Generalized pairwise comparisons for right-censored time to event outcomes

Brice Ozenne[1,2], Julien Peron[3], Marc Buyse[4]

[1] Neurobiology Reasearch Unit, Rigshospitalet, Copenhagen, Denmark.
[2] University of Copenhagen, Section of Biostatistics, Copenhagen, Denmark.
[3] Hospices Civils de Lyon, Service de Biostatistiques, Lyon, France.
[4] International Drug Development Institute (IDDI), San Francisco, USA.

SAfJR - April 25, 2019

Clinical trials in oncology

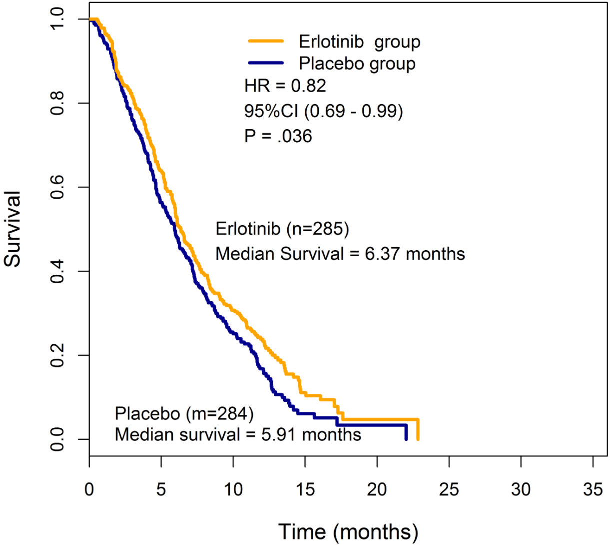
Efficacy/safety can reveal opposite effects of the treatment:

- longer survival
- serious but non-lethal adverse effects

However, efficacy and safety outcomes are usually analyzed and **reported independently**

- *efficacy*: using log-rank test
- *safety*: using Fisher's exact test

Example - Moore et al. (2007)



Example - Moore et al. (2007)

Worst grade related adverse event	Erlotinib group (n=282)	Placebo group (n=280)
Grade 1	48 (17%)	69 (24.6%)
Grade 2	118 (41.8%)	89 (31.5%)
Grade 3	72 (25.5%)	47 (16.8%)
Grade 4	11 (3.9%)	6 (2.1%)
Grade 5	4 (1.4%)	3 (1.1%)

Clinical trials - handling multiple endpoints

I don't think there is a good objective approach.

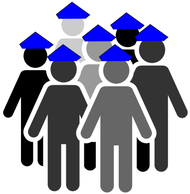
What about a good subjective approach?

Patient preference

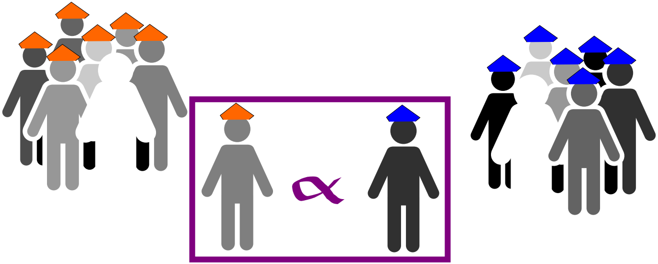
1. increase survival by at least 2 months
2. otherwise, least serious adverse events

Generalized pairwise comparisons (GPC)

Generalized pairwise comparisons (GPC)



Generalized pairwise comparisons (GPC)



GPC - 1 endpoint

Consider N patients divided into two groups:

- experimental group: m patients with response $(x_i)_{i \in \{1,\dots,m\}}$
- control group: n patients with response $(y_i)_{i \in \{1,\dots,n\}}$

Denote by $\tau \in \mathbb{R}^{+*}$ the smallest difference in response that is clinically relevant.

Our parameter of interest is the **net benefit**:

$$\Delta = \mathbb{P}[X \geq Y + \tau] - \mathbb{P}[Y \geq X + \tau]$$

Point estimation in GPC

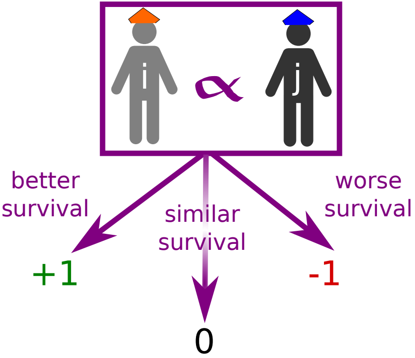
Defining $s_{ij} = \mathbb{1}_{x_i \geq y_j + \tau} - \mathbb{1}_{y_j \geq x_i + \tau}$ the score of the pair i, j:

$$\hat{\Delta} = \frac{1}{nm} \sum_{i=1}^{n} \sum_{j=1}^{m} s_{ij}$$

Point estimation in GPC

Defining $s_{ij} = \mathbb{1}_{x_i \geq y_j + \tau} - \mathbb{1}_{y_j \geq x_i + \tau}$ the score of the pair i, j:

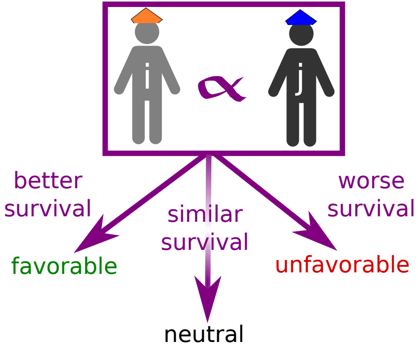
$$\hat{\Delta} = \frac{1}{nm} \sum_{i=1}^{n} \sum_{j=1}^{m} s_{ij}$$



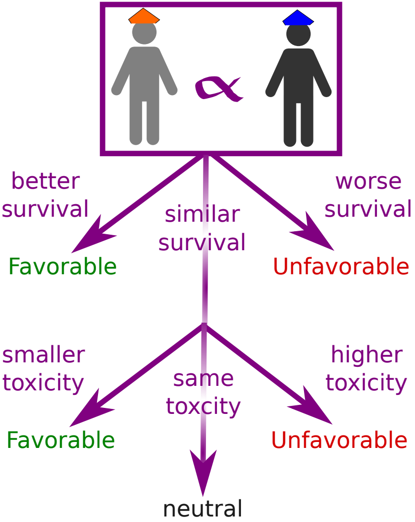
Point estimation in GPC

Defining $s_{ij} = \mathbb{1}_{x_i \geq y_j + \tau} - \mathbb{1}_{y_j \geq x_i + \tau}$ the score of the pair i, j:

$$\hat{\Delta} = \frac{1}{nm} \sum_{i=1}^{n} \sum_{j=1}^{m} s_{ij}$$



Generalization to multiple endpoints



i.e. $s_{ij} = \phi(\mathbf{x}_i, \mathbf{y}_j, \boldsymbol{\tau}) - \phi(\mathbf{x}_i, \mathbf{y}_j, \boldsymbol{\tau})$ where ϕ is a scoring rule.

GPC in presence of censoring

GPC in presence of censoring

We only observe:

- right-censored event times $(\tilde{x}_i)_{i\in\{1,\dots,m\}}$ and $(\tilde{y}_j)_{j\in\{1,\dots,n\}}$
- event type indicators $(\delta_{x,i})_{i\in\{1,\dots,m\}}$ and $(\delta_{y,j})_{j\in\{1,\dots,n\}}$.

How can we compute $s_{ij} = \mathbb{1}_{x_i \geq y_j+\tau} - \mathbb{1}_{y_j \geq x_i+\tau}$?

Gehan scoring rule

Set s_{ij} to 0 when the pair cannot be decidedly classified.

Gehan scoring rule

Set s_{ij} to 0 when the pair cannot be decidedly classified.

If $\delta_{x,i} = 0$ and $\delta_{y,j} = 1$:

Example 1: $\tilde{x}_i \geq \tilde{y}_j + \tau$

- $s_{ij} = 1$

Example 2: $\tilde{x}_i < \tilde{y}_j + \tau$

- $s_{ij} = 0$: uninformative pair

Gehan scoring rule

Set s_{ij} to 0 when the pair cannot be decidedly classified.

If $\delta_{x,i} = 0$ and $\delta_{y,j} = 1$:

Example 1: $\tilde{x}_i \geq \tilde{y}_j + \tau$

- $s_{ij} = 1$

Example 2: $\tilde{x}_i < \tilde{y}_j + \tau$

- $s_{ij} = 0$: uninformative pair

⚠ uninformative pairs bias the estimation of Δ toward the null (Buyse, 2008)

Peron scoring rule - Péron et al. (2018)

Compute

- $p_{ij}^{f} = \mathbb{P}\left[X \geq Y + \tau \middle| X \geq \tilde{x}_i, X \geq \tilde{y}_j, \delta_{x,i}, \delta_{y,j}\right]$
- $p_{ij}^{uf} = \mathbb{P}\left[Y \geq X + \tau \middle| X \geq \tilde{x}_i, X \geq \tilde{y}_j, \delta_{x,i}, \delta_{y,j}\right]$
- $s_{ij} = p_{ij}^{f} - p_{ij}^{uf}$

Peron scoring rule - Péron et al. (2018)

Compute

- $p_{ij}^{f} = \mathbb{P}\left[X \geq Y + \tau \middle| X \geq \tilde{x}_i, X \geq \tilde{y}_j, \delta_{x,i}, \delta_{y,j}\right]$
- $p_{ij}^{uf} = \mathbb{P}\left[Y \geq X + \tau \middle| X \geq \tilde{x}_i, X \geq \tilde{y}_j, \delta_{x,i}, \delta_{y,j}\right]$
- $s_{ij} = p_{ij}^{f} - p_{ij}^{uf}$

If $\delta_{x,i} = 0$ and $\delta_{y,j} = 1$

- $p_{ij}^{f} = \min\left(\frac{S_X(\tilde{y}_j + \tau)}{S_X(\tilde{x}_i)}, 1\right)$

where S_x is the survival in the experimental group.

Peron scoring rule - Péron et al. (2018)

Compute

- $p_{ij}^{f} = \mathbb{P}\left[X \geq Y + \tau \middle| X \geq \tilde{x}_i, X \geq \tilde{y}_j, \delta_{x,i}, \delta_{y,j}\right]$
- $p_{ij}^{uf} = \mathbb{P}\left[Y \geq X + \tau \middle| X \geq \tilde{x}_i, X \geq \tilde{y}_j, \delta_{x,i}, \delta_{y,j}\right]$
- $s_{ij} = p_{ij}^{f} - p_{ij}^{uf}$

If $\delta_{x,i} = 0$ and $\delta_{y,j} = 1$

- $p_{ij}^{f} = \min\left(\frac{S_X(\tilde{y}_j + \tau)}{S_X(\tilde{x}_i)}, 1\right)$

where S_x is the survival in the experimental group.

Example 1: $\tilde{x}_i \geq \tilde{y}_j + \tau$

- $s_{ij} = 1$

Example 2: $\mathbb{P}\left[X \geq \tilde{y}_j + \tau | X \geq \tilde{x}_i\right] = 1$

- $s_{ij} = 1$

Implementation

Package BuyseTest (available on CRAN):

- S_X and S_Y estimated using Kaplan-Meier

Example of code:

```
ff <- group ~ tte(survival, censoring = event, threshold = 2) +
     cont(toxicity)

e.Peron <- BuyseTest(formula = ff,
                     data = dt.fol,
                     scoring.rule = "Peron")
```

Implementation

Example of output:

```
summary(e.Gehan)
```

```
        Generalized pairwise comparisons with 2 prioritized endpoints

[...]
> treatment groups: gemcitabine (control) vs. folfirinox (treatment)
[...]
endpoint threshold  total favorable unfavorable neutral uninf  delta  Delta
survival         2 100.00     44.74       20.97   12.50 21.79 0.2377 0.2377
toxicity     1e-12  34.29     14.50        8.53   11.27  0.00 0.0597 0.2975
```

```
summary(e.Peron)
```

```
        Generalized pairwise comparisons with 2 prioritized endpoints

[...]
endpoint threshold  total favorable unfavorable neutral uninf  delta  Delta
survival         2 100.00     56.88       26.49   16.61  0.02 0.3039 0.3039
toxicity     1e-12  16.63      6.53        4.41    5.69  0.00 0.0212 0.3251
```

Limitations & perspectives

The Peron scoring rule requires a consistent estimator for the survival

- e.g. at $\tilde{y}_j + \tau$: may not be available
- remaining uninformative pairs

Ideas:

- lower and upper bound for p^f and p^{uf} (implemented)
- parametric model for S_X and S_Y
- inverse probability of censoring weights

Reference I

Buyse, M. (2008). Reformulating the hazard ratio to enhance communication with clinical investigators. *Clinical Trials*, 5(6):641.

Moore, M. J., Goldstein, D., Hamm, J., Figer, A., Hecht, J. R., Gallinger, S., Au, H. J., Murawa, P., Walde, D., Wolff, R. A., et al. (2007). Erlotinib plus gemcitabine compared with gemcitabine alone in patients with advanced pancreatic cancer: a phase iii trial of the national cancer institute of canada clinical trials group. *Journal of clinical oncology*, 25(15):1960–1966.

Péron, J., Buyse, M., Ozenne, B., Roche, L., and Roy, P. (2018). An extension of generalized pairwise comparisons for prioritized outcomes in the presence of censoring. *Statistical methods in medical research*, 27(4):1230–1239.