STATISTICS IN MEDICINE
Statist. Med. 2003; **22**:169–186 (DOI: 10.1002/sim.1425)

Non-inferiority trials: design concepts and issues – the encounters of academic consultants in statistics

Ralph B. D'Agostino Sr.[1,2,*,†], Joseph M. Massaro[1,2,3] and Lisa M. Sullivan[1,3]

[1] *Boston University Statistics and Consulting Unit, 111 Cummington Street, Boston, MA 02215, U.S.A.*
[2] *Harvard Clinical Research Institute, 930 Commonwealth Avenue, Boston, MA 02215, USA*
[3] *Boston University, Biostatistics, 715 Albany Street, Boston MA 02118, USA*

SUMMARY

Placebo-controlled trials are the ideal for evaluating medical treatment efficacy. They allow for control of the placebo effect and are most efficient, requiring the smallest numbers of patients to detect a treatment effect. A placebo control is ethically justified if no standard treatment exists, if the standard treatment has not been proven efficacious, there are no risks associated with delaying treatment or escape clauses are included in the protocol. Where possible and justified, they should be the first choice for medical treatment evaluation. Given the large number of proven effective treatments, placebo-controlled trials are often unethical. In these situations active-controlled trials are generally appropriate. The non-inferiority trial is appropriate for evaluation of the efficacy of an experimental treatment versus an active control when it is hypothesized that the experimental treatment may not be superior to a proven effective treatment, but is clinically and statistically not inferior in effectiveness. These trials are not easy to design. An active control must be selected. Good historical placebo-controlled trials documenting the efficacy of the active control must exist. From these historical trials statistical analysis must be performed and clinical judgement applied in order to determine the non-inferiority margin M and to assess assay sensitivity. The latter refers to establishing that the active drug would be superior to the placebo in the setting of the present non-inferiority trial (that is, the constancy assumption). Further, a putative placebo analysis of the new treatment versus the placebo using data from the non-inferiority trial and the historical active versus placebo-controlled trials is needed. Useable placebo-controlled historical trials for the active control are often not available, and determination of assay sensitivity and an appropriate M is difficult and debatable. Serious consideration to expansions of and alternatives to non-inferiority trials are needed. Copyright © 2003 John Wiley & Sons, Ltd.

KEY WORDS: control group; clinical trial; placebo control; active control; equivalence; non-inferiority; assay sensitivity

1. INTRODUCTION

The randomized clinical trial (RCT) is one of the most important advances in the twentieth century [1–3]. Its importance grew as evidence-based medicine became the norm for establishing efficacy of drugs, biologics and medical devices. In the early 1900s the efficacy of

* Correspondence to: Ralph B. D'Agostino, Boston University Statistics and Consulting Unit, 111 Cummington Street, Boston, MA 02215,U.S.A.
† E-mail: Ralph@bu.edu

Received October 2002
Accepted October 2002

medical treatments was based on anecdotal evidence, often gathered on one or several patients (medical reports and case series). Some treatments had profound effects such that evidence based on few patients was convincing (for example, penicillin). In general this was not the case. Later, more rigorous studies followed in which several patients were given the same treatment and evaluated. Many of these studies, however, were uncontrolled. Bradford Hill pointed out the problems of these and set the stage for RCTs in the medical arena [4]. Others illustrated the importance of RCTs and the potential deception of uncontrolled clinical trials by contrasting the 'positive results' reported in uncontrolled trials versus RCTs [5–7]. Spilker gave a review in four major clinical areas: psychiatry; depression; respiratory distress, and rheumatoid arthritis [5]. In each area, a substantially higher proportion of positive findings were reported in uncontrolled trials as compared to RCTs. For example, in psychiatric therapy trials, 83 per cent of uncontrolled trials reported positive findings, as compared to only 25 per cent of RCTs [6]. In rheumatoid arthritis trials, 62 per cent of uncontrolled trials reported positive findings, as compared to only 25 per cent of RCTs [7]. The RCT can distinguish the effects of a medical treatment from other effects, such as spontaneous changes in the course of the disease, the body's natural healing, improvement due to participating in a study (that is, the placebo effect), and biases in observation and measurement. Few now doubt the virtues of RCTs for assessing medical treatment efficacy.

The United States' Food and Drug Administration (FDA) emphasizes the need for RCTs for medical treatment (drugs, biologics and devices) approval. For example, the Code of Federal Regulations (CFR) Title 21, Part 314, outlines the procedures for applications to the FDA for approval to market new drugs and Section 126 outlines the criteria of 'adequate and well-controlled' studies [8]. Focus is on the RCT. The same emphasis holds in the international setting. The International Conference on Harmonisation (ICH) is attempting to consolidate procedures for the registration of pharmaceuticals in the European Union, Japan and the United States. The ICH E9 guidance document discusses statistical principles for clinical trials [9]. The ICH E10 guidance document discusses the selection of appropriate controls in clinical trials [10, 11]. The latter document describes five types of controls (placebo, no treatment, dose–response, active and historical), and outlines the advantages and disadvantages of each. The first four controls are concurrent controls. These controls in randomized clinical trials are preferable to historical controls as patients for both the test and control treatments are drawn from the same population and studied under similar conditions, thereby minimizing bias in the comparison. Of all the possible RCTs, to many the ideal is the placebo-controlled RCT.

In the absence of effective treatments, placebo-controlled RCTs are uncontroversial. When, however, a proven effective treatment exists, the ethics of the placebo-controlled trials are questionable. In this setting, the attacks against placebo-controlled trials are many and substantial [12–15]. Of most importance is the Declaration of Helsinki [16]. Article II.3 of this states 'In any medical study, every patient – including those of a control group, if any – should be assured of the best proven diagnostic and therapeutic method. This does not exclude the use of inert placebo studies where no proven diagnostic or therapeutic methods exists'. Many interpret this to mean that when an effective treatment exists the use of a placebo is unethical and should not be included in a RCT. Others, including prestigious groups such as the American Medical Association and the World Health Organization, leave room for the possible use of placebo-controlled RCTs under certain circumstances (see Section 2) [17–21].

The active-controlled trial has been one response to the attack on placebo-controlled trials. Here the new experimental treatment is compared to a proven active control treatment. The

new treatment may not be superior to the active treatment in terms of efficacy, but it may be equivalent. Borrowing ideas from the field of bioequivalency, medical researchers including clinicians and statisticians developed equivalency trials with their design issues and the necessary statistical testing procedures [22–27]. Upon further clarification of the issues, it became clear that what was desired were non-inferiority trials (or more precisely, non-inferiority active-controlled RCTs), even if the term 'equivalency trials' is often used. The objective of a non-inferiority clinical trial is to establish that the effect of the new treatment, when compared to the active control, is not below some pre-stated non-inferiority margin.

The designing, implementation and analysis of non-inferiority trials have presented substantial challenges and issues for the pharmaceutical, biologics and medical device industries. The FDA and its scientists are well aware of these [11, 28, 29]. In our roles as academic consultants, industry sponsors are constantly seeking advice to decide when a non-inferiority trial is warranted, to clarify for them the unique design concepts and the issues involved, to help design, implement and perform the trial and ultimately to aid in the analysis and interpretation of the study. In this paper we focus on the *design concepts and issues* involved. We illustrate these with real world examples, many that we have encountered.

In Section 2 we review the usefulness of the placebo-controlled trial and the situations where they may be justified, even when proven active treatments exist. Section 3 discusses two major issues in active-controlled non-inferiority trials: (i) the statistical hypotheses and tests involved in a non-inferiority trial and (ii) the selection of the non-inferiority margin. The latter includes discussion of clinical meaningfulness, assay sensitivity (which relates to establishing that the active treatment and in turn the experimental treatment would have been superior to placebo had a placebo been used in the trial), and the fear of what is called 'biocreep'. Section 4 concerns the putative placebo analysis as a means of establishing that the new treatment is superior to placebo. Section 5 deals with selecting the appropriate sample to use for the statistical analysis. In Section 6 we discuss the role of interim analysis. Then in Section 7 we expand the non-inferiority trial to consider safety issues and also review some alternatives to non-inferiority trials. Finally, in Section 8 we give a brief closing discussion and some recommendations.

2. PLACEBO-CONTROLLED TRIALS

An appropriate control group is always essential and, when feasible, a placebo control is optimal. Figures 1 and 2 demonstrate the problem when a study does not contain a placebo control. The comparison of the active control C with the test treatment T in Figures 1 and 2 indicates that the two treatments are similar. However, if a placebo group is not included in the study, then one can never be sure if the new treatment is better than the placebo, as Figure 1 indicates, or not different from the placebo, as Figure 2 indicates. Figure 1 corresponds to both C and T being effective, Figure 2 to neither being effective.

Historically, a placebo control group was the usual optimal control group for establishing efficacy of an experimental treatment. It has been the basis for many FDA approvals. Superiority of the experimental treatment over placebo in two well controlled and performed RCTs justified approval. At times it was essential to establish that the trial had sensitivity (or sometimes called assay sensitivity) and an active control was added as, for example, in analgesic studies [30, 31]. Here the comparison of the active control to the placebo was an

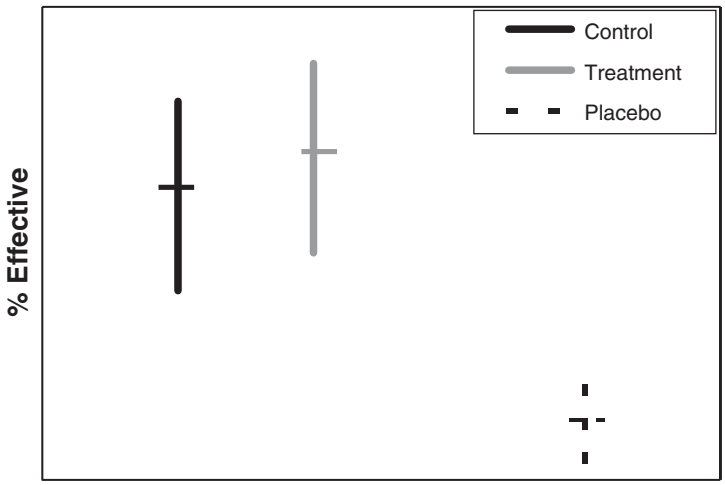

Figure 1. Comparison of test treatment (T) with active control (C) and unobserved placebo (P) (T and C superior to P).

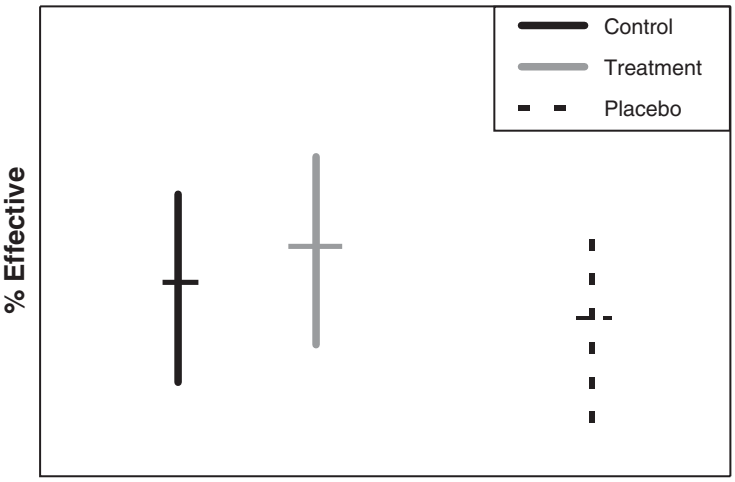

Figure 2. Comparison of test treatment (T) with active control (C) and unobserved placebo (P) (T and C not superior to P).

essential component of the analysis. The comparison of the active control to the experimental treatment was not required. The ideal was a study with a placebo, an active control and an experimental treatment.

Now with the large array of proven effective treatments, ethical considerations cast doubts on the appropriateness of using a placebo control. Dose response trials are possible alternatives, but they also raise ethical problems since the low dose may not be any different than a placebo. *So when is a placebo control justified in the presence of proven active treatments?* We agree with Ellenberg and Temple [21]. 'that placebo controls are ethical when delaying

or omitting available treatment has no permanent adverse consequences for the patient and as long as patients are fully informed about the alternatives'. We also believe escape clauses should be included in the protocol.

An active control arm may be included in the RCT, but the active control is there for reasons such as assay sensitivity. It is not necessary for comparison with the experimental treatment. Thus for many over-the-counter drug situations such as pain, headaches, upset stomach and the treatment of the common cold, placebo-controlled trials are ethical. Ellenberg and Temple [20, 21] discuss numerous prescription drug situations involving, for example, antidepressants and short term trials (such as some anti-hypertensive trials), and *settings where the available* '*effective treatment*' *may not be uniformly accepted as standard treatment* and so placebo-controlled trials are justified.

3. ACTIVE-CONTROLLED TRIALS/NON-INFERIORITY TRIALS

Now let us move to the situation where the placebo control is considered unethical or for some other reason is deemed inappropriate. This leads us to active-controlled trials in which the experimental treatment is compared directly to a proven effective active control. If the sponsor believes the experimental treatment is superior to the active control, then a standard superiority trial with the objective of showing that the experimental treatment is statistically and clinically superior to the active control is appropriate.

What, however, if anticipated superiority is not the case? Then a non-inferiority trial (that is, a trial with the objective of showing that the experimental treatment is statistically and clinically not inferior to the active control) may be appropriate. A sponsor of an experimental treatment may logically decide to conduct a non-inferiority trial even when he believes the active control's efficacy cannot be surpassed. Why? The new product may offer safety advantages. For example, a new anti-infective product may produce no resistant bacteria, a new respiratory distress product for premature infants may be synthetic as opposed to animal derived and pose less risk, a new asthma treatment inhaler may have no chlorofluorocarbons in contrast to the standard product [23]. In the case of HIV treatments, new products may have simpler regimens promoting adherence and potentially reducing resistance. It is even possible that costs, marketing and potential profits are the underlying reasons. For example, the costs of the new product may be less expensive or the sponsor may have better access to the markets.

3.1. Statistical algorithm for assessing non-inferiority

The statistical algorithms for assessing non-inferiority (and equivalency) are in Blackwelder's paper [22]. We give a brief summary here and in Table I. Let T and 'Test' represent the value of the efficacy variable for the new (experimental) treatment. Similarly let C and 'Control' and P and 'Placebo' represent the values of the efficacy variable for the active control and placebo, respectively. Further, say we have a trial where higher values of this efficacy variable are desirable. The standard null and alternative hypotheses for proving non-inferiority are

$$H_0\colon\ C - T \geqslant M\ (C \text{ is superior to } T)$$
$$H_1\colon\ C - T < M\ (T \text{ is not inferior to } C)$$

DOCKET ALARM

Explore Litigation Insights

Docket Alarm provides insights to develop a more informed litigation strategy and the peace of mind of knowing you're on top of things.

Real-Time Litigation Alerts

Keep your litigation team up-to-date with **real-time alerts** and advanced team management tools built for the enterprise, all while greatly reducing PACER spend.

Our comprehensive service means we can handle Federal, State, and Administrative courts across the country.

Advanced Docket Research

With over 230 million records, Docket Alarm's cloud-native docket research platform finds what other services can't. Coverage includes Federal, State, plus PTAB, TTAB, ITC and NLRB decisions, all in one place.

Identify arguments that have been successful in the past with full text, pinpoint searching. Link to case law cited within any court document via Fastcase.

Analytics At Your Fingertips

Learn what happened the last time a particular judge, opposing counsel or company faced cases similar to yours.

Advanced out-of-the-box PTAB and TTAB analytics are always at your fingertips.

API

Docket Alarm offers a powerful API (application programming interface) to developers that want to integrate case filings into their apps.

LAW FIRMS
Build custom dashboards for your attorneys and clients with live data direct from the court.

Automate many repetitive legal tasks like conflict checks, document management, and marketing.

FINANCIAL INSTITUTIONS
Litigation and bankruptcy checks for companies and debtors.

E-DISCOVERY AND LEGAL VENDORS
Sync your system to PACER to automate legal marketing.