Non-Inferiority Trials:
What are they and why are they so difficult?

David J. Cohen, M.D., M.Sc.

Director of Cardiovascular Research
Saint-Luke's Mid America Heart Institute

Professor of Medicine
University of Missouri-Kansas City

Disclosures

Grant Support/Drugs

- Eli Lilly - Schering Plough
- Baxter Pharmaceuticals - The Medicines Co.

Grant Support/Devices

- Cordis - Boston Scientific
- Edwards Lifesciences

Consulting/Advisory Boards

- Medtronic
- Cordis

DJC: 4/09

Terminology

Active Control Trial

- Clinical trial where the comparator strategy is an active (i.e., known effective) strategy
- Typically chosen when a placebo-control trial is thought to be either unethical, infeasible, or both

Superiority Trial

- Prototypical clinical trial where the goal is to demonstrate that the new treatment is better than placebo or standard therapy

Terminology -2

Non-Inferiority Trial

- Trial design where the goal is to show that the new therapy is not worse than standard therapy by some tolerable margin (e.g., 30-day mortality difference no greater than 1%)

Equivalence Trial

- Trial design where the goal is to determine whether the outcomes of 2 therapies are within some acceptable range of one another (e.g., 30-day mortality within $\pm$ 1%)

Why perform a non-inferiority trial?

- Placebo control trial unethical but still want to demonstrate that the new treatment is better than nothing ("putative placebo") approach
- New therapy may offer important advantages over currently available effective therapies
  - *Improved safety*
  - *Better tolerability/fewer side effects*
  - *Ease of use (2nd generation DES, QD drug, etc.)*
  - *Less expensive*
  - *Increased market competition (?)*

Non-Inferiority Trials- Cardiology Examples

Trial	Year	Condition	Control	Endpoint
GUSTO 3	1997	STEMI	r-tPA	30-d mortality
TARGET	2001	PCI	Abciximab	D/MI/U-TVR
REPLACE-2	2003	PCI	UFH + 2b3a	D/MI/U-TVR
FIRE	2003	SVG PCI	Guardwire	D/MI/U-TVR
SPACE	2006	Carotid Dz	CEA	D/MI/Ipsi-Stroke
ACUITY	2006	NSTE-ACS	UFH + 2b3a	D/MI/U-TVR/bleed
ENDEAVOR 4	2006	PCI/DES	Taxus DES	TVF
SYNTAX	2008	3vd or LM dz	CABG	D/MI/Stroke/Revasc

How can you prove equivalence?

Statistical Testing: Superiority Trial

H_0 (Null Hypothesis)

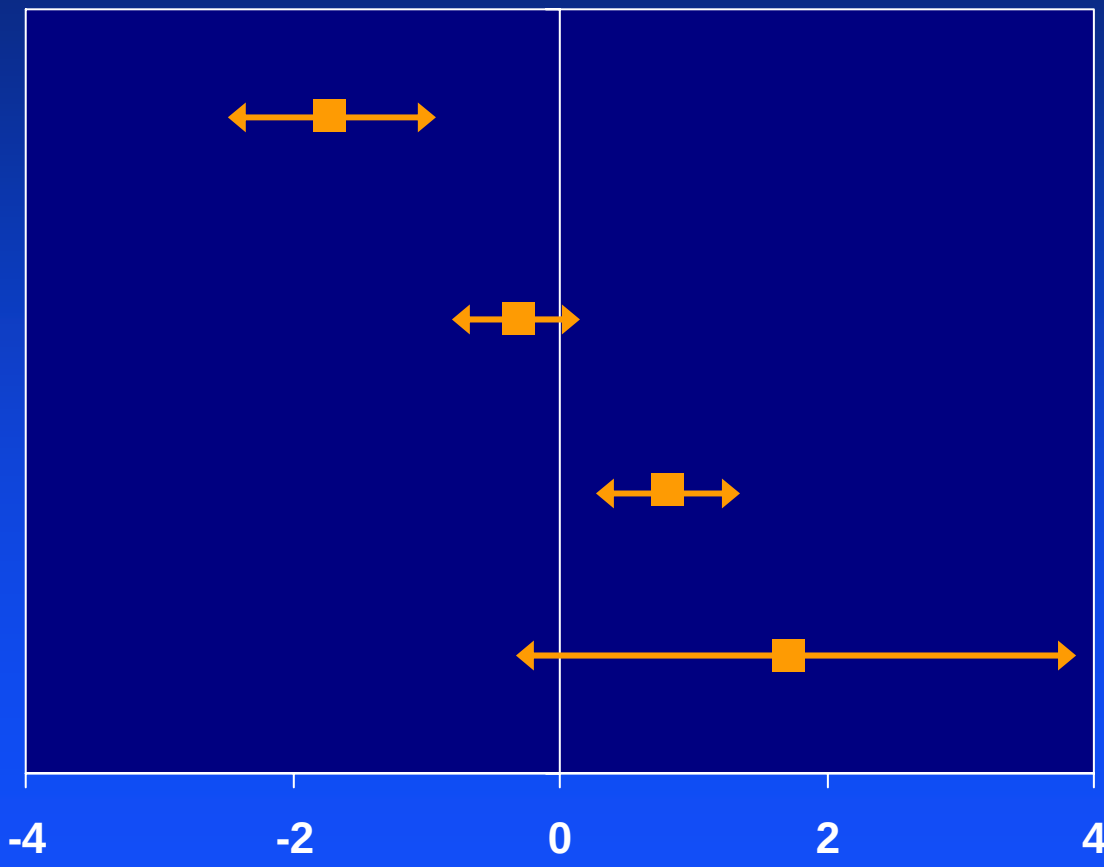
$E_T = E_S$

H_a (Alternate Hypothesis)

$E_T \neq E_S$

Application: If we can reject the null hypothesis (with 95% certainty), this represents strong evidence that the 2 treatments are not equivalent (and that one or the other is superior)

Statistical Concepts: Superiority Trial



Interpretation

Inferior

Uncertain

Superior

Uncertain

Statistical Testing: Non-Inferiority Trial

H_0 (Null Hypothesis)

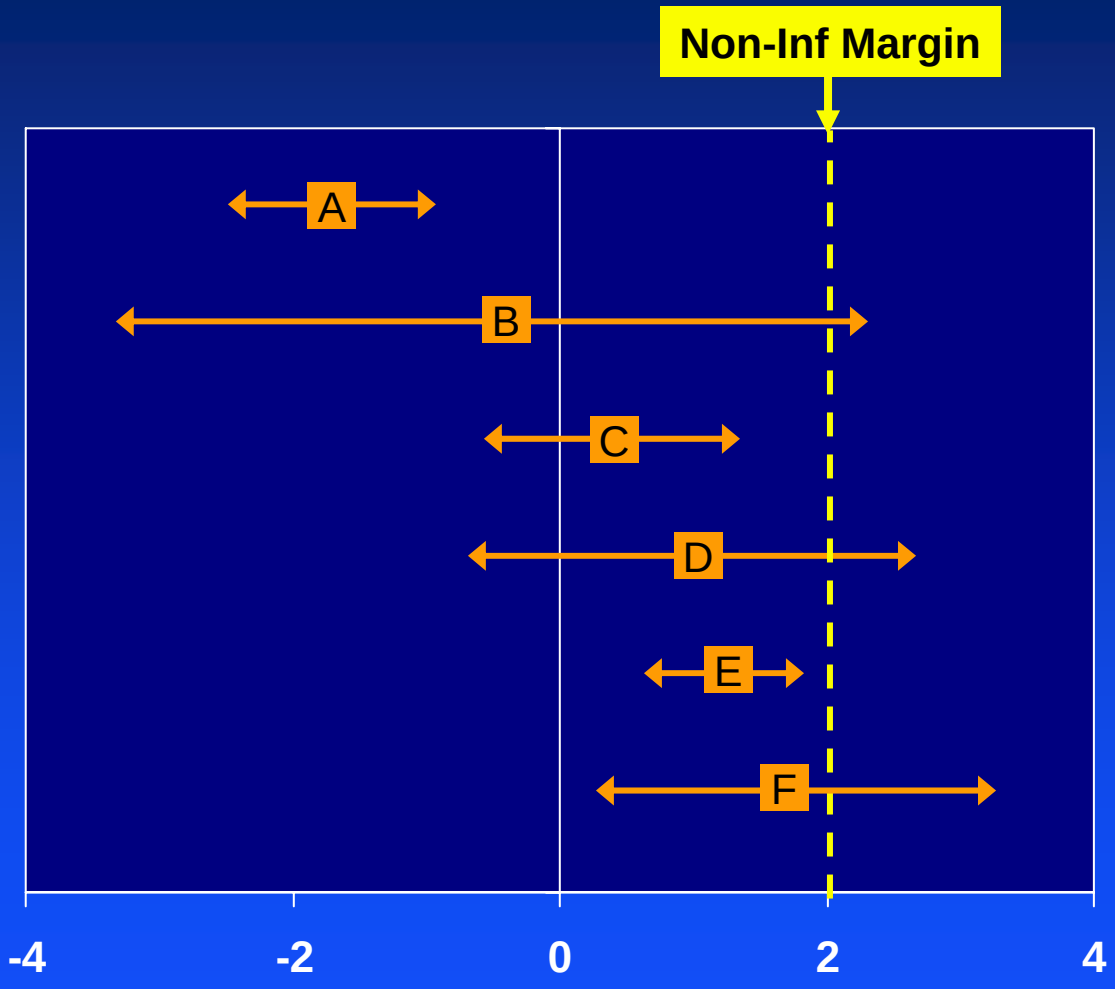
$E_T - E_S \geq \delta$

H_a (Alternate Hypothesis)

$E_T - E_S < \delta$

Application: If we can reject the null hypothesis (with 95% certainty), this provides strong evidence that the test treatment is not worse than the standard treatment by δ (the non-inferiority margin)

Statistical Concepts: Non-Inferiority Trial



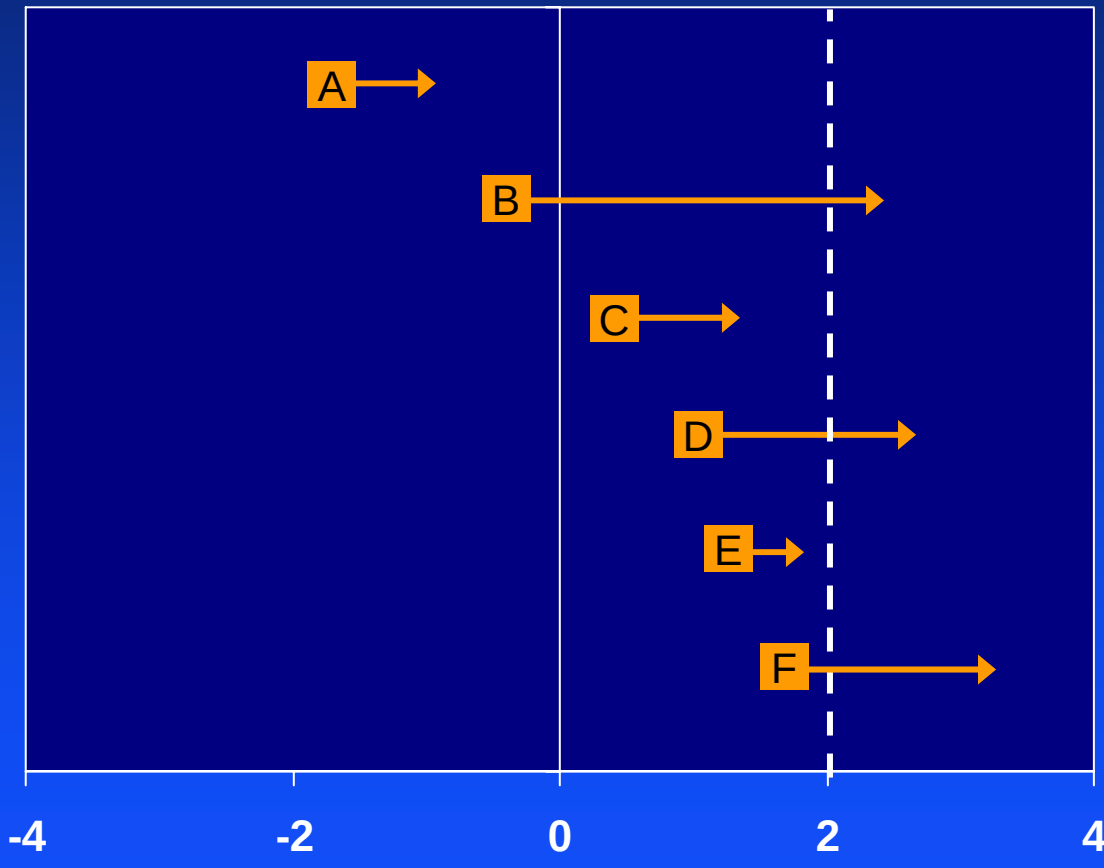
Interpretation

A	Non-inferior
B	Not non-inferior
C	Non-inferior
D	Not non-inferior
E	Non-inferior
F	Not non-inferior

2-sided test
Alpha = 5%

Statistical Concepts: Non-Inferiority Trial

Upper 1-sided confidence limit (97.5 percentile)



Difference in Primary Endpoint vs. Control

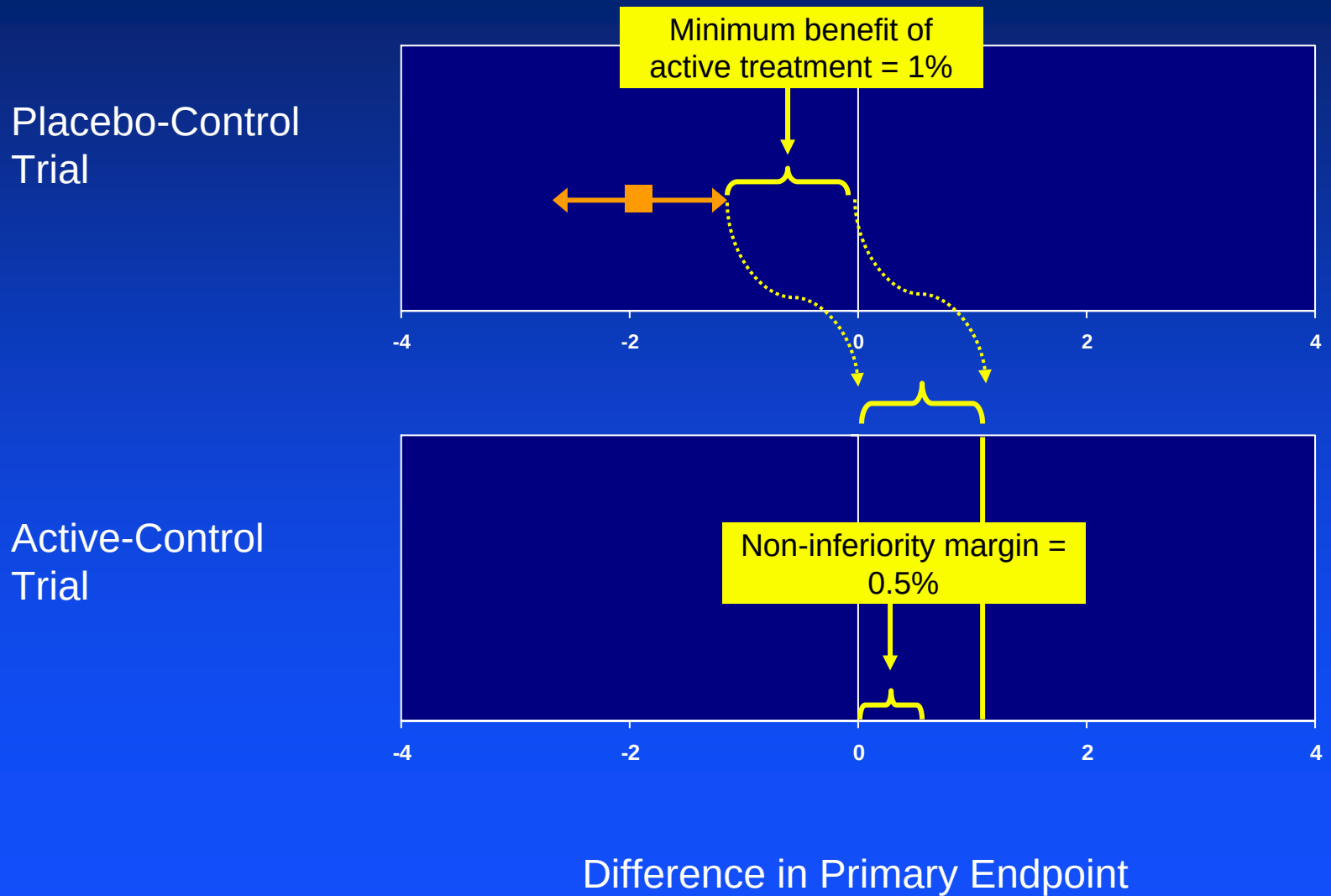
	Interpretation
A	Non-inferior
B	Not non-inferior
C	Non-inferior
D	Not non-inferior
E	Non-inferior
F	Not non-inferior

1-sided test
Alpha = 2.5%

Selecting a non-inferiority margin

- Critical to pre-specify the non-inferiority margin to avoid Type I error (false positive results)
- Potential approaches
  - Clinical rationale → expert opinion ("what is the maximum difference you would tolerate?")
  - Regulatory rationale → based on previous trials
  - Statistical + Clinical rationale → designed to preserve some minimum proportion of benefit vs. placebo ("putative placebo" approach)
- Rule of thumb: Margin cannot be greater than the smallest effect size that the active comparator would be expected to have vs. placebo

Selecting a Non-Inferiority Margin



Does the non-inferiority margin matter?

GUSTO 3

The New England Journal of Medicine

A COMPARISON OF RETEPLASE WITH ALTEPLASE FOR ACUTE MYOCARDIAL INFARCTION

THE GLOBAL USE OF STRATEGIES TO OPEN OCCLUDED CORONARY ARTERIES (GUSTO III) INVESTIGATORS*

ABSTRACT

Background Reteplase (recombinant plasminogen activator), a mutant of alteplase tissue plasminogen activator, has a longer half-life than its parent molecule and produced superior angiographic results in pilot studies of acute myocardial infarction. In this large clinical trial, we compared the efficacy and safety of these two thrombolytic agents.

Methods A total of 15,059 patients from 807 hospitals in 20 countries who presented within 6 hours after the onset of symptoms with ST-segment elevation or bundle-branch block were randomly assigned in a 2:1 ratio to receive reteplase, in two bolus doses of 10 MU each given 30 minutes apart, or an accelerated infusion of alteplase, up to 100 mg infused over a period of 90 minutes. The primary hypothesis was that mortality at 30 days would be significantly lower with reteplase.

Results The mortality rate at 30 days was 7.47 percent for reteplase and 7.24 percent for alteplase (adjusted P=0.54; odds ratio, 1.03; 95 percent confidence interval, 0.91 to 1.18). The 95 percent confidence interval for the absolute difference in mortality rates was −1.1 to 0.66 percent. Stroke occurred in 1.64 percent of patients treated with reteplase and in 1.79 percent of those treated with alteplase (P=0.50). The respective rates of the combined end point of death or nonfatal, disabling stroke were 7.89 percent and 7.91 percent (P=0.97; odds ratio, 1.0; 95 percent confidence interval, 0.88 to 1.13).

Conclusions As compared with an accelerated infusion of alteplase, reteplase, although easier to administer, did not provide any additional survival benefit in the treatment of acute myocardial infarction. Other results, particularly for the combined end point of death or nonfatal, disabling stroke, were remarkably similar for the two plasminogen activators.

(N Engl J Med 1997;337:1118-23.)

farcted vessel 90 minutes after therapy, as determined angiographically, but this was achieved in only 54 percent of patients.[3,4] Accordingly, a major goal of myocardial reperfusion therapy is to improve this rate of early fibrinolysis.

Recombinant plasminogen activator (reteplase) is a mutant of wild-type tissue plasminogen activator that lacks the finger, epidermal growth factor, and kringle-1 domains.[5] The slower clearance resulting from these changes in the molecule allows reteplase to be given as a bolus. In two angiographic trials, reteplase compared favorably with alteplase with regard to enhanced patency of the infarct-related vessel and the incidence of bleeding complications.[6,7] In a previous randomized comparison with streptokinase, treatment with reteplase resulted in an absolute 0.5 percent reduction in mortality at 30 days and a 1.0 percent reduction at 6 months, which, although not statistically significant, established the safety profile of the drug.[8] In the present trial, we tested the primary hypothesis that the mortality rate 30 days after acute infarction would be significantly lower with reteplase than with alteplase.

METHODS

Patient Population

Patients of any age who presented after 30 minutes of continuous symptoms but within 6 hours after the onset of symptoms of acute myocardial infarction and who had, on the basis of 12-lead electrocardiography, ST-segment elevation of at least 1 mm in two or more limb leads, ST-segment elevation of at least 2 mm in the precordial leads, or bundle-branch block were considered eligible. The exclusion criteria included active bleeding, a history of stroke or central nervous system damage, recent major surgery, systolic blood pressure greater than 200 mm Hg or diastolic blood pressure greater than 110 mm Hg at any time after arrival, recent noncompressible vascular puncture, or concomitant use of an oral anticoagulant with an international normalized ratio greater than 2. All patients provided informed consent for partic-

NEJM 1997;337:1118-23

COBALT

The New England Journal of Medicine

A COMPARISON OF CONTINUOUS INFUSION OF ALTEPLASE WITH DOUBLE-BOLUS ADMINISTRATION FOR ACUTE MYOCARDIAL INFARCTION

THE CONTINUOUS INFUSION VERSUS DOUBLE-BOLUS ADMINISTRATION OF ALTEPLASE (COBALT) INVESTIGATORS*

ABSTRACT

Background Accelerated infusion of alteplase (tissue plasminogen activator) over a period of 90 minutes induces more rapid lysis of coronary-artery thrombi than a 3-hour infusion. With two bolus doses of alteplase, further shortening the duration of administration, complete reperfusion was achieved in more than 85 percent of the patients in initial angiographic studies. We tested the hypothesis that double-bolus alteplase is at least as effective as accelerated infusion.

Methods In 398 hospitals, 7169 patients with acute myocardial infarction were randomly assigned to weight-adjusted, accelerated infusion of 100 mg of alteplase or to a bolus of 50 mg of alteplase over a period of 1 to 3 minutes followed 30 minutes later by a second bolus of 50 mg (or 40 mg for patients who weighed less than 60 kg). The primary end point was death from any cause at 30 days. The trial was stopped prematurely because of concern about the safety of the double-bolus injection.

Results Thirty-day mortality was higher in the double-bolus group than in the accelerated-infusion group: 7.98 percent as compared with 7.53 percent. The absolute difference was 0.44 percent, with a one-sided 95 percent upper boundary of 1.49 percent, which exceeded the prespecified upper limit of 0.40 percent to indicate equivalence in 30-day mortality between the two regimens. The respective rates of any stroke and of hemorrhagic stroke were 1.92 and 1.12 percent after double-bolus alteplase, as compared with 1.53 and 0.81 percent after an accelerated infusion of alteplase (P=0.24 and P=0.23, respectively).

Conclusions Double-bolus alteplase was not shown to be equivalent, according to the prespecified criteria, to accelerated infusion with regard to

Occluded Coronary Arteries (GUSTO) trial,[3] whereas a three-hour infusion of alteplase or duteplase (without intravenous heparin) was not superior to streptokinase in the second Gruppo Italiano per lo Studio della Sopravvivenza nell'Infarto Miocardico,[4] the International Study Group,[5] and the third International Study of Infarct Survival trials.[6] Double-bolus administration of alteplase (two bolus doses given 30 minutes apart) is a further shortening of the duration of administration. The rationale for testing brief infusions or bolus administration of fibrinolytic agents includes the observations that the generation of thrombin, associated with the use of these agents, is less pronounced with short infusions[7] and that the fibrinolytic effects of alteplase on thrombi are sustained after its clearance from the circulation.[8] In two early angiographic studies, high rates of patency were observed after double-bolus administration of alteplase, with no excess of bleeding complications.[9,10] In the largest of these studies, grade 3 flow rates (according to the Thrombolysis in Myocardial Infarction [TIMI] classification) were observed at 60 and 90 minutes in more than 85 percent of patients.[10]

Thus, besides offering the advantage of ease of use, double-bolus alteplase might be at least as effective as the accelerated infusion as used in the GUSTO I study. The current trial was performed to test this hypothesis.

METHODS

Organization of the Study

A total of 398 centers in 26 countries participated in the trial. For each patient, after inclusion and exclusion criteria were

NEJM 1997;337:1124-30

Impact of the Margin

GUSTO 3

Superiority Trial

- N = 15,059 pts with STEMI
- RPA vs. tPA
- Powered to detect 20% relative reduction in mortality

30-day mortality

RPA = 7.47%
tPA = 7.24%
} P_{diff}=0.54

95% CI = - 0.66% to 1.1%

Conclusions

- RPA and tPA are "similar" (margin of ~1%)

COBALT

Non-Inferiority Trial

- N = 7169 pts with STEMI
- Double bolus tPA vs. tPA
- Non-inferiority margin = 0.4% (lower bound of GUSTO benefit)

30-day mortality

RPA = 7.98%
tPA = 7.53%
} P_{diff}=0.53

95% upper CI = 1.49%

Conclusions

- Double bolus tPA "not equivalent" to accelerated infusion tPA

Practical Issues with Non-Inferiority Trials

Sample Size

- Often assumed that a non-inferiority trial must have a smaller sample size than a similar superiority trial→ under most circumstances, this is only the case when the non-inferiority margin is too large
- In general, use of a proper non-inferiority margin leads to very large sample sizes
- Implication: Only use a non-inferiority design when you think that the experimental therapy cannot beat the active control in a fair superiority trial
  - *From a practical perspective, it is often difficult to prove non-inferiority if the experimental therapy is even "slightly worse"*

Choice of the Non-Inferiority Margin

- Even with a well-justified margin (based on multiple previous placebo controlled trials), a claim of non-inferiority always tends to be less impactful
- Trial results always somewhat “unsatisfying” and subject to considerable post-hoc criticism

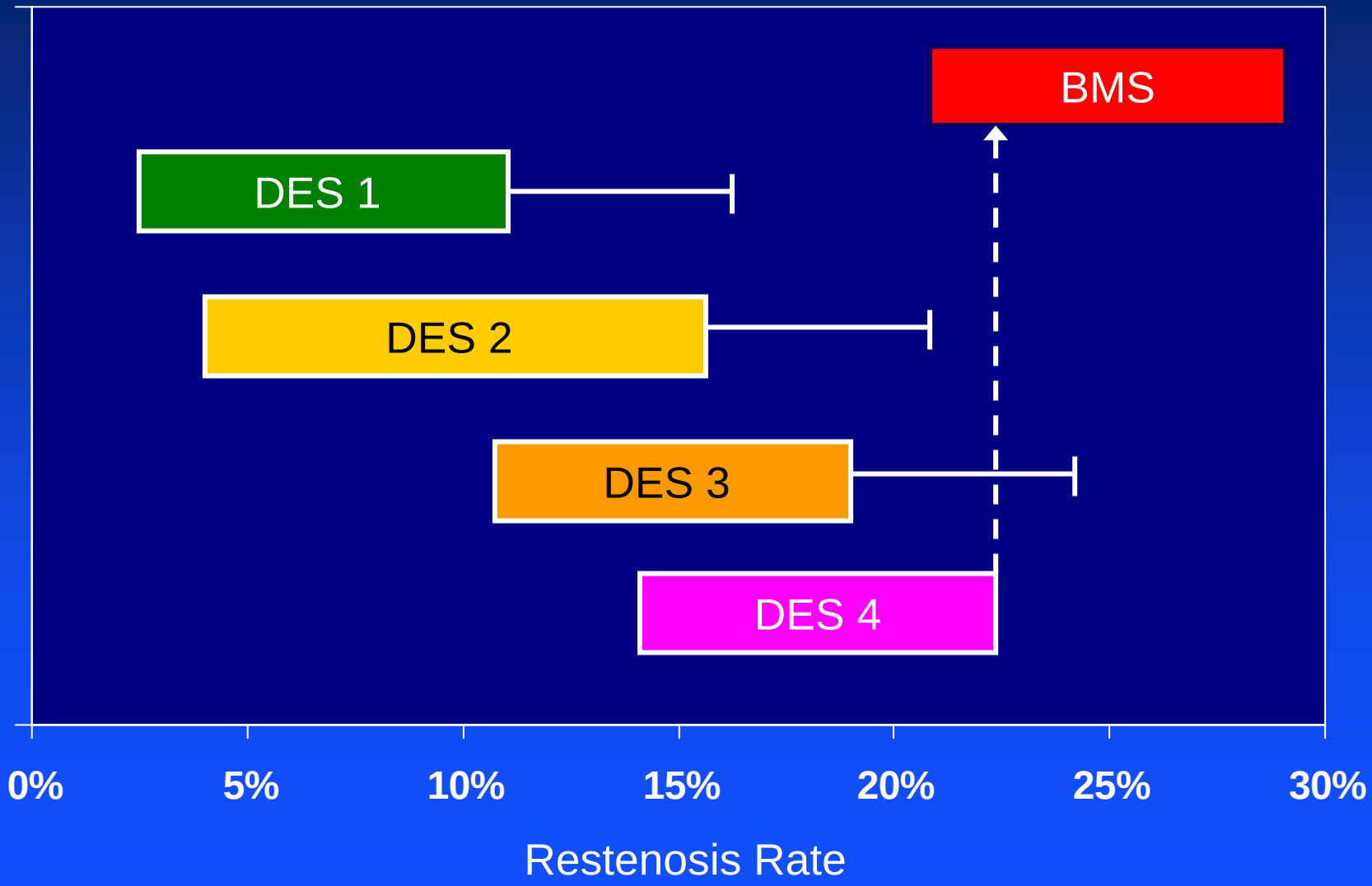
Assay Sensitivity

- Definition: Property of a clinical trial defined as the ability to distinguish an effective treatment from a less effective or ineffective treatment
- General concept: Would the current trial have detected a benefit vs. placebo if a placebo group had been included?
  - *If not, one cannot be confident that the 2 active treatments aren't just "equally ineffective"*

Assay Sensitivity- continued

- Key assumptions
  - *Active treatment clearly beneficial vs. placebo*
  - *Current study involves similar pts to previous trials, similar event rates, and similar background therapies*
- These assumptions cannot be readily verified in the trial→ inherent limitation of non-inferiority trials
- General techniques to improve assay sensitivity
  - *Blinded assessments of objective endpoints*
  - *Precise measurement techniques*
  - *Trial conditions similar to previous placebo-control trials*

Comparator Creep



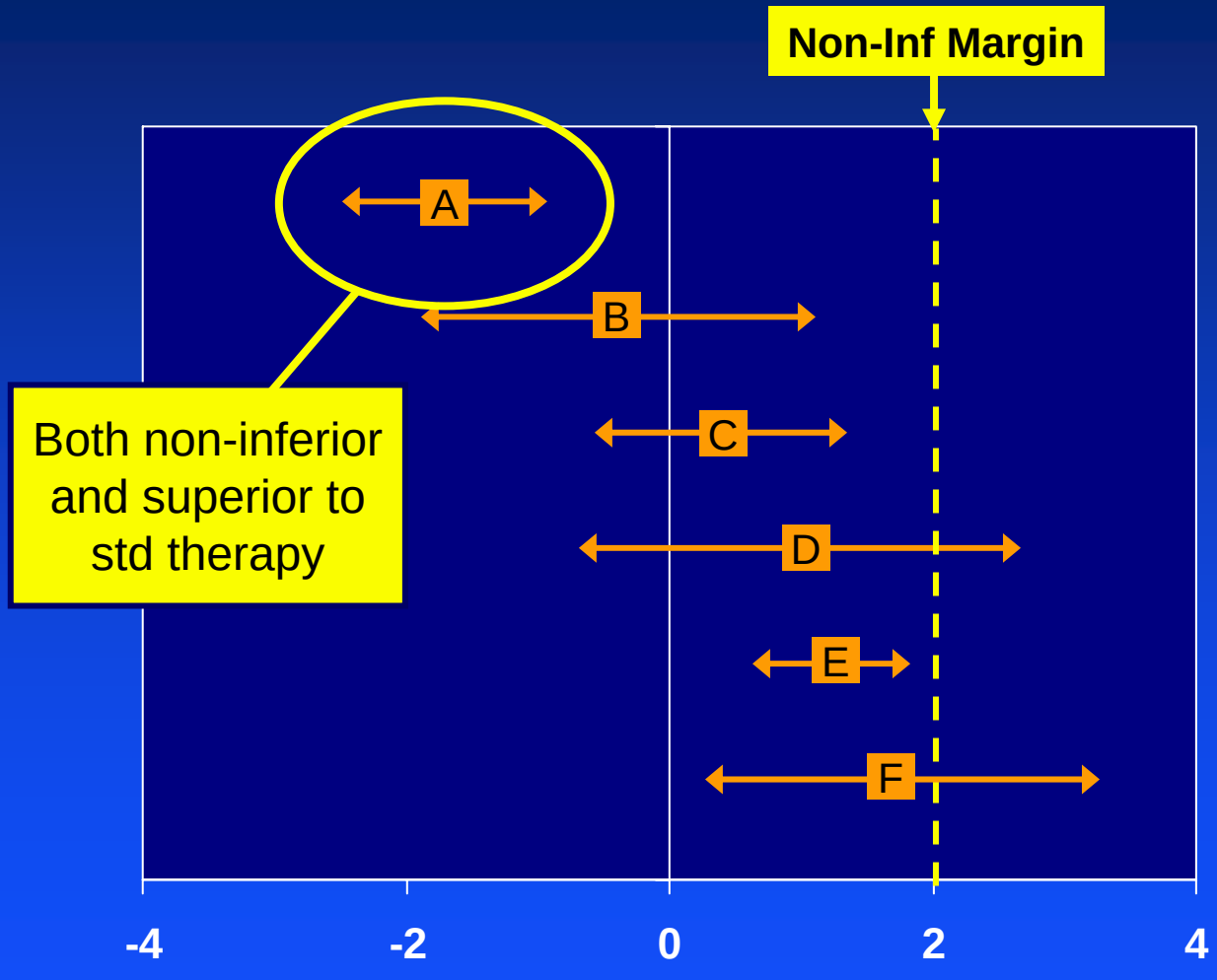
Data Quality/Trial Conduct

- In a superiority trial, the sponsor has a strong incentive to minimize errors in trial conduct and measurement, since these errors would bias the trial results toward the null
- In a non-inferiority trial, the incentives are reversed
- Key issues in trial conduct that can bias toward a finding of “non-inferiority”
  - *Enrolling a low risk population → unlikely to benefit from either treatment*
  - *Non-compliance and crossover*
  - *Insensitive outcome measure*
  - *Loss to follow-up*

Can we infer superiority?

- Yes… if the 95% confidence interval for the treatment benefit excludes both the non-inferiority margin and 0, this would generally be considered sufficient to reject the hypothesis of no difference
- But… the opposite is not true. If a superiority trial fails to reject the null hypothesis, one cannot infer non-inferiority
- Implication: If you want to have your cake and eat it, too→ design a proper non-inferiority trial

Statistical Concepts: Non-Inferiority Trial



	Interpretation
A	Non-inferior
B	Non-inferior
C	Non-inferior
D	Not non-inferior
E	Non-inferior
F	Not non-inferior

2-sided test
Alpha = 5%

Final Thoughts

- Given recent improvements in CV outcomes, in many cases new therapies may have only marginal benefits over existing therapy; as a result, there is increasing emphasis on non-inferiority trial designs
- Non-inferiority trials are not simply underpowered superiority trials→ in general, non-inferiority trials are more challenging to design, conduct, and interpret
- Choice of the non-inferiority margin is critical and, ideally, should be based on preservation of a relevant proportion of the benefit of the active comparator in previous studies

Final Thoughts- 2

- In order to demonstrate assay sensitivity, it is important to replicate the conditions of previous trials as closely as possible and to use highly sensitive and reliable measures of clinical benefit
  - *Remember that assay sensitivity is an assumption based on study design and external factors, and cannot be proven directly with trial data*
- With careful attention to these details, we can use non-inferiority trials as an effective tool to advance clinical science without sacrificing patient safety or important regulatory principles