Non-Inferiority designs: Not your grandma’s randomized trial

Daniella Zipkin MD
Associate Professor of Medicine
Duke University Department of Medicine
Objectives

• Describe the ways in which non-inferiority trials differ from traditional trials
• Define non-inferiority margin
• Assess validity of a non-inferiority trial
HARRY SCORES RECORD-BREAKING 8TH PERSONAL ‘NO WORSE’ AT THE PHARMA OLYMPICS!
Comparative Effectiveness

- Direct comparisons of active treatments
- Non-inferiority study designs are increasingly common as placebo recedes as the most appropriate comparator
Non-inferiority margin

AKA: How much worse is ‘not worse than’?
COULD BE ANYTHING.

WAY TOO GENERAL PRACTITIONER.
NON-INFERIORITY MARGIN

Attempting to find a newer intervention “not worse than” an existing intervention

Bonus Points: What happens if we narrow this margin? Need to recruit MORE patients, or recruit FEWER?
Non-inferiority Margin

- Historical trials – active control vs. placebo
- Some proportion of the minimum benefit seen vs. placebo
- Tighter margin = more subjects needed to maintain power (wide margin “easier” to prove)

Oral anticoagulants for preventing stroke in patients with non-valvular atrial fibrillation and no previous history of stroke or transient ischemic attacks (Review)

Aguilar MI, Hart R
Traditional comparative study

Superiority established

Superiority not established

Treatment difference, RR
Warfarin Superior to Placebo

Superiority established

Treatment difference, RR
Warfarin Superior to Placebo

Superiority established

Treatment difference, RR
Placebo Inferior to Warfarin

Treatment difference, RR
Placebo Inferior to Warfarin

Treatment difference, RR
Apixaban versus Warfarin in Patients with Atrial Fibrillation

Christopher B. Granger, M.D., John H. Alexander, M.D., M.H.S., John J.V. McMurray, M.D., Renato D. Lopes, M.D., Ph.D., Elaine M. Hylek, M.D., M.P.H., Michael Hanna, M.D., Hussein R. Al-Khalidi, Ph.D., Jack Ansell, M.D., Dan A. Atar, M.D., Alvaro Avezum, M.D., Ph.D., M. Cecilia Bahit, M.D., Rafael Diaz, M.D., J. Donald Easton, M.D., Justin A. Ezekowitz, M.B., B.Ch., Greg Flaker, M.D., David Garcia, M.D., Margarida Geraldes, Ph.D., Bernard J. Gersh, M.D., Sergey Golitsyn, M.D., Ph.D., Shinya Goto, M.D., Antonio G. Hermosillo, M.D., Stefan H. Hohnloser, M.D., John Horowitz, M.D., Puneet Mohan, M.D., Ph.D., Petr Jansky, M.D., Basil S. Lewis, M.D., Jose Luis Lopez-Sendon, M.D., Prem Pais, M.D., Alexander Parkhomenko, M.D., Freek W.A. Verheugt, M.D., Ph.D., Jun Zhu, M.D., and Lars Wallentin, M.D., Ph.D., for the ARISTOTLE Committees and Investigators*
The primary noninferiority hypothesis required that apixaban preserve at least 50% of the relative reduction in the risk of stroke or systemic embolism associated with warfarin (62%) in six previous, major randomized, controlled trials. This hypothesis provided a lower 95% confidence interval of 1.88 for the relative risk with placebo as compared with warfarin, and one half of this value was 1.44 (or 1.38 on a log scale). We estimated that with the occurrence of the primary outcome in 448 patients, the study would have 90% power to ensure that the upper boundary of the 99% confidence interval for the relative risk would be less than 1.44 and that the upper boundary of the 95% confidence interval for the relative risk would be less than 1.38, on the assumption that apixaban and warfarin had identical effects. On the basis of the overall event rate during the trial, we planned to recruit 18,000 pa-
Placebo Inferior to Warfarin

Treatment difference, RR

P
Add in Superiority?

- If testing for non-inferiority is a one-tailed comparison, you can then perform superiority testing

- “You haven’t spent all your tails yet”…
Non-inferiority Margin

Then ask about superiority

Is non-inferiority met?

0%

100%

PLACEBO

DRUG A

DRUG B
Assessing bias in non-inferiority trials
Non-inferiority: Assumptions

- **Assay sensitivity**: active control would have been superior to placebo if placebo were used.
- **Constancy**: the historical difference between active control and placebo is assumed to hold now.
- **Variability**: if estimates of historical benefit over placebo vary, use the smallest

Make sure active control had a fair shake!

- Dosing, time in therapeutic range – was it administered optimally?
- Number of outcome events comparable to prior
What happens when...

• Patients are randomized
• Allocation is concealed
• Groups are treated equally throughout
• Subjects and investigators are blinded
• Outcome assessors are blinded
• Follow up is complete
• Intention to treat principles are upheld

• GROUPS LOOK MORE SIMILAR!
What happens when...

- Not enough outcomes emerge (power is not maintained...)
- GROUPS LOOK MORE SIMILAR!
<table>
<thead>
<tr>
<th>True Hypothesis</th>
<th>TRUTH: $A &gt; B$</th>
<th>TRUTH: $A = B$</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>TRIAL: $A &gt; B$</strong></td>
<td>True Pos</td>
<td>False Pos ($\alpha$, type I error)</td>
</tr>
<tr>
<td><strong>TRIAL: $A = B$</strong></td>
<td>False Neg ($\beta$, type II error)</td>
<td>True Neg</td>
</tr>
</tbody>
</table>

Null Hypothesis = no difference

Power in Superiority Trial = $1 - \beta$
Power in a Non-inferiority Trial

- **Null Hypothesis** = Standard treatment group has fewer bad outcomes than experimental; Standard treatment is better

- **Alternative Hypothesis** = Experimental treatment is better or only slightly worse (within the margin)
Power in a Non-inferiority Trial

• “The power to find non-inferiority within the specified margin”
Power in a Non-inferiority Trial

- If power isn’t maintained (too few events for instance), but non-inferiority is still found, look at the event rates and decide if it still seems non-inferior to you.
- Always examine both ends of the confidence interval.
One more example
Short-term vs Conventional Glucocorticoid Therapy in Acute Exacerbations of Chronic Obstructive Pulmonary Disease
The REDUCE Randomized Clinical Trial

Jörg D. Leuppi, MD, PhD
Philipp Schuetz, MD
Roland Bingisser, MD
Michael Bodmer, MD
Matthias Briel, MD
Tilman Drescher, MD
Ursula Duerring, RN
Christoph Henzen, MD
Yolanda Leibbrandt, RN
Sabrina Maier, RN
David Miedinger, MD, PhD
Beat Müller, MD
Andreas Scherr, MD
Christian Schindler, PhD
Rolf Stoeckli, MD
Sebastien Viatte, MD, PhD
Christophe von Carnier, MD
Michael Tumml, MD
Jonas Rutishauser, MD

Importance International guidelines advocate a 7- to 14-day course of systemic glucocorticoid therapy in acute exacerbations of chronic obstructive pulmonary disease (COPD). However, the optimal dose and duration are unknown.

Objective To investigate whether a short-term (5 days) systemic glucocorticoid treatment in patients with COPD exacerbation is noninferior to conventional (14 days) treatment in clinical outcome and whether it decreases the exposure to steroids.

Design, Setting, and Patients REDUCE (Reduction in the Use of Corticosteroids in Exacerbated COPD), a randomized, noninferiority multicenter trial in 5 Swiss teaching hospitals, enrolling 314 patients presenting to the emergency department with acute COPD exacerbation, past or present smokers (≥20 pack-years) without a history of asthma, from March 2006 through February 2011.

Interventions Treatment with 40 mg of prednisone daily for either 5 or 14 days in a placebo-controlled, double-blind fashion. The predefined noninferiority criterion was an absolute increase in exacerbations of at most 15%, translating to a critical hazard ratio of 1.515 for a reference event rate of 50%.

Main Outcome and Measure Time to next exacerbation within 180 days.

Results Of 314 randomized patients, 289 (92%) of whom were admitted to the hospital, 311 were included in the intention-to-treat analysis and 296 in the per-protocol analysis. Hazard ratios for the short-term vs conventional treatment group were 0.95 (90% CI, 0.70 to 1.32; P = .005 for noninferiority) in the intention-to-treat analysis and 0.95 (90% CI, 0.68 to 1.26; P = .005 for noninferiority) in the per-protocol analysis, meeting our noninferiority criterion. In the short-term group, 56 patients (35.9%) reached the primary end point; 57 (36.8%) in the conventional group. Estimates of reexacerbation rates within 180 days were 37.2% (95% CI, 29.5% to 44.9%) in the short-term; 38.4% (95% CI, 30.6% to 46.3%) in the conventional, with a difference of −1.2% (95% CI, −12.2% to 9.8%) between the short-term and the conventional. Among patients with a reexacerbation, the median time to event was 43.5 days (interquartile range [IQR], 13 to 118) in the short-term and 29 days (IQR, 16 to 85) in the conventional. There was no difference between groups in time to death, the combined end point of exacerbation, death, or both and recovery of lung function. In the conventional group, mean cumulative prednisone dose was significantly higher (793 mg [95% CI, 710 to 876 mg] vs 379 mg [95% CI, 311 to 446 mg], P <.001), but treatment-associated adverse reactions, including hyperglycemia and hypertension, did not occur more frequently.

Conclusions and Relevance In patients presenting to the emergency department with acute exacerbations of COPD, 5-day treatment with systemic glucocorticoids was noninferior to 14-day treatment with regard to reexacerbation within 6 months of follow-up but significantly reduced glucocorticoid exposure. These findings support the use of a 5-day glucocorticoid treatment in acute exacerbations of COPD.

Trial Registration isrctn.org Identifier: ISRCTN19646069

JAMA. 2013;309(21):2223-2231
Published online May 21, 2013. doi:10.1001/jama.2013.5023
www.jama.com
Statistical Analysis

We used a modified Delphi technique to define noninferiority regarding the primary end point. Based on the judgement of 11 board-certified specialists, we defined a 15% absolute difference in the percentage of patients with a re-exacerbation during the 6 months of follow-up as the clinically tolerable upper limit. Based on previously published data, we assumed that approximately 50% of patients would experience an exacerbation during follow-up. Therefore, according to our noninferiority definition, the true proportion of patients under experimental (5 days) treatment experiencing a COPD exacerbation must not exceed 65%, which translates to a critical hazard ratio (HR) of 1.515 based on an exponential proportional hazards survival model.\textsuperscript{14}

During the study, observed loss to follow-up was much lower (<4%) than expected (20%). Assuming 5% loss to follow-up, an $\alpha$ error of 5%, and a power of 85%, we needed to recruit 150 patients in each study group.
<table>
<thead>
<tr>
<th>Primary End Point</th>
<th>Event Frequencies, No. (%)</th>
<th>Hazard Ratio (90% CI)</th>
<th>P Value&lt;sup&gt;a&lt;/sup&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Conventional Treatment (n = 155)</td>
<td>Short-term Treatment (n = 156)</td>
<td></td>
</tr>
<tr>
<td>Reexacerbations</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Intention to treat</td>
<td>57 (36.8)</td>
<td>56 (35.9)</td>
<td>0.95 (0.70-1.29)</td>
</tr>
<tr>
<td>Per protocol</td>
<td>57 (38.3)</td>
<td>54 (36.7)</td>
<td>0.93 (0.68-1.26)</td>
</tr>
<tr>
<td>Subgroup analyses&lt;sup&gt;b&lt;/sup&gt;</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GOLD grade</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1 and 2&lt;sup&gt;c&lt;/sup&gt;</td>
<td>6 (33.3)</td>
<td>6 (26.1)</td>
<td>0.73 (0.28-1.88)</td>
</tr>
<tr>
<td>3&lt;sup&gt;c&lt;/sup&gt;</td>
<td>19 (35.9)</td>
<td>15 (33.3)</td>
<td>0.93 (0.52-1.67)</td>
</tr>
<tr>
<td>4&lt;sup&gt;c&lt;/sup&gt;</td>
<td>31 (39.7)</td>
<td>34 (40.5)</td>
<td>0.99 (0.66-1.49)</td>
</tr>
<tr>
<td>Glucocorticoid pretreatment</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>13 (46.4)</td>
<td>16 (45.7)</td>
<td>0.93 (0.50-1.72)</td>
</tr>
<tr>
<td>No</td>
<td>44 (35.8)</td>
<td>40 (33.3)</td>
<td>0.88 (0.61-1.26)</td>
</tr>
</tbody>
</table>

Abbreviation: GOLD, Global Initiative for Chronic Obstructive Lung Disease.

<sup>a</sup>P value for noninferiority.

<sup>b</sup>Analyses were intention to treat. There was no evidence of heterogeneity in hazard ratios across subgroups (for GOLD grade, P = .82; for glucocorticoid pretreatment status, P = .93).

<sup>c</sup>Airflow limitation according to GOLD chronic obstructive pulmonary disease grading: 1, mild; 2, moderate; 3, severe; 4, very severe.
Minimizing bias in non-inferiority:

- Plan the POWER calculation stringently
- Account for drop-outs stringently
- Use more than one outcome assessor
- Add a “per protocol” analysis to your intention to treat analysis? If it matches ITT results, it strengthens non-inferiority

Per-Protocol Analysis:

Event Rates:

Intention to treat:
- Intervention: \( \frac{6}{16} = 0.37 \)
- Control: \( \frac{5}{16} = 0.31 \)

Per protocol:
- Intervention: \( \frac{6}{10} = 0.60 \)
- Control: \( \frac{5}{12} = 0.41 \)

Image credit: Eric Wei, MD
Let’s look at another example together: POET trial