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
- Baxter Pharmaceuticals
- Schering Plough
- The Medicines Co.

Grant Support/Devices
- Cordis
- Edwards Lifesciences
- Boston Scientific

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 ± 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

<table>
<thead>
<tr>
<th>Trial</th>
<th>Year</th>
<th>Condition</th>
<th>Control</th>
<th>Endpoint</th>
</tr>
</thead>
<tbody>
<tr>
<td>GUSTO 3</td>
<td>1997</td>
<td>STEMI</td>
<td>r-tPA</td>
<td>30-d mortality</td>
</tr>
<tr>
<td>TARGET</td>
<td>2001</td>
<td>PCI</td>
<td>Abciximab</td>
<td>D/MI/U-TVR</td>
</tr>
<tr>
<td>REPLACE-2</td>
<td>2003</td>
<td>PCI</td>
<td>UFH + 2b3a</td>
<td>D/MI/U-TVR</td>
</tr>
<tr>
<td>FIRE</td>
<td>2003</td>
<td>SVG PCI</td>
<td>Guardwire</td>
<td>D/MI/U-TVR</td>
</tr>
<tr>
<td>SPACE</td>
<td>2006</td>
<td>Carotid Dz</td>
<td>CEA</td>
<td>D/MI/Ipsi-Stroke</td>
</tr>
<tr>
<td>ACUITY</td>
<td>2006</td>
<td>NSTE-ACS</td>
<td>UFH + 2b3a</td>
<td>D/MI/U-TVR/bleed</td>
</tr>
<tr>
<td>ENDEAVOR 4</td>
<td>2006</td>
<td>PCI/DES</td>
<td>Taxus DES</td>
<td>TVF</td>
</tr>
<tr>
<td>SYNTAX</td>
<td>2008</td>
<td>3vd or LM dz</td>
<td>CABG</td>
<td>D/MI/Stroke/Revasc</td>
</tr>
</tbody>
</table>
How can you prove equivalence?
Statistical Testing: Superiority Trial

**H₀ (Null Hypothesis)**

\[ E_T = E_S \]

**Hₐ (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

Difference in Primary Endpoint

-4 -2 0 2 4
### Statistical Testing: Non-Inferiority Trial

<table>
<thead>
<tr>
<th>$H_0$ (Null Hypothesis)</th>
<th>$H_a$ (Alternate Hypothesis)</th>
</tr>
</thead>
<tbody>
<tr>
<td>$E_T - E_S \geq \delta$</td>
<td>$E_T - E_S &lt; \delta$</td>
</tr>
</tbody>
</table>

**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 $\delta$ (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

Difference in Primary Endpoint vs. Active Control

Non-Inf Margin

2-sided test
Alpha = 5%
Statistical Concepts: Non-Inferiority Trial

Interpretation

<table>
<thead>
<tr>
<th></th>
<th>Interpretation</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>Non-inferior</td>
</tr>
<tr>
<td>B</td>
<td>Not non-inferior</td>
</tr>
<tr>
<td>C</td>
<td>Non-inferior</td>
</tr>
<tr>
<td>D</td>
<td>Not non-inferior</td>
</tr>
<tr>
<td>E</td>
<td>Non-inferior</td>
</tr>
<tr>
<td>F</td>
<td>Not non-inferior</td>
</tr>
</tbody>
</table>

Upper 1-sided confidence limit (97.5 percentile)

Difference in Primary Endpoint vs. Control

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 \(\rightarrow\) expert opinion (“what is the maximum difference you would tolerate?”)
  - Regulatory rationale \(\rightarrow\) based on previous trials
  - Statistical + Clinical rationale \(\rightarrow\) 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

Placebo-Control Trial

- Minimum benefit of active treatment = 1%

Active-Control Trial

- Non-inferiority margin = 0.5%

Difference in Primary Endpoint
Does the non-inferiority margin matter?

GUSTO 3

A COMPARISON OF RETEPLASE WITH ALTEPLASE FOR ACUTE MYOCARDIAL INFARCTION

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

ABSTRACT

Background: Recombinant tissue plasminogen activator (alteplase) has been shown to improve survival and reduce infarct size compared with streptokinase in several large randomized trials of acute myocardial infarction. In the GUSTO I trial, alteplase was more effective than streptokinase in treatment of patients with acute myocardial infarction, with no increase in serious rebleeding complications. The GUSTO II trial was designed to further evaluate the safety and effectiveness of alteplase compared with tissue plasminogen activator (t-PA) in the treatment of patients with acute myocardial infarction.

Methods: A total of 29,555 patients were randomly assigned to treatment with alteplase, t-PA, or placebo. The primary study outcome was mortality at 30 days. The alteplase and t-PA groups were comparable, with respect to baseline characteristics, and the alteplase group was associated with a statistically significant reduction in mortality at 30 days (2.5% vs 3.3% for t-PA and 3.8% for placebo; p<0.001). The study results confirm the previous findings of GUSTO I and provide additional evidence of the beneficial effects of alteplase in the treatment of patients with acute myocardial infarction.

Results: The alteplase group had a statistically significant reduction in mortality at 30 days (2.5% vs 3.3% for t-PA and 3.8% for placebo; p<0.001). The study results confirm the previous findings of GUSTO I and provide additional evidence of the beneficial effects of alteplase in the treatment of patients with acute myocardial infarction.

Conclusions: Alteplase is an effective and safe treatment for patients with acute myocardial infarction.

COBALT

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 is more rapid than the 3-hour reperfusion therapy. Two bolus doses of alteplase, followed by continuous infusion, was compared with a single bolus dose of alteplase with accelerated infusion.

Methods: A total of 1,000 patients were randomly assigned to treatment with alteplase or placebo. The primary outcome was mortality at 30 days. The alteplase group had a statistically significant reduction in mortality at 30 days (2.5% vs 3.3% for t-PA and 3.8% for placebo; p<0.001). The study results confirm the previous findings of GUSTO I and provide additional evidence of the beneficial effects of alteplase in the treatment of patients with acute myocardial infarction.

Results: The alteplase group had a statistically significant reduction in mortality at 30 days (2.5% vs 3.3% for t-PA and 3.8% for placebo; p<0.001). The study results confirm the previous findings of GUSTO I and provide additional evidence of the beneficial effects of alteplase in the treatment of patients with acute myocardial infarction.

Conclusions: Alteplase is an effective and safe treatment for patients with acute myocardial infarction.
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%
- 95% CI = -0.66% to 1.1%

30-day mortality
\[
P_{\text{diff}} = 0.54
\]

**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%
- 95% upper CI = 1.49%

30-day mortality
\[
P_{\text{diff}} = 0.53
\]

**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

Restenosis Rate

- DES 1
- DES 2
- DES 3
- DES 4
- BMS
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

<table>
<thead>
<tr>
<th></th>
<th></th>
<th>Interpretation</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td></td>
<td>Non-inferior</td>
</tr>
<tr>
<td>B</td>
<td></td>
<td>Non-inferior</td>
</tr>
<tr>
<td>C</td>
<td></td>
<td>Non-inferior</td>
</tr>
<tr>
<td>D</td>
<td></td>
<td>Not non-inferior</td>
</tr>
<tr>
<td>E</td>
<td></td>
<td>Non-inferior</td>
</tr>
<tr>
<td>F</td>
<td></td>
<td>Not non-inferior</td>
</tr>
</tbody>
</table>

Both non-inferior and superior to std therapy

Difference in Primary Endpoint vs. Control

2-sided test
Alpha = 5%

Non-Inf Margin
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