# Non-Inferiority Trials in Surgical Oncology

Philipp Fueglistaler, MD,[1] Michel Adamina, MD,[1,2] and Ulrich Guller, MD, MHS[1,2]

[1]Department of Surgery, Divisions of General Surgery and Surgical Research, University of Basel, Basel, Switzerland
[2]Department of Surgery, University of Toronto, Toronto, Canada

The classical randomized controlled clinical trial is designed to prove superiority of an investigational therapy over an established therapy or placebo (here referred to as "superiority trial"). Although the randomized controlled superiority trial has its well-grounded role, clinical trials of non-inferiority are equally important in the advance of medical science. Non-inferiority trials test whether a new intervention is as good as a standard treatment with respect to curing the illness (e.g., overall survival) while offering other benefits over the standard therapy, such as lower toxicity, better side-effect profile, improved ease of administration, or reduced costs. The evaluation of non-inferiority is critical in many settings. In surgical oncology, for instance, treatments often combine advantages (e.g., survival benefit) with disadvantages (e.g., high post-operative morbidity due to extensive surgery, considerable toxic effects of an aggressive chemotherapy regimen). The various aspects of different therapeutic strategies may make a treatment decision difficult, requiring a non-inferiority trial to quantify risks and benefits. However, despite their great importance in clinical cancer research, the concept, design, and objectives of non-inferiority trials remain poorly understood in the surgical community. The goal of this review is to discuss the principles, strengths, and challenges of non-inferiority trials and introduce this highly relevant topic to the surgical reader, using examples from the field of surgical oncology.

**Key Words:** Non-inferiority trial—Surgical oncology—Statistics—Ethics—Assay sensitivity—Biocreep.

Traditional clinical trials compare a new treatment with a placebo or with the standard treatment with the aim of demonstrating superiority of the new treatment (referred to as "superiority trials"). Efficacy is most reliably demonstrated in this setting if a statistically significant difference is found. However, superiority trials are not always best suited to answer the clinical research question at hand.

For instance, in surgical oncology, clinical trials frequently have several endpoints. One endpoint, often overall survival, is designated as the primary endpoint. However, there are other relevant outcomes such as post-operative quality of life, short- and long-term morbidity, cosmesis, and costs, all of which are important for both patients and health care providers. For instance, a small decrease in overall survival may be justified—or even desired by patients—if there is a corresponding large gain in post-operative quality of life. In this setting, a non-inferiority trial may be the most appropriate study design.

Non-inferiority trials are performed to assess whether a new, investigational treatment or surgical procedure has equivalent (non-inferior) efficacy in treating the illness (e.g., similar overall survival) to an established treatment while offering other, relevant advantages. In surgical oncology, we are challenged by the rapid pace of change in surgical techniques. These new surgical techniques may be less invasive, more convenient and faster to perform, less expensive, or associated with an improved post-operative quality of life, compared to the existing procedure. However, prior to abandoning the standard tech-

nique, it is essential to assess whether the new procedure has equivalent (or non-inferior) efficacy regarding overall survival.

Moreover, if an effective therapy already exists, performing a placebo-controlled randomized clinical trial may pose ethical questions.[1–5] In this setting, it is appropriate either to conduct a superiority trial comparing a new treatment to a standard treatment, or to design a non-inferiority trial to compare a new therapeutic approach to an established standard therapy, assuming the new therapy to be equivalent (non-inferior) to the standard therapy with respect to efficacy while having other significant advantages.

## STATISTICS IN THE DESIGN OF NON-INFERIORITY TRIALS

### Non-Inferiority Margin ("delta", $\delta$)

From a statistical point of view, proving equivalence is like squaring the circle: it is impossible.[3,6–8] If one were to try to prove "equivalence", the sample sizes for such a trial would be infinitely large.[9] Thus, the investigators must specify a non-inferiority margin ("equivalence margin" or "delta", $\delta$).[6] This margin states how similar the performance of a new intervention must be relative to the standard treatment to be considered "non-inferior" or "equivalent". Hence, delta can be described as the chosen definition of inferiority in a non-inferiority trial. It is critically important that the non-inferiority margin be specified prior to the initiation of the study (a priori hypothesis).[10] The choice of delta will be the chief determinant of the sample size.[11] The sample size is proportional to $1/delta^2$: the smaller delta, the larger the sample size (and vice versa).

Choosing the appropriate non-inferiority margin for a given study is a challenging but essential task.[12] The first step is to estimate the expected effect of the standard therapy. Some investigators have suggested that the non-inferiority margin should be smaller than the expected effect size of the standard treatment to ensure that the investigational treatment has at least some efficacy.[2,10,12] However, in many contexts this premise is not sufficient and investigators require a greater percentage of efficacy of the investigational treatment compared to the standard one. The choice of a non-inferiority margin usually depends on all of the following factors: (a) known effect of standard therapy over placebo, (b) severity of the disease under investigation, (c) toxicity concerns, and (d) primary study endpoint.[13] Therefore, choosing

the appropriate delta represents a difficult task that requires extensive knowledge and input from both statisticians and clinicians.[10,14] The non-inferiority or equivalence margin may reflect the least clinically relevant difference. A new therapy may then be accepted in a non-inferiority trial if the difference between investigational and standard treatment is smaller than delta. This precludes the possibility that there exists a clinically relevant difference between the two treatments.[15] Ideally, a smaller delta should be selected if a more severe condition is being studied and/or if mortality is the primary endpoint. However, there is no true consensus for selection of non-inferiority margins[9], which may vary even among trials in similar therapeutic areas.

### Research Hypotheses in Non-Inferiority Trials

Traditional randomized clinical trials aim to prove a clinically relevant difference between standard and investigational therapy (alternative hypothesis, Ha), while the null hypothesis (H0) states that no difference between the treatments under investigation exists.[19] Importantly, a traditional superiority trial failing to reject the null hypothesis of no difference between treatments does not demonstrate equivalence of the treatments.

In the design of a non-inferiority trial, the null hypothesis (H0) is that the treatment in arm 2 (investigational treatment) is worse than the treatment in arm 1 (standard treatment) by a difference equal to or greater than the non-inferiority margin (Table 1).[20] If, however, the difference between the investigational and the standard treatment is smaller than the pre-specified equivalence margin (delta, $\delta$) or if the investigational treatment is superior to the standard treatment, the null hypothesis (H0) can be rejected, providing evidence supporting the adoption of the new treatment (Table 1).

Some statisticians and researchers differentiate between "non-inferiority" and "equivalence trial" designs.[16,17,18] The former refers to a one-sided trial design (difference in outcome < delta), whereas the latter has a two-sided study design (difference in outcome between −delta and +delta). Because of their intrinsic similarities in design and reporting and for the sake of simplicity, this educational review will not differentiate between non-inferiority and equivalence trials.

### Type I and Type II Error

A type I error (synonym: alpha) is a false-positive result[21] regardless of the type of study. In a non-

**TABLE 1.** *Research hypotheses in non-inferiority and superiority clinical trials*

| Non-inferiority trial | Conventional superiority trial |
|---|---|
| H0: $E_s - E_i \geq$ delta (inferiority) | H0: $E_s \geq E_i$, or H0: $E_s - E_i \geq 0$ |
| Ha: $E_s - E_i <$ delta (non-inferiority) | Ha: $E_s < E_i$, or Ha: $E_s - E_i < 0$ |

$E_s$, efficacy of standard treatment; $E_i$, efficacy of investigational treatment; delta: non-inferiority margin.

The null hypothesis (H0) of a non-inferiority trial states that the absolute difference between standard and investigational treatment exceeds the non-inferiority margin (delta). The alternative hypothesis (Ha) states that the absolute difference is smaller than delta.

**TABLE 2.** *Type I and II errors*

| Non-inferiority trial | Conventional superiority trial |
|---|---|
| Type I error $= \alpha$ (false positive) | Type I error $= \alpha$ (false positive) |
| $\rightarrow$ False conclusion that $E_S - E_i <$ delta | $\rightarrow$ False conclusion that $E_S < E_i$, or $\rightarrow$ False conclusion that $E_s - E_i < 0$ |
| Type II error $= \beta$ (false negative) | Type II error $= \beta$ (false negative) |
| $\rightarrow$ False conclusion that $E_S - E_i \geq$ delta | $\rightarrow$ False conclusion that $E_S \geq E_i$, or $\rightarrow$ False conclusion that $E_s - E_i \geq 0$ |

$E_s$, efficacy of standard treatment; $E_i$, efficacy of investigational treatment; delta: non-inferiority margin.

A type I error (false-positive result) in a non-inferiority trial is to falsely conclude that the new therapy is equivalent to, or superior than the standard therapy. A type II error (false-negative result) in a non-inferiority trial is to falsely conclude that the new therapy is inferior to the standard therapy.

inferiority trial a type I error is to falsely conclude that the new therapy is equivalent (= non-inferior) to or even superior than the standard therapy when this is not the case. Conversely, a type II error (beta, false-negative result)[21] in a non-inferiority trial is to falsely conclude that the new therapy is not equivalent (= inferior) to the standard therapy when both treatments are in fact equally efficient (Table 2).[6,19] In a non-inferiority trial, committing a type I error (obtaining a false-positive result) puts future patients at risk of having a poorer outcome (e.g., overall survival) as a result of the use of the new therapy which—in reality—is inferior to the standard treatment.[22] The risk of obtaining a false-positive result can be minimized either by setting the type I error probability below the standard of 0.05 (e.g., $\alpha = 0.01$) or by choosing a very small non-inferiority margin (delta). Such precautionary methods may, however, lead to very large sample sizes[6,11,19,23] potentially associated with enormous costs and efforts and thereby threatening the feasibility of the study.

If a false-negative result (type II error) occurs in a non-inferiority trial, future patients will not benefit from a new therapeutic approach with equal efficacy but other significant advantages (e.g., less invasive surgical procedure, less toxic chemotherapy regimen). The risk of obtaining false negative result can be minimized either by setting the type II error probability below the standard of 0.10–0.20 (e.g., $\beta = 0.05$) or by choosing a larger non-inferiority margin.

Importantly, although the chosen non-inferiority margin influences the probability of committing type I and II errors, the selected delta must be set based on meaningful clinical criteria, and should not be driven by statistical constraints.

## Sample Size

Sample size computations are crucial in the design of clinical trials and help to estimate the needed number of subjects for a given study. The importance for appropriate sample size computations is obvious: if the sample size is too large, resources will be wasted and the study may even be ethically questionable. Conversely, if the sample size is too small, the study will be underpowered thereby hindering researchers from drawing meaningful conclusions.[12,19,24]

In a conventional superiority trial, the power of a study is the probability of finding a statistically significant result for a specified true difference in the overall population.[19,21] An underpowered study due to a too small sample size may result in a statistically non-significant difference (a false-negative result), even if a clinically relevant difference exists, and hence may lead to erroneous conclusions.[19,24,25]

In a non-inferiority trial, the power of a study is the probability of finding equivalence (non-inferiority) between standard and investigational treatment if in fact the two therapies are equivalent in the overall population. Power is related to the alternative hypothesis, i.e., generally "no difference" in a non-inferiority trial. Furthermore, power is associated with the probability of committing a type II error ($\beta$, false-negative result): power $= 1 - \beta$. The higher the power, the lower the risk of committing a type II error. Conversely, an underpowered study will increase the risk of committing a type II error and may bias the study toward the null hypothesis (a false-negative result), even if the two treatments are equivalent.[3,26] It is therefore absolutely critical to perform sample size computations early in the process of designing the study. Only thoroughly performed sample size estimates and consecutively

resulting adequately large patient samples will prevent from obtaining false-negative findings, and hence from potentially hindering patients of the benefits of the investigational treatment.

The sample size in a non-inferiority trial depends, among other parameters, on the accepted rate of false-positive results (chosen type I error, alpha), the accepted rate of false-negative results (chosen type II error, beta), and the selected non-inferiority margin (delta). Sample size, alpha, beta, and delta are the four important parameters in the design of a non-inferiority trial. Investigators can specify only three of these parameters and subsequently determine the fourth. Generally, a margin of non-inferiority (delta) is first determined, which must be based on the expected efficacy of both standard and investigational treatments. The accepted rate of false-positivity (alpha) is then chosen as well as the power of the study (complementary to beta). Based on these parameters an adequate sample size can be calculated.

## CHALLENGES AND CAVEATS IN THE CONDUCT AND INTERPRETATION OF NON-INFERIORITY TRIALS

### Assay Sensitivity

Assay sensitivity refers to the ability of a study to differentiate between active and inactive treatments.[27] In a superiority trial, which aims to prove superiority of an investigational treatment over an established treatment or placebo, there is compelling evidence of the efficacy of this treatment if a statistically significant difference is found.[15] In this setting, no information from outside of this trial is needed to conclude efficacy.[27] Conversely, a non-inferiority trial may not address the efficacy of the treatments under investigation, but it does assess whether any difference in efficacy between the treatments is significantly less than the chosen delta (equivalence or non-inferiority margin). Hence, the investigator must assume that the standard treatment was in fact effective in the performed study to conclude efficacy of the investigational treatment.[15] This assumption of efficacy can be made if the standard treatment consistently demonstrated efficacy for the condition in question in previous placebo-controlled studies.[2,12,15,27] Assay sensitivity represents a critical issue in non-inferiority trials and efficacy of the investigational treatment cannot be demonstrated if the assumption of the assay sensitivity is impossible.[11,27]

The challenge of assay sensitivity could be removed if a three-arm (standard treatment vs. investigational treatment vs. placebo) non-inferiority trial were designed.[11] The alternative hypothesis would then be that the new, investigational treatment is non-inferior to the standard treatment but performs significantly better than placebo. However, the use of surgical placebos (sham surgery) is highly controversial and thus, surgical placebos are rarely used.[25]

### Improper Study Design, Conduct, or Interpretation

Improper study design and conduct (e.g., non-compliance of patients, drop-outs, inclusion of ineligible patients, etc.) is likely to decrease an existing difference between two treatments.[12,23] In a traditional superiority trial, investigators are keen in avoiding such "sloppiness",[3] in order to preclude the possible blurring of an existing difference. Furthermore, a failure to find a statistically significant difference between standard and investigational therapy might be due to an underpowered study design, to the occurrence of too few events, or to an imprecise measurement of the outcome (large variability).[19,24,26]

Conversely, in a non-inferiority trial, improper study design or conduct may bias the results towards a finding of equivalence (the alternative hypothesis).[23] Non-inferiority trials are particularly sensitive to protocol deviations. For instance, deviations from the inclusion criteria, from the intended treatment (crossover), or from the treatment schedule may result in smaller differences between standard and investigational treatment and hence make a conclusion of non-inferiority more likely. Thus, the investigators may wrongly conclude that the new investigational treatment is non-inferior to the standard treatment (false-positive result), whereas in fact, the standard treatment is superior with respect to efficacy compared with the new treatment. Future patients who undergo the new treatment may be at risk of having a poorer outcome due to this false conclusion. Thus, it is critically important that non-inferiority trials be performed with greatest attention to methodological rigor and study conduct.[12] Indeed, the existent literature documents that only a few trials concluding no difference between two treatments under investigation were actually adequately designed and conducted as non-inferiority trials with sufficient power. Greene et al.[9] recently reviewed 88 published controlled clinical trials and only found 23% of studies that reported equivalence actually setting an equivalence boundary and confirming it statistically. In addition, only 33% of reports calculated a sample size necessary to

confirm their results prior to conducting the trial. The authors suggest that many investigators conclude similarity based upon inappropriate statistical tests or inadequate sample sizes, possibly leading to unreliable and potentially dangerous information being employed in clinical practice.[9]

### "Intention-to-Treat" Versus "Per Protocol" Analysis

Intention-to-treat represents a central tenet of performing randomized clinical superiority trials. Intention-to-treat refers to the axiom that patients enrolled in a randomized clinical trial are analyzed according to their initial arm assignment, regardless whether or not they received the assigned treatment. It is well known that some patients cross over from the assigned treatment arm to the other arm. In a conventional superiority trial, an intention-to-treat analysis usually decreases the difference in the outcome between the study groups and provides a conservative approach to analysis by shifting results towards the null hypothesis (the hypothesis of no difference). Similarly, in a non-inferiority trial an intention-to-treat analysis may decrease observed differences between the compared arms. This will, however, bias the results towards the alternative hypothesis. Therefore, we share the opinion that a non-inferiority trial should be evaluated both by "intention-to-treat" as well as "per protocol" analyses.[2,3] If non-inferiority is found in both analyses, we can be reasonably certain that both treatments are in fact similar if study design and conduct were appropriate.

### Bio-Creep

Bio-creep refers to the phenomenon that the efficacy of investigational treatment could degrade over time as they are repeatedly compared to less efficacious treatments.[2] Although bio-creep is mostly used in the context of drug studies, it may also occur in surgical clinical trials. For instance, a standard surgical procedure (A) with an efficacy of 75% was originally determined to be superior to an initial procedure (0), which had an efficacy rate of 55%. In a non-inferiority trial, a new, less-invasive procedure (B) was compared to the standard procedure (A) and was found to be non-inferior with a delta less than 20% and efficacy rates of 60% (procedure B) and 75% (procedure A). Another investigational surgical procedure (C) was later found to be non-inferior to procedure (B) with a delta less than 20% and efficacy rates of 45% (procedure C) and 60% (procedure B). Thus, by these criteria, one could erroneously assume
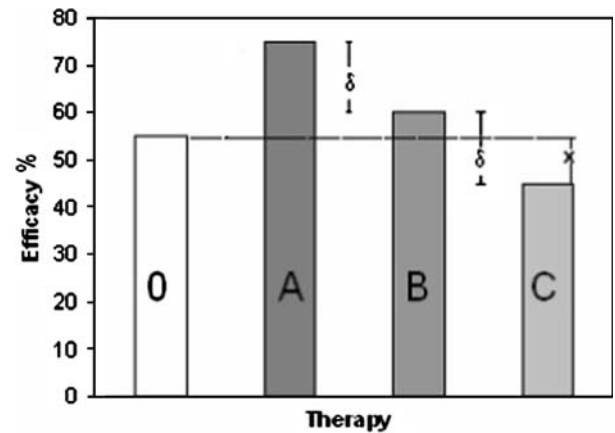


**FIG. 1.** Biocreep in a non-inferiority trial. $A$ = Standard surgical procedure (efficacy of 75%), originally superior to initial procedure (0, efficacy of 55%). $B$ = New therapy (efficacy 60%) non-inferior compared to $A$ with a delta less than 20%. Another therapy $C$ (efficacy 45%) non-inferior to therapy $B$ with a delta less than 20%. Erroneous conclusion that $C = B = A$, which is superior to 0. In fact, the efficacy of therapy $C$ is worse than 0 by a difference $x$.

procedure (C) = procedure (B) = standard procedure (A), which was superior to the initial procedure (0), when in fact the efficacy of procedure C may be worse than historically demonstrated for procedure (0) (Fig. 1). This phenomenon is termed "biocreep" and underscores the cardinal importance of selecting an appropriate comparator therapy and non-inferiority margin for a non-inferiority trial.[2]

### Ethical Considerations in Performing a Non-Inferiority Trial

The unique design of non-inferiority trials also involves particular human ethics considerations.[28] Based upon the uncertainty principle, a patient should only be entered into a clinical trial if it is unclear which treatment might be more beneficial for the patient.[28,29] It is critically important to carefully select the investigational treatment and endpoints for a non-inferiority trial. Since the null hypothesis in a non-inferiority trial requires investigators to assume that the investigational treatment is inferior to the standard treatment, it has been suggested that this information must be shared with potential subjects, possibly affecting their willingness to participate in the trial.[28]

## EXAMPLES OF NON-INFERIORITY TRIALS IN SURGICAL ONCOLOGY

### Example 1

A landmark study of a comparison between two surgical therapies for colon cancer will serve us as a

first example of a non-inferiority trial in surgical oncology:[30,31] in this multicenter trial, 872 patients with histologically proven adenocarcinoma of the colon were randomized to undergo either open colectomy or laparoscopic-assisted colectomy for colon cancer. There is a growing body of evidence that laparoscopic colon surgery is associated with better cosmesis, reduced incidence of early infectious wound complications as well as late incisional hernia formations, decreased post-operative pain, shorter hospital stay, and faster recovery.[32–38]

On the basis of all these advantages laparoscopic colon surgery would seem preferable to open colectomy. However, until recently, it was unknown whether laparoscopic surgery was as good (non-inferior) to open surgery with respect to disease-free and overall survival. Thus, the aim of this trial was to evaluate if disease-free and overall survival are equivalent (non-inferior) after laparoscopic compared to the conventional open approach. The trial was designed with a power of 81% (type II error probability: 19%) to declare that the laparoscopic procedure is inferior if indeed it were associated with at least a 23% increase (non-inferiority margin, delta) of cancer recurrence. Alpha (type I error probability) was set at 5%.

After a median follow-up of 4.4 years, both disease-free and overall survival were similar in both arms, suggesting that the laparoscopic-assisted colectomy is a viable alternative with respect to efficacy (disease-free and overall survival) while having significant advantages over the conventional approach.

### Example 2

A multicenter randomized clinical trial in surgical oncology serves as another excellent example of a non-inferiority trial.[39] This study enrolled patients with cancer of the distal half of the stomach. Patients were randomized to undergo either total gastrectomy (arm 1) or subtotal gastrectomy (arm 2). The primary endpoint of the study was overall survival at 5-year follow-up.

This non-inferiority trial aimed at clearing a decade-long controversy whether to perform a total or subtotal gastrectomy in patients with distal gastric cancer. Previously, a single randomized controlled trial addressed this relevant research question in 1989.[40] However, this previous study was methodologically suboptimal (e.g., underpowered trial, lacking control for prognostic variables). A total gastrectomy may reduce the likelihood of recurrence at the proximal resection margin, and may eliminate

the risk of developing a metachronous carcinoma in the gastric remnant, suggesting a superiority in terms of oncologic outcome. However, former non-randomized studies did not show a better overall survival in patients undergoing total gastrectomy compared to those having a subtotal gastrectomy.[41–44] Furthermore, it is well known that a total gastrectomy, which requires an esophageal anastomosis, is the more extensive operation and may be associated with an increased morbidity, nutritional deficiencies, a lacking preservation of physiologic functions of the gastric remnant, and a reduced quality of life.[45–47] Therefore, it would be of great clinical and potentially economic benefit if the more extensive total gastrectomy could be replaced by a less traumatic procedure, such as the distal gastrectomy in patients with distal gastric cancer. Hence, a non-inferiority trial was designed to compare 5-year overall survival rates between patients undergoing total and subtotal gastrectomy. Total gastrectomy was taken as a reference treatment, with an expected 5-year survival of 50%. The null hypotheses (H0) stated that the 5-year survival for subtotal gastrectomy was 40% or less. Conversely, the alternative hypothesis (Ha) stated equivalent 5-year overall survival rates for total and subtotal gastrectomies, with a non-inferiority margin of 10%, a type I error ($\alpha$) probability set at 0.05, and a type II error ($\beta$) probability set at 0.2. Power calculations based on these assumptions resulted in a total sample size of 600 patients. A total of 618 patients were recruited, hence this trial was adequately powered. The estimated 5-year overall survival was 62.4% in the total gastrectomy group and 65.3% in the subtotal gastrectomy group. Therefore, the null hypothesis was rejected and the alternative hypothesis of equivalence of both treatment options was accepted. This non-inferiority trial considerably contributed to our current surgical practice of performing subtotal gastrectomy in patients with distal gastric cancer, and helped to improve the post-operative quality of life for many of these patients.

### CONCLUSIONS

The non-inferiority study in surgical oncology is a unique research tool to conduct a clinical trial demonstrating non-inferiority for investigational treatments. This particular study design can provide evidence that the investigational treatment is as efficacious as the standard treatment with respect to the primary endpoint (e.g., overall survival), yet has fewer side effects, is less expensive, easier to administrate, or

is associated with improved post-operative quality of life. However, the design and study conduct of non-inferiority trials is associated with particular challenges and caveats. Particular attention to rigorous methodology must be paid when planning, performing, or interpreting results of a non-inferiority trial.

We hope that the present overview clarifies some of the most important aspects of non-inferiority trials and helps the surgical oncologist in the interpretation and implementation of their results.

## QUALITY OF A NON-INFERIORITY TRIAL—A CHECKLIST FOR THE SURGEON

- Is the study designed as a non-inferiority trial? Remember: superiority trials that fail to reach statistical significance do not prove equivalence.
- Is the clinical effectiveness of the standard treatment well established (assay sensitivity)? Remember: the standard therapy in a non-inferiority trial must perform as expected from previous clinical trials.
- Are type I error (alpha, false positive) and type II error (beta, false negative) stated in the Method section?
- Is a non-inferiority margin (delta) defined in the Method section?
- Is there a sample size calculation based on pre-stated alpha, beta, and non-inferiority margin (delta)?
- Is the chosen non-inferiority margin smaller than the least clinically meaningful difference? Would the investigational treatment be acceptable if its efficacy was within the boundaries of the chosen delta?
- Does the study reach its target accrual? Is there a considerable gap between computed sample size and enrolled number of patients?
- Was the data analysis performed according to the intention-to-treat principle or is it a per protocol analysis? Remember that both evaluations are important in a non-inferiority trial and should result in similar conclusions for robust evidence of equivalence.

## REFERENCES

1. World Medical Association Declaration of Helsinki Ethical principles for medical research involving human subjects. *JAMA* 2000; 284:3043–45.
2. D'Agostino RB Sr, Massaro JM, Sullivan LM. Non-inferiority trials: design concepts and issues—the encounters of academic consultants in statistics. *Stat Med* 2003; 22:169–86.
3. Jones B, Jarvis P, Lewis JA, Ebbutt AF. Trials to assess equivalence: the importance of rigorous methods. *BMJ* 1996; 313:36–9.
4. Rothman KJ, Michels KB. The continuing unethical use of placebo controls. *N Engl J Med* 1994; 331:394–98.
5. Taubes G. Use of placebo controls in clinical trials disputed. *Science* 1995; 267:25–6.
6. Blackwelder WC. "Proving the null hypothesis" in clinical trials. *Control Clin Trials* 1982; 3:345–53.
7. Cullen M, Stenning S. Clinical trials with moving targets: a commentary on a non-inferiority trial in testicular cancer. *Lancet Oncol.* 2004; 5:129–32.
8. Friedman LM. Basic study design. In: Friedman LM, Furberg CD, DeMets DL, (eds). Fundamentals of clinical trials. 3rd ed. Berlin Heidelberg New York: Springer (1998) pp 94–129.
9. Greene WL, Concato J, Feinstein AR. Claims of equivalence in medical research: are they supported by the evidence?. *Ann Intern Med* 2000; 132:715–22.
10. Wiens BL. Choosing an equivalence limit for noninferiority or equivalence studies. *Control Clin Trials* 2002; 23:2–14.
11. Pigeot I, Schafer J, Rohmel J, Hauschke D. Assessing non-inferiority of a new treatment in a three-arm clinical trial including a placebo. *Stat Med* 2003; 22:883–99.
12. Siegel JP. Equivalence and noninferiority trials. *Am Heart J* 2000; 139:S166–70.
13. Laster LL, Johnson MF. Non-inferiority trials: the ,at least as good as' criterion. *Stat Med* 2003; 22:187–200.
14. U.S. Department of Health and Human Services, Food and Drug Administration, Center for Drug Evaluation and Research, Center for Biologics Evaluation and Research. Guidance for Industry. E10 Choice of Control Group and Related Issues in Clinical Trials, section 1.5.1.1 entitled "Historical Evidence of Sensitivity to Drug Effects and Choosing the Non-inferiority Margin". 2001. 12-7-2006.
15. Rothmann M, Li N, Chen G, Chi GY, Temple R, Tsou HH. Design and analysis of non-inferiority mortality trials in oncology. *Stat Med* 2003; 22:239–64.
16. Gotzsche PC. Lessons from and cautions about noninferiority and equivalence randomized trials. *JAMA* 2006; 295:1172–1174.
17. Le Henanff A, Giraudeau B, Baron G, Ravaud P. Lessons from and cautions about noninferiority and equivalence randomized trials. *JAMA* 2006; 295:1147–1151.
18. Piaggio G, Elbourne DR, Altman DG, Pocock SJ, Ewans S. Lessons from and cautions about noninferiority and equivalence randomized trials. *JAMA* 2006; 295:1152–1160.
19. Guller U, Oertli D. Sample size matters: a guide for surgeons. *World J Surg* 2005; 29:601–5.
20. Blackwelder WC, Chang MA. Sample size graphs for "proving the null hypothesis". *Control Clin Trials* 1984; 5:97–105.
21. Guller U, Delong ER. Interpreting statistics in medical literature: a vade mecum for surgeons. *J Am Coll Surg* 2004; 198:441–58.
22. Guller U, Blumenstein BA. Trends in clinical trials in surgical oncology: implications for outcomes research. *Clin Ther* 2003; 25:684–98.
23. Browner WS, Newman TB, Cummings SR, Hulley SB. Estimating sample size and power: the nitty-gritty. In: Hulley SB, Newman TB, Cummings SR, Hearst N, Grady S, Browner WS, (eds). Designing clinical research. 2nd ed. Philadelphia: Lippincott Williams and Wilkins (2001) pp. 65–91.
24. Dunnett CW, Gent M. Significance testing to establish equivalence between treatments, with special reference to data in the form of 2 × 2 tables. *Biometrics* 1977; 33:593–602.
25. Guller U. Surgical outcomes research based on administrative data: inferior or complementary to prospective randomized clinical trials?. *World J Surg* 2006; 30:255–66.
26. Cummings SR, Grady D, Hulley SB. Designing an experiment: clinical trials I. In: Hulley SB, Newman TB, Cummings SR, Hearst N, Grady S, Browner WS, (eds). Designing clinical

research, 2nd ed. Philadelphia: Lippincott Williams and Wilkins (2001) pp. 143–55.

27. Temple R, Ellenberg SS. Placebo-controlled trials and active-control trials in the evaluation of new treatments. Part 1: ethical and scientific issues. *Ann Intern Med* 2000; 133:455–63.

28. Djulbegovic B, Clarke M. Scientific and ethical issues in equivalence trials. *JAMA* 2001; 285:1206–08.

29. Peto R, Baigent C. Trials: the next 50 years. Large scale randomised evidence of moderate benefits. *BMJ* 1998; 317:1170–71.

30. The Clinical Outcomes of Surgical Therapy Study Group A comparison of laparoscopically assisted and open colectomy for colon cancer. *N Engl J Med* 2004; 350:2050–59.

31. Nelson H, Weeks JC, Wieand HS. Proposed phase III trial comparing laparoscopic-assisted colectomy versus open colectomy for colon cancer. *J Natl Cancer Inst Monogr* 1995; 51–6.

32. Abraham NS, Young JM, Solomon MJ. Meta-analysis of short-term outcomes after laparoscopic resection for colorectal cancer. *Br J Surg* 2004; 91:1111–24.

33. Braga M, Vignali A, Gianotti L, et al. Laparoscopic versus open colorectal surgery: a randomized trial on short-term outcome. *Ann Surg* 2002; 236:759–66.

34. Braga M, Frasson M, Vignali A, Zuliani W, Civelli V, Di Carlo V. Laparoscopic vs. open colectomy in cancer patients: long-term complications, quality of life, and survival. *Dis Colon Rectum* 2005; 48:2217–23.

35. Dunker MS, Stiggelbout AM, van Hogezand RA, Ringers J, Griffioen G, Bemelman WA. Cosmesis and body image after laparoscopic-assisted and open ileocolic resection for Crohn's disease. *Surg Endosc* 1998; 12:1334–40.

36. Guller U, Jain N, Hervey S, Purves H, Pietrobon R. Laparoscopic vs open colectomy: outcomes comparison based on large nationwide databases. *Arch Surg* 2003; 138:1179–86.

37. Lumley J, Stitz R, Stevenson A, Fielding G, Luck A. Laparoscopic colorectal surgery for cancer: intermediate to long-term outcomes. *Dis Colon Rectum* 2002; 45:867–72.

38. Seshadri PA, Poulin EC, Schlachta CM, Cadeddu MO, Mamazza J. Does a laparoscopic approach to total abdominal colectomy and proctocolectomy offer advantages?. *Surg Endosc* 2001; 15:837–42.

39. Bozzetti F, Marubini E, Bonfanti G, Miceli R, Piano C, Gennari L. Subtotal versus total gastrectomy for gastric cancer: five-year survival rates in a multicenter randomized Italian trial. Italian Gastrointestinal Tumor Study Group. *Ann Surg* 1999; 230:170–8.

40. Gouzi JL, Huguier M, Fagniez PL, et al. Total versus subtotal gastrectomy for adenocarcinoma of the gastric antrum. A French prospective controlled study. *Ann Surg* 1989; 209:162–6.

41. Harrison LE, Karpeh MS, Brennan MF. Total gastrectomy is not necessary for proximal gastric cancer. *Surgery* 1998; 123:127–30.

42. McNeer G, Bowden L, Booner RJ, McPeak CJ. Elective total gastrectomy for cancer of the stomach: end results. *Ann Surg* 1974; 180:252–6.

43. Launois B, Cardin JL, Bardaxoglou E, et al. Management of cancer of the stomach: total gastrectomy versus sub-total gastrectomy. *Hepatogastroenterology* 1991; 38:45–52.

44. Gennari L, Bozzetti F, Bonfanti G, et al. Subtotal versus total gastrectomy for cancer of the lower two-thirds of the stomach: a new approach to an old problem. *Br J Surg* 1986; 73:534–8.

45. Davies J, Johnston D, Sue-Ling H, et al. Total or subtotal gastrectomy for gastric carcinoma? A study of quality of life. *World J Surg* 1998; 22:1048–55.

46. Jentschura D, Winkler M, Strohmeier N, Rumstadt B, Hagmuller E. Quality-of-life after curative surgery for gastric cancer: a comparison between total gastrectomy and subtotal gastric resection. *Hepatogastroenterology* 1997; 44:1137–42.

47. Bozzetti F. Total versus subtotal gastrectomy in cancer of the distal stomach: facts and fantasy. *Eur J Surg Oncol* 1992; 18:572–9.