A mixed approach for proving non-inferiority with respect to binary endpoints

Valentin Rousson and Burkhard Seifert

Biostatistics Unit
Institute for Social and Preventive Medicine
University of Zurich

rousso@ifspm.uzh.ch

Berne, ROeS 2007
Comparing treatments in a randomized clinical trial (RCT)

In a RCT, a new treatment should be

- **Significantly better than a placebo**

- **Statistically non-inferior to an established treatment** (active control)

**Significantly better**: Better (beyond chance)

**Statistically non-inferior**: Not much worse (beyond chance)

Can be assessed by comparing the **left bound of a 95% CI** for the relative effect of the new treatment with a **non-inferiority margin** (see e.g. Blackwelder, *Controlled Clinical Trials*, 1982)
Non-inferiority vs. equivalence testing

- Relative effect
- Equivalence domain
- Non-inferiority
- Equivalence
- No conclusion

- ns
- ns
- ns
- sig
- sig
Measuring a relative effect for a binary endpoint

Let $p_X$ be the proportion of success for the active control

Let $p_Y$ be the proportion of success for the new treatment

These are measures of “absolute effects”

To measure a relative effect, one should compare both proportions

Candidates are:

• The difference $D = p_Y - p_X$

• The ratio $R = \frac{p_Y}{p_X}$

• The odds-ratio $OR = \frac{p_Y(1 - p_X)}{p_X(1 - p_Y)}$
### Possible values

<table>
<thead>
<tr>
<th></th>
<th>$D$</th>
<th>$R$</th>
<th>$OR$</th>
</tr>
</thead>
<tbody>
<tr>
<td>General</td>
<td>$(-1; 1)$</td>
<td>$(0; +\infty)$</td>
<td>$(0; +\infty)$</td>
</tr>
<tr>
<td>Given $p_X$</td>
<td>$(-p_X; 1-p_X)$</td>
<td>$(0; 1/p_X)$</td>
<td>$(0; +\infty)$</td>
</tr>
<tr>
<td>$p_X = 0.2$</td>
<td>$(-0.2; 0.8)$</td>
<td>$(0; 5)$</td>
<td>$(0; +\infty)$</td>
</tr>
<tr>
<td>$p_X = 0.5$</td>
<td>$(-0.5; 0.5)$</td>
<td>$(0; 2)$</td>
<td>$(0; +\infty)$</td>
</tr>
<tr>
<td>$p_X = 0.8$</td>
<td>$(-0.8; 0.2)$</td>
<td>$(0; 1.25)$</td>
<td>$(0; +\infty)$</td>
</tr>
</tbody>
</table>

⇒ Domain of possible values of $OR$ independent from the absolute efficacy of treatments (see e.g. Garrett, *Statistics in Medicine*, 2003)

⇒ The non-inferiority margin can be chosen independently from the absolute efficacy of treatments (for example $\varepsilon_{OR} = 0.5$)
Proportion of success and proportion of failure

Comparing proportions of failure \(q_X = 1 - p_X\) and \(q_Y = 1 - p_Y\) yields

- \(D' = q_X - q_Y = (1 - p_X) - (1 - p_Y) = D\)

- \(R' = \frac{q_X}{q_Y} = \frac{1 - p_X}{1 - p_Y} \neq R\)

- \(OR' = \frac{q_X (1-q_Y)}{q_Y (1-q_X)} = \frac{(1-p_X)p_Y}{(1-p_Y)p_X} = OR\)

Example: \(p_X = 0.8,\ p_Y = 0.9,\ q_X = 0.2\) and \(q_Y = 0.1\)

- \(D = D' = 0.1\)

- \(R = 1.125\) and \(R' = 2\)

- \(OR = OR' = 2.25\)
Relationship between $D$ and $OR$

• $OR$ and $D$ are not one-to-one related

Examples:

If $p_X = 0.8$, $OR = 0.5$ corresponds to $D = -0.13$ ($p_Y = 0.67$)

If $p_X = 0.5$, $OR = 0.5$ corresponds to $D = -0.17$ ($p_Y = 0.33$)

• But given $p_X$, they are related as

$$D = \frac{p_X(1-p_X)(OR-1)}{1 + p_X(OR-1)}$$

Example: If one knew that $p_X = 0.8$, it is the same to show $OR \geq 0.5$ than to show $D \geq -0.13$

Result: It is statistically easier to show the latter than to show the former if $p_X = p_Y$ is large!
Planning a non-inferiority study

• One usually assumes \( OR = 1 \) and aims to show \( OR \geq \varepsilon_{OR} \)

Assuming a value for \( p_X = p_Y \), one should enroll per group

\[
n_{OR} = \frac{2(z_\alpha + z_\beta)^2}{p_X(1 - p_X)(\log(\varepsilon_{OR}))^2}
\]

• But given \( p_X \), one may equivalently assume \( D = 0 \) and aims to show \( D \geq \varepsilon_D \) with

\[
\varepsilon_D = \frac{p_X(1 - p_X)(\varepsilon_{OR} - 1)}{1 + p_X(\varepsilon_{OR} - 1)}
\]

Here, one should enroll per group

\[
n_D = \frac{2(z_\alpha + z_\beta)^2 p_X(1 - p_X)}{\varepsilon_D^2}
\]

Result: \( n_D \leq n_{OR} \) if \( p_X \geq \frac{1}{1 - \varepsilon_{OR}} + \frac{1}{\log(\varepsilon_{OR})} \) (Rousson and Seifert, 2007)
## Sample size comparison ($n_{OR}/n_D$)

<table>
<thead>
<tr>
<th>$1 - \beta$</th>
<th>$p_X = p_Y$</th>
<th>0.43</th>
<th>0.5</th>
<th>$\varepsilon_{OR}$</th>
<th>0.55</th>
<th>0.8</th>
</tr>
</thead>
<tbody>
<tr>
<td>80%</td>
<td>0.5</td>
<td>89/99</td>
<td>131/142 (139)</td>
<td>176/187</td>
<td>1262/1272</td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.6</td>
<td>92/88</td>
<td>137/129 (126)</td>
<td>184/173</td>
<td>1314/1267</td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.7</td>
<td>105/84</td>
<td>156/127 (126)</td>
<td>210/174</td>
<td>1502/1383</td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.8</td>
<td>138/90</td>
<td>205/142 (145)</td>
<td>275/199</td>
<td>1971/1731</td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.9</td>
<td>245/128</td>
<td>364/212 (226)</td>
<td>489/305</td>
<td>3503/2932</td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.95</td>
<td>464/214</td>
<td>688/365 (401)</td>
<td>925/535</td>
<td>6638/5421</td>
<td></td>
</tr>
<tr>
<td>90%</td>
<td>0.5</td>
<td>119/133</td>
<td>175/190 (186)</td>
<td>236/250</td>
<td>1689/1703</td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.6</td>
<td>123/117</td>
<td>183/172 (169)</td>
<td>245/231</td>
<td>1759/1696</td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.7</td>
<td>141/112</td>
<td>209/170 (169)</td>
<td>280/232</td>
<td>2010/1851</td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.8</td>
<td>185/120</td>
<td>274/190 (193)</td>
<td>368/266</td>
<td>2638/2317</td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.9</td>
<td>328/171</td>
<td>486/283 (299)</td>
<td>654/409</td>
<td>4690/3926</td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.95</td>
<td>622/287</td>
<td>921/488 (531)</td>
<td>1238/717</td>
<td>8886/7257</td>
<td></td>
</tr>
</tbody>
</table>

(In parentheses are the $n_D$ obtained using the formula of Farrington and Manning, *Statistics in Medicine*, 1990)
Mixed approach to show non-inferiority

1. Define the non-inferiority margin for $OR$, for example $\varepsilon_{OR} = 0.5$

2. Assume a proportion of success for the active control, for example $p_X = 0.8$

3. Using the one-to-one relationship above, calculate the corresponding non-inferiority margin for $D$, for example

$$\varepsilon_D = \frac{0.8(1 - 0.8)(0.5 - 1)}{1 + 0.8(0.5 - 1)} = -0.13$$

4. Calculate a 95% CI for $D$ and check that its lower bound is larger than $\varepsilon_D$ ($-0.13$ in that example)
**Relationship between $\varepsilon_D$ and $\varepsilon_{OR}$ given $p_X$**

<table>
<thead>
<tr>
<th>$p_X$</th>
<th>$\varepsilon_{OR}$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0.43</td>
</tr>
<tr>
<td>0.5</td>
<td>−0.199</td>
</tr>
<tr>
<td>0.55</td>
<td>−0.205</td>
</tr>
<tr>
<td>0.6</td>
<td>−0.208</td>
</tr>
<tr>
<td>0.65</td>
<td>−0.206</td>
</tr>
<tr>
<td>0.7</td>
<td>−0.199</td>
</tr>
<tr>
<td>0.75</td>
<td>−0.187</td>
</tr>
<tr>
<td>0.8</td>
<td><strong>−0.168</strong></td>
</tr>
<tr>
<td>0.85</td>
<td>−0.141</td>
</tr>
<tr>
<td>0.9</td>
<td>−0.105</td>
</tr>
<tr>
<td>0.95</td>
<td>−0.059</td>
</tr>
</tbody>
</table>

Most of these margins lie between $−0.1$ and $−0.2$ and become wider for decreasing efficiencies, as recommended by the FDA (1997)
Approximate 95% CI for $D$ and $OR$

Let $\hat{p}_X$ and $\hat{p}_Y$ be empirical proportions of success from samples of size $n$, and let $\hat{D} = \hat{p}_Y - \hat{p}_X$ and $\hat{OR} = \hat{p}_Y (1 - \hat{p}_X) / (\hat{p}_X (1 - \hat{p}_Y))$

$$\hat{D} \pm 1.96 \sqrt{\frac{\hat{p}_X (1 - \hat{p}_X)}{n}} + \frac{\hat{p}_Y (1 - \hat{p}_Y)}{n}$$  \hspace{1cm} (1)

$$\log(\hat{OR}) \pm 1.96 \sqrt{\frac{1}{\hat{p}_X (1 - \hat{p}_X) n} + \frac{1}{\hat{p}_Y (1 - \hat{p}_Y) n}}$$  \hspace{1cm} (2)

• Conventional $OR$ approach: Uses $\varepsilon_{OR}$ and (2)

• Mixed approach: Uses $\varepsilon_D(\varepsilon_{OR}, p_X)$ and (1)
Power comparison \((n = 150 \text{ per group})\)
Example

- Non-inferiority margin $\varepsilon_{OR} = 0.5$
  
  Assumed value for $p_X = 0.8$
  
  Corresponds to non-inferiority margin $\varepsilon_D = -0.13$

- Data: $n = 150$ patients per group
  
  $\hat{p}_X = 125/150 = 0.833$ and $\hat{p}_Y = 121/150 = 0.807$
  
  Corresponds to $\hat{D} = -0.027$ and $\hat{OR} = 0.83$

- Conventional $OR$ approach:
  
  95% CI for $OR$ is $[0.46; 1.50] \Rightarrow$ One fails to show non-inferiority

- Mixed approach:
  
  95% CI for $D$ is $[-0.11; 0.06] \Rightarrow$ One has shown non-inferiority
Conditional power comparison \((n = 150 \text{ per group})\)
Generalization to an ordinal endpoint

Let $X$ and $Y$ be outcomes of active control and new treatment on an ordinal scale (e.g. failure, moderate success, big success)

One usually measures relative effect by

$$
\theta_1 = \Pr\{X < Y\} + \frac{\Pr\{X = Y\}}{2}
$$

or

$$
\theta_2 = \frac{\Pr\{X < Y\}}{1 - \Pr\{X = Y\}}
$$

Equivalently, one can use

$$
\theta_1^* = 2\theta_1 - 1 = \Pr\{X < Y\} - \Pr\{X > Y\}
$$

or

$$
\theta_2^* = \frac{\theta_2}{1 - \theta_2} = \frac{\Pr\{X < Y\}}{\Pr\{X > Y\}}
$$
Generalization of $D$ and $OR$

For a binary endpoint with $p_X$ and $p_Y$, one has

$$\Pr\{X < Y\} = (1 - p_X)p_Y \text{ and } \Pr\{X = Y\} = (1 - p_X)(1 - p_Y) + p_Xp_Y$$

Thus,

$$\theta_1^* = \Pr\{X < Y\} - \Pr\{X > Y\} = (1 - p_X)p_Y - (1 - p_Y)p_X = p_Y - p_X = D$$

and

$$\theta_2^* = \frac{\Pr\{X < Y\}}{\Pr\{X > Y\}} = \frac{(1 - p_X)p_Y}{(1 - p_Y)p_X} = OR$$

Remark: $D$ and $OR$ can further be generalized as

$$D = \Pi_c - \Pi_d \quad \text{(Kendall's $\tau_a$, Biometrika, 1938)}$$

and

$$OR = \frac{\Pi_c}{\Pi_d} \quad \text{(General OR of Agresti, Biometrics, 1980)}$$

where $\Pi_c$ and $\Pi_d$ are probabilities of concordance/discordance
Generalization of the mixed approach

1. Define a non-inferiority margin $\varepsilon_{OR}$

2. Assume a distribution $p_X$ for $X$ (for example $p_X = (0.1, 0.2, 0.7)$)

3. Calculate a corresponding non-inferiority margin $\varepsilon_D$ based on $\varepsilon_{OR}$ and $p_X$ (no unique solution, see Rousson and Seifert, 2007)

4. Calculate a 95% CI for $D$ (see Edwardes, *Biometrics*, 1995) and check that its lower bound is larger than $\varepsilon_D$
## Power comparison (3 possible outcomes, \( n = 150 \) per group)

<table>
<thead>
<tr>
<th>( p_2 )</th>
<th>0.1</th>
<th>0.2</th>
<th>0.3</th>
<th>0.4</th>
<th>0.5</th>
<th>0.6</th>
<th>0.7</th>
<th>0.8</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.1</td>
<td>72/53</td>
<td>84/68</td>
<td>90/80</td>
<td>92/87</td>
<td>92/92</td>
<td><strong>90/93</strong></td>
<td><strong>84/93</strong></td>
<td>72/89</td>
</tr>
<tr>
<td>0.2</td>
<td>84/69</td>
<td>91/81</td>
<td>94/89</td>
<td>95/93</td>
<td>94/95</td>
<td><strong>91/95</strong></td>
<td><strong>84/93</strong></td>
<td></td>
</tr>
<tr>
<td>0.3</td>
<td>89/79</td>
<td>94/88</td>
<td>96/93</td>
<td>95/95</td>
<td>94/96</td>
<td></td>
<td><strong>89/94</strong></td>
<td></td>
</tr>
<tr>
<td>0.4</td>
<td>91/86</td>
<td>94/92</td>
<td>95/95</td>
<td>95/95</td>
<td>92/94</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.5</td>
<td>92/90</td>
<td>94/93</td>
<td>94/94</td>
<td>92/92</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.6</td>
<td>90/91</td>
<td>91/92</td>
<td>89/88</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.7</td>
<td><strong>84/88</strong></td>
<td>84/84</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.8</td>
<td><strong>71/75</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>