Interpreting Non-inferiority Clinical Trials

Jaekyu Shin, PharmD, BCPS
Associate Professor of Clinical Pharmacy
University of California San Francisco
I have no conflict of interest related to this presentation.
Learning Objectives

• By the end of this presentation, participants should be able to

1. Define superiority, non-inferiority, and equivalence trials.
2. Compare purpose, sample size, margin, null hypothesis, and statistical analysis plan between superiority and non-inferiority trials.
3. Critically evaluate results of superiority and non-inferiority trials.
Which of the Following Is a Superiority Trial?

1. A trial designed to show that treatment A is better than treatment B.
2. A trial designed to show that treatment A is not worse than treatment B.
3. A trial designed to show that treatment A is equal to treatment B.
4. A trial designed to show that treatment A is not better than treatment B.
Definition

- **Superiority trial**
  - To test whether treatment A is better than treatment B \((A > B)\).

- **Non-inferiority trial**
  - To test whether treatment A is not worse than treatment B \((A \geq B)\).

- **Equivalence trial**
  - To test whether treatment A is equal to treatment B \((A = B)\).
Rivaroxaban (Xarelto)

• Oral factor Xa inhibitor
• For the treatment of non-valvular atrial fibrillation (afib), and treatment and prophylaxis of venous thromboembolism (VTE)
Two Clinical Trials on Rivaroxaban

**ATLAS ACS 2–TIMI 51**
- *Superiority* trial
- Recent acute coronary syndrome (ACS)
- Riva vs. placebo
- 3 arms:
  - Riva 2.5 mg BID
  - Riva 5 mg BID
  - Placebo
- Primary outcome: death from CV causes, MI, stroke

**ROCKET-AF**
- *Non-inferiority* trial
- Non-valvular atrial fibrillation (afib)
- Riva vs. warfarin
- 2 arms:
  - Riva 20 mg daily
  - Warfarin (target INR 2-3)
- Primary outcome: stroke or systemic embolism

NEJM 2012;366:9; NEJM 2011;365:883
Why Did These Two Trials Have the Different Trial Designs?

1. The study populations were different.
2. The comparison groups were different.
3. The number of groups in comparison was different.
4. The primary outcomes were different.
Situations to Use a Non-inferiority Trial

• New treatment has other advantages than primary outcomes
  – Route of administration
  – Side effects
  – Dosing schedule
  – Monitoring schedule
  – Duration of therapy
Importance of Null Hypothesis

- Statistical tests test null hypothesis

- If the probability that the observation is true under the null hypothesis is too small, then reject the null hypothesis and accept the alternative hypothesis
Hypothesis of a Superiority Trial

• $H_0$
  – There is no difference in the primary outcome between treatments A and B ($A = B$)

• $H_A$
  – There is a difference in the primary outcome between treatments A and B ($A \neq B$)
ROCKET-AF is a non-inferiority clinical trial comparing rivaroxaban with warfarin in patients with non-valvular atrial fibrillation. The study outcome was composite of stroke or systemic embolism. Which of the following is the null hypothesis of the trial?

1. There is no difference in the rate of stroke or systemic embolism between rivaroxaban and warfarin.
2. Rivaroxaban has a lower rate of stroke or systemic embolism than warfarin.
3. Rivaroxaban has a higher rate of stroke or systemic embolism than warfarin.
4. Rivaroxaban has a rate of stroke or systemic embolism similar to that of warfarin.
True or False?

• Non-inferiority trials require smaller sample sizes than superiority trials.
Number of Participants

ATLAS ACS 2–TIMI 51
• A total of 15,526 patients underwent randomization

ROCKET-AF
• A total of 14,264 patients underwent randomization
### Sample Size Calculation

<table>
<thead>
<tr>
<th>ATLAS ACS 2–TIMI 51</th>
<th>ROCKET-AF</th>
</tr>
</thead>
<tbody>
<tr>
<td>A total of 983 primary outcome events would provide a <strong>power</strong> of ~ 96% to detect a 22.5% <strong>relative reduction</strong> between the combined-dose group receiving rivaroxaban and the placebo group with a two-sided <strong>type I error rate</strong> of 0.05.</td>
<td>A minimum of 363 primary outcome events would provide a <strong>power</strong> of 95% to calculate a <strong>noninferiority margin</strong> of 1.46 with a one-sided <strong>alpha level</strong> of 0.025.</td>
</tr>
</tbody>
</table>
Sample size depends on power, type I error rate, effect size, and underlying event rate.
**ATLAS-ACS 2 – TIMI 51**  
(A Superiority Trial)

- Results: primary outcome

<table>
<thead>
<tr>
<th>Riva (combined) (N=10,229)</th>
<th>Placebo (N=5,133)</th>
<th>HR (95% CI)</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>626 (8.9%)</td>
<td>376 (10.7%)</td>
<td>0.84 (0.74-0.96)</td>
<td></td>
</tr>
</tbody>
</table>
**ROCKET-AF**  
(A Non-inferiority Trial)

- Results: primary outcome*

<table>
<thead>
<tr>
<th></th>
<th>Riva (N=7,081)</th>
<th>Warfarin (N=7,090)</th>
<th>HR (95% CI)</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>269 (3.8%)</td>
<td>306 (4.3%)</td>
<td>0.88 (0.75-1.03)</td>
<td></td>
</tr>
</tbody>
</table>

*Intention-to-treat analysis
• To determine statistical significance, how do we use the confidence interval?
Determining Statistical Significance: Superiority Trial

Reference line (1 or 0)

A better

B better

Interpretation
Determining Statistical Significance: Non-inferiority Trial

Reference line (Margin)

A not worse

A worse

Interpretation

JAMA 2006;295:1152; JAMA 2012;308:2694
ROCKET-AF
(A Non-inferiority Trial)

- A minimum of 363 primary outcome events would provide a power of 95% to calculate a noninferiority margin of 1.46 with a one-sided alpha level of 0.025.

<table>
<thead>
<tr>
<th>Riva (N=7,081)</th>
<th>Warfarin (N=7,090)</th>
<th>HR (95% CI)</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>269 (3.8%)</td>
<td>306 (4.3%)</td>
<td>0.88 (0.75-1.03)</td>
<td>&lt;0.001</td>
</tr>
</tbody>
</table>
Apixaban vs. Enoxaparin after Knee Replacement

- Apixaban would be noninferior to enoxaparin with respect to the primary outcome, with a noninferiority margin in which the upper limit of the 95% CI for relative risk did not exceed 1.25.

<table>
<thead>
<tr>
<th></th>
<th>Apix (N=1,157)</th>
<th>Enox (N=1,130)</th>
<th>HR (95% CI)</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>104 (9.0%)</td>
<td>100 (8.8%)</td>
<td>1.02 (0.78-1.32)</td>
<td></td>
</tr>
</tbody>
</table>

NEJM 2009;361:594
Noninferiority Trials and Clinical Significance

• A minimum of 363 primary outcome events would provide a power of 95% to calculate a noninferiority margin of 1.46 with a one-sided alpha level of 0.025.

• Riva considered as not worse than warfarin even if riva may have a 46% increased risk of primary outcome events.

• Do you agree?
Noninferiority Trials and Clinical Significance

<table>
<thead>
<tr>
<th></th>
<th>Warfarin (N=7,090)</th>
<th>HR (95% CI)</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Riva (N=7,081)</td>
<td>269 (3.8%)</td>
<td>0.88 (0.75-1.03)</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td></td>
<td>306 (4.3%)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Do you consider an up to 3% increase in relative risk as not worse than warfarin?
Noninferiority Trials and Clinical Significance

- RE-COVER: Dabigatran vs. warfarin for VTE tx
  - We tested for noninferiority by comparing the upper boundary of the 95% confidence interval for the hazard ratio with the predefined margin of 2.75 and the upper boundary of the 95% confidence interval for the difference in risk with the predefined margin of 3.6 percentage points.

<table>
<thead>
<tr>
<th></th>
<th>Dabi (N=1,274)</th>
<th>Warfarin (N=1,265)</th>
<th>Risk difference (95% CI)</th>
<th>HR (95% CI)</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>30 (2.4%)</td>
<td>27 (2.1%)</td>
<td>0.3 (-0.8-1.5)</td>
<td>1.10 (0.65-1.84)</td>
<td>&lt;0.001</td>
</tr>
</tbody>
</table>

NEJM 2009;361:2341
Superiority Trials and Clinical Significance

- A total of 983 primary outcome events would provide a power of ~ 96% to detect a **22.5% relative reduction** between the combined-dose group receiving rivaroxaban and the placebo group.

<table>
<thead>
<tr>
<th></th>
<th>Riva (combined) (N=10,229)</th>
<th>Placebo (N=5,133)</th>
<th>HR (95% CI)</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>626 (8.9%)</td>
<td>376 (10.7%)</td>
<td>0.84 (0.74-0.96)</td>
<td>0.008</td>
<td></td>
</tr>
</tbody>
</table>
Analysis Plans Used in RCTs

- Intention to treat
  - Analyze as randomized
- Per-protocol (as treated)
  - Analyze as treatment received
## Per-protocol Analysis in ROCKET-AF

### Table 2. Primary Endpoint of Stroke or Systemic Embolism.

<table>
<thead>
<tr>
<th>Study Population</th>
<th>Rivaroxaban</th>
<th></th>
<th>Warfarin</th>
<th></th>
<th>Hazard Ratio (95% CI)</th>
<th>P Value</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>No. of Patients</td>
<td>No. of Events</td>
<td>Event Rate</td>
<td>No. of Patients</td>
<td>No. of Events</td>
<td>Event Rate</td>
</tr>
<tr>
<td>Per-protocol, as-treated population‡</td>
<td>6958</td>
<td>188</td>
<td>1.7</td>
<td>7004</td>
<td>241</td>
<td>2.2</td>
</tr>
<tr>
<td>Safety, as-treated population</td>
<td>7061</td>
<td>189</td>
<td>1.7</td>
<td>7082</td>
<td>243</td>
<td>2.2</td>
</tr>
<tr>
<td>Intention-to-treat population§</td>
<td>7081</td>
<td>269</td>
<td>2.1</td>
<td>7090</td>
<td>306</td>
<td>2.4</td>
</tr>
<tr>
<td>During treatment</td>
<td>188</td>
<td>1.7</td>
<td></td>
<td>240</td>
<td>2.2</td>
<td>0.79 (0.66–0.96)</td>
</tr>
<tr>
<td>After discontinuation</td>
<td>81</td>
<td>4.7</td>
<td></td>
<td>66</td>
<td>4.3</td>
<td>1.10 (0.79–1.52)</td>
</tr>
</tbody>
</table>
Why was the per-protocol analysis used to analyze the data to test non-inferiority?

1. It was an error. The intention-to-treat analysis should have been used.

2. The per protocol analysis may decrease the risk of type I error compared with the intention-to-treat analysis.

3. The per protocol analysis may decrease the risk of type II error compared with the intention-to-treat analysis.
Choosing an Analysis Plan

• Statistical tests test whether the observation is true under the null hypothesis
• Type I error
  – False positive
  – Most serious error
• Choose an analysis plan less likely to make false positive
Analysis Plan Less Likely to Make False Positive

<table>
<thead>
<tr>
<th>Superiority</th>
<th>Non-inferiority</th>
</tr>
</thead>
<tbody>
<tr>
<td>$H_0$</td>
<td>$H_0$</td>
</tr>
<tr>
<td>- There is no difference in primary outcomes between treatments A and B ($A = B$)</td>
<td>- Treatment A is worse than treatment B ($A &lt; B$)</td>
</tr>
<tr>
<td>- Analysis plan less likely to make false positive:</td>
<td>- Analysis plan less likely to make false positive:</td>
</tr>
</tbody>
</table>
## Per-protocol and ITT Analyses in ROCKET-AF

### Table 2. Primary End Point of Stroke or Systemic Embolism.

<table>
<thead>
<tr>
<th>Study Population</th>
<th>Rivaroxaban</th>
<th>Warfarin</th>
<th>Hazard Ratio (95% CI)</th>
<th>P Value</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>No. of Patients</td>
<td>No. of Events</td>
<td>Event Rate no./100 patient-yr</td>
<td>No. of Patients</td>
</tr>
<tr>
<td>Per-protocol, as-treated population</td>
<td>6958</td>
<td>188</td>
<td>1.7</td>
<td>7004</td>
</tr>
<tr>
<td>Safety, as-treated population</td>
<td>7061</td>
<td>189</td>
<td>1.7</td>
<td>7082</td>
</tr>
<tr>
<td>Intention-to-treat population§</td>
<td>7081</td>
<td>269</td>
<td>2.1</td>
<td>7090</td>
</tr>
<tr>
<td>During treatment</td>
<td>188</td>
<td>1.7</td>
<td></td>
<td>240</td>
</tr>
<tr>
<td>After discontinuation</td>
<td>81</td>
<td>4.7</td>
<td></td>
<td>66</td>
</tr>
</tbody>
</table>
Testing Superiority

• Once non-inferiority is established, then superiority can be tested.

<table>
<thead>
<tr>
<th>Study Population</th>
<th>Rivaroxaban</th>
<th>Warfarin</th>
<th>Hazard Ratio (95% CI)</th>
<th>P Value</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>No. of Patients</td>
<td>No. of Events</td>
<td>Event Rate</td>
<td>No. of Patients</td>
</tr>
<tr>
<td>Per-protocol, as-treated population</td>
<td>6958</td>
<td>188</td>
<td>1.7</td>
<td>7004</td>
</tr>
<tr>
<td>Safety, as-treated population</td>
<td>7061</td>
<td>189</td>
<td>1.7</td>
<td>7082</td>
</tr>
<tr>
<td>Intention-to-treat population</td>
<td>7081</td>
<td>269</td>
<td>2.1</td>
<td>7090</td>
</tr>
<tr>
<td>During treatment</td>
<td>188</td>
<td>1.7</td>
<td></td>
<td>240</td>
</tr>
<tr>
<td>After discontinuation</td>
<td>81</td>
<td>4.7</td>
<td></td>
<td>66</td>
</tr>
</tbody>
</table>
Why was as-treated analysis used for safety analysis?

Table 2. Primary End Point of Stroke or Systemic Embolism.*

<table>
<thead>
<tr>
<th>Study Population</th>
<th>Rivaroxaban</th>
<th>Warfarin</th>
<th>Hazard Ratio (95% CI)</th>
<th>P Value</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>No. of Patients</td>
<td>No. of Events</td>
<td>Event Rate no./100 patient-yr</td>
<td>No. of Patients</td>
</tr>
<tr>
<td>Per-protocol, as-treated population‡</td>
<td>6958</td>
<td>188</td>
<td>1.7</td>
<td>7004</td>
</tr>
<tr>
<td>Safety, as-treated population</td>
<td>7061</td>
<td>189</td>
<td>1.7</td>
<td>7082</td>
</tr>
<tr>
<td>Intention-to-treat population†</td>
<td>7081</td>
<td>269</td>
<td>2.1</td>
<td>7090</td>
</tr>
<tr>
<td>During treatment</td>
<td>188</td>
<td>1.7</td>
<td>240</td>
<td>2.2</td>
</tr>
<tr>
<td>After discontinuation</td>
<td>81</td>
<td>4.7</td>
<td>66</td>
<td>4.3</td>
</tr>
</tbody>
</table>
You are developing a new oral drug for the treatment of type 2 diabetes. Given the cardiovascular safety concern of antidiabetic drugs on the market (e.g., rosiglitazone), you want to test whether your drug does not increase the risk of MI, stroke, or cardiovascular death compared with placebo. Which of the following design is best to test the cardiovascular safety of your drug?

1. Superiority trial
2. Non-inferiority trial
3. Equivalence trial
If you are designing a noninferiority trial to test cardiovascular safety of your new antidiabetic drug, what would be your margin compared with placebo?

1. 10% (i.e., hazard ratio up to 1.10)
2. 20%
3. 30%
4. 40%
5. 50%
6. None of the above.
Effect of Sitagliptin on Cardiovascular Outcomes in Type 2 Diabetes

We assessed the primary noninferiority hypothesis by determining whether the upper boundary of the two-sided 95% confidence interval of the hazard ratio for the risk of the primary composite cardiovascular outcome did not exceed 1.30 in the sitagliptin group, as compared with the placebo group, in the per-protocol population,
Summary

• When evaluating a non-inferiority trial,
• Consider what advantages other than efficacy the new treatment has over the standard treatment.
• Check which analysis plan (i.e., ITT vs. per-protocol) was used.
• Use the non-inferiority margin to determine statistical significance of results
• Compare safety results with efficacy results and potential advantages of the new treatment
• Determine clinical significance
1. Write down the course code. Space has been provided in the daily program-at-a-glance sections of your program book.

2. To claim credit: Go to www.cshp.org/cpe before December 1, 2016.