Generalized pairwise comparisons for precision medicine

Tomasz Burzykowski
IDDI & Hasselt University, Belgium
tomasz.burzykowski@iddi.com
Personalizing Treatment Choices

• **Evidence-based medicine**
  – (Meta-analyses of) randomized control trials
  – Subgroup analyses, if appropriate

• **Precision medicine**
  “Giving the right treatment to the right patient at the right time”

• **Personalized medicine**
  – Precision medicine with personalized/patient-centric choices for therapeutic decisions
Consider the following results

A patient might reason:

- Taking combination, I’m more likely to live longer (by how much?)
- Taking combination, I’m more likely to have grade 3/4 adverse events (AEs)
- I’m willing to experience AEs for a survival benefit of at least \( m \) months...

Limitations of Standard Analyses

- A single (primary) endpoint drives decision-making
- Other endpoints are analyzed descriptively
- Safety informally balanced against efficacy, resulting in debatable risk / benefit analyses
- Patient preferences are not formally taken into account
Generalized pairwise comparisons of prioritized outcomes in the two-sample problem

Marc Buyse\textsuperscript{a,b,*†}
Randomized Trial

Let $X_i$ be the outcome of $i^{th}$ subject in $T$ ($i = 1, \ldots, n$)

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

Let $X_i$ and $Y_j$ be the observed values of a continuous outcome.

- $X_i > Y_j$ (WIN)
- $X_i < Y_j$ (LOSS)
- $X_i = Y_j$ (TIE)
- $X_i$ or $Y_j$ missing (uninformative) (?)

Buyse, Stat Med 2010;29:3245
Net Benefit

- Let $p_{ij}$ be equal to
  - 1 if the pair is a win
  - -1 if the pair is a loss
  - 0 if the pair is a tie/?

- Then
  \[ U = \sum_{ij} p_{ij} / (nm) \]
  - A generalization of the Wilcoxon-Mann-Whitney test-statistic

- And (if no missing data)
  \[ \Delta = E(U) = P(X>Y)-P(X<Y) = P(T \text{ better})-P(C \text{ better}) \]
# Illustration

<table>
<thead>
<tr>
<th>TREATMENT GROUP (T)</th>
<th>CONTROL GROUP (C)</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>1</td>
</tr>
<tr>
<td>5</td>
<td>3</td>
</tr>
<tr>
<td>6</td>
<td>3</td>
</tr>
<tr>
<td>9</td>
<td>7</td>
</tr>
<tr>
<td>11</td>
<td>9</td>
</tr>
<tr>
<td>12</td>
<td>9</td>
</tr>
</tbody>
</table>
TIES

TREATMENT GROUP (T) versus CONTROL GROUP (C)

<table>
<thead>
<tr>
<th>TREATMENT GROUP (T)</th>
<th>CONTROL GROUP (C)</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>1</td>
</tr>
<tr>
<td>5</td>
<td>3</td>
</tr>
<tr>
<td>6</td>
<td>3</td>
</tr>
<tr>
<td>9</td>
<td>7</td>
</tr>
<tr>
<td>11</td>
<td>9</td>
</tr>
<tr>
<td>12</td>
<td>9</td>
</tr>
</tbody>
</table>

4 TIES
Losses

TREATMENT GROUP (T)

CONTROL GROUP (C)

3
5
6
9
11
12

9 LOSSES
The probability of a patient having a better outcome

- if on treatment is 0.64
- if on control is 0.25

The "net benefit" is 0.39

Note: the "win ratio" is 0.64 / 0.25 = 2.56

---

**Net Benefit**

<table>
<thead>
<tr>
<th>Ties</th>
<th>Wins</th>
<th>Losses</th>
<th>Net benefit</th>
</tr>
</thead>
<tbody>
<tr>
<td>4 / 36 = 0.11</td>
<td>23 / 36 = 0.64</td>
<td>9 / 36 = 0.25</td>
<td>0.64 − 0.25 = 0.39</td>
</tr>
</tbody>
</table>

Pocock et al. Eur Heart J 2012; 33: 176
Now let $X_i$ and $Y_j$ be the observed values of any outcome measure (continuous, time-to-event, binary, categorical, …)
## Time to Event

<table>
<thead>
<tr>
<th>$X_i$ censored</th>
<th>$Y_j$ censored</th>
<th>$X_i &gt; Y_j$</th>
<th>$X_i &lt; Y_j$</th>
<th>$X_i = Y_j$</th>
</tr>
</thead>
<tbody>
<tr>
<td>No</td>
<td>No</td>
<td>Win</td>
<td>Loss</td>
<td>Tie</td>
</tr>
<tr>
<td>Yes</td>
<td>No</td>
<td>Win</td>
<td>?</td>
<td>?</td>
</tr>
<tr>
<td>No</td>
<td>Yes</td>
<td>?</td>
<td>Loss</td>
<td>?</td>
</tr>
<tr>
<td>Yes</td>
<td>Yes</td>
<td>?</td>
<td>?</td>
<td>?</td>
</tr>
</tbody>
</table>
## Thresholds of Clinical Relevance

| $X_i$ censored | $Y_j$ censored | $X_i-Y_j > m$ | $X_i-Y_j < -m$ | $|X_i-Y_j| \leq m$ |
|----------------|----------------|---------------|----------------|-----------------|
| No             | No             | Win           | Loss           | Tie             |
| Yes            | No             | Win           | ?              | ?               |
| No             | Yes            | ?             | Loss           | ?               |
| Yes            | Yes            | ?             | ?              | ?               |
Net Benefit – Proportional Hazards

**KAPLAN-MEIER CURVES**

**NET BENEFIT OF AT LEAST m MONTHS**

Péron et al, JAMA Oncol 2016;2:901
There is a 13% net probability that survival will be longer on T than C.

The “net benefit” of T is 13%
Net Benefit – Proportional Hazards

KAPLAN-MEIER CURVES

NET BENEFIT OF AT LEAST m MONTHS

There is an 8% net benefit of at least 10 months in favor of T

Péron et al, JAMA Oncol 2016;2:901
Power – Proportional Hazards

Power of several tests in the proportional hazards scenario

- GPC, no censoring
- Log-Rank test, no censoring
- GPC, censoring = 20%
- Log-Rank test, censoring = 20%

Threshold of clinical pertinence
Net Benefit – Delayed Difference

KAPLAN-MEIER CURVES

NET BENEFIT OF AT LEAST \( m \) MONTHS

Example: immunotherapy for advanced solid tumors

Péron et al, JAMA Oncol 2016;2:901
Power – Delayed Difference

Power of several tests in the delayed treatment effect scenario

Threshold of clinical pertinence

- GPC, no censoring
- Log-Rank test, no censoring
- GPC, censoring = 20%
- Log-Rank test, censoring = 20%
Net Benefit – Cure Rate

**KAPLAN-MEIER CURVES**

**NET BENEFIT OF AT LEAST m MONTHS**

Example: allografts in childhood tumors

Péron et al, JAMA Oncol 2016;2:901
Power – Cure Rate

Power of several tests in the cure rate scenario

Threshold of clinical pertinence

Power

GPC, no censoring
Log-Rank test, no censoring
GPC, censoring =20%
Log-Rank test, censoring=20%
Net Benefit

• Related to the ‘probabilistic index’, \( P(X > Y) \), [Accion et al. 2006, De Neve et al. 2013].

• The probabilistic index and related measures do not automatically generalize to other settings or different patient populations, where the variability of the outcome(s) of interest could be quite different [Senn 2011, Thas et al. 2012].

• These measures of benefit may be best seen as complementary to traditional (parametric) measures of benefit.
Prioritized Outcomes

Now let \((X_i, X'_i)\) and \((Y_j, Y'_j)\) be observed values of two outcome measures, with \(X\) \((Y)\) being prioritized over \(X'\) \((Y')\).

\[ X_i, X'_i \quad \text{pairwise comparison} \quad Y_j, Y'_j \]

<table>
<thead>
<tr>
<th>(X_i) vs. (Y_j)</th>
<th>(X'_i) vs. (Y'_j)</th>
<th>Pair is</th>
</tr>
</thead>
<tbody>
<tr>
<td>WIN</td>
<td>WIN</td>
<td>WIN</td>
</tr>
<tr>
<td>LOSS</td>
<td>LOSS</td>
<td>LOSS</td>
</tr>
<tr>
<td>TIE or ?</td>
<td>WIN</td>
<td>WIN</td>
</tr>
<tr>
<td>TIE or ?</td>
<td>LOSS</td>
<td>LOSS</td>
</tr>
<tr>
<td>TIE or ?</td>
<td>TIE or ?</td>
<td>TIE or ?</td>
</tr>
</tbody>
</table>

Buyse, Stat Med 2010;29:3245
Net Benefit

- Assume K outcomes
- Let $p_{ijk}$ ($k=1, ..., K$) be equal to
  - 1 if the pair is a win for the k-th outcome
  - -1 if the pair is a loss for the k-th outcome
  - 0 if the pair is a tie for the k-th outcome
- Define $u_{ijk} = I$(the pair is ? for the k-th outcome)
- Let $U(K) = \sum_{ij} \{p_{ij1}(1-u_{ij1}) + ... + p_{ijk}u_{ij1}...u_{ij,K-1}(1-u_{ijk})\}/(nm)$
- And $\Delta = E\{U(K)\}$
A Recent Example

Myeloablative Autologous Stem-Cell Transplantation for Severe Scleroderma


METHODS
We randomly assigned adults (18 to 69 years of age) with severe scleroderma to undergo myeloablative autologous stem-cell transplantation (36 participants) or to receive cyclophosphamide (39 participants). The primary end point was a global rank composite score comparing participants with each other on the basis of a hierarchy of disease features assessed at 54 months: death, event-free survival (survival without respiratory, renal, or cardiac failure), forced vital capacity, the score on the Disability Index of the Health Assessment Questionnaire, and the modified Rodnan skin score.

RESULTS
In the intention-to-treat population, global rank composite scores at 54 months showed the superiority of transplantation (67% of 1404 pairwise comparisons favored transplantation and 33% favored cyclophosphamide, P=0.01). In the per-protocol population
**BENEFIT Project**

- **Biostatistical Estimation of Net Effects For Individualization of Therapy**

- **Funds:** the Walloon Region, Biowin – the Health Cluster of Wallonia and Innoviris, the Brussels Institute for Research and Innovation.

- **Partners**
  - International Drug Development Institute (IDDI)
  - Bristol-Myers Squibb
  - European Organization for Research and Treatment of Cancer (EORTC)
  - Université Catholique de Louvain (UCL)
  - Université Claude Bernard Lyon 1 (Lyon, France)
BENEFIT Project: Goals

• Methods
  – Extensions of GPC (missing data, longitudinal, cross-over, ...)
  – “Optimal” GPC (censoring, risk/benefit, ...)
  – Comparisons with traditional methods
  – Use for trial design

• Applications (oncology, ophthalmology, ...)

• Software
  – Open
  – Proprietary (design, analysis, patient)
Net Benefit for Longitudinal Data

- Normally-distributed outcome $Y$
- Two measurements: “earlier” $Y_1$ and “later” $Y_2$
- $Y_2$ “primary”
- For uncorrelated $Y_1$ and $Y_2$

\[
\Delta = \theta_2 + \theta_1 \cdot \left\{ \Phi\left( \frac{\mu_{20} + \tau_2 - \mu_{21}}{\sigma_2 \sqrt{2}} \right) - \Phi\left( \frac{\mu_{20} - \tau_2 - \mu_{21}}{\sigma_2 \sqrt{2}} \right) \right\}
\]

\[
= \{P(T \text{ better for } Y_2) - P(C \text{ better for } Y_2)\} + \{P(T \text{ better for } Y_1) - P(C \text{ better for } Y_1)\} \cdot P(\text{tie on } Y_1)
\]

\[
= (\text{Net benefit for } Y_2) + (\text{Net benefit for } Y_1) \cdot P(\text{tie on } Y_2)
\]
Net Benefit Under MCAR Dropout

• Data for \( Y_1 \) complete
• Data for \( Y_2 \) missing completely at random in each treatment group
  – \( \omega_0 = P(Y_2 \text{ observed for control}), \omega_1 = P(Y_2 \text{ observed for treatment}) \)

• Then, for uncorrelated \( Y_1 \) and \( Y_2 \),
  \[
  \Delta_{\text{MCAR}} = \omega_0 \omega_1 \Delta + \theta_1 (1 - \omega_0 \omega_1)
  \]

• Hence, estimation ignoring missing data (even for MCAR) is biased!

• A corrected estimator obtained from
  \[
  \Delta = \frac{\Delta_{\text{MCAR}} - \theta_1 (1 - \omega_0 \omega_1)}{\omega_0 \omega_1}
  \]
Net Benefit Under Dropout/Correlation

- Formulae quickly complicate for correlated $Y_1$ and $Y_2$
  - MCAR:

$$\Delta = \theta_2 \Phi(\beta_0) \Phi(\beta_0 + \gamma) + \theta_1 (1 - \Phi(\beta_0) \Phi(\beta_0 + \gamma)) + \\
\phi(\frac{Y_{11} - \mu_{11}}{\sigma_1}) \left( BvN(h_{11}, \frac{y_{11} - \tau_1 - \mu_{10}}{\sigma_1}; \rho_3) - \\
BvN(h_{12}, \frac{y_{11} - \tau_1 - \mu_{10}}{\sigma_1}; \rho_3) \right) dy_{11} - \\
+ \frac{1}{\sigma_1} \int_{-\infty}^{\infty} \phi\left( \frac{y_{10} - \mu_{10}}{\sigma_1} \right) \left( BvN(h_{13}, \frac{y_{10} - \tau_1 - \mu_{11}}{\sigma_1}; -\rho_3) - \\
BvN(h_{14}, \frac{y_{10} - \tau_1 - \mu_{11}}{\sigma_1}; \rho_3) \right) dy_{10}$$

- Even more for MAR...
- Nevertheless, IPW estimators can be constructed
• GPCs are attractive
  – In terms of patient centricity:
    • “Net benefit”, a patient-relevant measure
    • Accommodate prioritized outcomes
  – In statistical terms:
    • Equivalent to standard non-parametric tests in simple cases
    • May have better power than, e.g., the logrank test
    • Allow for testing of clinically relevant differences
• GPCs require more fundamental research
  – In terms of theoretical properties:
    • sufficiency, completeness?
    • robustness to missing data
    • handling multiple relevance-thresholds
    • generalizability beyond the available sample?
    • …
  – In terms of applicability:
    • Disease domains where additional insight can be obtained?