Generalized pairwise comparisons in immuno-oncology

Marc Buyse, ScD
San Francisco, CA
Outline

- General setup
- Extensions of the Wilcoxon Mann-Whitney test
  - Generalized pairwise comparisons
  - The net chance of a better outcome
- Potential uses for immuno-oncology
  - Delayed treatment effects
  - Several prioritized outcomes
- Conclusions
- References
Let $X_i$ be the continuous outcome of the $i^{th}$ subject in $T$ 
\[ (i = 1, \ldots, n) \]

Let $Y_j$ be the continuous outcome of the $j^{th}$ subject in $C$ 
\[ (j = 1, \ldots, m) \]

The Mann-Whitney form of the Wilcoxon test

The Wilcoxon test statistic can be derived from all possible pairs of subjects, one from $T$ and one from $C$.

Let

$$U_{ij} = \begin{cases} 
+1 & \text{if } X_i > Y_j \\
-1 & \text{if } X_i < Y_j \\
0 & \text{otherwise}
\end{cases}$$

$$U = \frac{1}{m \cdot n} \sum_{i=1}^{n} \sum_{j=1}^{m} U_{ij}$$

$$W = m \cdot n \cdot (1 - U)/2$$

Generalization to any outcome measure

Let $X_i$ and $Y_j$ be observed outcomes for any outcome measure (continuous, time to event, binary, categorical, ...)

General measure of treatment effect

Define

\[
U_{ij} = \begin{cases} 
+1 & \text{if } (X_i, Y_j) \text{ pair is favorable} \\
-1 & \text{if } (X_i, Y_j) \text{ pair is unfavorable} \\
0 & \text{otherwise}
\end{cases}
\]

\[
U = \frac{1}{mn} \sum_{i=1}^{n} \sum_{j=1}^{m} U_{ij}
\]

\(U\) is a general measure of treatment effect called the «net chance of a better outcome» (\(\Delta\)). This measure is analogous to Pocock’s «win ratio» (on the absolute scale).

The net chance of a better outcome ($\Delta$)

$\Delta$ is a linear transformation of the probabilistic index $P(X > Y)$:

$$U = \Delta = 2 \cdot P(X > Y) - 1$$

<table>
<thead>
<tr>
<th>Situation</th>
<th>$P(X &gt; Y)$</th>
<th>$\Delta$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$T$ uniformly worse than $C$</td>
<td>0</td>
<td>-1</td>
</tr>
<tr>
<td>$T$ no different from $C$</td>
<td>0.5</td>
<td>0</td>
</tr>
<tr>
<td>$T$ uniformly better than $C$</td>
<td>1</td>
<td>+1</td>
</tr>
</tbody>
</table>

The empirical distribution of $\Delta$ (hence tests of significance and confidence intervals for $\hat{\Delta}$) can be obtained through re-randomization.

Generalization to *prioritized* outcomes, e.g. OS

<table>
<thead>
<tr>
<th>OS difference &gt; 1 year</th>
<th>OS difference &gt; 6 months</th>
<th>Pair is</th>
</tr>
</thead>
<tbody>
<tr>
<td>Favorable</td>
<td>-</td>
<td>Favorable</td>
</tr>
<tr>
<td>Unfavorable</td>
<td>-</td>
<td>Unfavorable</td>
</tr>
<tr>
<td>Neutral or ?</td>
<td>Favorable</td>
<td>Favorable</td>
</tr>
<tr>
<td>Neutral or ?</td>
<td>Unfavorable</td>
<td>Unfavorable</td>
</tr>
<tr>
<td>Neutral or ?</td>
<td>Neutral or ?</td>
<td>Neutral</td>
</tr>
</tbody>
</table>
The Net Chance of a Longer Survival as a Patient-Oriented Measure of Treatment Benefit in Randomized Clinical Trials

Julien Péron, MD, PhD; Pascal Roy, MD, PhD; Brice Ozanne, PhD; Laurent Roche, PhD; Marc Buyse, ScD

**IMPORTANCE** Time to events, or survival end points, are common end points in randomized clinical trials. They are usually analyzed under the assumption of proportional hazards, and the treatment effect is reported as a hazard ratio, which is neither an intuitive measure nor a meaningful one if the assumption of proportional hazards is not met.

Published online April 28, 2016.
Proportional hazards

Early difference

Example: cytotoxics
Delayed difference

Scenario 3: delayed survival difference

Example: immunotherapy for advanced solid tumors
Cure rate

Example: allografts in childhood tumors
Mixture of two populations

Example: predictive factor for targeted therapy
Power - proportional hazards
Power – delayed differences
Power – cure rate

![Graph showing the relationship between power and the threshold of clinical pertinence for different methods: GPC with no censoring, Log-Rank test with no censoring, GPC with 20% censoring, and Log-Rank test with 20% censoring. The graph plots power on the y-axis against the threshold of clinical pertinence on the x-axis.]
Generalization to several prioritized outcomes, e.g. OS and grade 3-4 toxicity

<table>
<thead>
<tr>
<th>OS difference</th>
<th>Worst grade toxicity</th>
<th>Pair is</th>
</tr>
</thead>
<tbody>
<tr>
<td>&gt; 2 months</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Favorable</td>
<td>-</td>
<td>Favorable</td>
</tr>
<tr>
<td>Unfavorable</td>
<td>-</td>
<td>Unfavorable</td>
</tr>
<tr>
<td>Neutral or ?</td>
<td>Favorable</td>
<td>Favorable</td>
</tr>
<tr>
<td>Neutral or ?</td>
<td>Unfavorable</td>
<td>Unfavorable</td>
</tr>
<tr>
<td>Neutral or ?</td>
<td>Neutral or ?</td>
<td>Neutral</td>
</tr>
</tbody>
</table>
Several prioritized outcomes

Assessing the benefit–risk of new treatments using generalised pairwise comparisons: the case of erlotinib in pancreatic cancer

J Péron*,1,2, P Roy1,2, K Ding3, W R Parulekar3, L Roche1,2 and M Buyse4
Several prioritized outcomes

Table 3. Main analysis of the benefit–risk balance of erlotinib and gemcitabine combination

<table>
<thead>
<tr>
<th>Priority</th>
<th>Proportion of pairs (%)</th>
<th>Difference</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Erlotinib &gt; placebo</td>
<td>Placebo &gt; erlotinib</td>
</tr>
<tr>
<td>OS (threshold = 2 months)</td>
<td>37.0</td>
<td>32.3</td>
</tr>
<tr>
<td>Worst related AE grade</td>
<td>7.5</td>
<td>15.7</td>
</tr>
<tr>
<td>Overall</td>
<td>44.5</td>
<td>48.1</td>
</tr>
</tbody>
</table>

Abbreviations: > = better than; AE = adverse events; Δ[erlotinib] = proportion in favour of the erlotinib group; OS = overall survival.

Ref: Peron et al. BJC 2015.
Several prioritized outcomes

First priority: Overall survival
Second priority endpoint: Worst grade of at least possibly related adverse events
Generalized Pairwise Comparisons

- are equivalent to standard non-parametric tests in simple cases
- naturally lead to a patient-relevant general measure of treatment effect, $\Delta$
- may have better power than the logrank test (e.g. for delayed treatment benefits)
- allow for testing of differences thought to be clinically relevant
- allow for any number of prioritized outcomes of any type to be analyzed simultaneously
References


