Non-inferiority studies: non-sense or sense?

Prof. Emmanuel Lesaffre
Department of Biostatistics, Erasmus MC, Rotterdam, the Netherlands
L-Biostat, K.U.Leuven, Leuven, Belgium

VOLUME 27 · NUMBER 17 · JUNE 10 2009

JOURNAL OF CLINICAL ONCOLOGY

ORIGINAL REPORT

Phase III Noninferiority Trial Comparing Irinotecan With Oxaliplatin, Fluorouracil, and Leucovorin in Patients With Advanced Colorectal Carcinoma Previously Treated With Fluorouracil: N9841

George P. Kim, Daniel J. Sargent, Michelle R. Mahoney, Kendrith M. Rowland Jr, Philip A. Philip, Edith Mitchell, Abraham P. Mathews, Tom R. Fitch, Richard M. Goldberg, Steven R. Alberts, and Henry C. Pitot

Purpose

The primary goal of this multicenter phase III trial was to determine whether overall survival (OS) of fluorouracil (FU) -refractory patients was noninferior when treated with second-line infusional fluorouracil, leucovorin, and oxaliplatin (FOLFOX4; arm B) versus irinotecan (arm A). Cross-over to the other treatment on disease progression was mandated.

Rheumatology 2007;46:496–507
Advance Access publication 27 August 2006
doi:10.1093/rheumatology/kel296

Efficacy and safety of etoricoxib 30 mg and celecoxib 200 mg in the treatment of osteoarthritis in two identically designed, randomized, placebo-controlled, non-inferiority studies

C. O. Bingham III, A. I. Sebba[1], B. R. Rubin[2], G. E. Ruoff[3], J. Kremer[4], S. Bird[5], S. S. Smugar[5], B. J. Fitzgerald[5], K. O'Brien[5] and A. M. Tershakovec[5]

Objective. To compare the efficacy of etoricoxib 30 mg with the generally maximum recommended dose of celecoxib, 200 mg, in the treatment of osteoarthritis (OA) in two identically designed studies.

Methods. Two multi-centre, 26-week, double-blind, placebo-controlled, non-inferiority studies were conducted, enrolling patients who were prior non-steroidal anti-inflammatory drug (NSAID) or acetaminophen users. There were 599 patients in study 1 and 608 patients in study 2 randomized 4:4:1:1 to etoricoxib 30 mg qd, celecoxib 200 mg qd or one of two placebo groups for 12 weeks. After 12 weeks, placebo patients were evenly distributed to etoricoxib or celecoxib based on their initial enrollment randomization schedule. The primary hypothesis was that etoricoxib 30 mg would be at least as effective as celecoxib 200 mg for the time-weighted average change from baseline over 12 weeks for Western Ontario and McMaster (WOMAC) Pain Subscale, WOMAC Physical Function Subscale and Patient Global Assessment of Disease Status. Active treatments were also assessed over the full 26 weeks. Adverse experiences were collected for safety assessment.

Results. In both studies, etoricoxib was non-inferior to celecoxib for all three efficacy outcomes over 12 and 26 weeks; both were superior to placebo ($P < 0.001$) for all three outcomes in each study over 12 weeks. The safety and tolerability of etoricoxib 30 mg qd and celecoxib 200 mg qd were similar over 12 and 26 weeks.

Conclusions. Etoricoxib 30 mg qd was at least as effective as celecoxib 200 mg qd and had similar safety in the treatment of knee and hip OA; both were superior to placebo.

1. Review of study designs

- **Type of null-hypothesis:**
  - Superiority
  - Equivalence
  - Non-inferiority
- Here, only **two-group** comparisons: **E(experimental)** & **C(ontrol)**

1. Review of study designs

- **Superiority trial:** most classical design

  Prove that **E is better than C**

- **Equivalence trial:** showing bio-equivalence

  Show that **E is equivalent to C**

- **Non-inferiority trial:** popular in active –controlled trials

  Show that **E is not (much) worse than C**

2. Superiority trial

- **Example I: GUSTO-I study** (NEJM, 1993)
  - Comparison of two thrombolytic drugs: **SK (C)** and **rt-PA (E)**
  - Primary endpoint = 30-day mortality (binary)
  - SK: **10,370** patients rt-PA: **10,348** patients
  - **SK: 7.4%** **rt-PA: 6.3%**
  - **H_0: $\Delta = 0$**
  - Chi-square test: 8.94 **P=0.0028**
  - 95% C.I. for Δ: **[0.36%, 1.73%]**

2. Superiority trial

- **Example II: part of GUSTO-I**
  - Comparison of two thrombolytic drugs: **SK** and **rt-PA**
  - Primary endpoint = 30-day mortality (binary)
  - SK: **1,000** patients rt-PA: **1,000** patients
  - **SK: 7.4%** **rt-PA: 6.3%**
  - **H_0: $\Delta = 0$**
  - Chi-square test: 0.79 **P=0.37**
  - 95% C.I. for Δ: **[-1.21%, 3.21%]**

2. Superiority trial

- **Conclusions:**
  - Ex I: **significant result** $\Rightarrow$ SK & rt-PA have **different effect**
  - Ex II: non-significant result $\Rightarrow$ SK & rt-PA have **same effect???**
  - Can we conclude for example II that SK & rt-PA are NOT different?

*Non-significant result **DOES NOT** imply that two treatments are equally good(bad)*

*One can **NEVER** prove that two treatments are equally good (bad)*

2. Superiority trial

- **Classical result:**
  - **P < 0.05 ⇔ 95% C.I. does NOT include Δ = 0**
  - **P ≥ 0.05 ⇔ 95% C.I. includes Δ = 0**
  - **Two-sided 95% C.I. ⇔ 1-sided 97.5% C.I.**
- Classical test **= superiority** test
- Classical trial **= superiority** trial
- Assumed Δ (in sample size calculations) for superiority trial = Δ_S

2. Superiority trial

Rheumatology 2007;46:496–507
Advance Access publication 27 August 2006
doi:10.1093/rheumatology/kel296

Efficacy and safety of etoricoxib 30 mg and celecoxib 200 mg in the treatment of osteoarthritis in two identically designed, randomized, placebo-controlled, non-inferiority studies

C. O. Bingham III, A. I. Sebba[1], B. R. Rubin[2], G. E. Ruoff[3], J. Kremer[4], S. Bird[5], S. S. Smugar[5], B. J. Fitzgerald[5], K. O'Brien[5] and A. M. Tershakovec[5]

Objective. To compare the efficacy of etoricoxib 30 mg with the generally maximum recommended dose of celecoxib, 200 mg, in the treatment of osteoarthritis (OA) in two identically designed studies.

Methods. Two multi-centre, 26-week, double-blind, placebo-controlled, non-inferiority studies were conducted, enrolling patients who were prior non-steroidal anti-inflammatory drug (NSAID) or acetaminophen users. There were 599 patients in study 1 and 608 patients in study 2 randomized 4:4:1:1 to etoricoxib 30 mg qd, celecoxib 200 mg qd or one of two placebo groups for 12 weeks. After 12 weeks, placebo patients were evenly distributed to etoricoxib or celecoxib based on their initial enrollment randomization schedule. The primary hypothesis was that etoricoxib 30 mg would be at least as effective as celecoxib 200 mg for the time-weighted average change from baseline over 12 weeks for Western Ontario and McMaster (WOMAC) Pain Subscale, WOMAC Physical Function Subscale and Patient Global Assessment of Disease Status. Active treatments were also assessed over the full 26 weeks. Adverse experiences were collected for safety assessment.

Results. In both studies, etoricoxib was non-inferior to celecoxib for all three efficacy outcomes over 12 and 26 weeks; both were superior to placebo ($P < 0.001$) for all three outcomes in each study over 12 weeks. The safety and tolerability of etoricoxib 30 mg qd and celecoxib 200 mg qd were similar over 12 and 26 weeks.

Conclusions. Etoricoxib 30 mg qd was at least as effective as celecoxib 200 mg qd and had similar safety in the treatment of knee and hip OA; both were superior to placebo.

2. Superiority (trial)

TABLE 2a. Primary end points: analysis of TWA change from baseline (flare/randomization visit) averaged over weeks 2, 4, 8 and 12 (mITT Population), study 1

		Baseline		Treatment		*vs* Celecoxib 200 mg[a]			*vs* Placebo		
	N	Mean	s.d.	Mean	s.d.	Difference	95% CI	*P*	Difference	95% CI	*P*
WOMAC Pain Subscale											
Etoricoxib 30 mg	228	67.4	16.2	39.6	22.9	−3.12	(−7.02, 0.77)	0.116	−15.07	(−19.72, −10.41)	<0.001
Celecoxib 200 mg	236	67.5	16.3	42.8	22.9				−11.95	(−16.57, −7.32)	<0.001
Placebo	126	66.6	16.2	54.2	24.6						
WOMAC Physical Function Subscale											
Etoricoxib 30 mg	228	65.5	17.6	42.2	22.9	−1.74	(−5.53, 2.05)	0.367	−12.86	(−17.40, −8.31)	<0.001
Celecoxib 200 mg	236	66.6	17.9	44.6	23.2				−11.11	(−15.63, −6.59)	<0.001
Placebo	125	64.7	18.0	54.6	23.9						
Patient Global Assessment of Disease Status (PGADS)											
Etoricoxib 30 mg	228	72.2	17.6	41.3	22.7	−4.05	(−8.11, 0.02)	0.051	−16.44	(−21.31, −11.57)	<0.001
Celecoxib 200 mg	236	71.2	16.3	45.0	23.1				−12.39	(−17.23, −7.56)	<0.001
Placebo	126	69.1	18.1	56.7	23.6						

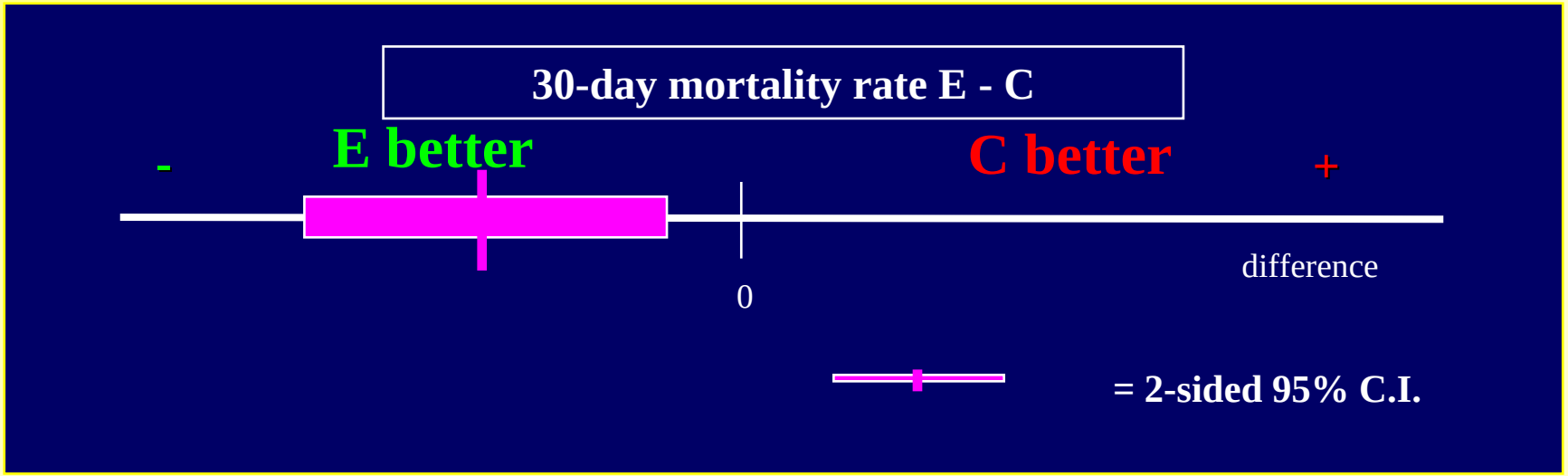
Superiority tests

2. Superiority trial

Aim superiority trial: Show that E is better than C

How?

Show that 95% C.I. does not contain 0 = significant at 0.05



$H_0: \Delta = 0$ $H_a: \Delta \neq 0$ at α

In fact only interested in:

$H_0: \Delta \geq 0$ $H_a: \Delta < 0$ at $\alpha/2$

Take home message 1

- A **non-significant** result **NEVER** implies that the 2 treatments are **EQUALLY GOOD**
- One can **NEVER** prove that 2 treatments are **EQUALLY GOOD**
- If we **believe** that 2 treatments are **EQUALLY GOOD**, then **another design** is needed $\Rightarrow$ **EQUIVALENCE TRIAL**

3. Equivalence trial

Aim: Prove that E is equally good as C

- **BUT, this can NEVER be done in practice**

=> Practical definition of **"equally good"** is needed

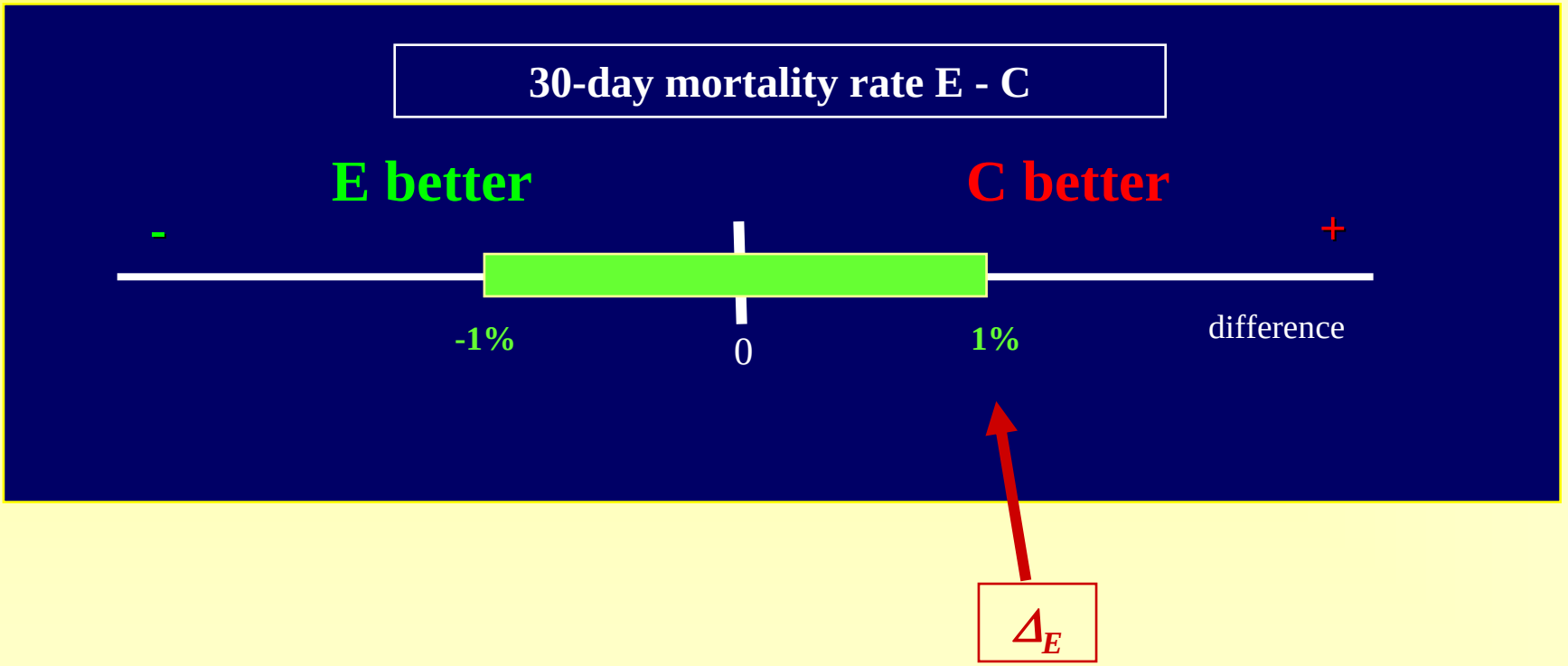
- Possible practical definition:

  2 treatments **do not differ** in effect
  more than a clinically justified value Δ_E

  => Specify interval of **clinical equivalence**

3. Equivalence trial

Interval of clinical equivalence

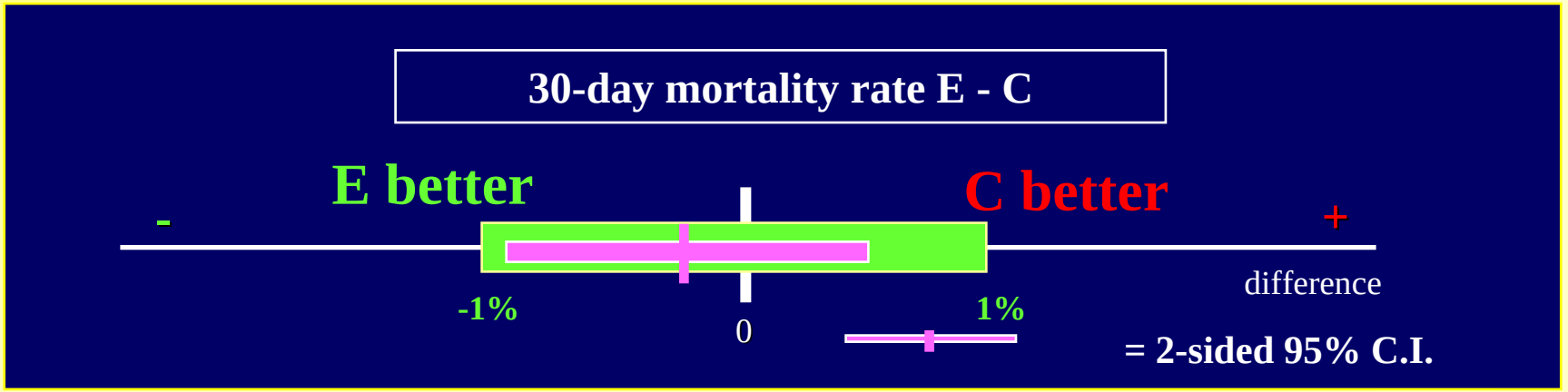


3. Equivalence trial

Aim equivalence trial: Show that E & C are clinically equivalent

How?

Show that 95% C.I. is INSIDE in interval of therapeutic equivalence



H_0: $\Delta > 1\%$ or $\Delta < -1\%$

H_a: $-1\% < \Delta < 1\%$

3. Equivalence trial

- **Clinical equivalence is often not the aim of a RCT**

- Most equivalence trials = **bioequivalence trials**
  to compare a **generic drug** with an **original drug**
  to show that they have the "same" **PK profile**
  $\Rightarrow$ PK variables C_{max}, C_{min} and AUC must be "close"

- In bioequivalence trials, often **90% CI is used**

- Δ_E = value such that: **"patient will not detect any change in effect when replacing one drug by the other"**

- **Noninferiority trials** (next) are often (wrongly) called equivalence trials

Take home message 2

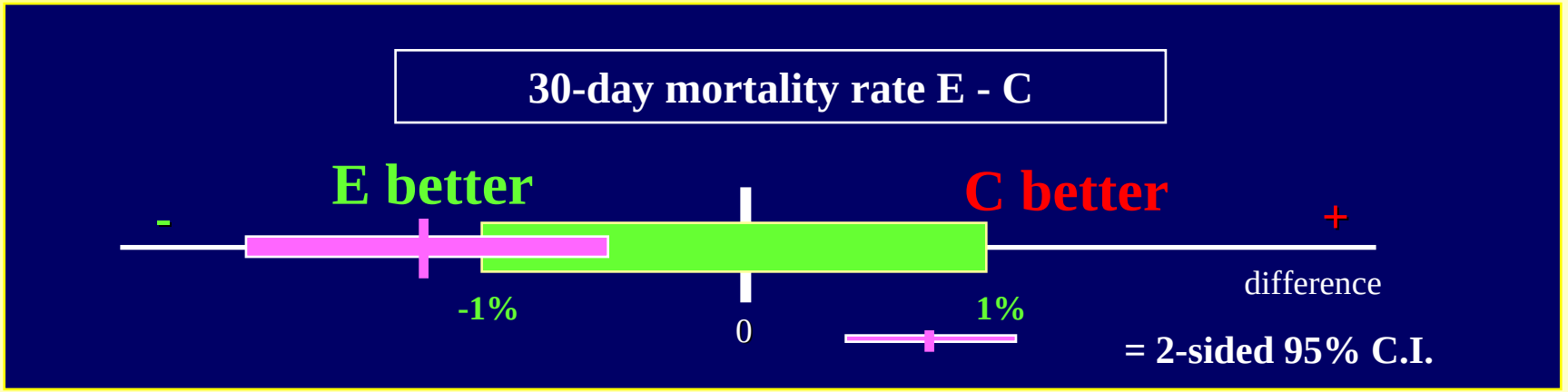
- If **you wish to prove** that two treatments are **EQUALLY GOOD** perform an **equivalence trial**

- **If significant**, then only proven that two treatments are **ROUGHLY** equally good (or bad)

4. Non-inferiority trial

Introduction

- Equivalence trials are **not appropriate** for therapeutic trials, e.g. if E is clearly superior to C then “equivalence” does not hold.



4. Non-inferiority trial

Introduction

- **What to do?**

⇒ **Prove that E is NOT worse than C ?**

? E better than C (superiority, but not believed)

? E equal to C (not possible to prove)

⇒ Prove that E is **NOT MUCH worse** than C**!**

➢ Specify a **margin (upper bound)** of what can be tolerated

➢ = **interval of clinical non-inferiority**

4. Non-inferiority trial

Introduction

Rheumatology 2007;46:496–507
Advance Access publication 27 August 2006
doi:10.1093/rheumatology/kel296

Efficacy and safety of etoricoxib 30 mg and celecoxib 200 mg in the treatment of osteoarthritis in two identically designed, randomized, placebo-controlled, non-inferiority studies

C. O. Bingham III, A. I. Sebba[1], B. R. Rubin[2], G. E. Ruoff[3], J. Kremer[4], S. Bird[5], S. S. Smugar[5], B. J. Fitzgerald[5], K. O'Brien[5] and A. M. Tershakovec[5]

Objective. To compare the efficacy of etoricoxib 30 mg with the generally maximum recommended dose of celecoxib, 200 mg, in the treatment of osteoarthritis (OA) in two identically designed studies.

Methods. Two multi-centre, 26-week, double-blind, placebo-controlled, non-inferiority studies were conducted, enrolling patients who were prior non-steroidal anti-inflammatory drug (NSAID) or acetaminophen users. There were 599 patients in study 1 and 608 patients in study 2 randomized 4:4:1:1 to etoricoxib 30 mg qd, celecoxib 200 mg qd or one of two placebo groups for 12 weeks. After 12 weeks, placebo patients were evenly distributed to etoricoxib or celecoxib based on their initial enrollment randomization schedule. The primary hypothesis was that etoricoxib 30 mg would be at least as effective as celecoxib 200 mg for the time-weighted average change from baseline over 12 weeks for Western Ontario and McMaster (WOMAC) Pain Subscale, WOMAC Physical Function Subscale and Patient Global Assessment of Disease Status. Active treatments were also assessed over the full 26 weeks. Adverse experiences were collected for safety assessment.

Results. In both studies, etoricoxib was non-inferior to celecoxib for all three efficacy outcomes over 12 and 26 weeks; both were superior to placebo ($P < 0.001$) for all three outcomes in each study over 12 weeks. The safety and tolerability of etoricoxib 30 mg qd and celecoxib 200 mg qd were similar over 12 and 26 weeks.

Conclusions. Etoricoxib 30 mg qd was at least as effective as celecoxib 200 mg qd and had similar safety in the treatment of knee and hip OA; both were superior to placebo.

4. Non-inferiority trial

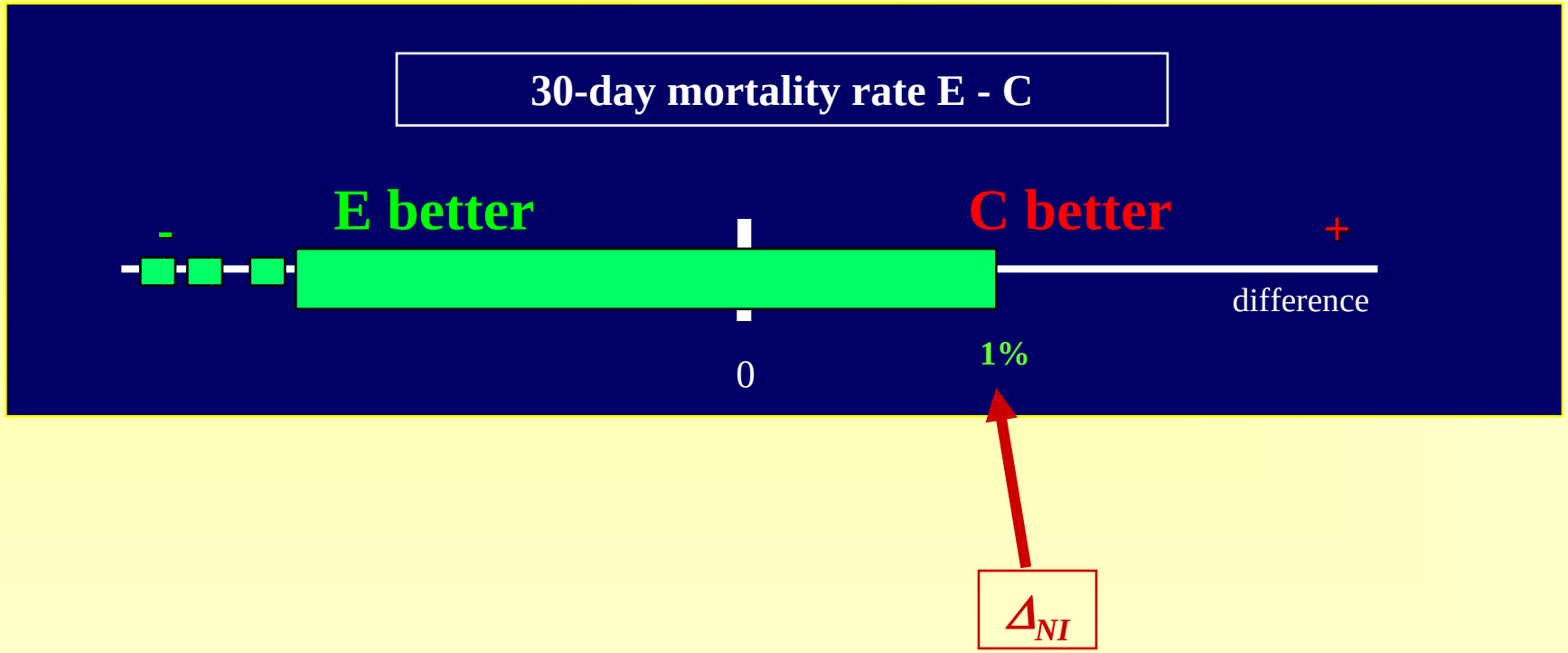
Introduction

- **Showing non-inferiority can be of interest because of:**
  - **Not ethically possible** to do a **placebo-controlled trial**
  - E is **not expected to be better** than C on primary efficacy endpoint, but is **better on secondary endpoints**
  - E is **not expected to be better** than C on primary efficacy endpoint, but is **safer**
  - E is **not expected to be better** than C on primary efficacy endpoint, but is **cheaper to produce or easier to administer**
  - E is **not expected to be better** than C on primary efficacy endpoint in clinical trial, but **compliance will be better** outside the clinical trial and hence **efficacy better outside the trial**

4. Non-inferiority trial

Interval of clinical non-inferiority

Interval of clinical non-inferiority



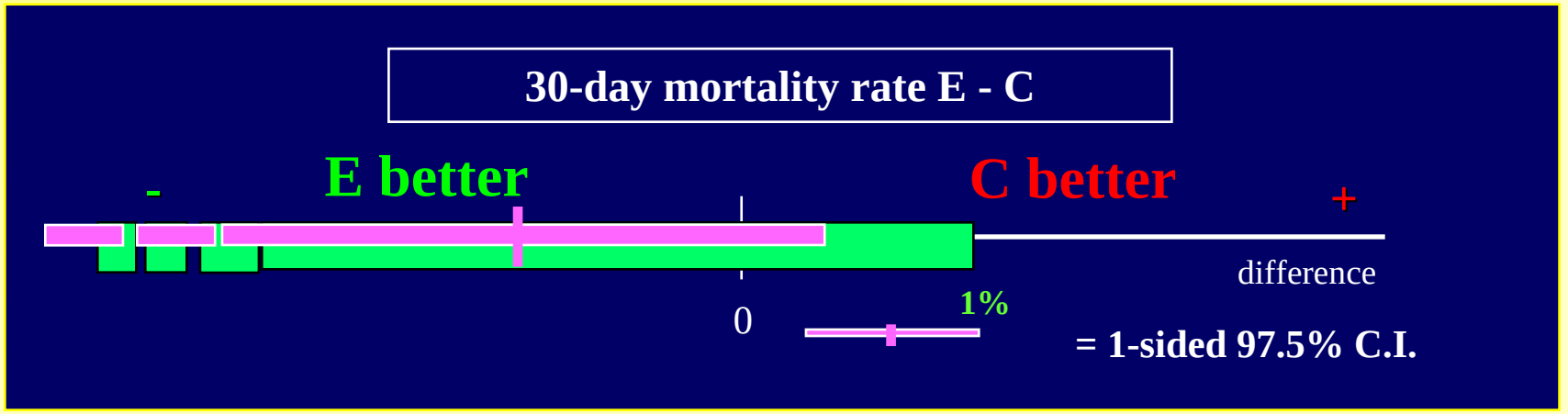
4. Non-inferiority trial

Interval of clinical non-inferiority

Aim non-inferiority trial:
Show that E is not (much) inferior to C

How?

Show that 95% C.I. is inside interval of therapeutic equivalence



H_0: $\Delta > 1\%$

H_a: $\Delta < 1\%$

4. Non-inferiority trial

Bingham et al.

With 200 patients each in the etoricoxib and celecoxib groups and 100 patients in the placebo group, each study provided an overall power of at least 87% to satisfy the primary hypothesis of non-inferiority between actives, and of actives demonstrating superiority over placebo. This assumes no difference between actives for the three co-primary end points, and active-placebo differences of −11.1, −10.2 and −11.5 mm [illegible]MAC physical function and PGADS TWA change from baseline, respectively, with standard deviations of 20.5, 20.1 and 22.0, respectively. To satisfy the primary hypothesis, the following were required: (i) the upper bound of the 95% confidence intervals (CIs) for difference between active treatments (etoricoxib 30 mg–celecoxib 200 mg) was not >10 mm with respect to the TWA change from baseline over 12 weeks for the three primary end points; and (ii) etoricoxib 30 mg qd was superior ($P \leq 0.05$) to placebo for the TWA change from baseline over 12 weeks for these end points.

Non-inferiority comparison

Superiority tests

4. Non-inferiority trial

Bingham et al.

10mm

TABLE 2a. Primary end points: analysis of TWA change from baseline (flare/randomization visit) averaged over weeks 2, 4, 8 and 12 (mITT Population), study 1

		Baseline		Treatment		*vs* Celecoxib 200 mg[a]			*vs* Placebo		
	N	Mean	s.d.	Mean	s.d.	Difference	95% CI	*P*	Difference	95% CI	*P*
WOMAC Pain Subscale											
Etoricoxib 30 mg	228	67.4	16.2	39.6	22.9	−3.12	(−7.02, 0.77)	0.116	−15.07	(−19.72, −10.41)	<0.001
Celecoxib 200 mg	236	67.5	16.3	42.8	22.9				−11.95	(−16.57, −7.32)	<0.001
Placebo	126	66.6	16.2	54.2	24.6						
WOMAC Physical Function Subscale											
Etoricoxib 30 mg	228	65.5	17.6	42.2	22.9	−1.74	(−5.53, 2.05)	0.367	−12.86	(−17.40, −8.31)	<0.001
Celecoxib 200 mg	236	66.6	17.9	44.6	23.2				−11.11	(−15.63, −6.59)	<0.001
Placebo	125	64.7	18.0	54.6	23.9						
Patient Global Assessment of Disease Status (PGADS)											
Etoricoxib 30 mg	228	72.2	17.6	41.3	22.7	−4.05	(−8.11, 0.02)	0.051	−16.44	(−21.31, −11.57)	<0.001
Celecoxib 200 mg	236	71.2	16.3	45.0	23.1				−12.39	(−17.23, −7.56)	<0.001
Placebo	126	69.1	18.1	56.7	23.6						

Non-inferiority comparison

Superiority P-values

4. Non-inferiority trial

Conclusion

- **Non-inferiority trials are quite simple**
  - **Compare two treatments with a 95% CI**
  - **If 95% CI is below (above) $\Delta_{NI} \Rightarrow$ E is <u>non-inferior</u> to C**
  - **If 95% CI is NOT below (above) $\Delta_{NI} \Rightarrow$ E is <u>NOT non-inferior</u> to C**
  - **Simple!**
  - **Simple?**
  - **How to determine the margin?**

The NEW ENGLAND JOURNAL of MEDICINE

ORIGINAL ARTICLE

A Comparison of Laparoscopically Assisted and Open Colectomy for Colon Cancer

The Clinical Outcomes of Surgical Therapy Study Group*

ABSTRACT

BACKGROUND

Minimally invasive, laparoscopically assisted surgery was first considered in 1990 for patients undergoing colectomy for cancer. Concern that this approach would compromise survival by failing to achieve a proper oncologic resection or adequate staging or by altering patterns of recurrence (based on frequent reports of tumor recurrences within surgical wounds) prompted a controlled trial evaluation.

METHODS

We conducted a noninferiority trial at 48 institutions and randomly assigned 872 patients with adenocarcinoma of the colon to undergo open or laparoscopically assisted colectomy performed by credentialed surgeons. The median follow-up was 4.4 years. The primary end point was the time to tumor recurrence.

4. Non-inferiority trial

Example non-inferiority

RESULTS

At three years, the rates of recurrence were similar in the two groups — 16 percent among patients in the group that underwent laparoscopically assisted surgery and 18 percent among patients in the open-colectomy group (two-sided P=0.32; hazard ratio for recurrence, 0.86; 95 percent confidence interval, 0.63 to 1.17). Recurrence rates in surgical wounds were less than 1 percent in both groups (P=0.50). The overall survival rate at three years was also very similar in the two groups (86 percent in the laparoscopic-surgery group and 85 percent in the open-colectomy group; P=0.51; hazard ratio for death in the laparoscopic-surgery group, 0.91; 95 percent confidence interval, 0.68 to 1.21), with no significant difference between groups in the time to recurrence or overall survival for patients with any stage of cancer. Perioperative recovery was faster in the

4. Non-inferiority trial

Motivation clinical boundary?

The NEW ENGLAND JOURNAL of MEDICINE

CORRESPONDENCE

TO THE EDITOR: Nelson et al. report on the Clinical Outcomes of Surgical Therapy (COST) trial, which compared laparoscopically assisted colectomy with open colectomy for colon cancer (May 13 issue).[1] Unfortunately, the methods described in their article fail to support the claim of noninferiority[2] for several reasons. First, the authors do not explicitly define a noninferiority boundary. Second, the statistical methods described in the article are those of a failed superiority trial rather than a noninferiority trial. Third, one approach to demonstrating noninferiority is to show that the upper limit of the one-sided 95 percent confidence interval for the hazard ratio is less than the noninferiority boundary. We have calculated that this value is 1.16 for the risk of death and 1.11 for the risk of recurrence with laparoscopic treatment.

4. Non-inferiority trial

Determination of margin (Δ_{NI})

- **Two ways to choose the margin Δ_{NI}:**
  - **Direct comparison (clinical reasoning)**: E ⇔ C
    - Δ_{NI} is determined on **clinical reasoning**
  - **Indirect comparison (putative placebo)**: E ⇔ P(lacebo) via C
    - Δ_{NI} is determined on **statistical reasoning**
  - **Combination:** E ⇔ C & E ⇔ P via C
    - Δ_{NI} is determined on **clinical** & **statistical** reasoning

4. Non-inferiority trial

Example SK versus rt-PA

- **Thrombolytic example:**
  - Choice of Δ_{NI} **= 1%** can be driven by **different reasonings** (clinical, statistical, clinical & statistical)
  - **Clinical:** 1% = largest difference **without causing concern** ← *Rarely used*
  - **Statistical:** 1% = difference => safely conclude **E better than P**
  - **Combined:** 1% = difference => safely conclude **E better than P & without causing concern** ← *Often used*

4. Non-inferiority trial

Margin determined clinically

- **Determine Δ_{NI} clinically:**
  - **Consensus on Δ_{NI}?** Not easy when YOU are performing the first non-inferiority study in that therapeutic domain (e.g. malaria study)
  - **Possible to find a clinically acceptable Δ_{NI}?** Difficult to justify (purely on clinical grounds) in a mortality trial (e.g. ASSENT II study)

*Establishing margin on **purely** clinical arguments is difficult*

4. Non-inferiority trial

Margin determined clinically & statistically

- **Determine Δ_{NI}**
  - Determine difference of C ⇔ P by e.g. a meta-analysis
    => **2% better**
  - Determine **95% C.I. around 2%** equal to, say, **[1.7%, 2.3%]**
  - Then E can be **at most 1.7% worse** than C to guarantee (with 95% confidence) that E is **better than P**
  - Thus **Δ_{NI} < 1.7%**

- Check if 1.7% is clinically acceptable, if not ⇒ lower Δ_{NI} (to say 1%)

4. Non-inferiority trial

Choice of Active Control

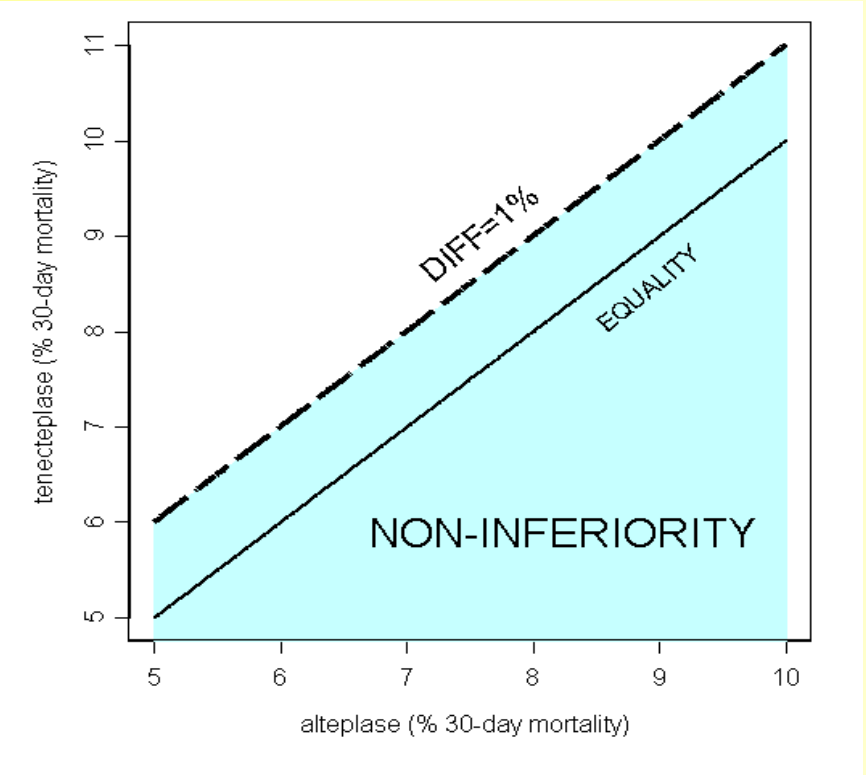
Some considerations

- **C** must have a **well-established, predictable, quantifiable effect**
  - Multiple placebo-controlled RCTs must be available
  - If not, then there is always the risk that E cannot be "proven" better as P
- **Constancy assumption**
  - C ⇔ P effect remains the same

4. Non-inferiority trial

ASSENT II study-1

- **ASSENT II** (one of the 1st NI trials in the area)
  RCT comparing single-bolus tenecteplase (E)
  with accelerated infusion of alteplase (C) in acute m.i.
- Primary endpoint = 30-day mortality
- When Δ_{NI} **= 1%**, region of non-inferiority

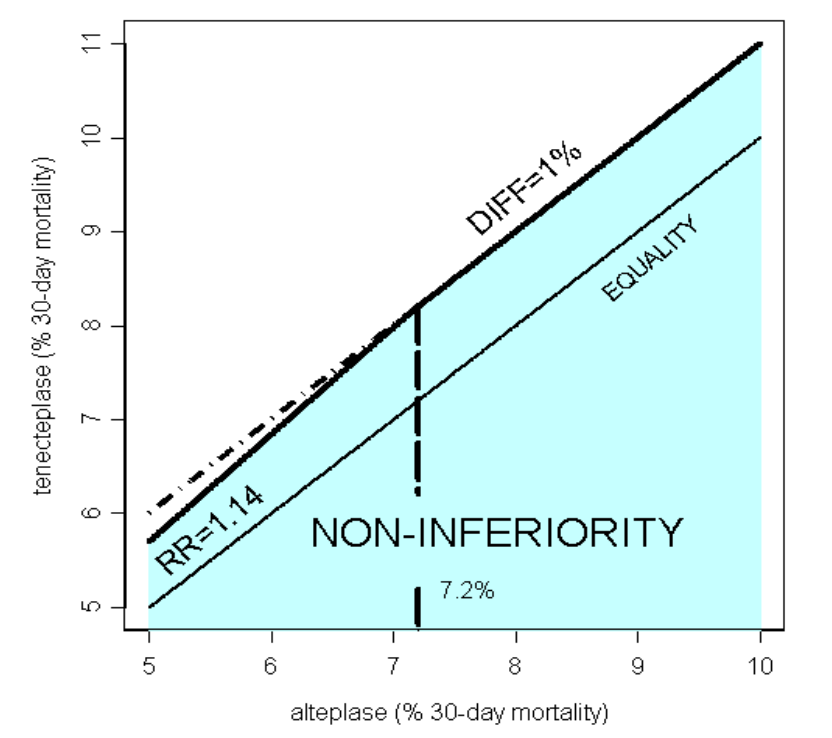


4. Non-inferiority trial

ASSENT II study-2

- **Problem** with non-inferiority region:
  for small mortality rates under alteplase,
  the **allowable relative risk is too high**.

- Let NI margin depend on true alteplase result:
  change-point at 7.2% (GUSTO III study)



4. Non-inferiority trial

ASSENT II study-3

Questions

- How was **absolute difference = 1%** chosen?
- How was **rr = 1.14** determined?

Determination of margin in collaboration with FDA

- Δ_{NI} **= 1%:** because of GUSTO-1 trial:
  alteplase was 1% better than SK and SK has proved to be better than placebo + taking 90% confidence intervals into account
- rr_{NI} **= 1.14**, a result of the Fibrinolitics Therapy Trialists meta-analysis showing the effect of SK + effect of rt-PA versus SK from GUSTO-1 study + taking 90% confidence intervals into account.

4. Non-inferiority trial

ASSENT II study-4

Results:

- **90% C.I.** was used instead of **95% C.I.** (early NI trial)
- Endpoint = 30-day mortality
- **E** (tenecteplase): **6.16%**
  C (alteplase): **6.18%**
- rr = **0.997**
- 90% C.I. = **[0.904, 1.101]**

Conclusion: **E not-inferior to C**

4. Non-inferiority trial

Malaria study-1

An **open randomized multi-centre clinical trial** in Africa, comparing 3 artemisinin-based combination treatments:

(1) ASMP (fixed dose over 3 days)

(2) ASMP (fixed dose over 24 hours)

(3) Artemether-Lumefantrine (AL) (fixed dose over 3 days) on **Plasmodium falciparum malaria**

ASMP fixed dose over 24 hours is easier to administer

Main objectives

1. To test the hypothesis that ASMP as fixed dose administered over 24 hours is **not inferior** in efficacy to the same drug administered over 3 days, measured by the primary endpoint: PCR corrected ACPR on day 28.
2. To test the hypothesis that ASMP as fixed dose is **not inferior** in efficacy to AL as follows ……

4. Non-inferiority trial

Malaria study-2

For the first non-inferiority analysis:

H_0: True proportion of cured patients treated with ASMP on 3 days

- True proportion of cured patients treated with ASMP on 24 hours $\geq$ **6 %**

The corresponding alternative hypothesis is:

H_a: True proportion of cured patients treated with ASMP on 3 days

- True proportion of cured patients treated with ASMP on 24 hours **< 6 %**

In early studies with the combination AL, recrudescence of malaria on day 28 was found to be low and varies between 0 and 5%. Re-infection is however sometimes rather high and can vary from 1 – 20 %, particularly in areas with high malaria transmission pressure (Mutabingwa et al., 2005). In some more recent studies, recrudescence was found to be 6 and 8 % respectively (Falade, 2005 and Martensson, 2005).

We conclude that, taking into account the studies obtaining a recrudescence of 0 to 5% and the studies mentioning a ***recrudescence of 6-8%, a non-inferiority interval bounded by 6% can be motivated.*** *Although, the* ***exact choice of the clinical difference*** *is* ***difficult*** *to make.*

4. Non-inferiority trial

ITT or PP analysis?

- **Analysis population**
  - **Superiority trial:**
    - Standard analysis is based on **ITT** (intention-to-treat) population
    - Reason = because of **conservative** effect of **ITT approach**
  - **Non-inferiority trial:**
    - ITT analysis is **NOT conservative**: dropouts and bad conduct of the study push the results of the 2 arms towards each other
    - **PP (per-protocol) analysis** is **preferred** but does **not** provide the ultimate answer
    - **Pragmatic approach**: do **PP & ITT analysis**

4. Non-inferiority trial

Sample size calculations

- **Sample size calculations**
  - **Superiority trial:** **n** depends on (among other things) on
    Δ_S = the clinically important difference
  - **NI trial:** **n** depends on (among other things) on
    Δ_{NI} = the upper-bound for non-inferiority
  - When $\Delta_S = \Delta_{NI}$ the **sample sizes are equal**
  - Δ_S for a **superiority trial** must be greater than Δ_{NI} in a **NI trial**

    $\Rightarrow$ **sample size of NI trial > > sample size of superiority trial**

4. Non-inferiority trial

Sample size calculations

- **Non-inferiority & superiority in the same trial**
  - Applied to the **same population** (ITT or PP)

    Non-inferiority & superiority are tested both at 0.05 (**no penalty**) because of **Closed Testing Principle**

  - When non-inferiority is applied to PP & superiority to ITT

    First non-inferiority & then superiority: **no penalty**

    First superiority & then non-inferiority: **penalty** (multiplicity adjustment)

4. Non-inferiority trial

CONSORT guidelines

Reporting of Noninferiority and Equivalence Randomized Trials

An Extension of the CONSORT Statement

Gilda Piaggio, PhD
Diana R. Elbourne, PhD
Douglas G. Altman, DSc
Stuart J. Pocock, PhD
Stephen J. W. Evans, MSc
for the CONSORT Group

The CONSORT (Consolidated Standards of Reporting Trials) Statement, including a checklist and a flow diagram, was developed to help authors improve their reporting of randomized controlled trials. Its primary focus was on individually randomized trials with 2 parallel groups that assess the possible superiority of one treatment compared with another but is now being extended to other trial designs. Noninferiority and equivalence trials have methodological features that differ from superiority trials and present particular difficulties in design, conduct, analysis, and interpretation. Although the rationale for such trials occurs frequently, those designed and described specifically as noninferiority or equivalence trials appear less commonly in the medical literature. The quality of reporting of those that are published is often inadequate. In this article, we present an adapted CONSORT checklist for reporting noninferiority and equivalence trials and provide illustrative examples and explanations for those items amended from the original CONSORT checklist. The intent is to improve reporting of noninferiority and equivalence trials, enabling readers to assess the validity of their results and conclusions.

JAMA. 2006;295:1152-1160 www.jama.com

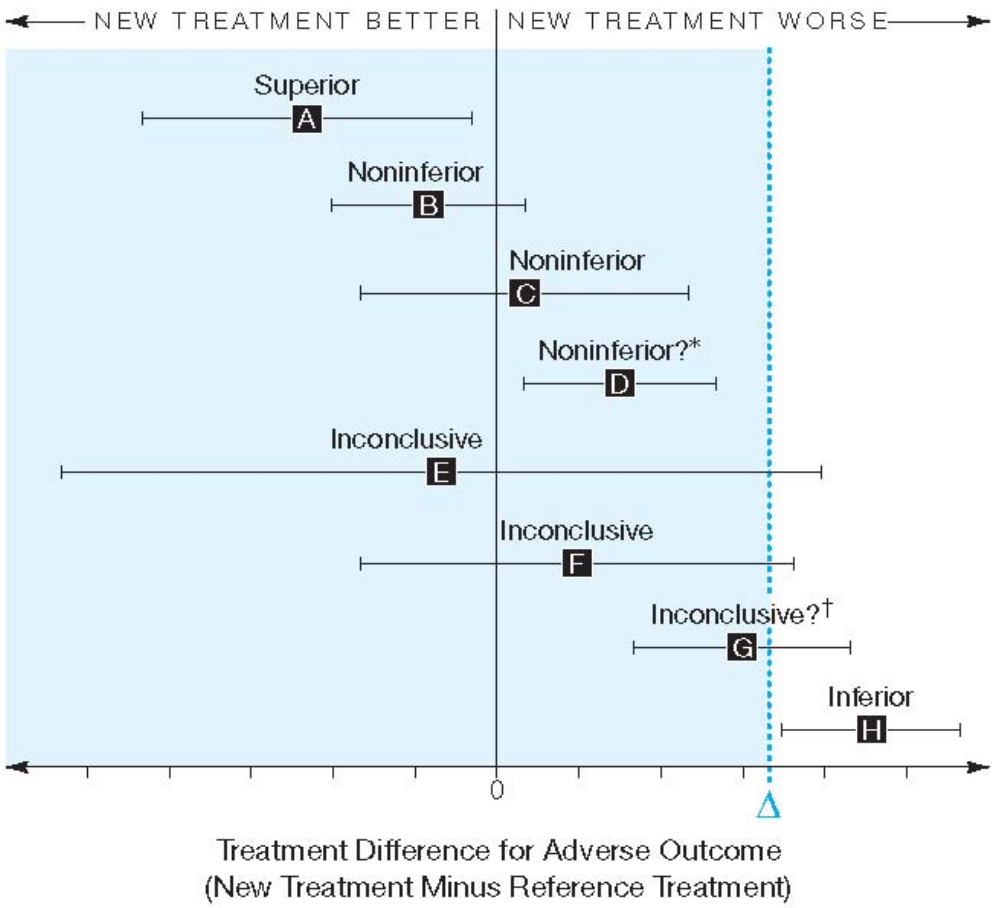
THE CONSOLIDATED STANDARDS of Reporting Trials (CONSORT) statement was developed to alleviate the problem of inadequate reporting of randomized controlled trials (RCTs),[1-4] which has been associated with biased treatment effects.[5-7] The statement comprises evidence-based recommendations for reporting RCTs, including a flowchart of participants through the trial.

CONSORT's primary focus is on parallel group trials [1-3] aiming to identify

4. Non-inferiority trial

CONSORT guidelines

Figure. Possible Scenarios of Observed Treatment Differences for Adverse Outcomes (Harms) in Noninferiority Trials



A bit of mathematics

- $A = B \;\&\; B = C \Rightarrow A = C$
- $A < B \;\&\; B < C \Rightarrow A < C$

- $A \approx B \;\&\; B \approx C \Rightarrow A \approx C$???
- A ni B & B ni C $\Rightarrow$ A ni C ???
- A ni B & B ni C $\Rightarrow$ A ni P ??? (biocreep)

Take home messages

- **Non-significant result** with a superiority trial is **NOT** a proof of equality
- Goals for the three designs are different:
  - **Superiority trial**: (say) E is <u>better</u> than C
  - **Equivalence trial**: E is <u>not too different</u> from C
  - **Non-inferiority trial**: E is <u>not much</u> worse than C
- (Equivalence and) non-inferiority depend on **choices of the trialist**:
  - Interval of clinical (equivalence) non-inferiority
  - 90% $\Leftrightarrow$ 95% C.I.
- **NI trials**
  - **Make life complicated** $\Rightarrow$ if possible use placebo-controlled RCT
  - **Unethical ?** (Garattini & Bertele, The Lancet, 2007)

NI trials unethical?

- If no better performance of E versus C is aimed at, **is trial ethical? (Garattini & Bertele, The Lancet, 2007)**

- Some arguments:
  - Always try to reformulate non-inferiority design into superiority design ... if necessary change set up of study or endpoint ...
  - Which boundaries to take?
  - Commercial aims ≠ patients' interests
  - What to write in informed consent?

    ... study is potentially less efficacious but costs less ... for the company ... ?

- Lively discussion after appearance of the Lancet paper ...

Tips for reading (NI trials)

- **Look** carefully at the definition of non-inferiority. This is of **crucial importance** for the appreciation of the result.
- Check if **definition of non-inferiority is well justified** for a clinical viewpoint.
- When **comparing** non-inferiority studies, check that **definition of NI** is the **same**
- **Check the conduct of the trial:** All aspects which reduce the quality of the trial will help “showing” not-inferiority!
- Non-inferiority **CAN NOT** be defined/claimed a posteriori!