What are non-inferiority drug trials?

Trepo, Elodie Andree Anais

UNIVERSITY OF ZAGREB
SCHOOL OF MEDICINE

Elodie Trepo

What are non-inferiority drug Trials?

GRADUATE THESIS

Zagreb, 2016

This graduation thesis was made at department of pharmacology mentored by Professor Vladimir Trkulja and was submitted for evaluation of academic year 2015/2016.

Abbreviations :

RCT: randomized control trial

NI : non-inferiority

FDA : Food and Drug Administration

NCE : new chemical entities

IND : Investigational New Drug application

NDA : new drug application

CI: confidence interval

Table of Contents:

Summary:

Typically, clinical trials compare a new product or therapy with another that already exists to determine if the new one is as successful as, or better than, the existing one. In some studies, participants may be assigned to receive a placebo. Comparing a new product with a placebo can be the fastest and most reliable way to demonstrate the new product's therapeutic effectiveness. However, placebos are not used if a patient would be put at risk, particularly in the study of treatments for serious illnesses by not having effective therapy. Most of non-inferiority studies compare new products with an approved therapy. The active-controlled trial with a non-inferiority design has gained popularity in recent years. They have methodological challenges, especially in determining the non-inferiority margin. Regulatory guidelines provide some general statements on how an NI trial should be conducted. In this narrative review based on scientific literature and papers, but also publications about clinical trials, we will define NI trial concepts, explain the main methodology of NI margin determination with its limitations and the potential margin of improvement.

Key words: randomized controlled trial, non-inferiority, active-control, non-inferiority margin

1. Introduction

1.1 Generalities about clinical trial

A clinical trial is an interventional research on a human subject. They are designed to answer specific questions about biomedical or behavioral interventions, including new treatments such as novel vaccines or drugs. (1) Clinical trials are the gold standard in order to confirm drug efficacy. They allow rigorous scientific evaluation of treatment strategies and validation of patient care. Considering the complexity of pathophysiology, pharmacology, the probabilistic outcome in medicine, and because of bias in observational studies, it makes clinical trials the rational basis for physicians to provide credible evidence and draw information used to adapt their therapeutic practice.

1.2 Randomization and blinding

Randomization is the process by which two or more alternative treatments are assigned to volunteers by chance rather than by choice. This is done to avoid any bias with investigators assigning volunteers to one group or another. The results of each treatment are compared at specific points during a trial, which may last for years. When one treatment is found superior, the trial is stopped so that the fewest volunteers receive the less beneficial treatment. RCT are almost always defined as single or double blinded trials. Blind studies are designed to prevent members of the research team or study participants from influencing the results. Indeed, the participants do not know which medicine is being used . This allows scientifically accurate conclusions. In **single-blind studies**, only the patient is not told what is being administered. In a **double-blind study**, only the pharmacist knows; members of the research team are not told which patients are getting which medication, so that their observations will not be biased.

If medically necessary, however, it is always possible to find out what the patient is taking. (2)

When the aim of the randomized controlled trial (RCT) is to show that the one treatment is superior to another the test is called a superiority trial and the associated statistical test is a superiority test. A non significant superiority test is often mistaken as a proof that two treatments are not different. But proving that two treatments are equal is impossible with statistical tools. The closest possibility is to prove that two treatments are equivalent, meaning are not too different in characteristics, where 'not too different' is defined in a clinical matter.

1.3 Introduction to NI trials

Relative to development of medical treatments, it is becoming increasingly difficult to develop more powerful drugs, thusly the pharmaceutical companies are looking for treatments that have approximately the same efficacy, or a little worse, but demonstrate better qualities in other aspect. This type of RCTs aims to show that an experimental treatment is not inferior to the control treatment. Such trials are called non-inferiority trials. (3) The concept of NI trials started in 1970, based on the methodology of bioequivalence trials. In some circumstances, for example trials involving serious outcomes such as mortality, it is unethical to assign patients to a placebo. The term 'active control trial' refers to trials in which the control treatment employed is an active one. Most NI trials are using active control that is why they became increasingly popular in 1990s after the introduction of regulatory guidelines about the use of active- controlled trials. This is shown by a major increase of publications on NI trials since the first guideline. (4) A research on Pubmed including the terms 'non-inferiority', with 'randomized controlled trial' and 'humans' not 'bioequivalent' revealed only 1 publication in 1998 and the number increased to 189 in 2015, with a peek level in 2014 of 230 publications.

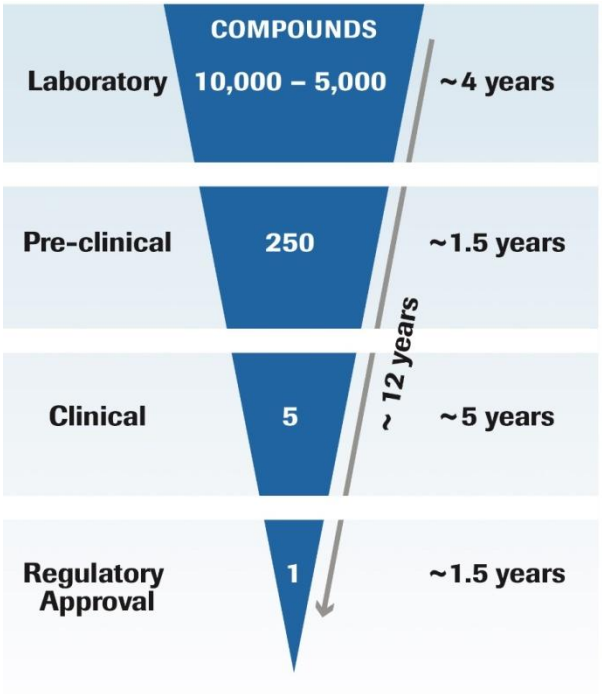
Table 1: Pubmed research [(non-inferior) or (noninferior) and (randomized controlled trial) and (humans) not (bioequivalence)] *http://ncbi.nlm.nih.gov accessed april 2016*.

Year	Count
2016 (not complete)	11
2015	189
2014	230
2013	205
2012	187
2011	146
2010	122
2009	108
2008	84
2007	58
2006	44
2005	30
2004	14
2003	10
2002	7
2001	2
2000	2
1999	1
1998	1

This illustrates the growing interest in NI trials as well as the increased need for readers and clinicians to understand the concept of this methodology.

First we will start by describing where NI stands with the general process of drug development, from its discovery until commercialization. In order to understand details about NI trials methodology we will explain in a second time what the different types of trial are and what are their differences. After the review of methodological and statistical concepts of NI margin we will discuss the quality of non-inferiority trials.

2. General overview of the regulatory clinical drug development



[figure 1: Drug DevelopmentvProcess .

www.roche.com/understanding_clinical_trials.pdf accessed May 2016].

Research and development aims to prevent and treat disease. The end result a small capsule seems so simple but the process for developing a new drug is anything but that. Development can take up to 20 years and cost 1 billion dollars. Several stages and team work involving the government, universities and pharmaceutical companies are required to reach the finish line.

2.1 Discovery and development

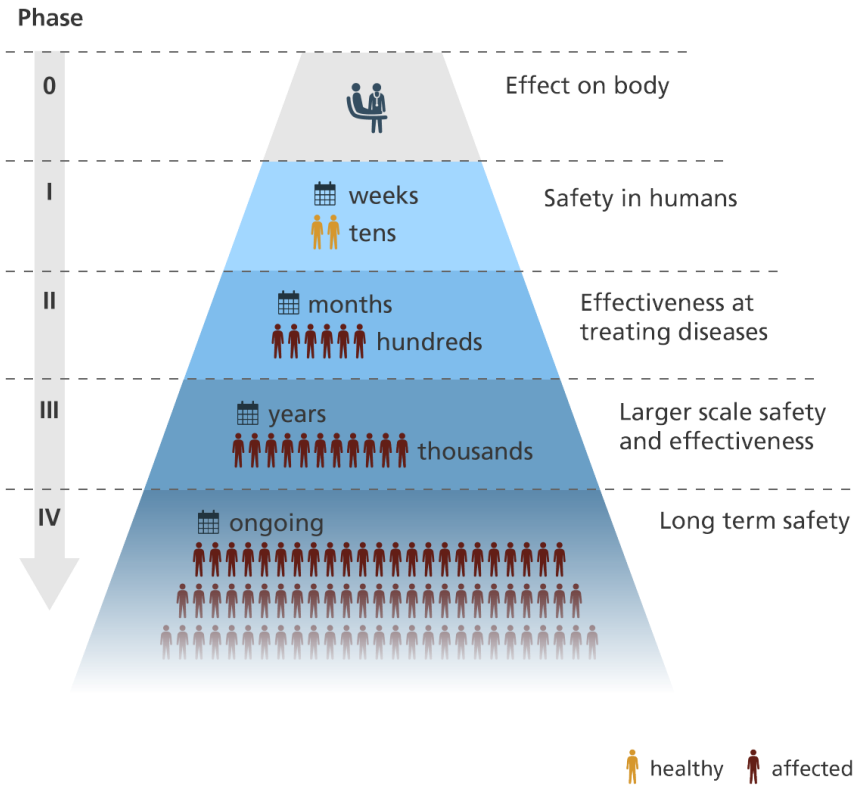
Research for a new drug begins in the laboratory. National Institute of Health, labs around the world work to discover fundamental knowledge about diseases. They aim to identify drug targets, mostly genes and proteins that play a crucial role in development or appearance of a disease. Scientists then investigate on how they interfere with these targets to control and eliminate a disease. They test thousands of compounds to see if they have an effect, but very small part will turn out to really have one. After the list of potential drug candidates is done, research on absorption, distribution and metabolism of the compound starts. Once researchers identify a promising compound for development they conduct experiments to gather information on potential benefits and mechanisms of action, best dosages, side effects, how it affects different groups of people and how it interacts with other drugs.

2.2 Pre-clinical stage

Researchers must find out if the drug has potential to be dangerous before testing a drug on people. They are two types of pre-clinical research: in vitro and in vivo. The new compound undergoes laboratory and animal testing to answer basic question about safety.
GLP which stands for '**Good Laboratory Practices**' are required by FDA and are defined in medical product development regulations. They set basic requirements for study conduct, personnel, facilities, equipment, written protocols, operating procedure, study reports and a

system of quality assurance oversight for each study to help assure the safety of FDA-regulated product. (5) Pre-clinical studies are usually not very large. Nevertheless, they must provide detailed information on **toxicity levels and doses**. If the results are encouraging, the principal investigators will apply for an approval to start clinical experiments. Those approvals are issued by special governmental agencies, such as European Agency for the Evaluation of Medical Products and US FDA. After these pre-clinical tests, researchers decide which compound is safe for further testing on people. Pre-clinical research is necessary but is not a substitute for studies on how drugs interact in a human body.

2.3. Clinical phases



[Figure 2 : A flow diagram to show the different phases in a clinical trial. www.yourgenome.org accessed May 2016]

Clinical research, also called trials, refer to studies on people which make sure they are safe and effective. Before starting the tests, researchers must design the trial, so it follows a specific protocol. They have to decide who will participate, how people will be part of the study, how long will the studies last, decide on a control group, drug administration, which data will be collected and how it will be analyzed. Before starting clinical research, drug developers must submit an **'Investigational New Drug' (IND) application** to FDA, including animal study data and toxicity, manufacturing information, clinical protocols, data from any prior human research and information about the investigator. The FDA team that consists of a statistician, pharmacologist, medical officer, project manager, pharmakineticist, chemist and microbiologist has 30 days to review the IND submission. They can either approve the beginning of a trial or place a clinical hold if the participants are exposed to significant risk, if the investigators are not qualified, or if the IND application lacks information about the risks.

Clinical trial has 3 phases. **The first phase** or '**safety'** is a small trial usually enrolls 20 to 100 **healthy volunteers**. During the first phase, researchers want to determine if a potential drug is safe for human use, and which doses can be tolerated by a human organism. Also pharmacokinetics is being carefully monitored as well as side effects.

Phase two '**protocol'** involves 50 to 500 **people with the disease** that is to be treated with a potential drug. The purpose of that phase is to determine its **effectiveness** and to further evaluate its safety. In other words, investigators would like to know if treatment works well enough to be tested in larger phase 3 trials, as well as the sample size needed for the final phase, optimal dose, and optimal treatment protocol. (6) Side effects are also studied, although they were tested in phase 1, there might be some more that doctors didn't know

about.

Phases three are randomized controlled trials on larger patient groups, from 300 to 3000 volunteers who have the disease. They aim to confirm previous findings, especially **drug effectiveness**, by **comparing it with standard or equivalent treatments** and collect information that will allow the experimental drug or treatment to be used safely. The decision to move ahead with this study is major one because the costs are considerably higher than for the two earlier phases combined. A drug identified as effective and safe in phase 2 may not enter phase 3 for a number of reasons. If it has insufficient efficacy when compared with its competitors, or when the side effects exceed the risk profile required to proceed.(7)

This phase represents a critical part of a drug's clinical testing cycle and is the one that we will focus on because **NI trials play a major role here**. The results of this phase will show the FDA investors and doctors if the drug really works.

2.4.FDA review and pharmacovigilance

When a drug is satisfactory after phase three, the company can file an application to market the drug. It is called a **NDA: New Drug Application**. NDA tells the full story of the drug in order to show that it is safe and effective. The NDA must include all information from pre-clinical to phase 3 along with proposed labeling, safety updates, information about drug abuse, patient information, any data from outside study and directions for use. If it is not complete the FDA will refuse the NDA. The review team has 6 to 10 months to make a decision. Each member will review the application concerning his section. The inspector will travel to the study site for inspection. In the end the manager will assemble all documents in order to create the record for FDA review. Complete information on drug safety is impossible

despite rigorous steps in process of drug development. FDA reviews report can decide to add cautious about dosage or usage if reviews report problems with prescription.

Phase 4 clinical trial represents a post marketing surveillance trial, also called **pharmacovigilance**. These ongoing reviews of drugs are required by regulatory authorities or sponsoring companies for competitive reasons. The goal of pharmacovigilance is to detect complications, long term side effects that were not described during phases 1 to 3 trials. In that case the drug will be taken out of the market.

3. Three types of hypothesis tested in clinical trials

RCT are essential to assess and compare the effectiveness of treatments. Superiority testing are typically used in comparative trials. However, superiority is getting more difficult to show, and becomes less important as margins of improvement decrease to clinically irrelevant levels. Alternative methods to compare groups in RCT are equivalence and non-inferiority. (8)

3.1 Inequality trials

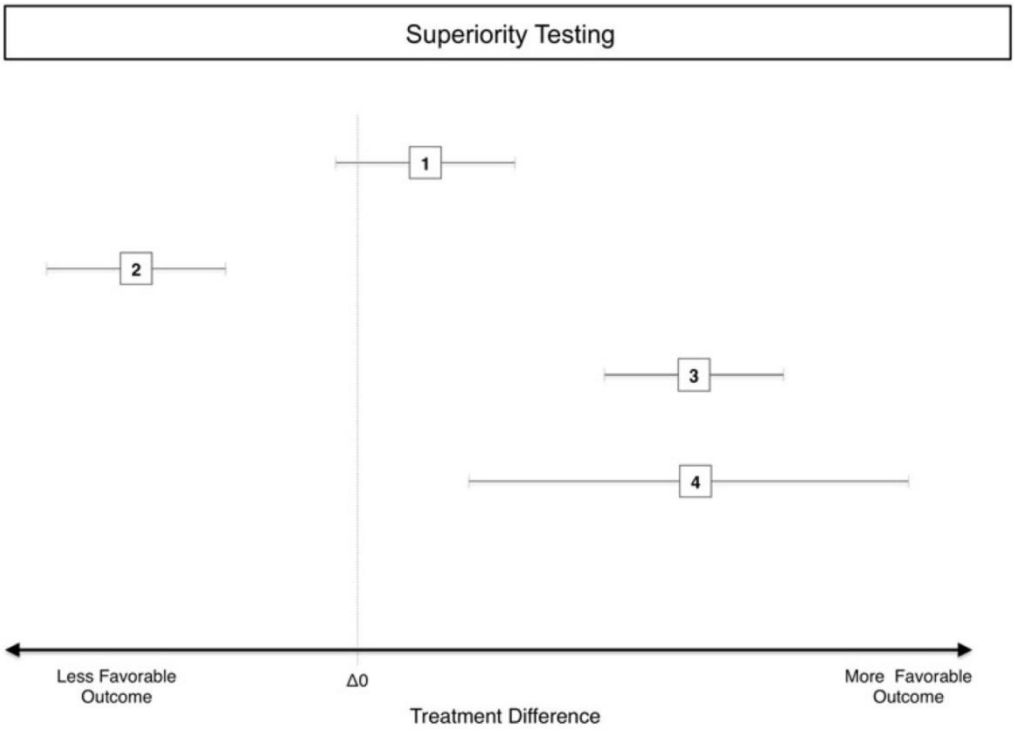
3.1.1 Definitions

Inequality trials, also called **superiority trials** started in 1900s, while William Gosset encountered the problem of having to find the best barley to brew Guinness beer. (9) His work led the way for the establishment of two indices related to superiority testing: p values and confidence intervals. P value is the probability that the differences observed between two or more groups occurred by chance if there really was no difference between the groups, usually set at 5% as a threshold. .Whereas the confidence interval represents the interval around an

observed parameter guaranteed to include the true value to some level of confidence. The true value can be expected to be within that interval with 95% confidence. (10) The CI should not include 0 for continuous values or 1 for ratios, both of which are consistent with no difference.

3.1.2 Superiority trial : looking for better

A superiority trial aims to demonstrate the superiority of a new treatment compared to an established therapy or placebo. Its means that the outcome of the patient after receiving the medication is better than with the old one, or with placebo. It can be for example decreased risk of clot with anti-coagulant therapy, a better surgical procedure with smaller amount of blood loss, etc. The goal of this type of trial is to test if the hypothesis of superiority is true or not. In order to do so, the first step in a superiority trial is to set the null hypothesis called Ho, in opposition with the alternative hypothesis (Ha) that we are trying to prove. The null hypothesis states that there is no association between the predictor and outcome values. In other words appropriate statistical tests needed to assess superiority should be performed, with the null hypothesis being: the difference between treatments is equal to zero, and the alternative hypothesis: treatment are different, the difference between treatments is not equal to zero. The rejection of the null hypothesis is in foundation of the methodological assessment of superiority. (11) The extra effect of the new therapy compared to the reference therapy is called the **Least Relevant Difference**, often written as **delta**.



[Figure 3: diagram illustrating the principle of superiority testing. The effects of an active control and a new treatment are given as 95% CIs of the difference between treatments, measured along the x-axis. DELTA 0 line represents a least relevant difference between groups of 0. (8)]

We can see that trial n°1 confidence interval includes zero so there is no difference between groups. Trial n°2 lies below zero line of less favorable outcome, implying a significantly worse, inferior result than the treatment is it compared to. Trial n°3 is significantly better, superior results, show by a 95% CI that lies entirely on the favorable side. Furthermore, the CI is shorter than with treatment n°4, therefore the differences between treatments in trial n°3 are more precise than in n°4, which also shows superiority.

3.1.3 Type 1 and 2 errors, limitations

The trial should demonstrate as precisely as possible the true difference in effects between treatments. However, the result may deviate from the true difference and give erroneous results because of the random variation. For example if the null hypothesis H0 were true, it is still possible that the trial in some case would show a difference. This is called **type 1 "false positive"** error, which would introduce an ineffective therapy. On the other hand if the alternative hypothesis of the delta difference were true, the trial can fail to demonstrate a difference. This type of error is called **type 2 "false negative"**, and rejects an effective therapy. This is why the investigator needs to specify how large risks of type 1 and type 2 errors would be acceptable for the trial. Because of limitation in patient number and resource, some small risk in error is tolerated. Most often type 1 error, that occurs with **probability alpha,** is specified to 5% and type 2 error that occurs with **probability beta**, 10 to 20%. 1 – beta is power, the probability of saying that there is a relationship or difference when there is one. It is the probability of confirming the theory correctly, so a trial designer would generally want this to be as large as possible in order to be confident in detecting a hypothesized difference in treatment effects. (12)

Superiority trials have been the standard for some time, but the success of new drugs that are always better than the previous ones creates a ceiling effect. That is why equivalence and non-inferiority trials are very useful when the superiority of one medication over another is neither expected or worse than the comparator. In general, superiority trials are used when new advances in treatment therapy, or effect of active control is small.

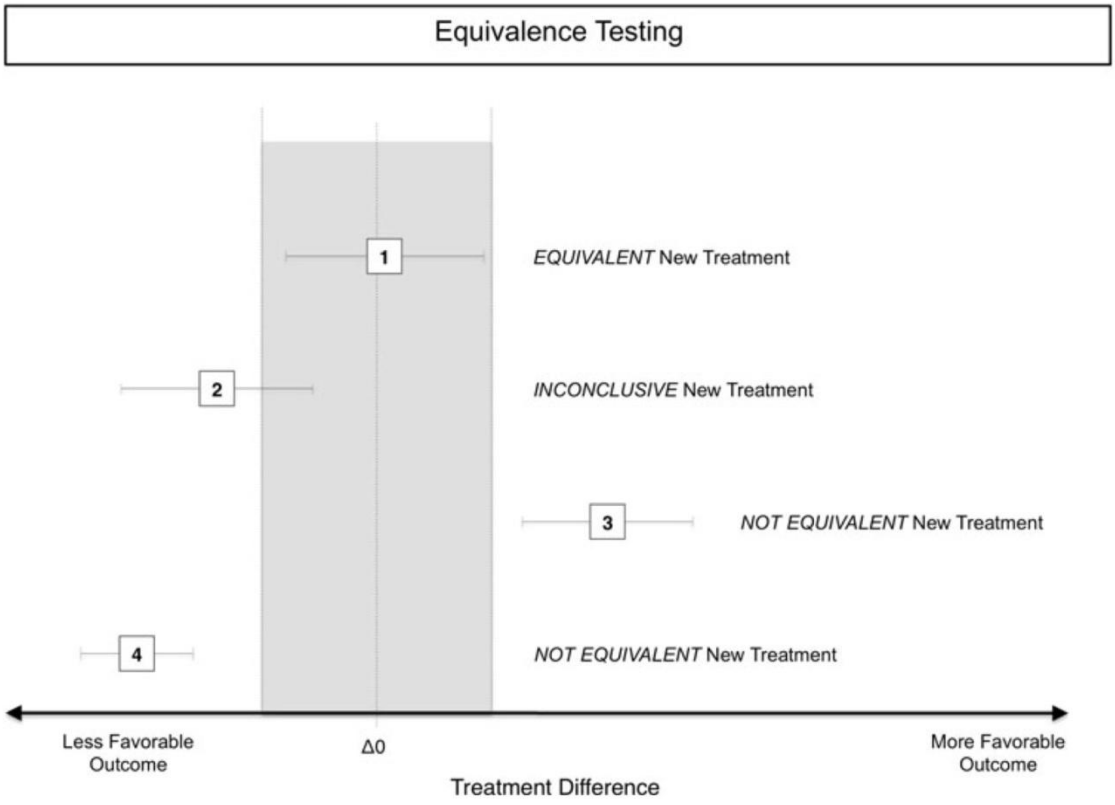
3.2 Equivalence trials

3.2.1 Pharmaceutical, clinical and bioequivalence

Most of the methodology of equivalence trials comes from studies of **pharmaceutical bioequivalence**. Drug products are defined as bioequivalent if they display comparable bioavailability, or absorbed percentage available at action site, when studied under similar experimental conditions. For drug products to be considered **pharmaceutical equivalents** they need to contain the same active ingredients, the same dosage form, route of administration and are identical in strength or concentration. Also they are formulated to contain the same amount of active ingredient in the same dosage form and to meet the same applicable standards: strength, quality, purity and identity. They may differ in characteristics such as shape, configuration, release mechanisms, packaging, excipients including colors, flavors, preservatives, expiration time, and, with certain limits, labeling. (13) We can say that pharmaceutical equivalence is a theoretical equivalence. On the other hand, **therapeutic equivalence** can be defined as "practical equivalence". Firstly the drugs need to be pharmaceutical equivalent but also are expected to have the same clinical effect and safety profile when administered to patients. In other words, pharmaceutical equivalent that have been shown to be bioequivalent and the same by other determinations of clinical effect and safety profile are therapeutic equivalents. Therapeutic equivalents are then expected to produce identical drug concentration time profiles and therapeutic response when administered under the same conditions. As an example, a new nasal anti-congesting drug that is equivalent to the standard one would have to contain the same active ingredient (either pseudoephedrine or phenylephrine) to be pharmacologically equivalent, but also clinically interchangeable and lessen the mucous membrane secretions and open the airway as much as the standard one.

3.2.2 Equivalence margin definition

"As much as" is defined as equivalence margin including lower and upper limits, we say that **equivalence testing is a two-sided test**. An equivalent margin is estimated and added to either side of the active treatment, and the effect of the new treatment is tested against this range.



[Figure 4: The effects of an active control and a new treatment are given. The grey area is the equivalence margin. Only treatment 1 lies within the equivalent margin and is equivalent to the new treatment. (8)]

A new treatment is equivalent only if it is no better and no worse, both within a margin (delta) than the active control. (8) This true two-sided equivalence approach is more common in pharmacokinetics, in which a difference in either direction from the reference treatment is of

importance. Whereas non-inferiority approach is much more common for therapeutic or prophylactic trials. (14) The value and impact of the study depend on how well the equivalence margin can be justified in terms of relevant evidence and clinical considerations. Regulatory issues have to be considered also. It is usually based on the margin of superiority of the standard treatment against placebo, estimated from previous studies. It must be stressed that this value should be determined before the data is recorded. This is essential to maintain the type 1 error at the desired level.

3.2.3 Equivalence after insignificant superiority?

Also, it is not possible to conclude on equivalence if superiority failed to be demonstrated. On one hand the alternative and null hypothesis were not defined in the same manner. The alternative hypothesis intended to show a difference. A non significant result only implies that equality cannot be ruled out. On the other hand, the margin of equivalence is not considered and previously defined, so the concept of equivalence is not well defined. (15) It is often though that equivalence is better than non-inferiority, because of the positive ring of being the same rather than non inferior, but this actually inverts the real situation where a non-inferiority treatment has potential for superiority, whereas an equivalent treatment, by definition, cannot be better than the active control.

3.3 Non inferiority trials

3.3.1 Active control instead of placebo

The use of RCT designed to directly compare a test treatment and placebo is the most straightforward strategy to discriminate between effective and ineffective therapy. However, such a trial is not always a viable clinical development pathway for ethical reasons. Withholding an effective drug from a patient by using a placebo may lead to serious

complications or death. (16) The alternative is to compare the test treatment with an efficacious standard or **active control treatment** and demonstrate that the test is not inferior to the active control by a pre-determined margin instead of directly showing the superiority of the test treatment over placebo. The non-inferiority becomes the goal of the trial. (17)

3.3.2 Superiority in secondary end points

NI trials are used when the new treatment has technical similarities with the existing, or if the active control has moderate to significant effect, or specific safety problems. Another reason for choosing non-inferiority over superiority designs is that a new treatment may not be better in the primary end point but **better in secondary end points**. The new design may be designed to be safer with **fewer complications** or at least less severe. Not only side effects of treatment such as nausea, vomiting, headaches can be improved, but also for example the pain duration after an intervention is crucial for the patient. Of course less pain after surgery can't outweigh the primary end points of the surgery in itself, but might be a strong argument. (8) When we talk about secondary end points, it can also be the **overall efficiency**. The **cost** of a treatment is not negligible. A treatment that is less expensive is more desirable. Given the strong and strongly growing emphasis on value-based healthcare and cost-effectiveness, these are important findings and might be included in some form in society guidelines and health policy regulations in the near future.

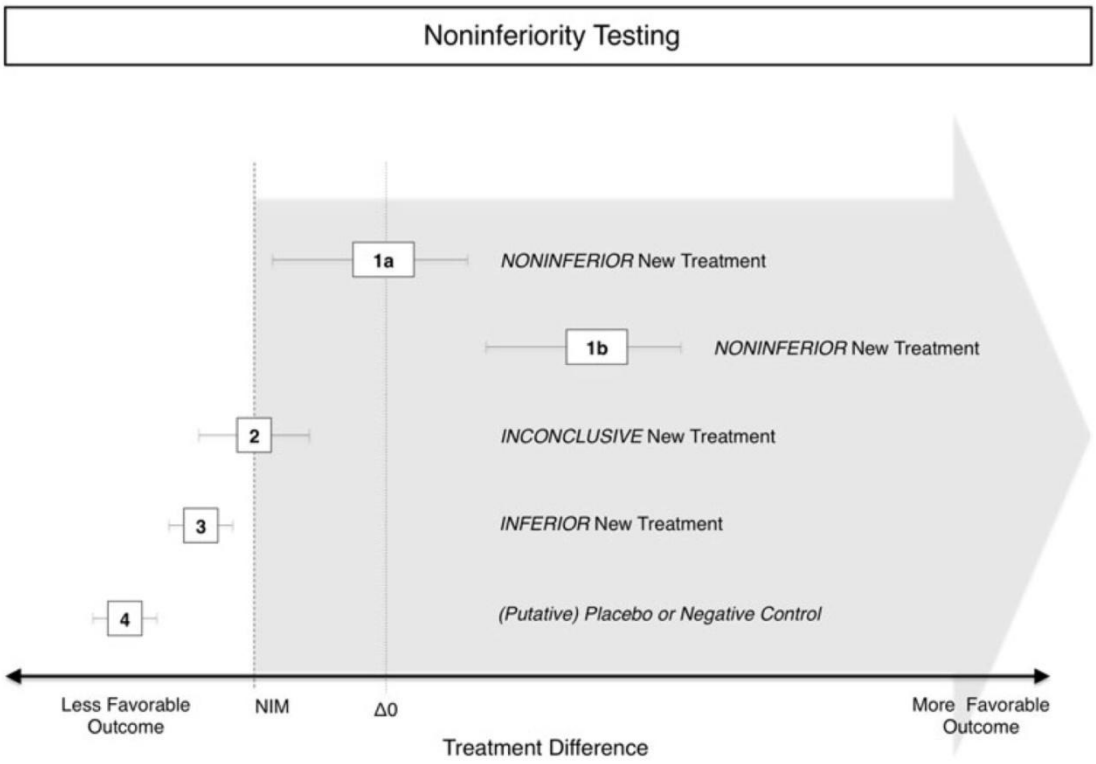
3.3.3. Superior is also non-inferior

Non-inferiority testing is desirable as it also allows the potential to establish that it is better. Non-inferiority assessment is **one-sided testing**, meaning that it does not allow the possibility that the new treatment is worse (with a NI margin) than the active control, but better is also non-inferior, and non-inferiority testing **does not exclude establishing superiority**. NI testing can be complemented by superiority testing in one study without the need for adjusting

for multiple testing or loss of power of validity. **The reciprocal is however not true**. If the trial fails to show that it is superior and that there is a significant difference it doesn't mean that there is no difference at all. It might be that the **power of the trial** wasn't strong enough to show this difference between two treatments, or that the sample population was too small. In NI studies, the alternative hypothesis is that the experiment therapy is inferior to the standard therapy. This comes from a null hypothesis that stated that the experimental treatment is equal to or better than the control treatment. As we saw previously the alternative and null hypothesis of superiority testing are different, so the experimental method is not adequate. The prerequisite for NI trials also are more demanding than superiority. They require much larger sample sizes than superiority studies because the typical sizes of the anticipated margin in NI trials are much smaller than what would be considered a clinically meaningful difference between groups in superiority studies.

3.3.4. Non-inferiority margin definition

A non-inferiority trial starts with defining "no worse (within a margin)" or non-inferior before the beginning of the trial.



[Figure 5: diagram illustrating the principle of non-inferiority testing comparing a new treatment with an active control. An a priori defined NI margin (NIM) is added to the line of zero difference between treatments delta 0. (8)]

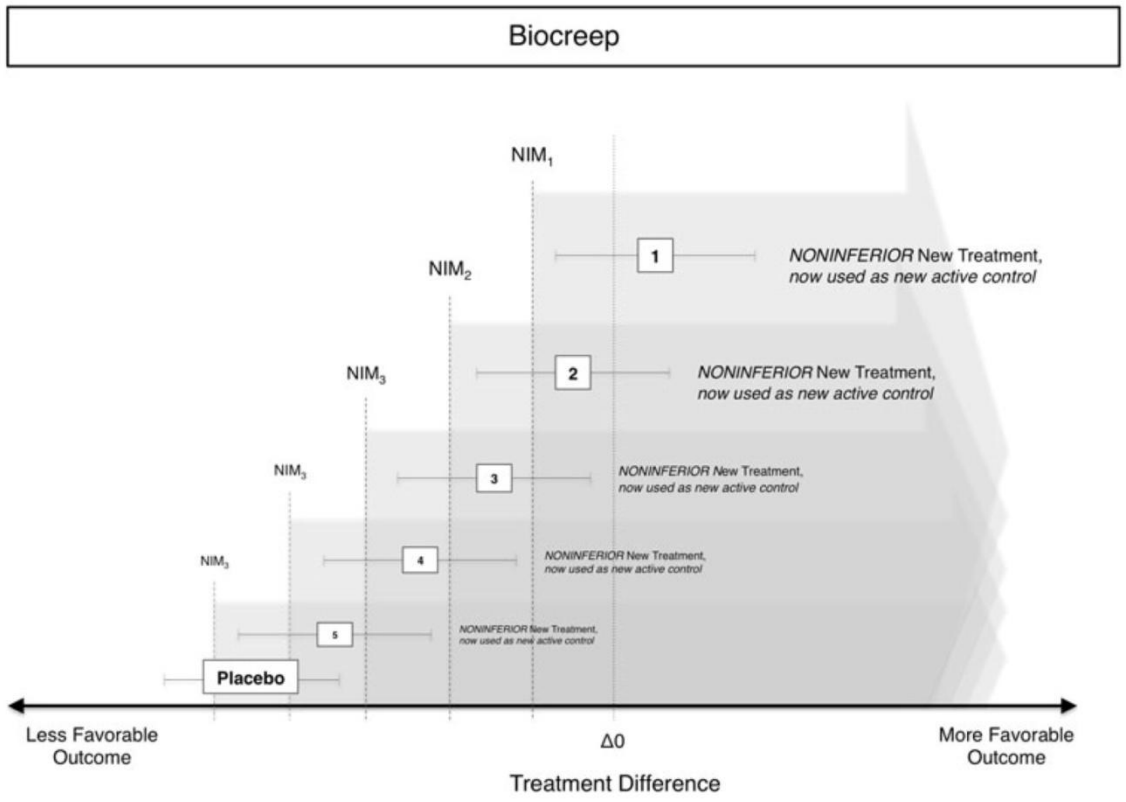
This requires the definition of an outcome and a threshold in that outcome, below which one would consider a new treatment to be inferior to an older one. (8) The modeling objectives of a NI trial can be classified into two categories, depending on whether or not placebo was used as a comparator in RCT in the past. When direct comparisons between placebo and active control are present in the historical trials, the ultimate goal is to predict the effect of the test treatment over placebo, or equivalently, determine the probability that the test treatment is better than the placebo as if the placebo is included in the NI trial. On the other hand, when only direct comparison among active control treatments are available, one needs to show that the test treatment is not inferior to the active control with a pre-specified margin. (16) This

margin should be determined during the design stage of a study before the actual experiment. Choosing the size of the margin is complex, and there is no explicit rule. This is one of the most difficult parts of the trial. How much efficiency on the result, such as mortality outcome, can we lose to get a better overall result, general status of the patient? The answer is very debatable. Usually findings from earlier studies and estimates of clinically relevant differences are combined. We will discuss later how it is determined in theory.

3.3.5 Biocreep

After a NI trial, a new therapy may be accepted as effective, even if its treatment effect is slightly smaller than the current standard. It is therefore possible that, after a series of trial where the new therapy is slightly worse than the preceding drugs, an ineffective or harmful therapy might be incorrectly declared efficacious; this is known as 'biocreep'. Several factors may influence the rate at which biocreep occurs, including the distribution of the effects of the new agents being tested and how they change over time, the choice of the active comparator, and changes in the effect of the active comparator from one trial to the next. (17) For example, if we accept that the new treatment B is not worse than 5% to the standard one A, and then comparing an even newer treatment C with treatment B again by 5% with treatment C, and then treatment D with C and so on. Although the NI margin of 5% has never been violated in individual comparisons, treatment Z is far from the effect of the original treatment A. The definition and use of such a margin might seem arbitrary to some, but it actually is more rigorous than a superiority design because it involves a predefined minimum difference and statistical testing for the specific difference, whereas superiority trials assess only the

significance but not the size of a difference.



[Figure 6: The reiterative use of non-inferior treatments as new active controls for the next study levels to an overall reduction of effectiveness from treatments 1 to 5 although the non-inferiority threshold was never violated in individual studies. The effect of placebo treatment is lower than the effect in the first, second, and third non-inferiority studies, but by the time of the fourth study, the non-inferiority threshold has crept to levels consistent with placebo treatment. (8)]

To put it in a nutshell, comparison of a new treatment with an active control rather than placebo, establishment of a new treatment with better secondary outcomes, or as the first step in testing superiority of a new treatment, all are reasons showing why NI trial are getting more important nowadays. Nevertheless, there is a potential weakness in NI testing, with possibility to flood the healthcare market with 'me too' procedures and products (18) that are

non-inferior than gold standard but do not add additional value, and risk of biocreep which highlight the fact that methodological rigor is even more important in NI than superiority test.

4. Main points about non-inferiority trials

4.1 Bias

In statistics bias means ‘a tendency of an estimate to deviate in one direction from a true value.’ (19) This systematic deviation from the true value can result in either underestimation or overestimation of the effects of an intervention. Because there is usually more interest in showing that a new intervention works than in showing that it does not work, biases in clinical trials most often lead to an exaggerate in the magnitude or importance of the effects of new interventions. The true effects of any health care intervention are unknown. But researchers try to anticipate, detect, quantify, and control bias to produce results from a sample of participants that can be generalized to the target population at large. Most bias occur during the actual course of a trial, from the allocation of participants to study groups, through the delivery of interventions, to the measurement of outcomes. Other types of bias can arise, however, even before the trial is carried out, in the choice of problem to study or type of research to use, or after the trial is carried out, in its analysis, and its publication. Bias can even be introduced by the person who is reading the report of a trial. (20) Some strategies must be settled to prevent as much as possible these errors that will eventually lead to a deviation of inferences or results from the truth. The study must be **prospective**, meaning that the population sample, strategy plan, criteria for judgment, margins, all had to be set before the start of the trial. The use of a **control group,** receiving placebo or standard treatment is also essential in order to assure the quality of the trial inference. (21) If for example the study on a new treatment for grasses allergy as 100% of cure, we will conclude on very high

efficacy. But if the the control group also shows the same results, then the conclusion will be different. A lot of external factors can create effects that can be confused with the effects of the treatment studied. Natural evolution of the disease, placebo effect, or the existence of concomitant treatment are a few examples. Only the use of control group can prevent inclusion of these **confounders,** defined as 'a variable that is related to both the exposure and outcome but is not measured or is not distributed equally between groups.' (22) So that the results showed in the study are only linked with the treatment given, the different groups should be chosen on the exact same basic criteria such as age, sex, height and weight, but also on the severity of the disease and should differ only on the treatment received. As we saw in the introduction, randomization is the only way for the groups to be truly comparable. Any other method would introduce a **selection bias**. The method in practice is done by a computer generating random numbers assigned to each patient corresponding to which treatment they will follow. (23). The randomization must be kept throughout the study. In order to do so, the patient and the researcher can't know the nature of the treatment. This is what we call a double-blinded study. Otherwise, **follow-up bias** may be introduced. It can be a modification in rhythm of visits to the patient, how well the examination will be performed and side-effect. Even unconsciously the researcher will be influenced in his perception of the patient. It can also be adding complementary exams, or simply change in therapeutic adherence of different patients if they will continue to take a pill knowing it contains no active substance. **Attrition bias** is induced by exclusion of patients during the trial. Various situations can lead to a premature drop-out of patients, such as serious side effects, a lack of therapeutic adherence, the absence of the patient to follow-up visits, intake of forbidden medication, or the inclusion of patients not meeting inclusion criteria of the trial. At first sight it could seem logical not to take into account these patient in the analysis, because if they didn't take the drug or didn't follow the procedures described in the beginning it wouldn't allow a proper result of the

efficacy of this drug. However, this loss of patients can create bias because their exclusion are usually not random but have a probability to depend on the treatment received and/or the evolution of the patient. If we exclude this patients, we might lose the comparability between the randomized groups. '**Intention to treat'** analysis is a strategy for the analysis of RCT that compares patients in the groups to which they were originally assigned. This interpreted as including all patients, regardless of whether they actually satisfied the entry criteria, the treatment actually received, and subsequently withdrawal or deviation from the protocol. (24) For non-inferior trials '**per protocole'**, where only patients who received a strict conform treatment are presented in the study, should also be present before concluding on the results. Indeed, the drop-out is potentially more important with patients receiving the reference treatment, which would give results in favor of non-inferiority. Also specific bias concerning NI trials are related to the reference treatment used. It might not be administered correctly either in too week dosage or too slowly, and decrease its efficacy compared to the new treatment. The reference treatment can also be administered to patients that are less receptive to the therapy, or easily interrupted because of side effects, or simply not be the best treatment that actually exist at the moment. As a conclusion we can say that results can be biased if the reference treatment is not maximized. Its loss of efficacy can be due to selection bias, where the choice of patients is not in favor of the standard treatment or to follow-up bias if dosage and administration are not optimal. These conditions would then create proof of non-inferiority, but the results of the clinical trial then should be discussed. This is why a quality critical analysis using proper methodology and statistic of publications or clinical trial results is essential, because this is what with allow the physician to evaluate them and decide to use them or not during his/her clinical practice.

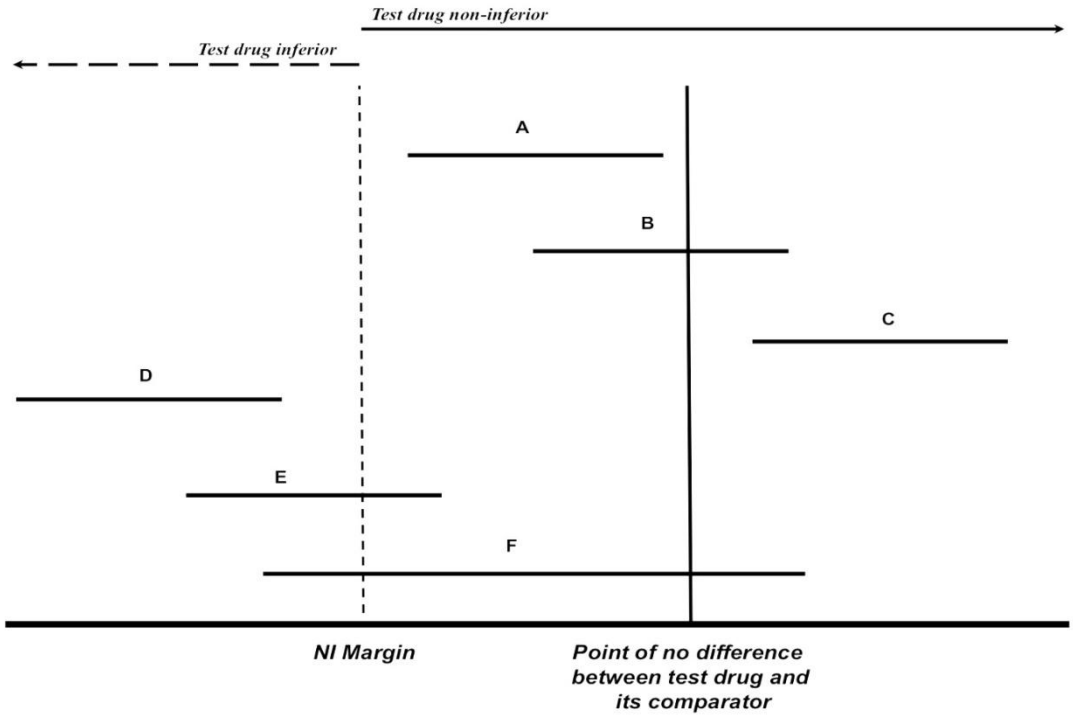
4.2. NI margin determination :

As we already defined what are NI margins, we will now focus on the concept of determination. It is emphasized and discussed in detail in the US Food and Drug Administration's new draft guideline for NI trials. (25) The guideline was composed based on previous guidelines and methodological publications on NI trials (26) published since the 1980s. Its focus is on showing indirect efficacy of the test drug compared with placebo. The method presented in the guidelines consists in determining two parameters M1 and M2. **M1** is the **statistical** part, an **objective parameter**. It represents the effect of the active control compared with placebo, assumed present in the NI trial. . The statistical margin for the effect between the new drug and active control is set at the upper boundary of the 95% confidence interval of difference between placebo and active control. This value is obtained from relevant previous placebo-controlled trial of the active comparator. The best would be a meta-analysis of several placebo-controlled trials for a better estimate of the active comparator. The second step is to calculate **M2 from M1** by choosing a certain amount of the effect to be preserved, called the **clinical relevance margin**. The draft FDA guideline implicitly recommends using a preserved-effect of 50% to determine M2. Choosing a higher percentage to be preserved results in a stricter or more conservative NI margin, meaning it is more difficult to conclude NI. For example, if it was concluded that it would be necessary for a test drug to preserve 75% of a mortality effect, M2 would be 25% of M1. The formula to calculate M2 for a risk difference is : (1-preserved effects)*-(M1). For the determination of this margin, it is the choice of a clinician, which makes it **subjective parameter.** It is related to how much of the treatment effected is judged necessary, a consideration that may reflect the seriousness of the outcome, the benefit of the active comparator and the relative safety profiles of the test drug and comparator. This factor has considerable practical implications. For example, in large cardiovascular studies, it is unusual to seek retention of more than 50% of the effect of the

control drug, even if this might be clinically reasonable, because doing so will usually cause the size of the study to become infeasible. (4) How investigators incorporate this clinical judgment remains unknown. These implicit clinical judgments might have been derived from clinical experience. However, these judgments remain subjective and different clinicians may propose contradicting judgments. That is why it is important to determine how this clinical judgment can be incorporated in the NI margin determination.

4.3. Interpretation of NI trials

The inference from the result of an NI trial is based on the CI of the treatment difference between the new drug and its comparator. NI is inferred when the CI is at the correct side and excludes the NI margin. (27). To illustrates this, we categorized the possible CIs in NI trials into six types as presented in Figure 7:



[Figure 7: The confidence interval categories and non-inferiority interpretation. *Horizontal line represents CI. The point-of-no difference is the point at which the estimate treatment difference between the new drug and comparator is neutral: zero for a difference in outcome or one for a ratio.(8)]*

Types A,B, and C can be defined as non-inferior , since their CI excludes the NI margin. In types D,E and F non-inferiority is not shown. Type C lies completely beyond the point-of-no difference line, would potentially demonstrate that the new drug is superior to its comparator. The switch from NI to superiority as we saw earlier is not excluded. But it is of course regulated by the Committee for Proprietary Product guidelines for example. (28) Type A also requires cautious interpretation. Although the lower limit lies above the NI margin, thereby showing NI, the upper limit lies below the point-of-no-difference, indicating that the new drug is actually statistically inferior to its comparator. However, the new drug can still be claimed

to be clinically non-inferior if the NI margin was determined on the basis of clinical relevance.

4.4. Sensitivity and constancy assumption

Assay sensitivity is defined as the ability of an RCT to distinguish an effective treatment from an ineffective treatment. A drug is considered effective if it shows a significant treatment effect as compared with placebo. In a superiority trial, a significant difference between two treatments directly confirms assay sensitivity. In contrast, a NI does not directly show the efficacy of both drugs as compared with placebo. A NI could mean that both drugs were effective, but it could also mean that both drugs were ineffective. One possible solution is to include a placebo arm to confirm that both the new drug and the comparator drug are better than placebo. When designing NI trials, other options should be considered before making the decision to omit a placebo arm. For example to assign fewer people in the placebo group and shorten the duration of treatment, or to create an adaptive trial design in which placebo non-responders can be reallocated. Another related assumption that can't be verified within the trial is the **constancy assumption**, which states that the effect of the active comparator versus placebo is present in the current trial. The determination of the NI margin directly relies on the size of the estimated treatment effect between the active comparator and the placebo. For the inference to be valid, it has to be assumed that this estimate is accurate for the trial at hand. And this can't be assessed completely objectively. However, it can be supported by a proper meta-analysis and by a demonstration of similarity between the current trial and the trials used for setting the margin. The constancy assumption also relies on the absence of any influence from a number of factors, such as changes in standard of care, which are not easily verifiable. The question, therefore, remains as to whether the NI trials and

placebo-controlled trials were similar enough. If a placebo are can't be included, the authors should discuss how they have arrived at the conclusion that the trial has assay sensitivity and provide data-driven as well as clinical reasons for assuming that the constancy assumption is true. Without these assessments, the reader can't reliably judge whether the conclusions from the trial are valid and relevant for treatment decisions. In order to increase objectivity, more guidance is needed to improve adequate and consistent determination of clinically acceptable NI margins.

5. Discussion.

5.1 Ethics

One of the first reasons why some people want to ban NI trials is that they are considered unethical, since they do not offer any possible advantage to present on future patients, and they disregard patient's interests in favor of commercial ones. They are believed to fail to meet the commitments of good clinical research: 'Ask an important question, and answer it reliably'. (29) But we clearly explained previously that even when a trial is set out to prove that a new drug has additional benefit a part of the study still has to assess non-inferiority of the primary outcome. Although one could claim that in any trial a superiority aim should be included, non-inferiority for other outcomes remains an important additional goal. Patients can be involved in clinical trials only if there is a reasonable potential advantage, for them or for future patients. The advantage could be increased efficacy, decreased toxicity, different toxic profiles, better compliance, longer duration of action, and other sizable factors. NI trials also allow patients and doctors to have the possibility of choosing among different drugs. This has a particular interest for patients who do not respond to standard treatment. In this case non-inferiority trials offer a very useful alternative. (4)

5.2 Methodological argument

As we saw in the previous section the determination factors in determining NI margins are assay sensitivity and constancy assumption which relies heavily on clinical evidence. The second argument in favor to ban NI trials is the **methodological argument**. The fact that a NI trial can't be objectively determined. It is true that interpretation and inference of NI trials are complicated, partly because of the incompleteness of the information. A research in PubMed in February 2009 that randomly selected 300 NI trial publication showed that <50% of the trials reported the method used to determine the NI margin, and <10% of the trials stated that the NI margin was a priori justified on the basis of clinical margin. They also found out that >8% were interpreted incorrectly, and <10% of them included placebo arms to ensure assay sensitivity or discuss the validity of constancy assumption. (3) The International Conference on Harmonization of Technical Requirements for Registration of Pharmaceuticals for Human Use, the European Medicines Agency, and the most recent draft of US Food and Drug Administration guidelines emphasize that determination of the NI margin should be based on both clinical and statistical margins. Meanwhile, it is hard to determine from the literature whether the trials actually followed the guidelines. So far, most of the efforts to overcome the methodological challenges in NI trials have concentrated on this issue, and better regulation is expected. Furthermore, even though the use of 'outdated' placebo-controlled trials data might not be avoidable, increasing knowledge on the evidence base of the drugs and the disease itself, the size of the estimated treatment effect between the active comparator and the placebo can be more accurately defined.

5.3. Subjectivity

Efforts to reduce subjectivity in clinical testing have been studied more extensively. They include the patient's perception that might have an importance in the results and has to be taken into account, and use of a systematic scoring system in defining a minimal clinically importance difference. Most, if not all of the regulations and guidelines focus on statistical methods. Unfortunately, the interpretation and reporting of trial results from the perspective of clinical importance has not received similar emphasis. This imbalance promotes the historical tendency to consider clinical trials results that are statistically significant as also clinically important, and conversely, those with statistically insignificant results as being clinically unimportant. (30)

The strongly criticized subjectivity in clinical judgment, this part that can't be defined in a statistical matter applicable in any trial also might prevent the drugs tested to gradually move to less effective treatments. In other words, perception of the investigators plays an important role in preventing biocreep. This degradation of the efficacy of the investigational treatment is theoretically not a mistake on the paper, only an outside judgment can prevent this cyclical phenomenon.(31)

5.4 Same drug, different results

One of the main problems encountered while using NI trial is the variability and results. Let's take the example of a NI trial designed to study apixan, a specific factor Xa inhibitor that may provide effective thromboprophylaxis with a lower risk of bleeding and improved ease of use than low-molecular-weight heparins such as enoxaparin. The study plan was based on the hypothesis that apixaban would be non-inferior to enoxaparin with respect to the primary

efficacy outcome which includes incidence in all VTE and death from any cause, with the use of a prespecified NI margin in which the upper limit of the 95% confidence interval for relative risk did not exceed 1,25 and for absolute risk 5,6 percentage points. Both criteria had to be met to establish non-inferiority. Results showed 9% incidence in primary outcome compared to 8,8% with enoxaparin. The criteria were not met because the relative risk interval was from 0,78 to 1.32, although the absolute risk was <5,6. The reason why the trial was expecting better outcome is that the results were expected to be much different for enoxaparin. Indeed, the assumptions made in establishing the criteria for NI and calculating the sample size were based on previous clinical trials. The judgment of outcome events in this trial was consistent with the trial were about 16% in the control group given enoxaparin. In the trial, the 8,8% incidence of the primary efficacy outcome in patients treated with enoxaparin was only 55% of the predicted rate. This made it difficult to meet the prespecified criteria for non-inferiority. (32) This is the case in many other studies on thromboprophylactic drugs and clinical trial as a whole.

Clinical trials are a very expensive undertaking, consuming a great deal of time and resources. To compare the efficacy of different drugs, dosages, surgeries or combinations of these treatments can cost over $5OO million and take many years, so it is of great importance that the design of the clinical trial gives a good chance of successfully demonstrating a treatment effect. Meanwhile, the benefit of the patient is and should always be the main goal, before the success of a drug development. This is why regulators need to have attention for unproved claims of additional benefit in NI trials to avoid misuse of the results of NI trials as a cover for unethical marketing. Last but not least we can say that there is still ample room to improve the determination of the NI trials and especially the NI margin. To support it, dialogue with

regulators to solve specific issues in NI trials could be improved, for example through scientific advice.

Acknowledgements

I would like to thank my mentor the Professor Vladimir Trkulja for his support as my academic supervisor. It was a pleasure to write this thesis. Through this work, I learned a lot about the world of industry, strategies, and how it is regulated. Also, as part of a career plan it helped me to know more about different jobs and what they really mean, hopefully guiding me to make good choices in the future.

References

(1): Wikipedia.org https://en.wikipedia.org/wiki/Clinical_trial. Accessed May 2016.

(2) http://health-information/nih-clinical-research-trials-you/basics Accessed May 2016

(3): Ryan M. Superiority, Equivalence, and Non-inferiority trials, bulletin of the NYU Hospital for joint disease 2008;66(2):150-4

(4) Grace Wangge. Non-inferiority trials: methodological and regulatory challenges. 2012. 150 pages.

(5) : United States Food and Drug Administration www.fda.gov . Accessed May 2016

(6) : Bilic-Zulle L, Dogas Z,Grcevic D, Hren D, Huic M, Ivanis A, Katavic V, Lukic I, Marusic A, Petrak J, Petrovecki M, Sambunjak D. Principles of Research in Medecine. Matko Marusic; 2007. 295pages.

(7) : Ronald P. Evens . Drug and Biological Development: from molecule to Product and Beyond.; Springer 2007, 383 pages.

(8): Patrick Vavken . Rationale for Methods of Superiority, Noninferiority, or Equivalence Designs in Orthopaedic, Controlled Trials; Clin Orthop Relat Res. 2011; 469:2645-2653

(9): Student. On testing varieties of cereal. Biometrika. 1923;15:271-293

(10): Essential Evidence-Based Medicine second edition, Dan Mayer, 2010, 442 pages.

(10): Erik Christensen Methodology of superiority vs. equivalence trials and non-inferiority trials, Journal of Hepatology. 46 .2007; 947-954

(11): Gonzalez C, Bolanos R, de Sereday M, Editorial on hypothesis and ojectives in clinical trials: superiority, equivalence and non-inferiority. Thromb J.2009; 7:3.

(12) : Tracy M, Methods of Sample Size Calculation for Clinical Trials www.theses.gla.ac.uk/ MSC_Feb_2009.pdf accessed May 2016

(13) : Shargel. Applied Biopharmaceutics and Pharmacokinetics 2nd edition : Appleton-Century-Crofts; 1985, p129-131

(14) : American Medical Association. Reporting of Noninferiority and Equivalence Randomized Trials, extension of the CONSORT 2010 Statement. JAMA.; 2010; Vol 308, No24

(15): Walker and Nowacki. Understanding Equivalence and Noninferiority Testing.,J Gen Intern Med. 2011; 26(2):192-196

(16): Lin, Gamalo-Siebers and Tiwari. Non-inferiority and networks: inferring efficacy from a web data. Wiley Online Library. 2015;10.1002

(17): Everson-Syewart S, Emerson SS. Bio-creep in non-inferiority clinical trials. Stat Med. 2010; 29: 2769-2780

(18): Pocock SJ. The pros and cons of noninferiority trials. Fundam Clin Pharmacol. 2003;17:483-490

(19): Merriam G and C. Webster's Third International Dictionary, Unabridged. 1976. 453pages.

(20) Owen R. Reader bias. Journal of American Medical Association 1982;247:2533-2534

(21): Chamers T.C.,Smith S.J, Blackburn B. A method for assessing the quality of a randomized clinical trial. Control Clin Trials 1981;2:31-49

(22): Ouyang E. and Rosario. Population Health and Epidemiology. Essential Med Notes 2015. P.7

(23): Altman D.G.,Bland J.M. How ro randomize? BMJ 1999;319:703-704.

(24): Hollis S. Campbell F. What is meant by intention to treat analysis? Survey of published randomized controlled trials BMJ 1999;319:670-674

(25): Center for Drug Evaluation and Research (CDER) Center for Biologics Evaluation and Research (CBER). 2010 DRAFT GUIDANCE: Guidance for industry non-inferiority clinical trials.

(26): Lange S, Freitag G. Choice of delta: requirements and reality, results of a systematic review. Biom J 2005 Fed;47(1):12-27;99-107.

(27) . Senn S. Active control equivalence studies. Statistical Issues in Drug Development, 2nd Edition Glasgow,UK:Wiley;2007.p.235-47

(28) Hung HMJ, Wang S, O'Neill R. Challenges and regulatory experiences with non-inferiority trial design without placebo arm. Biometrical Journal 2009;51(2):324-34

(29): Garattini S, Bertele V. Non-inferiority trials are unethical because they disregard patients' interests. The Lancet October 2007.

(30): Man-Son-Hing M, Laupacis A, Wells G, Determination of the Clinical Importance of Study Results. J Gen Intern Med. 2002 Jun; 17(6): 469-476.

(31): Beryl P, Vach W. Is there a danger of 'biocreep' with non-inferiority trials? Trials 2011 Dec 2013;12 Suppl 1:A29

(32): Lassen M, Raskob G, Gallus A, Pineo G, Chen D, Portman R. Apixaban or Enoxaparin. N Engl J Med 2009; incidence in all VTE and death 361:594-604

Biography

I was born in Lyon in 1991. After graduating from high school in Lyon I studied medicine in the same city for two years. It was called Grange Blanche initially before it fused with two other faculties forming Lyon-Est medical pole. In 2010 I transferred to the University of Zagreb in order to continue my studies in the English program. After completing his medical training in Zagreb, I will go to Lyon in order to attend a professional master in management and marketing of health industries, taking place in IMIS school which is part of IGS group.