Testing Superiority, Equivalence, and Non-Inferiority

Benoît Mâsse

Statistical Center for HIV/AIDS Research & Prevention
Seattle, WA USA

Consultation on Standards of Prevention in HIV Prevention Trials
Speke Resort Munyonyo, Kampala, Uganda
March 26-28, 2009
Active Control Superior

New Intervention Superior

Non-Inferiority Margin

Superiority

Non-inferiority

Equivalence

-Δ_M  0  +Δ_M

Difference in HIV Rates

(Adapted from Gomberg-Maitland et al, AHJ Vol. 146:3)
Superiority
Non-inferiority
Equivalence

Active Control
Superior

New Intervention
Superior

Difference in HIV Rates

(Adapted from Gomberg-Maitland et al, AHJ Vol. 146:3)
Non-Inferiority Margin

Superiority

Non-Inferiority

Equivalence

Active Control Superior

New Intervention Superior

\(-\Delta_M\) \hspace{1cm} 0 \hspace{1cm} +\Delta_M

Difference in HIV Rates

(Adapted from Gomberg-Maitland et al, AHJ Vol. 146:3)
(a) Non-inferiority shown
(b) Non-inferiority not shown
(c) Equivalence
(d) Equivalence
(e) Equivalence
(f) Equivalence not shown
(g) Superiority shown

Control Better $- \Delta$ True difference $0$ $+ \Delta$ New Intervention Better
## ICH E10: Choice of Control Groups & Related Issues in Clinical Trials

### Possible: If there is historical evidence of sensitivity to intervention effect
Are Non-Inferiority Trials Feasible?

Feasibility in terms of Sample Size

Sample size primarily driven by:

1) The anticipated effectiveness of the new intervention
2) The size of the non-inferiority margin: small margin requires large sample size
Anticipated Effectiveness of New Intervention?

- Difficult to Determine
  - Phase II HIV prevention trials typically do not provide ‘early’ estimates of effectiveness
  - In this case, for sample size calculations, we typically assume that the new intervention has the same effectiveness as the active control

- Unfortunately, this leads to larger sample sizes
Size of Non-Inferiority Margin ($M$)

No agreed way of determining $M$

- But always done conservatively

- For licensure, $M$ typically need to be negotiated with regulators
Size of Non-Inferiority Margin ($M$)

However, if the effectiveness of the active control is small to moderate (i.e. 30%-50%) $M$ will have to be quite small!
### EFFECTIVENESS

- **Active Control**
  - 60%

- **Placebo**
  - 0%

- **Non-Inferiority Margin (M)**
  - \(60\% - M\)

- **Amount of Effect to be Retained (R)**
  - \(60\%\)
## Non-Inferiority Trial Examples

<table>
<thead>
<tr>
<th>Active Control Effectiveness</th>
<th>Amount of Effect to be Retained</th>
<th>Number of HIV endpoints required*</th>
<th>Required resources</th>
</tr>
</thead>
<tbody>
<tr>
<td>60%</td>
<td>25%</td>
<td>~80</td>
<td>~Average size of current trials</td>
</tr>
<tr>
<td>35%</td>
<td>25%</td>
<td>~1500</td>
<td>~5-7 times the size of current large trials</td>
</tr>
<tr>
<td>30%</td>
<td>25%</td>
<td>~6500</td>
<td>~20 times the size of current large trials</td>
</tr>
</tbody>
</table>

*Assuming a 2-arm non-inferiority trial with 80% power and false-positive error rate of 2.5%
## Non-Inferiority Trial Examples

<table>
<thead>
<tr>
<th>Active Control Effectiveness</th>
<th>Amount of Effect to be Retained</th>
<th>Number of HIV endpoints required*</th>
<th>Required resources</th>
</tr>
</thead>
<tbody>
<tr>
<td>60%</td>
<td>40%</td>
<td>~190</td>
<td>~Average size of current large trials</td>
</tr>
</tbody>
</table>

*Assuming a 2-arm non-inferiority trial with 80% power and false-positive error rate of 2.5%
Size of Non-Inferiority Margin (M)

- 30-60 person-years per HIV endpoint
  - 1500 endpoints will require 45,000 to 90,000 person-years
  - 190 endpoints will require 6,000 to 12,000 p-y

- Small M leads to sample sizes in the range of +100,000 women !!!
What about the impact of the standard prevention package?

- If the prevention package is indeed effective then this will increase the sample size further more

=> If baseline incidence is 2% instead of 4%, the # of HIV endpoints remains the same but the # of person-years will double
Concluding Remarks

• Placebo or Active Controls?
  – Is ‘effective’ intervention accepted uniformly as standard treatment?
  – Will it ‘work’ in every populations?
  – Will past results be still valid?

• Placebo Controls: in un-tested populations
  – or where doubts exist
Concluding Remarks

• Active Control Trials:
  – Non-inferiority trials are more challenging
  – Non-inferiority trials likely to be not feasible if effectiveness of active control is in the 20%-50% range.

• Superiority trials: ‘Beat’ the active control
  – Could aimed at superiority first then settle for non-inferiority … that’s OK but need to define the non-inferiority margin a-priori.
### A Superiority Trial Example

<table>
<thead>
<tr>
<th>Active Control Effectiveness</th>
<th>Effectiveness of New Intervention to be Detected</th>
<th>Number of HIV endpoints required*</th>
<th>Required resources</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>~160</td>
<td>~Average size of current large trials</td>
</tr>
<tr>
<td>30%</td>
<td>55%</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*Assuming a 2-arm superiority trial with 80% power and false-positive error rate of 2.5%