once an approach is taken, the decision cannot be changed in post-hoc analysis. That is, the format of the hypotheses has to be tailored to the major aims of the study and must then be followed. An example of one innovative study in which the design combined a non-inferiority and superiority analysis is the Rivaroxaban versus Warfarin in Nonvalvular Atrial Fibrillation (ROCKET AF Study) which was a double-blind phase 3 study in more than 14,000 patients with atrial fibrillation. Patients were randomized to 20-mg rivaroxaban once daily (or 15 mg in patients with moderate renal impairment at screening) or to dose-adjusted warfarin (titrated to an international normalized ratio [INR] of 2.5). In the ROCKET-AF trial, patients were randomly assigned to receive either rivaroxaban or warfarin. In a per protocol, as-treated analysis, rivaroxaban was found to be noninferior to warfarin with respect to the primary end point of stroke or systemic embolism. As a pivotal trial for the new oral factor Xa inhibitor, rivaroxaban met its primary end point showing the drug was noninferior to warfarin. Disappointingly, however, in the same study the intention-to-treat superiority analysis failed to show the drug had an advantage, statistically, over warfarin. In an on-treatment analysis addressing the superiority question, however, rivaroxaban fared better, the rates of the composite major and non-major clinically relevant bleeding were comparable in the rivaroxaban- and warfarin-treatment arms [11].

Crossover Design

In crossover designs, both treatments (investigational and control) are administered sequentially to all subjects, and randomization occurs in terms of which treatment each patient receives first. In this manner each patient serves as his/her own control. The two treatments can be an experimental drug vs. placebo or an experimental drug compared to an active control. The value of this approach beyond being able to use each subject as their own control, centers on the ability (in general) to use smaller sample sizes. For example, a study that might require 100 patients in a parallel group design might require fewer patients in a crossover design. But like any decision made in clinical research there is always a 'price to pay.' For example, the washout time between the two treatments is arbitrary, and one has to assume that they have eliminated the likelihood of carryover effects from the first treatment period (plasma levels of the drug in question are usually used to determine the duration of the crossover period, but in some cases the tissue level of the drug is more important). Additionally, there is some disagreement as to which baseline period measurement, (the first baseline period or the second baseline period-they are almost always not the same) should be used to compare the second period effects.

N of 1 Trials

During a clinical encounter, the benefits and harms of a particular treatment are paramount; and, it is important to determine if a specific treatment is benefiting the patient or if a side effect is the result of that treatment. This is particularly a problem if adequate trials have not been performed regarding that treatment. Inherent to any study is the consideration of why a patient might improve as a result of an intervention. Of course, what is generally hoped for is that the improvement is the result of the intervention. However, improvement can also be a result of the disease's natural history, placebo effect, or regression to the mean (see Chap. 7). Clinically (in a practice setting), a response to a specific treatment is assessed by a trial of therapy, but this is usually performed without rigorous methodological standards so the results may be in question; and, this has led to the n of 1 trial (sometimes referred to as an RCT crossover study in single patients). In its usual form, n of 1 trials are randomized, double-blind, multiple crossover comparisons of an active drug against placebo in individual patients, and may be useful for determining individual treatment effects and as a tool to estimate heterogeneity of treatment effects in a population. An example of heterogeneity of treatment effects is the study by Pedro-Botet et al. [12]. Whereas the mean percent LDL-C response following 12 months of atorvastatin therapy (10 mg qd) was in the order of 35 %, the heterogeneity of effect is nicely portrayed in Fig. 4.2.

The requirements of the n of 1 design are: the patient receives active, investigational therapy during one period, and alternative therapy (e.g. placebo) during another

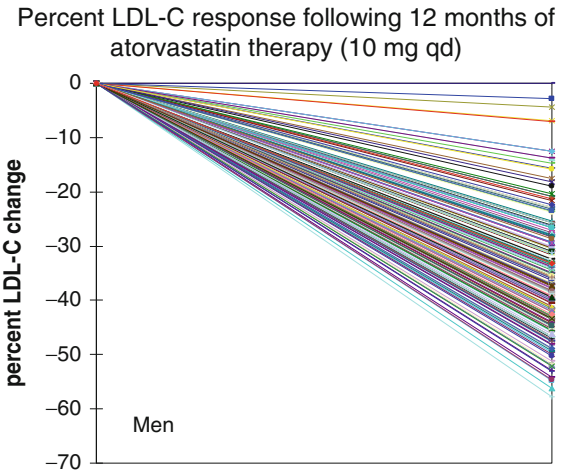

Fig. 4.2 The percent in LDL-C lowering in response to 12 months of Atorvastatin Therapy (10 mg/QD) (Pedro-Botet et al. [12])

period as would occur with typical crossover designs. As is also true of crossover designs, the order of treatment from one patient to another is randomly varied, and other attributes-blinding/masking, ethical issues, etc.- are adhered to just as they are in the classical RCT. In contrast to the typical crossover design however, at a pre-specified point (perhaps a given number of crossovers, or degree of improvement or deterioration) the patient's involvement in the study is stopped and their response held until all patients complete the trial.

There are at least three obvious sources of variability in clinical trials. Firstly, pure differences occur between patients: e.g. some are more seriously ill than others. Secondly, there is variability within patients: even given the same treatment they, or their measurements, may vary from time to time. Thirdly, some patients may react more favorably to a given treatment than other patients. The parallel group trial does not and cannot distinguish between types of variability; and, while the standard crossover trial will distinguish between the first type of variability and the other two it does not distinguish easily between the second and third. The n of 1 trial does address some of these issues in variability.

The n-of-1 trial does have some characteristics of the "playing the winner, dropping the loser" adaptive design (see below), but unlike this latter design, the patient in the n-of-1 trial may end the study (for that patient) when a pre-specified endpoint is reached. Some caveats to consider before designing an n-of-1 trial is that these trials are oriented towards symptomatic treatments that have rapid improvement upon treatment initiation, and rapid loss of efficacy upon therapy discontinuation. The use of this trial design is thus problematic when dealing with chronic disease therapies in which the acute response does not predict long term outcome, when the anticipated treatment effect is difficult to differentiate from random fluctuations of disease, and when treatment effects are small (i.e. hard to detect in an individual patient).

Table 4.2 Possible outcomes and stopping rules in N of 1 trials

Result	Continue	Stop
Benefit likely, harm unlikely	×	
Benefit possible, harm unlikely	×	
Benefit possible, harm possible		×
Benefit unlikely, harm unlikely		×
Benefit possible, harm possible		×
Inconclusive result		×

Adapted from: Mahon et al. [13]

An example of the n of 1 trial was reported by Mahon et al. [13] regarding the evaluation of the efficacy of theophylline for irreversible chronic airflow limitation. As these authors state; "*though the efficacy of theophylline for irreversible chronic airflow limitation has been established in conventional randomized controlled trials, its efficacy in individual patients is often in doubt*." Patients fulfilling the entry criteria for this trial (n = 31), were randomized by coin toss to either an n of 1 trial or standard treatment by a person unaware of their baseline characteristics. Some patients entered an open trial of theophylline that was given for 2 weeks at their previously used dose, and all patients were uncertain that theophylline was helpful while taking it openly. This was established by the patient not affirmatively answering the question, "Are you certain that theophylline is helping you?" Each patient was then randomized to a double-blind, multiple crossover comparison of theophylline vs. placebo and their results were compared to use of theophylline as standard therapy (administered according to published guidelines). For the n of 1 trial participant's, the order of theophylline and placebo were randomly determined and the physician monitoring the response was blinded as to treatment assignment. If deterioration occurred the patient was immediately switched to the other treatment, while if on the other hand the deterioration occurred during the second treatment period they were switched back to the first period treatment. In this way this study design of early switching or stopping treatment is designed to limit the ethical problem of a patient remaining symptomatic during alternate (particularly placebo) treatment.

Potentially, a number of different scenarios could occur as outlined in Table 4.2. The difference in theophylline use at 6 months between the n of 1 trial and standard practice groups – without significant changes in exercise capacity and quality of life – suggests that the suspected bias of standard practice towards unnecessary treatment is real, by virtue of the much greater use of theophylline among standard practice patients (difference 47 %).

In 2011, Gabler et al. [14] reviewed 108 n of 1 trials done between 1985 and 2010 on 2,154 participants, and concluded that n of 1 trails are a useful tool for enhancing therapeutic precision in a wide range of conditions, and should be conducted more often.

Factorial Designs

Many times it is possible to evaluate 2 or even 3 treatment regimens in one study. In the Physicians Health Study, for example, the effect of aspirin and beta carotene was assessed [15]. Aspirin was being evaluated for its ameliorating effect on myocardial

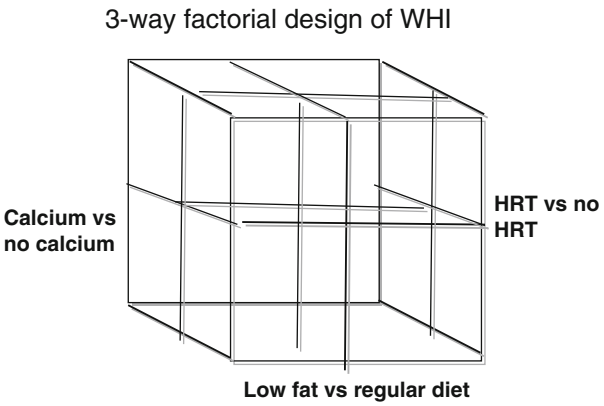

Fig. 4.3 Three-way Latin square design

infarction, and beta carotene on cancer. Subjects were randomized to 1 of 4 groups; placebo and placebo, aspirin and placebo, beta carotene and placebo, and aspirin plus beta carotene. In this manner, each drug could be compared to placebo, and any interaction of the two drugs in combination could also be evaluated. This type of design certainly can add to the efficiency of a trial, but this is counterbalanced by increased complexity in performing and interpreting the trial results. In addition, the overall trial sample size is increased (4 randomized groups instead of the usual 2), but the overall sample size is likely to be less than the total of two separate studies, one addressing the effect of aspirin and the other of beta carotene. In addition two separate studies would lose the ability to evaluate treatment interactions, if that is a concern. Irrespective, costs (if it is necessary to answer both questions) should be less with a factorial design compared to two separate studies, since recruitment, overhead etc. should be less. The Woman's Health Initiative is an example of a three-way factorial design [16]. In this study, hormone replacement therapy, calcium/vitamin D supplementation, and low fat diets were evaluated (see Fig. 4.3). Overall, factorial designs can be seductive but can be problematic, and it is best used for unrelated research questions, both as it applies to the intervention as well as the outcomes.

Case-Crossover Design

Case-crossover designs are a variant, having components of a crossover, and a case–control design. The case cross over design was first introduced by Maclure in 1991 [17]. It is usually applied to study transient effects of brief exposures on the occurrence of a 'rare' acute onset disease. The presumption is that if there are precipitating events preceding the outcome of interest, these events should be more frequent during the period immediately preceding the outcome, than at a similar period that is more distant from the outcome. For example, if physical and/or mental stress triggers sudden cardiac death (SCD), one should find that SCD occurred more frequently during or shortly after these stressors. In a sense, it is a way of assessing

whether the patient was doing anything unusual just before the outcome of interest. As mentioned above, case-crossover studies are related to a prospective crossover design in that each subject passes through both the exposure (in the case-crossover design this is called the hazard period) and 'placebo' (the control period). The case-cross over design is also related to a case–control study in that it identifies cases and then looks back for the exposure (but in contrast to typical case–control studies, in the case-crossover design the patient serves as their own control). Of course, one needs to take into account the times when the exposure occurs but is not followed by an event (this is called the exposure-effect period). The hazard period is defined empirically (one of this designs limitations, since this length of time may be critical yet somewhat arbitrary) as the time period before the event (say an hour or 30 min) and is the same time given to the exposure-effect period. A classic example of the case-crossover design was reported by Hallqvist et al., where the triggering of an MI by physical activity was assessed [18]. To study possible triggering of first events of acute myocardial infarction by heavy physical exertion, Halqvist et al. conducted a case-crossover analysis. Interviews were carried out in 699 myocardial infarction patients. The relative risk from vigorous exertion was 6.1 (95 % confidence interval: 4.2, 9.0), while the rate difference was 1.5 per million person-hours [18].

In review, the strengths of the case-crossover study design include using subjects as their own control (self matching decreases between-person confounding, although if certain characteristics change over time there can be individual confounding), and improved efficiency (since one is analyzing relatively rare events). In the example of the Halqvist study, although MI is common, MI just after physical exertion is not [18]. Weaknesses of the study design, besides the empirically determined time for the hazard period, include: recall bias, and that the design can only be applied when the time lag between exposure and outcome is brief and the exposure is not associated with a significant carryover effect.

Externally Controlled Trials (Before-After Trials)

Using historical control's as a comparator to the intervention is problematic, since the natural history of the disease may have changed over time, and certainly sample populations may have changed (e.g. greater incidence of obesity, more health awareness, new therapies, etc. now vs. the past). However, when an RCT with a concomitant control cannot be used (this can occur for a variety of reasons-see example below) there is a way to use a historical control that is not quite as problematic. Olson and Fontanarosa cite a study by Cobb et al. to address survival during out of hospital ventricular fibrillation [19]. The study design included a pre-intervention period (the historical control) during which emergency medical technicians (EMT) administered defibrillation as soon as possible after arriving on scene of a patient in cardiac arrest. This was followed by an intervention period where the EMT performed CPR for 90 s before defibrillation. In this way many of the problems of typical historical controls can be overcome in that in the externally controlled design, one can use the

same sites and populations in the 'control' and intervention groups as would be true of a typical RCT, it is just that the control is not concomitant. Another example is that of Sipilä et al. who assessed the impact of a guideline implementation intervention on antihypertensive drug prescribing; specifically, to assess the effects of a multifaceted (education, audit, and feedback, local care pathway) quality program. The proportions of patients receiving specific antihypertensive drugs and multiple antihypertensive drugs were measured before and after the intervention for three subgroups of hypertension patients: hypertension only, with coronary heart disease, and with diabetes.

Nonconventional Clinical Trial Designs

As the field of clinical trial methodology evolves, the need for alternative designs increases. This is reviewed by Howard [20] as it related to studies of stroke, but clearly it is not limited to that area. Howard outlined four such nonconventional approaches: dose selection trials; adaptive clinical trials; shift analysis; and Bayesian analysis.

Briefly, dose selection trials allow for dose adjustment as the trial proceeds primarily based upon the occurrence and frequency of any adverse effects at the dose being studied (unless the event rate is so low that it is not likely to be seen in a limited number of patients). The intent is to find the "optimal" dose (i.e. the highest potential dose that is associated with a low occurrence of adverse drug events). Adaptive clinical trials refers to a study design that is adjusted based upon data collected initially (sometimes confused with group- sequential studies) [21]. As Howard noted, "*Shih eloquently relates that group sequential methodology has the goal of saving lives or resources, whereas the adaptive clinical trial approach has the goal of saving the study*" [21].

"Shift analysis" allows for a reduction in sample size or gain in power, but further discussion is beyond the scope of this book. Bayesian analysis (Also see Chap. 14) is a potentially rapidly rising approach in clinical trials. Simplifying, the characteristic that defines any statistical approach is how it deals with uncertainty (see Chap. 18). The traditional approach to dealing with uncertainty is the frequentist approach, which deals with fixed sample sizes based upon prior data; but otherwise the information present from prior studies is not incorporated into the study being now implemented. That is, with the frequentist approach "the difference between treatment groups is assumed to be an unknown and fixed parameter". A Bayesian approach uses previous data to develop a prior distribution of potential differences between treatment groups and updates this data with that collected during the trial being performed to develop a posterior distribution (this is akin to the discussion in Chap. 14 that addresses pre and post test probability).

There are strong advocates of the frequentist and the Bayesian approach, which should indicate that neither is perfect and that one or the other may be preferable in certain situations. The argument then devolves to not which is better, but in which circumstance might one be preferable. Further discussion is also beyond this books scope, but should be of interest to the more advanced student.

Adaptive Designs

It is recognized that increased spending on biomedical research has not increased success rates of drug development-due to: diminished margin for improvement, chronic diseases are harder to study, rapidly escalating costs, and pharmaceutical company mergers that have decreased new-drug candidates. This has led to more innovative designs for evaluating drug efficacy. Adaptive designs give flexibility for identifying the optimal clinical benefit of a test treatment without ***"significantly"*** undermining the validity and integrity of the intended study. Some examples are the use of adaptive randomization; group sequential analysis (discussed in Chap. 9), and sample size re-estimation. Adaptive designs can be prospective (e.g. adaptive randomization, stopping a trail early due to safety, futility, or efficacy, dropping the loser (playing the winner); concurrent (e.g. modifying inclusion/exclusion criteria, modifying a dose/regimen and treatment duration); or, retrospective (e.g. changes in the statistical plan prior to database lock or unblinding of treatment codes). Whereas some adaptive changes require no or little statistical adjustment (e.g. dropping a treatment arm, modifying dosing paradigms, modifying randomization ratios; modifying subject selection, modifying visit schedules, or modifying study eligibility criteria), some do (e.g. requiring statistical adjustments, resizing a study, and allowing for the inclusion of subjects who participated in earlier drug development studies in a later development study – although this not generally recommended). What generally cannot be recommended in adaptive designs are: changes in the primary endpoint, and more than 1 adjustment to sample size.

One example of an adaptive design and to be contrasted to n of 1 trials (see above) is "playing the winner – dropping the loser", (this is an example of adaptive randomization). This design allows for dropping inferior treatment responses and adding additional arms, so it is useful, for example, in early drug development studies when there are uncertainties regarding dose levels. An example of how this is operationalized is starting out with a probability of 50 % randomization to both groups (allocation ratio of 1:1), and you randomize a patient to one of the treatments. If they do well, you increase the likelihood that the next randomization will be to the same group, the basic idea is to keep adjusting the likelihood randomization to a specific treatment group in order to increase the chances of the beneficial treatment going to the winner. For example, choose a staring base, say 20 subjects, and a 1:1 randomization scheme (10 A/20 A + B). One then randomizes a patient, and one assumes that they went to "A" and did well. Then the likelihood of the next patient being randomized to A would change from 50:50 to 52:48. Let's assume that the next patient despite increased odds of going to A in fact gets randomized to B and does poorly. We now further increase the likelihood going to "A" (say to 55 %). If a patient is randomized to B and does well, one adjusts the chances that the next patient will be randomized to B, and so on. Over time, if one group is doing better the likelihood of a patient being randomized to that group increases.

Registry-Based Randomized Clinical Trials (RRCT)

The Thrombus Aspiration in ST Elevation Myocardial Infarction in Scandinavia Trial (TASTE) was reported in 2013 and the study design caused a lot of excitement [22]. The TASTE trial enrolled ST elevation MIs as they entered a long-standing Swedish Web System registry for another goal. Based upon that registry (which provided comprehensive data collection and follow-up, TASTE built a web-based randomization that allocated 7,200 patients to either treatment by thrombus aspiration followed by PTCA or PCTA only. The enthusiasm about the design was that it allowed for completeness of follow-up at a lower cost, and no commercial involvement. As Laure and D'Agostino point out, "*with this clever design, which leveraged clinical information that was already being gathered for the registry and for other preexisting databases, the investigators were able to quickly identify potential participants, to enroll thousands of patients in little time, to avoid filling out long case report forms, to obtain accurate follow-up with minimal effort, and to report their findings, all for less than a typical RO1 grant*" [23]. They do go on to point out a number of potential problems with the RRCT, however, including the quality of the data, missing data, privacy, blinding etc. But, the RRCT potentially presents an alternative to the standard RCT in countries with large observational registry programs.

References

1. Breslin JE. Quote Me. Ontario: Hounslow Press; 1990.
2. Siegel JP. Equivalence and noninferiority trials. Am Heart J. 2000;139:S166–70.
3. Assay Sensitivity. In: Wikipedia. Available from: http://en.wikipedia.org/wiki/Assay_sensitivity
4. Snapinn SM. Noninferiority trials. Curr Control Trials Cardiovasc Med. 2000;1:19–21.
5. Walter SD. Choice of effect measure for epidemiological data. J Clin Epidemiol. 2000;53:931–9.
6. D'Agostino Sr RB, Massaro JM, Sullivan LM. Non-inferiority trials: design concepts and issues – the encounters of academic consultants in statistics. Stat Med. 2003;22:169–86.
7. Wiens BL, Zhao W. The role of intention to treat in analysis of noninferiority studies. Clin Trials. 2007;4:286–91.
8. Diamond GA, Kaul S. An Orwellian discourse on the meaning and measurement of noninferiority. Am J Cardiol. 2007;99:284–7.
9. Piaggio G, Elbourne DR, Altman DG, Pocock SJ, Evans SJ, for the CONSORT Group. Reporting of noninferiority and equivalence randomized trials: an extension of the CONSORT statement. JAMA. 2006;295:1152–60.
10. Jones B, Jarvis P, Lewis JA, Ebbutt AF. Trials to assess equivalence: the importance of rigorous methods. BMJ. 1996;313:36–9.
11. Patel MR, Mahaffey KW, Garg J, Pan G, Singer DE, Hacke W, et al. Rivaroxaban versus warfarin in nonvalvular atrial fibrillation. N Engl J Med. 2011;365:883–91. doi:10.1056/NEJMoa1009638.
12. Pedro-Botet J, Schaefer EJ, Bakker-Arkema RG, Black DM, Stein EM, Corella D, et al. Apolipoprotein E genotype affects plasma lipid response to atorvastatin in a gender specific manner. Atherosclerosis. 2001;158:183–93.
13. Mahon J, Laupacis A, Donner A, Wood T. Randomised study of n of 1 trials versus standard practice. BMJ. 1996;312:1069–74.

14. Gabler NB, Duan N, Vohra S, Kravitz RL. N-of-1 trials in the medical literature: a systematic review. Med Care. 2011;49:761–8. doi:10.1097/MLR.0b013e318215d90d.
15. Hennekens CH, Eberlein K. A randomized trial of aspirin and beta-carotene among U.S. physicians. Prev Med. 1985;14:165–8.
16. Rossouw JE, Anderson GL, Prentice RL, LaCroix AZ, Kooperberg C, Stefanick ML, et al. Risks and benefits of estrogen plus progestin in healthy postmenopausal women: principal results from the Women's Health Initiative randomized controlled trial. JAMA. 2002;288:321–33.
17. Maclure M. The case-crossover design: a method for studying transient effects on the risk of acute events. Am J Epidemiol. 1991;133:144–53.
18. Hallqvist J, Moller J, Ahlbom A, Diderichsen F, Reuterwall C, de Faire U. Does heavy physical exertion trigger myocardial infarction? A case-crossover analysis nested in a population-based case-referent study. Am J Epidemiol. 2000;151:459–67.
19. Olson CM, Fontanarosa PB. Advancing cardiac resuscitation: lessons from externally controlled trials. JAMA. 1999;281:1220–2.
20. Howard G. Nonconventional clinical trial designs: approaches to provide more precise estimates of treatment effects with a smaller sample size, but at a cost. Stroke. 2007;38:804–8.
21. Shih WJ. Group sequential, sample size re-estimation and two-stage adaptive designs in clinical trials: a comparison. Stat Med. 2006;25:933–41.
22. Frobert O, Lagerqvist B, Olivecrona GK, et al. Thrombus aspiration during ST-segment elevation myocardial infarction. N Engl J Med. 2013. [Epub ahead of print]. doi:10.1056/NEJMoa1308789.
23. Lauer MS, D'Agostino RB, Sr. The randomized registry trial – the next disruptive technology in clinical research? N Engl J Med. 2013. [Epub ahead of print]. doi:10.1056/NEJMp1310102.