Generalized Pairwise Comparison on Immuno-Oncology clinical trial data: a case study

Dr Julien PERON, Pr Delphine MAUCORT-BOULCH, Pr Pascal ROY,
Pr Marc BUYSE
November 2017
Department of Biostatistics HCL – LBBE UCBL
Department of Medical oncology HCL – LBBE UCBL

Lyon 1

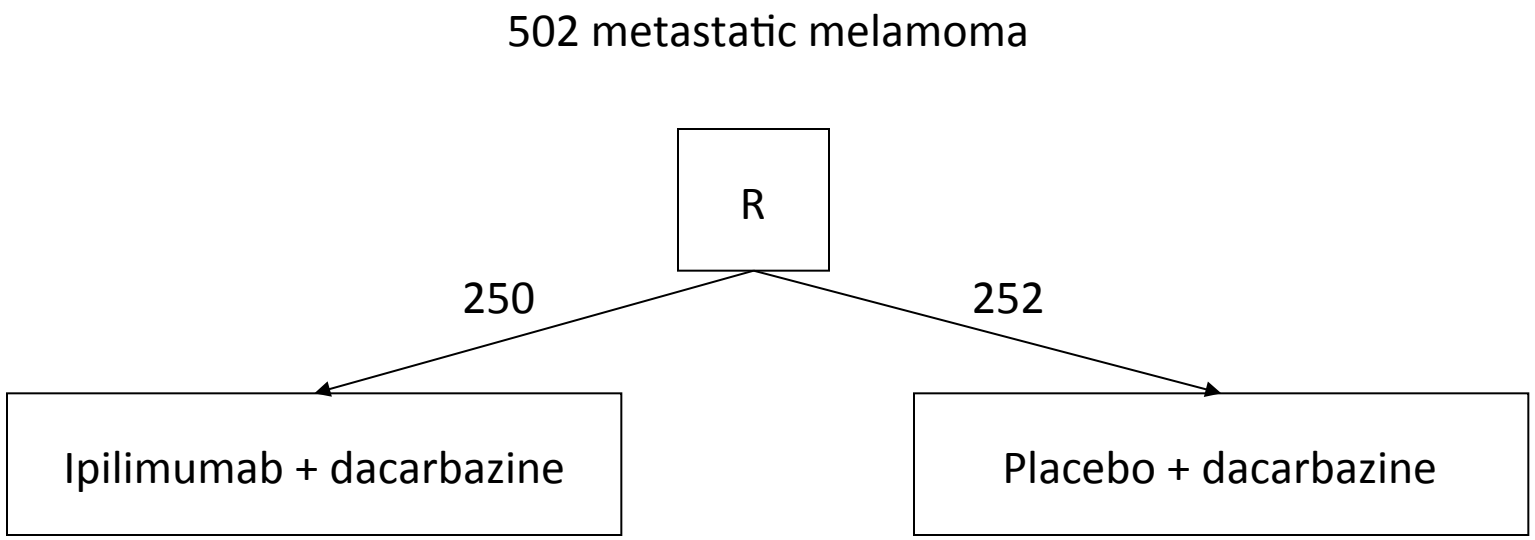
LBBE
BIOMETRIE ET BIOLOGIE EVOLUTIVE

iC HCL
INSTITUT DE CANCÉROLOGIE
DES HOSPICES CIVILS DE LYON

Hospices Civils de Lyon

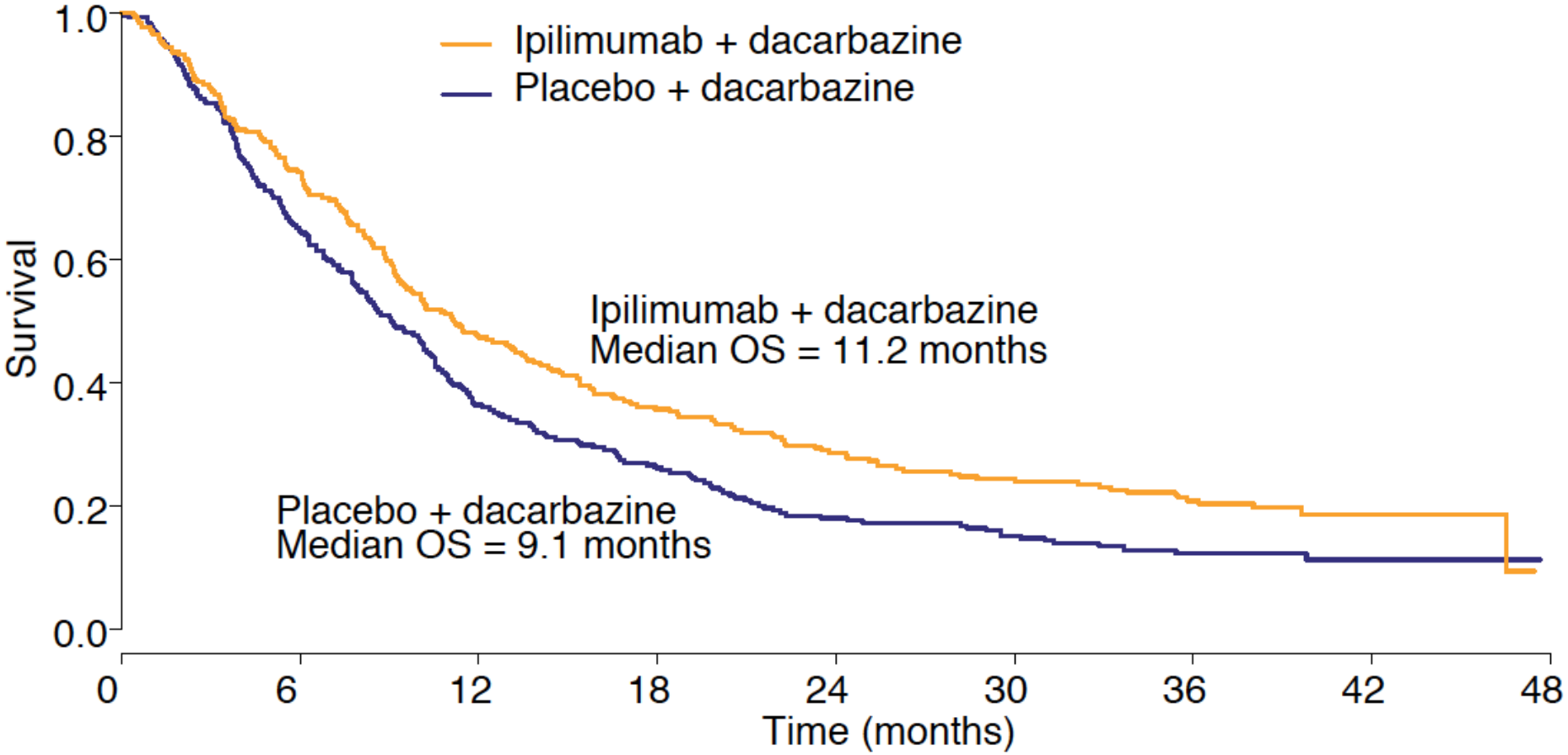
Case study

The CA184-024 trial



Robert et al. NEJM 2011

OS results in the CA184-024 trial

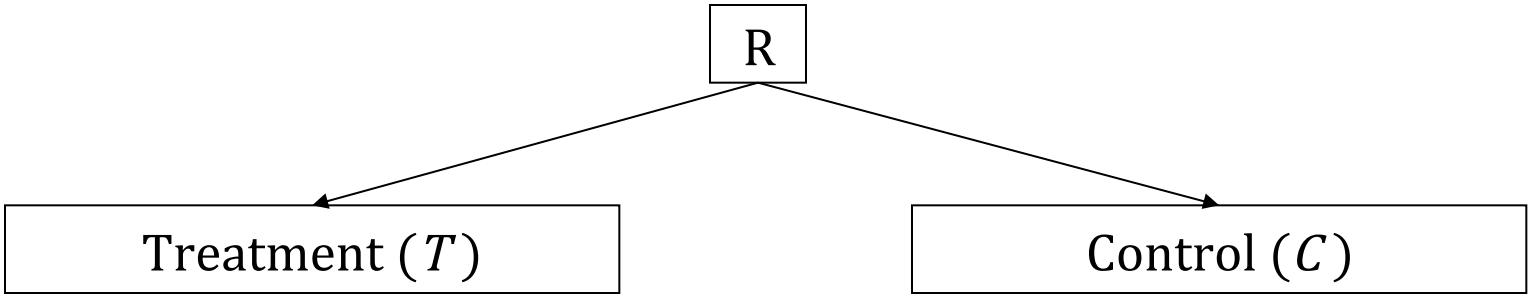


Pcb	252	160	89	64	44	37	26	7	0
Ipi	250	181	114	85	68	57	41	10	0

Outline

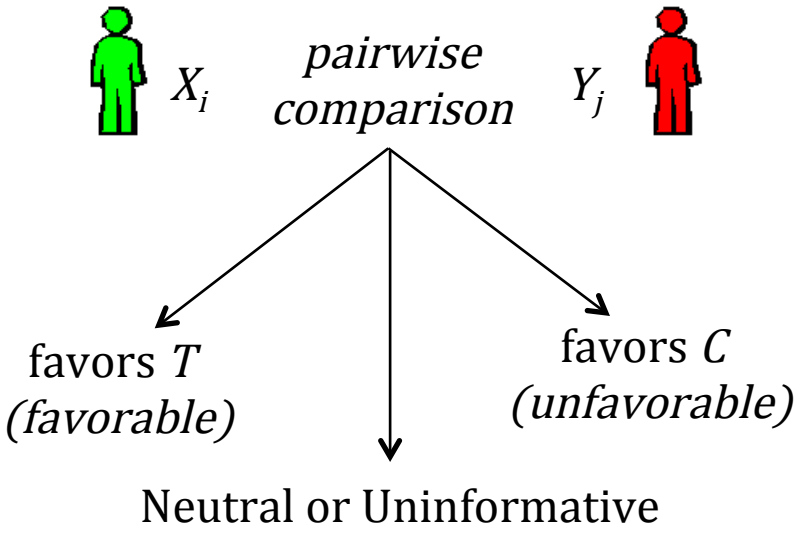
- The procedure of generalized pairwise comparisons
- A patient-oriented measure of treatment benefit
- Application on immuno-oncology trials
  - Simulation study
  - Illustration on an ipilimumab trial

Methods – Pairwise comparisons



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 $(j = 1. \ldots . m)$



Buyse M. stat in med 2010

Methods – Definition of thresholds

Continuous outcome

Pair	Rating
$x_i - y_j > \tau$	Favorable
$x_i - y_j < (-\tau)$	Unfavorable
$\mid x_i - y_j \mid \leq \tau$	Neutral
x_i or y_j missing	Uninformative

Buyse. stat in med 2010

Methods – Standard procedure for pairwise scoring

$$U_{ij} = \begin{cases} +1 \text{ when the pair } (X_i, Y_j) \text{ is favorable} \\ -1 \text{ when the pair } (X_i, Y_j) \text{ is unfavorable} \\ 0 \text{ otherwise} \end{cases}$$

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

Δ in named «net benefit»

An empirical distribution of Δ can be obtained by permutation

Buyse. stat in med 2010

Some notations

- x_i^0 and y_j^0 : time-to-event
- x_i and y_j : time-to-observation
- Event indicator :

$$\delta_i = \left.\begin{cases} 1 \; if \; x_i = x_i^0 \\ 0 \; if \; x_i < x_i^0 \end{cases}\right\} \text{in group T}$$

$$\varepsilon_j = \left.\begin{cases} 1 \; if \; y_j = y_j^0 \\ 0 \; if \; y_j < y_j^0 \end{cases}\right\} \text{in group C}$$

- Survival function:

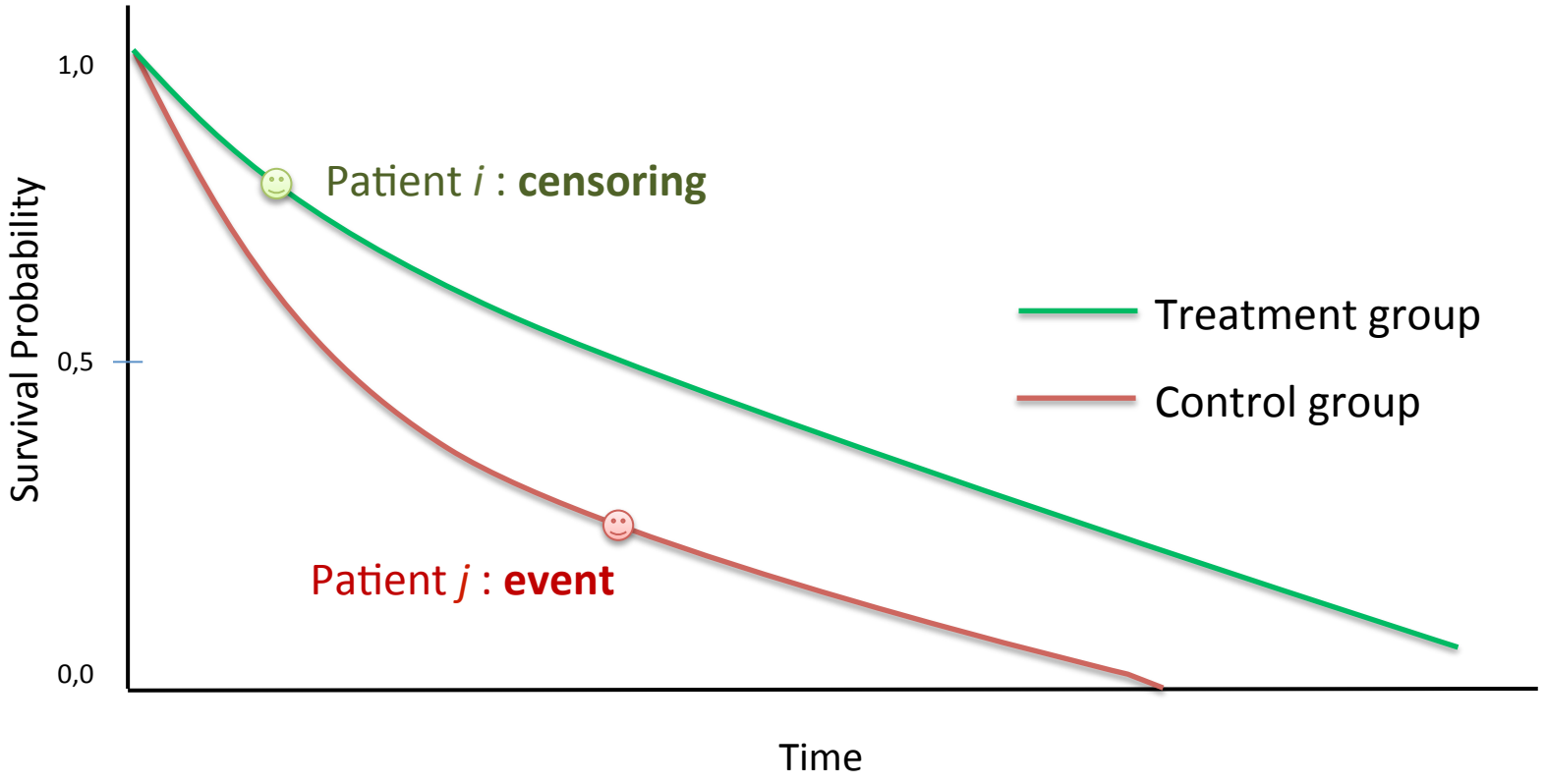
$$S_T(t) = \mathbb{P}\left[x_i^0 \geq t\right] \text{ and } S_C(t) = \mathbb{P}\left[y_j^0 \geq t\right]$$

The standard procedure to include time-to-event' outcome

$(\delta_i, \varepsilon_j)$	$x_i - y_j \geq \tau$	$x_i - y_j \leq -\tau$	$\lvert x_i - y_j \rvert < \tau$
$(1,1)$	Favorable	Unfavorable	Neutral
$(0,1)$	Favorable	Uninformative	Uninformative
$(1,0)$	Uninformative	Unfavorable	Uninformative
$(0,0)$	Uninformative	Uninformative	Uninformative

Buyse M. *Stat in med*, 2010

The standard procedure to include time-to-event outcome

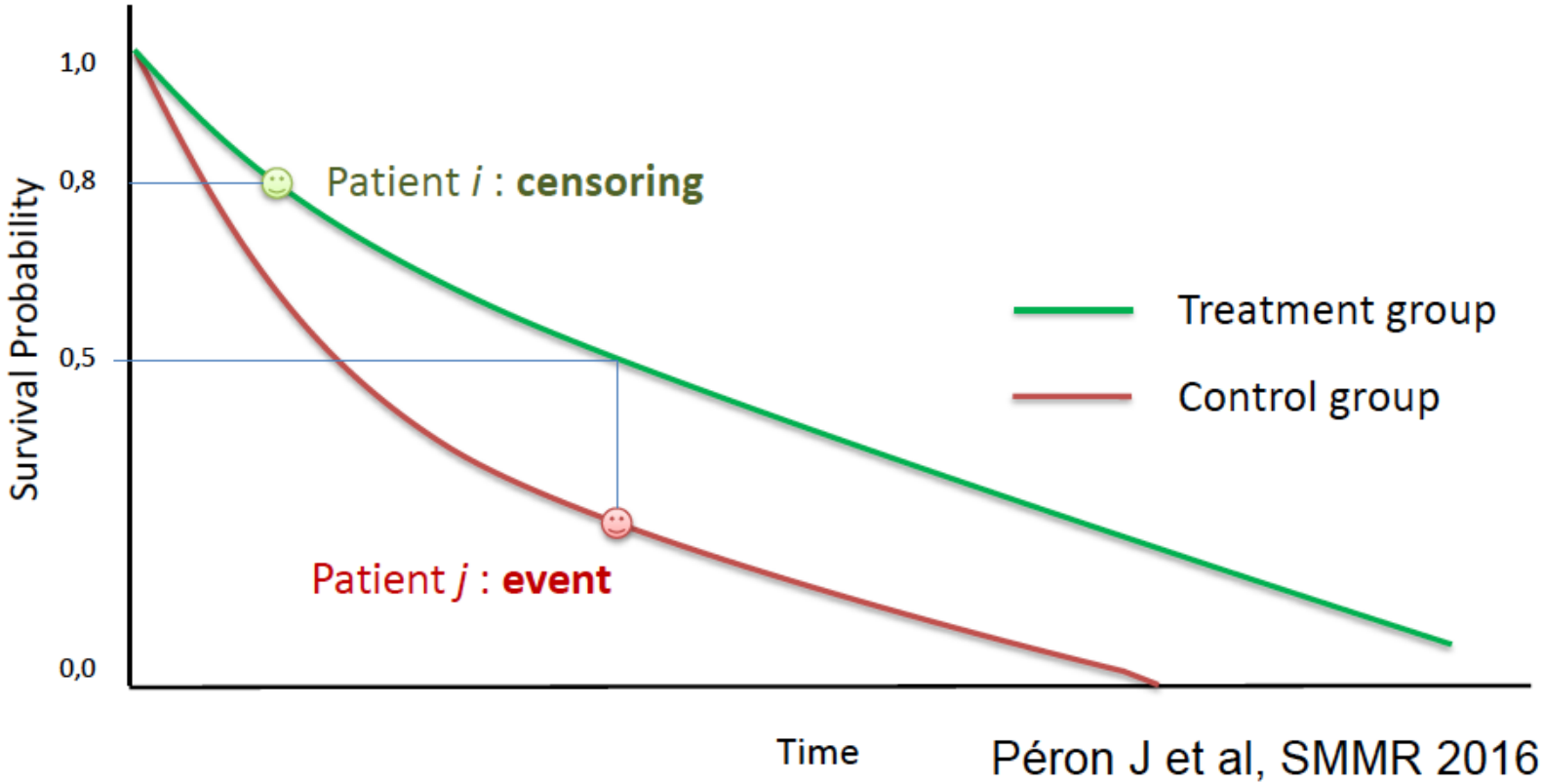


Gehan. *Biometrika,* 1965

The extended procedure taking into account 'non-informative' pairs

Based on the Kaplan-Meier estimate of the survival function

$$\mathbb{P}\left[\left(x_i^0 > y_j^0\right) \middle| \left(x_i^0 > x_i\right)\right] = \frac{\hat{S}_{Ttt}(y_j)}{\hat{S}_{Ttt}(x_i)} = \frac{0{,}5}{0{,}8}$$

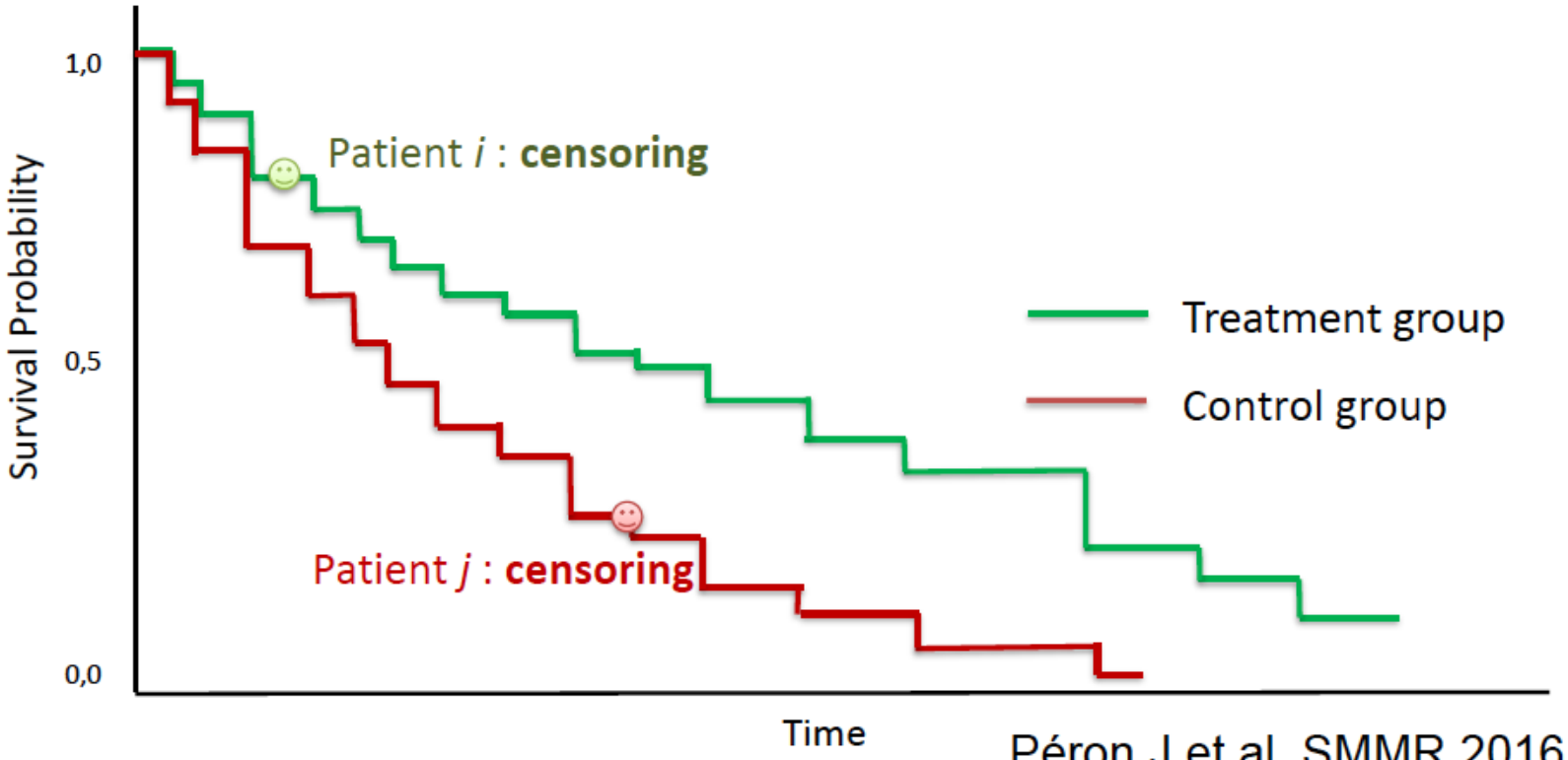


Péron J et al, SMMR 2016

The extended procedure taking into account 'non-informative' pairs

When the estimation of the survival function is discontinue :

$$\mathbb{p}\left[\left(x_i^0 > y_j^0\right) \middle| \left(x_i^0 > x_i\right), \left(y_j^0 > y_j\right)\right] = -\sum_{t>y_j}^{\infty} \frac{\hat{S}_{Ttt}(t)}{\hat{S}_{Ttt}(x_i)\hat{S}_{Ctrl}(y_j)} \cdot \left(\hat{S}_{Ctrl}(t^+) - \hat{S}_{Ctrl}(t^-)\right)$$



Péron J et al, SMMR 2016

The extended procedure taking into account 'non-informative' pairs

For pairs that can not be decidely classified because of censoring, we compute:

$$\mathbb{P}\left(x_i^0 > y_j^0 + \tau\right), \text{ et } \mathbb{P}\left(y_j^0 > x_i^0 + \tau\right), \text{ et } \mathbb{P}\left(|x_i^0 - y_j^0| < \tau\right)$$

The pairwise score is:

$$s_{ij} = \mathbb{P}\left(x_i^0 > y_j^0 + \tau\right) - \mathbb{P}\left(y_j^0 > x_i^0 + \tau\right)$$

The net benefit is then:

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

Achievements of the extended procedure

- Reduction of the Bias of in the presence of censored observations
  - Correction available (proportional hazards)

- Increased power of the permutation test compared to standard procedure
  - Proportional hazards and administrative censoring < 67% (B Efron, Stanford Univ, 1967)
  - Late treatment effect

Péron J et al, SMMR 2016

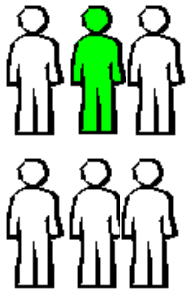
Outline

- The procedure of generalized pairwise comparisons
- A patient-oriented measure of treatment benefit
- Application on immuno-oncology trials
  - Simulation study
  - Illustration on an ipilimumab trial

The net benefit

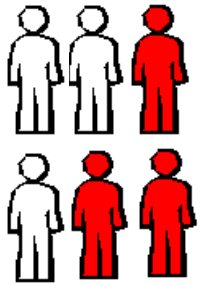
Probability for a random patient in the Treatment group to have a 'better outcome' than a random patient in the Control group …

Treatement group



$$\Delta = \mathbb{P}(\boldsymbol{X} > \boldsymbol{Y}) - \mathbb{P}(Y > X)$$

Control group

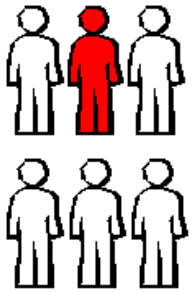


Buyse M. *Stat in med*, 2010

The net benefit

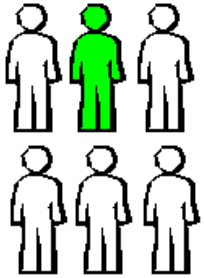
…minus the opposite probability.

Treatement group



$$\Delta = \mathbb{P}(X > Y) - \mathbb{P}(\boldsymbol{Y} > \boldsymbol{X})$$

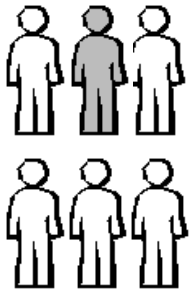
Control group



Buyse M. *Stat in med*, 2010

The net benefit

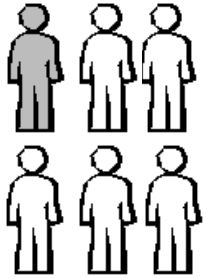
Treatement group



$\mathbb{P}(Y=X)$

$\Delta = \mathbb{P}(X>Y) - \mathbb{P}(Y>X)$

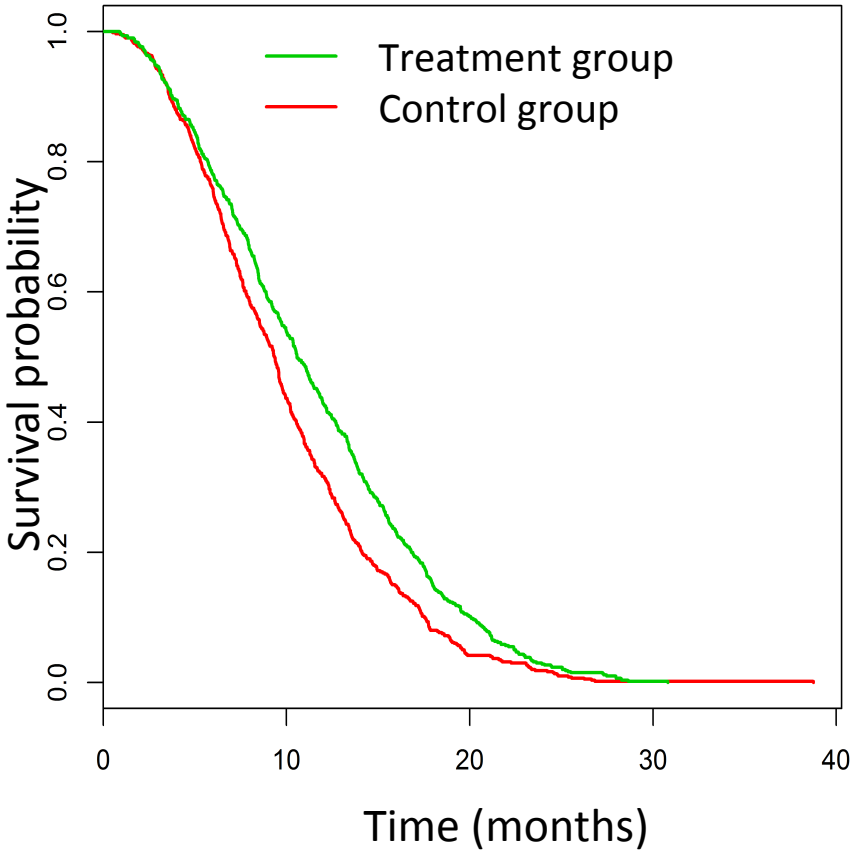
Control group

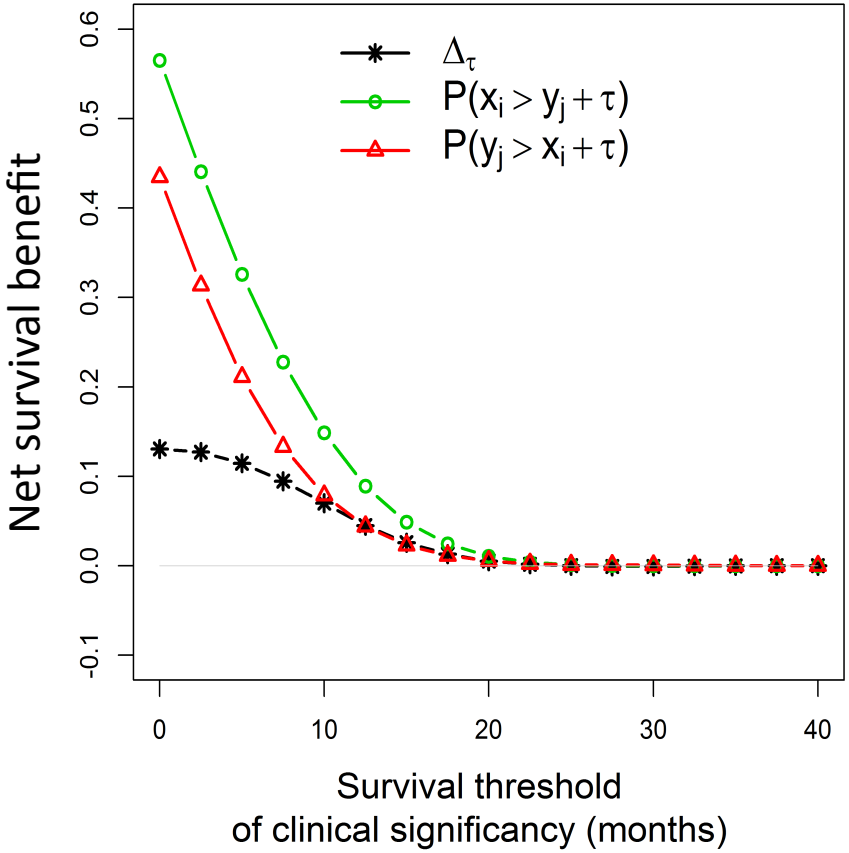


Buyse M. *Stat in med*, 2010

The net survival benefit

Proportional hazards

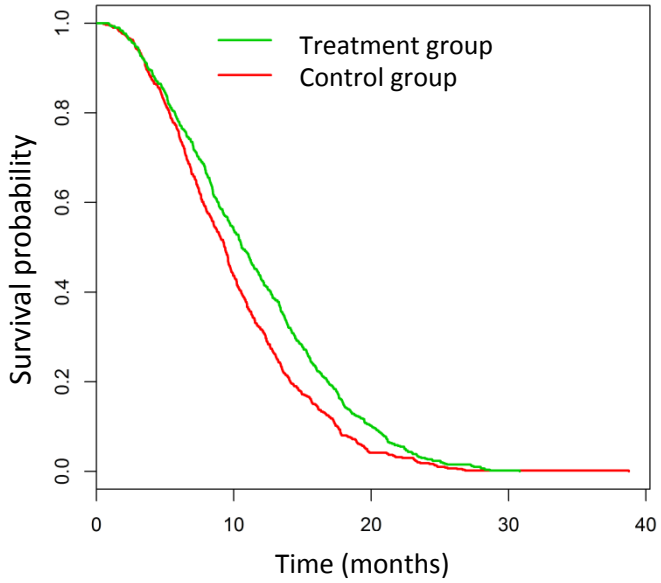


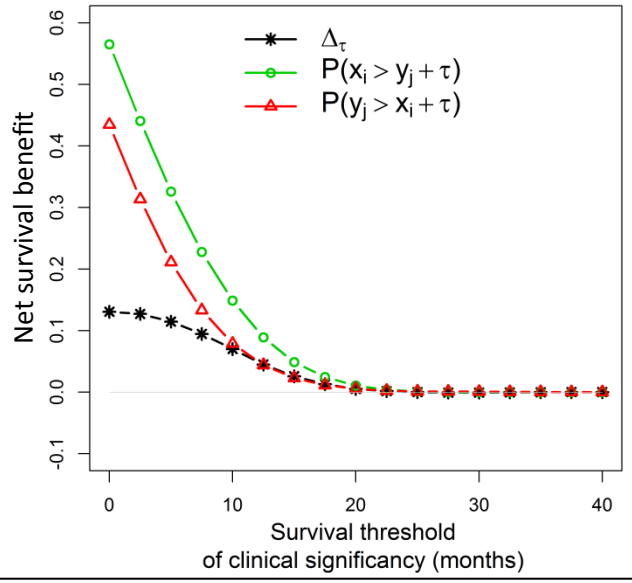


Péron et al, *JAMA oncology, 2016*

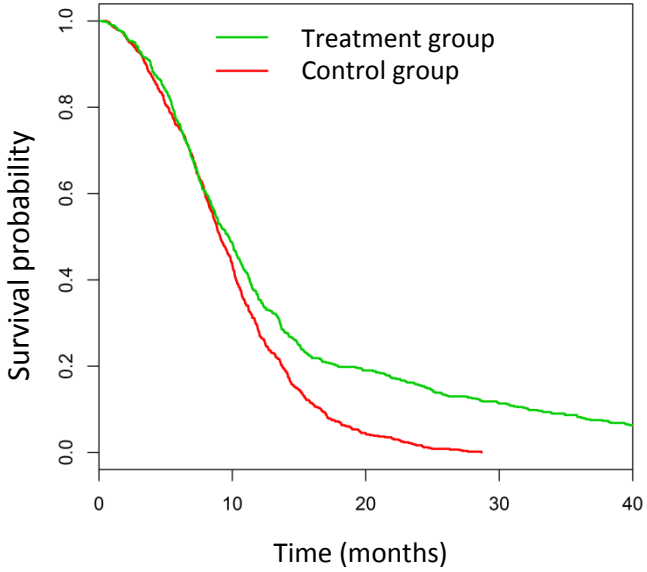
The net survival benefit

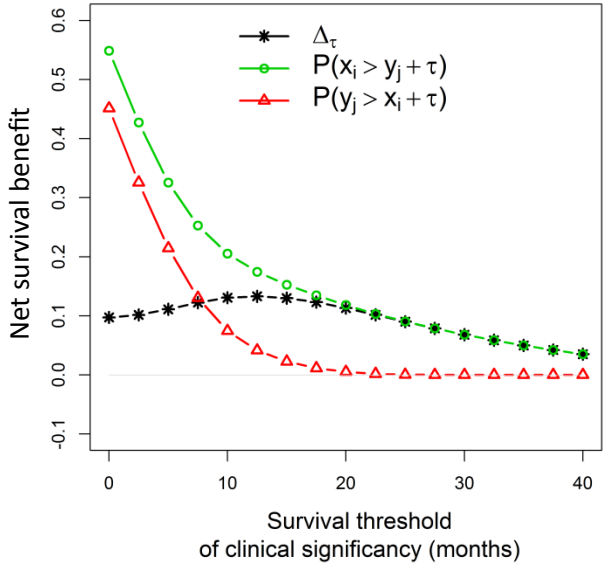
Proportional Hazards





Delayed treatment effect

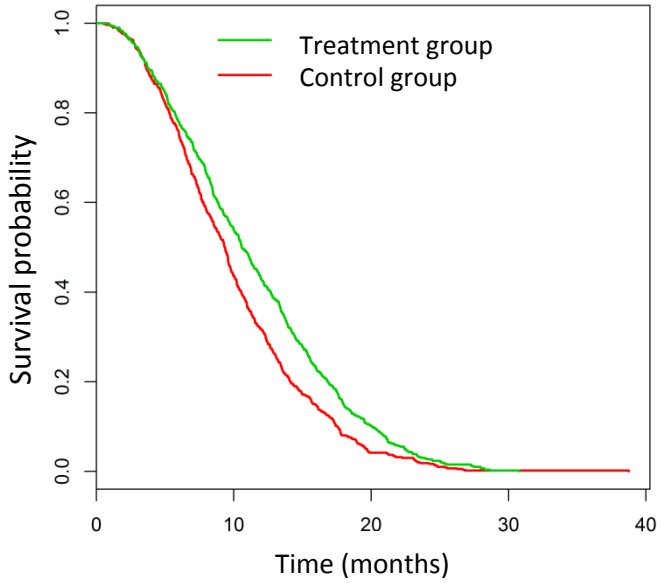


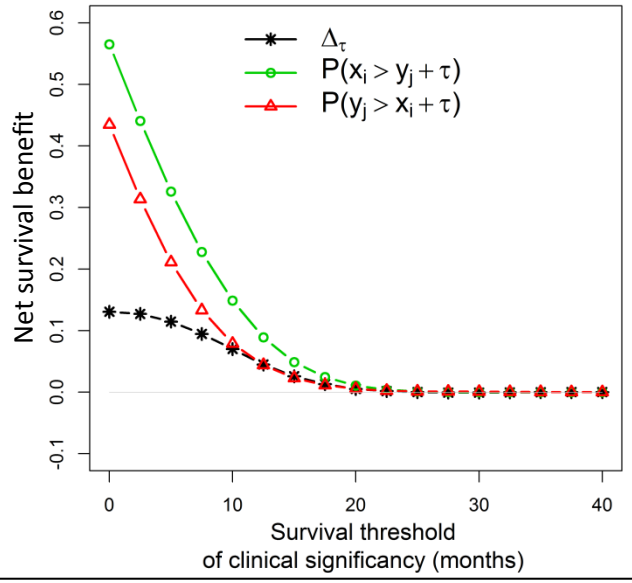


Péron et al, *JAMA oncology, 2016*

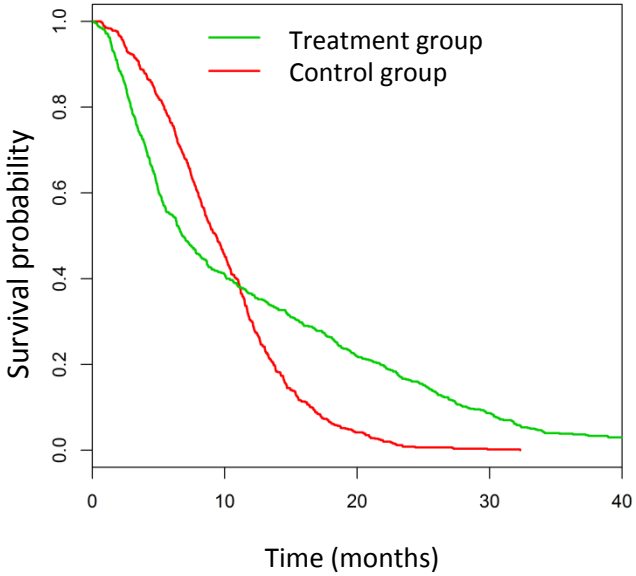
The net survival benefit

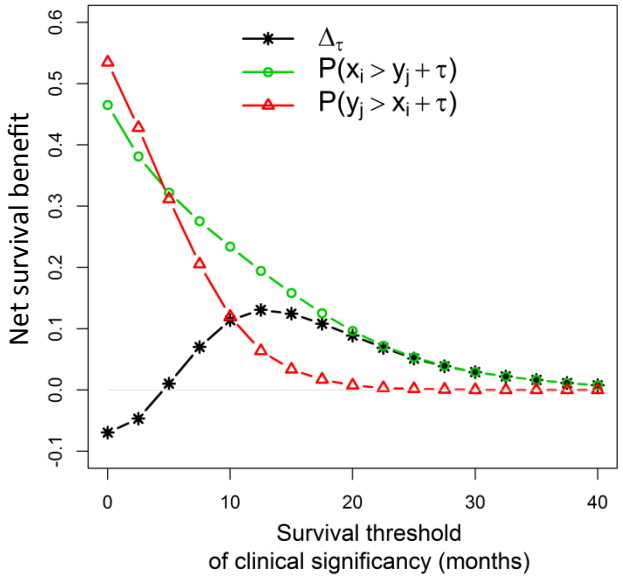
Proportional Hazards





Opposite hazards





Péron et al, *JAMA oncology, 2016*

Outline

- The procedure of generalized pairwise comparisons
- A patient-oriented measure of treatment benefit
- Application on immuno-oncology trials
  - Simulation study
  - Illustration on an ipilimumab trial

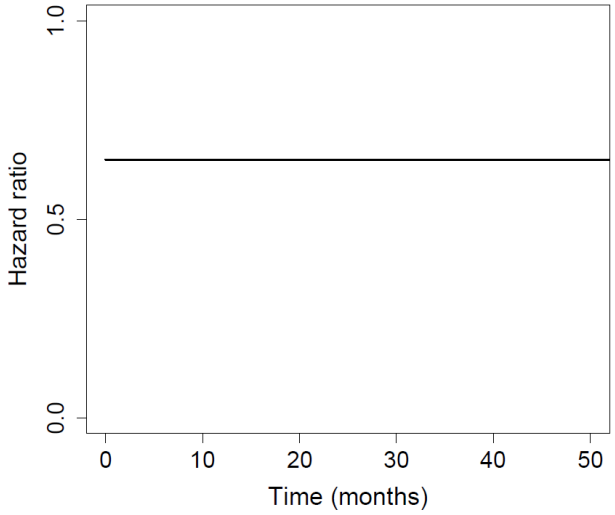
Simulation study - Design

- **Objective:** To assess the power of tests based on generalized pairwise comparisons for delayed treatment effect

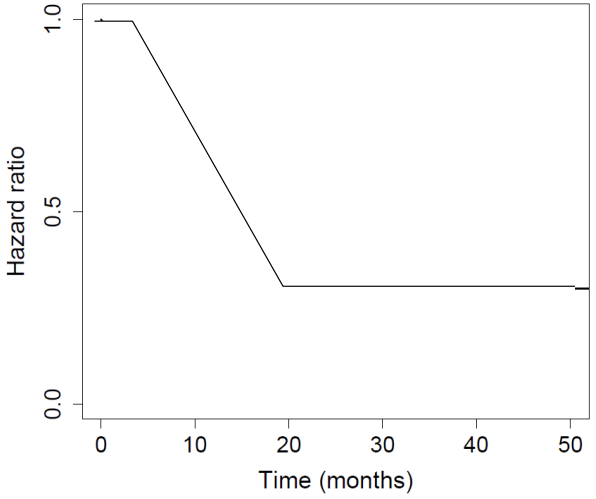
- Simulation of M = 1000 datasets with N = 200 patients
  - One time-to-event outcome

Simulation study - Design

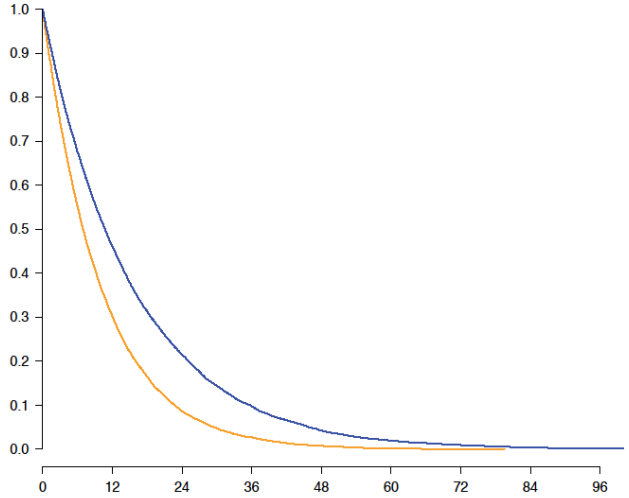
Scenario 1 : Proportional hazards



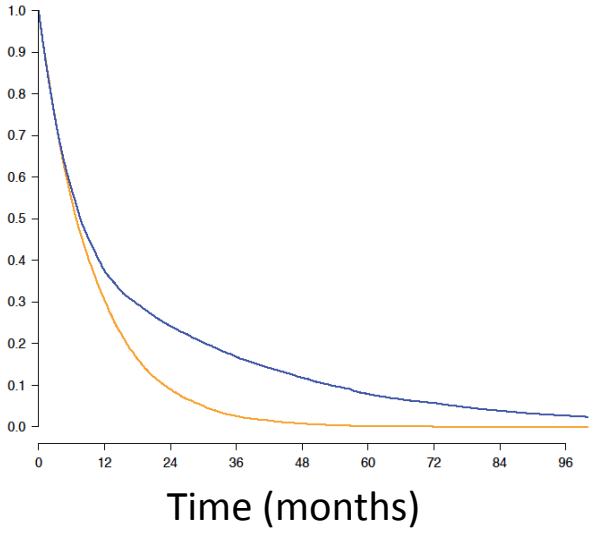
Scenario 2 : Late treatment effect



Survival



Survival

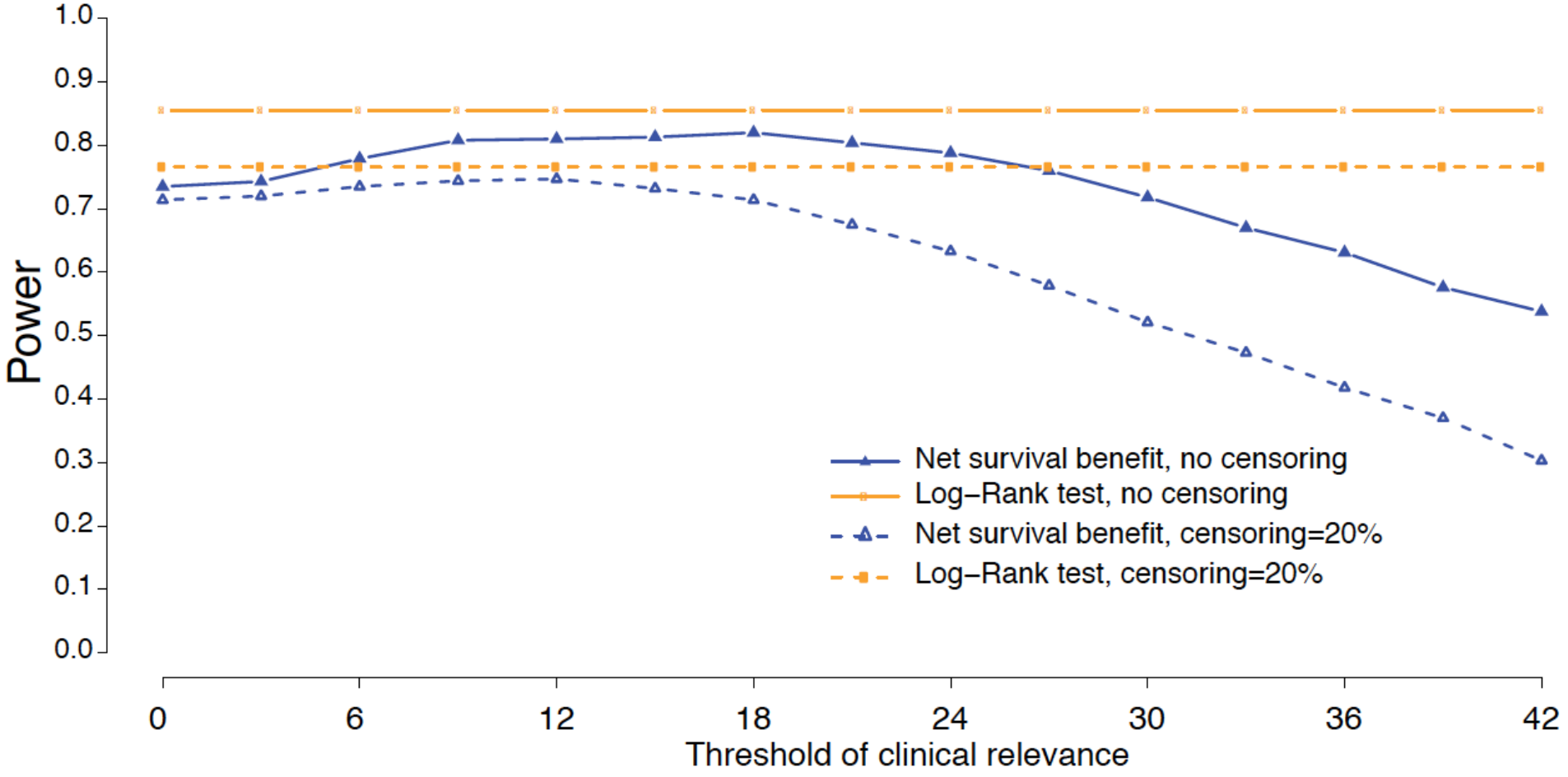


Simulation study - Design

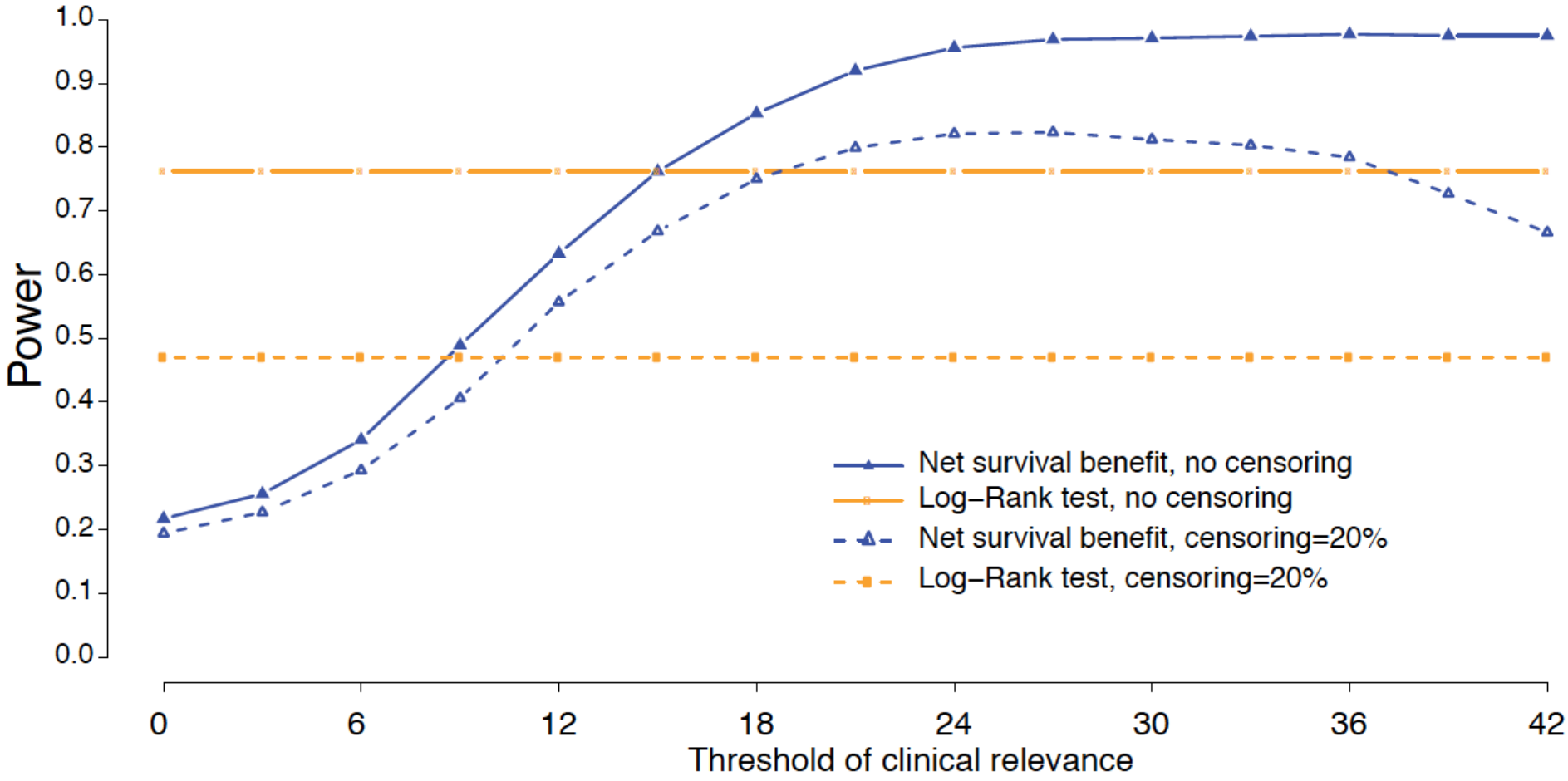
- Administrative censoring proportion
  - Uniform distribution
  - Between 0% and 20%

- For each simulated dataset
  - Estimation of the net survival benefit of at least τ months [0 to 42 months] (extended procedure)
  - Test of the null hypothesis (Permutation test, Log-Rank test)

Proportional Hazards - POWER



Delayed treatment effect - POWER



Conclusions of the simulation study

When a long-term survival benefit is expected

(anticancer immune therapy)

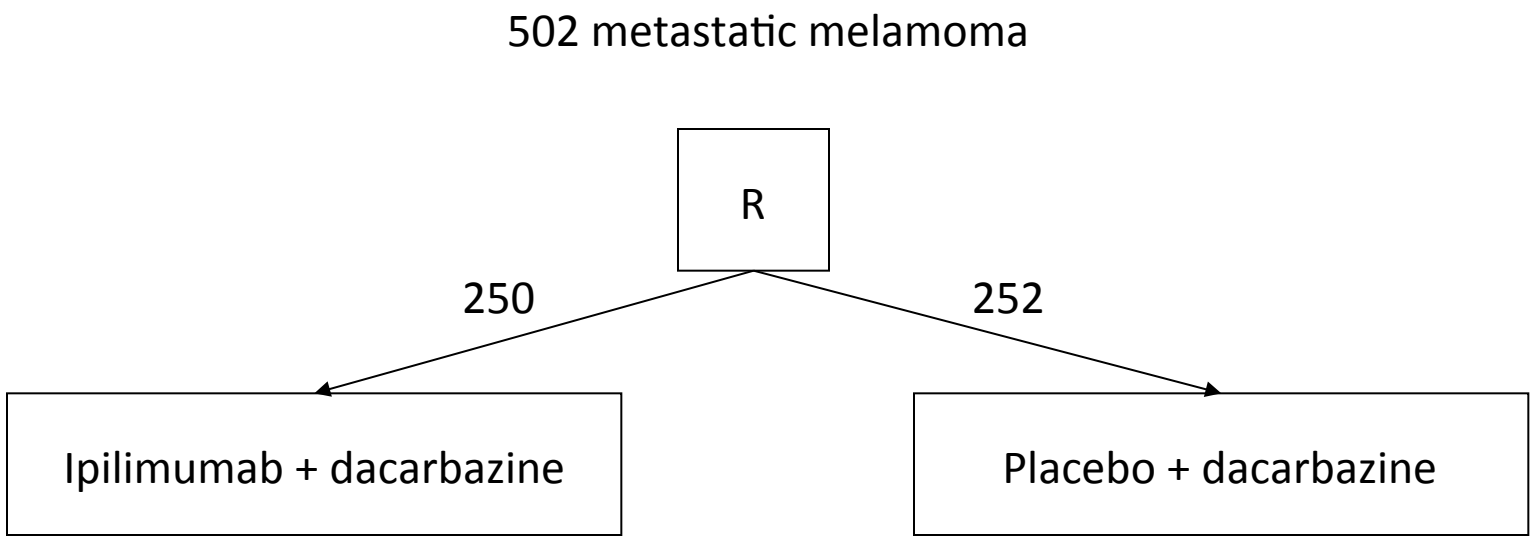
The net survival benefit is:

- Arguably more relevant than traditional methods ➔ focus on long term survival differences
- More powerful than traditional method

Outline

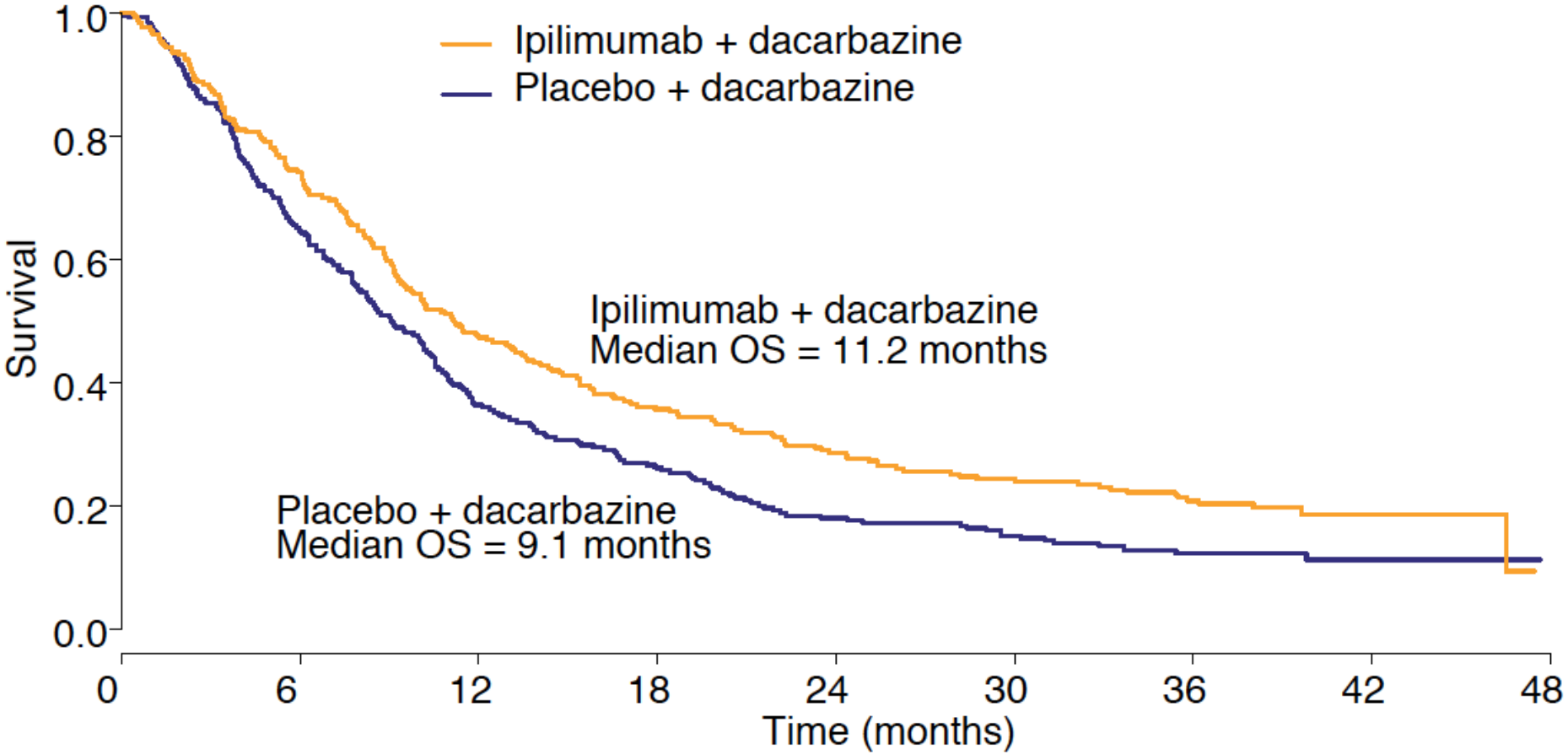
- The procedure of generalized pairwise comparisons
- A patient-oriented measure of treatment benefit
- Application on immuno-oncology trials
  - Simulation study
  - Illustration on an ipilimumab trial

The net survival benefit in the CA184-024 trial



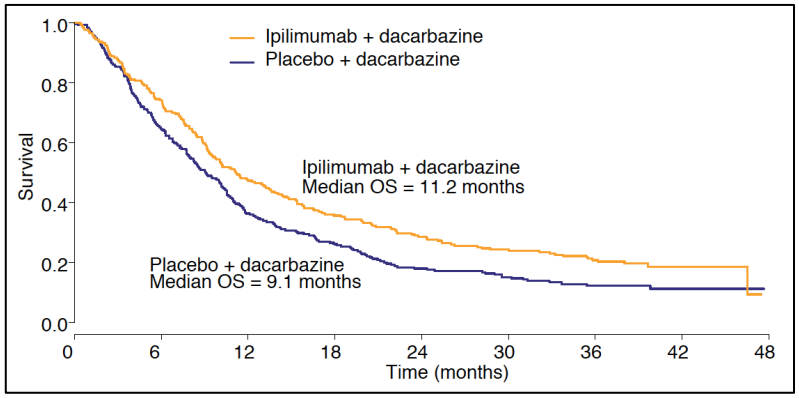
Robert et al. NEJM 2011

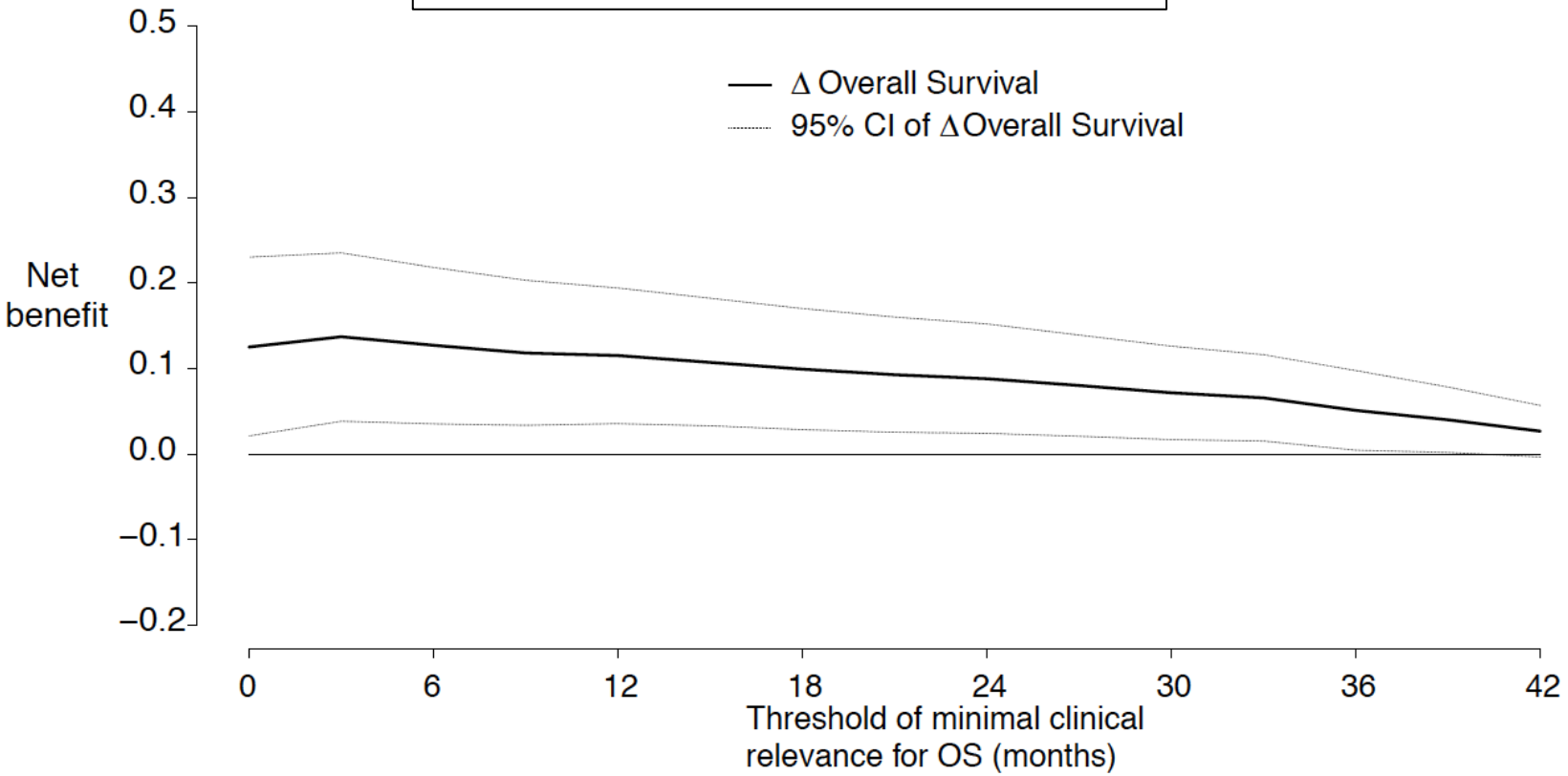
OS results in the CA184-024 trial



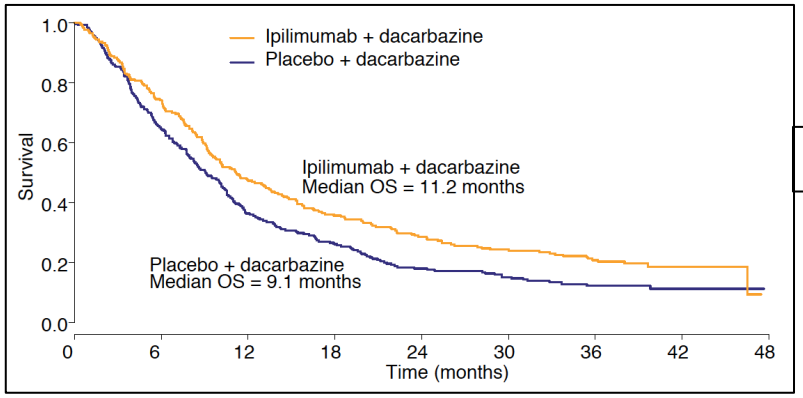
Pcb	252	160	89	64	44	37	26	7	0
Ipi	250	181	114	85	68	57	41	10	0

OS results in the CA184-024 trial

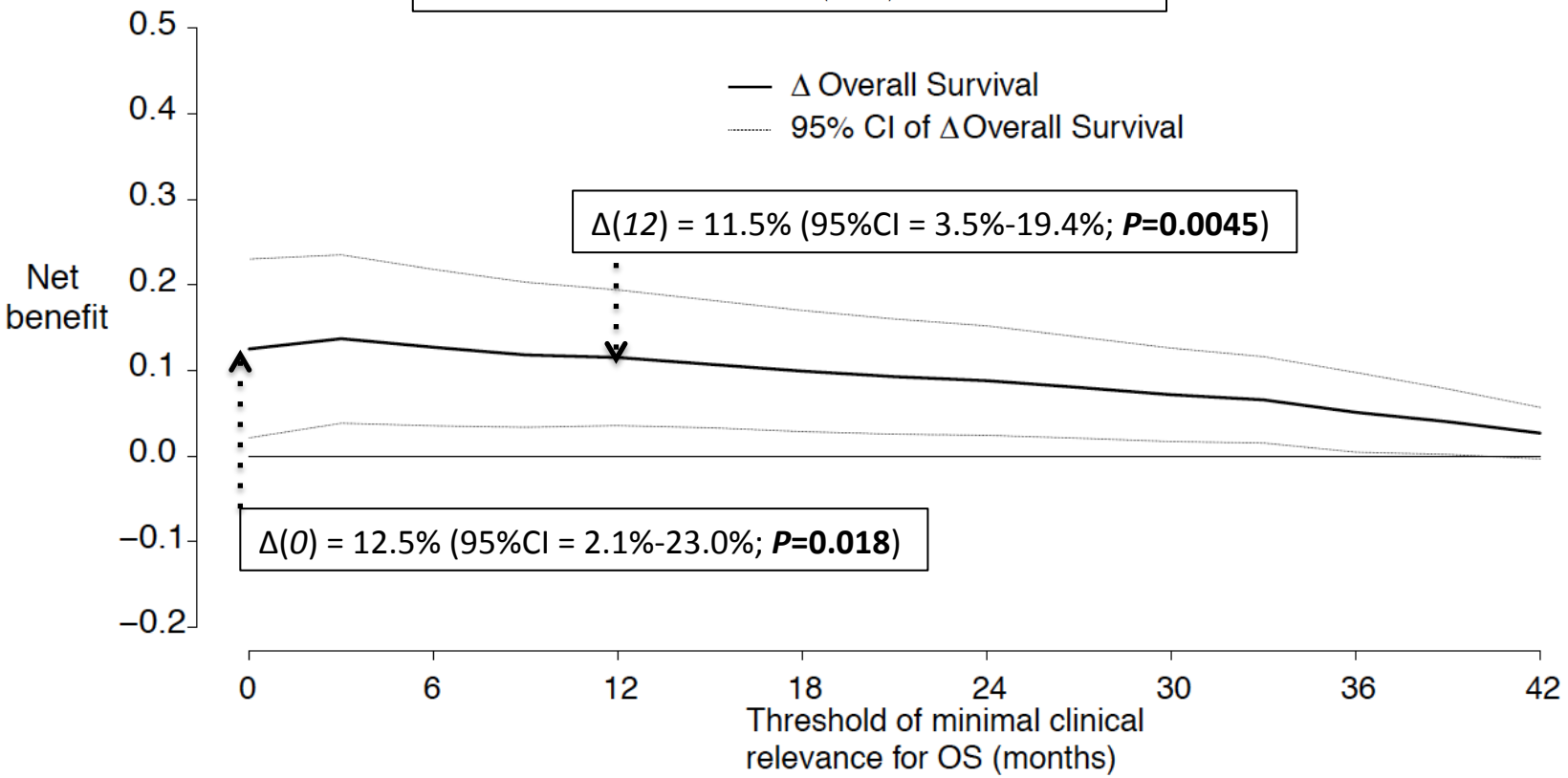




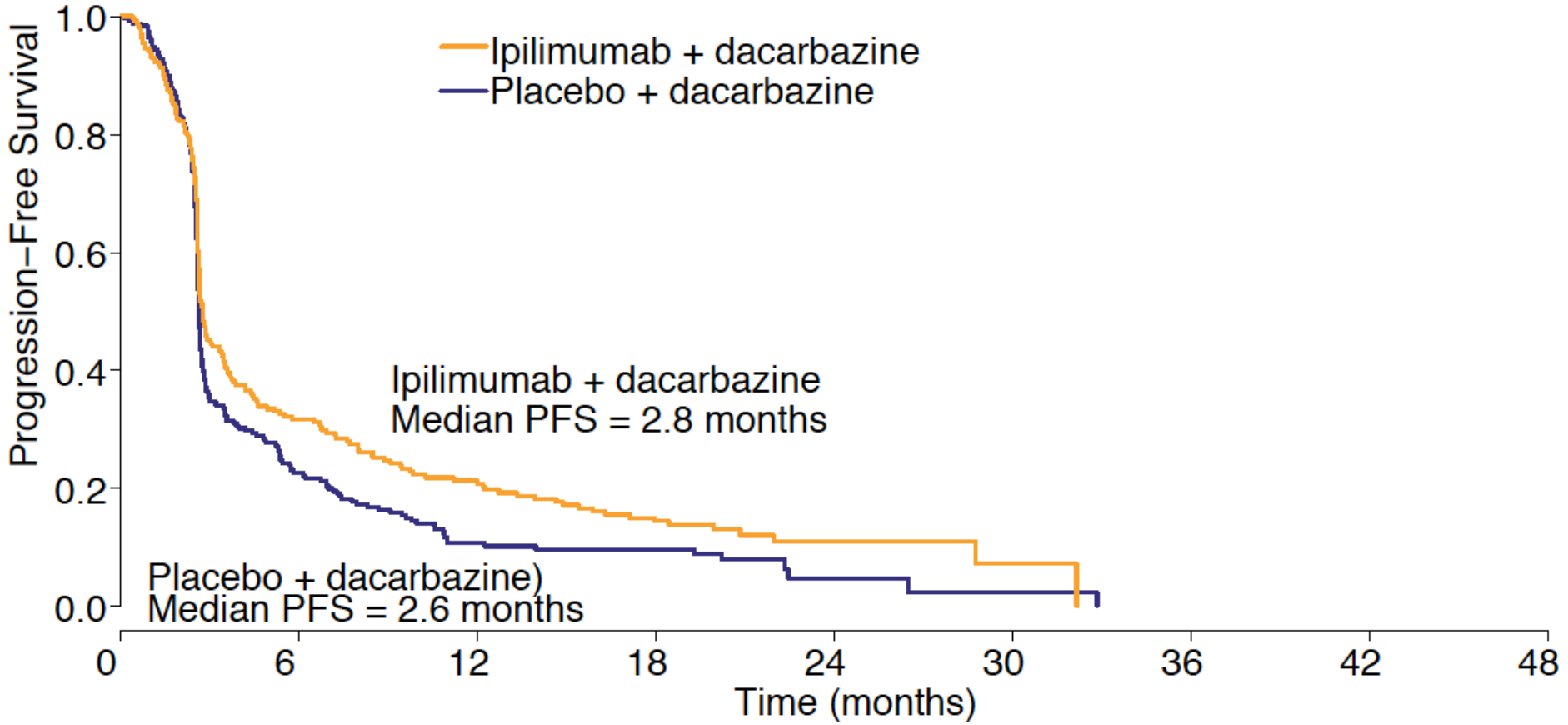
OS results in the CA184-024 trial



Log rank P = 0.0054

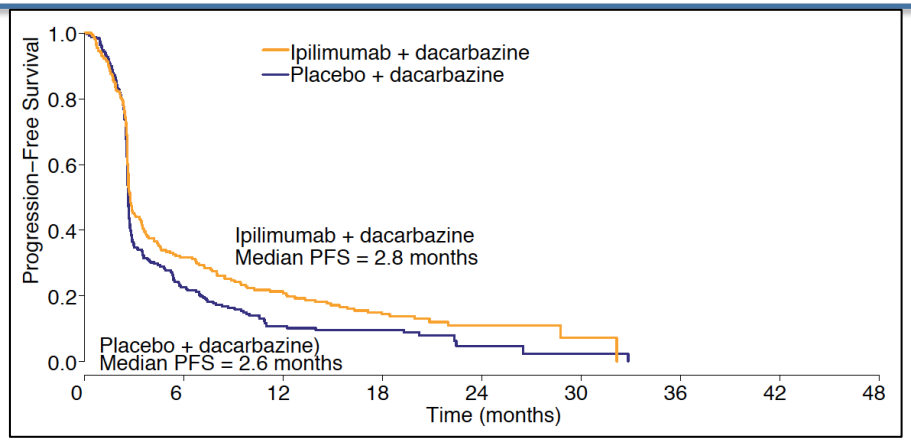


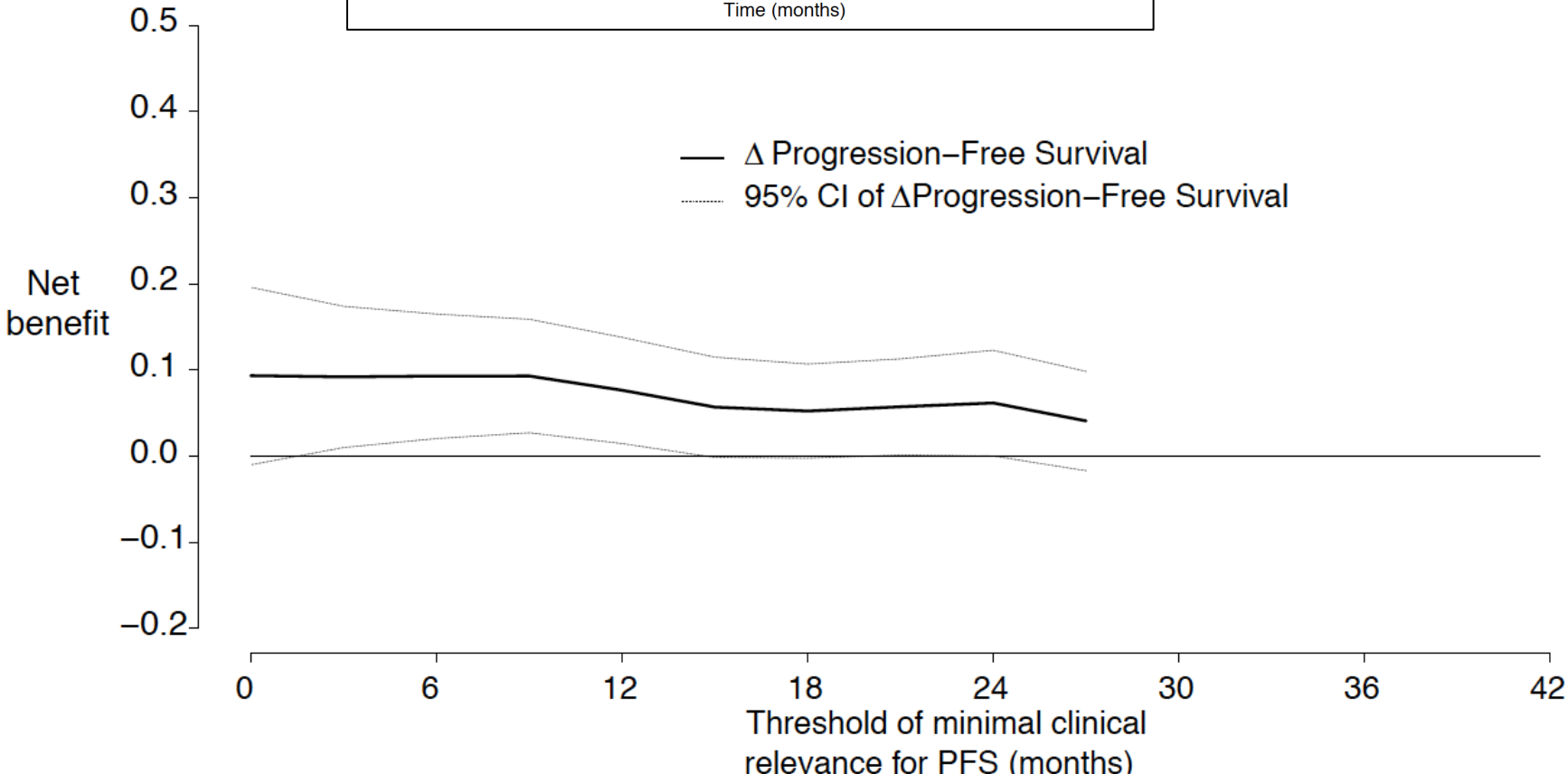
PFS results in the CA184-024 trial



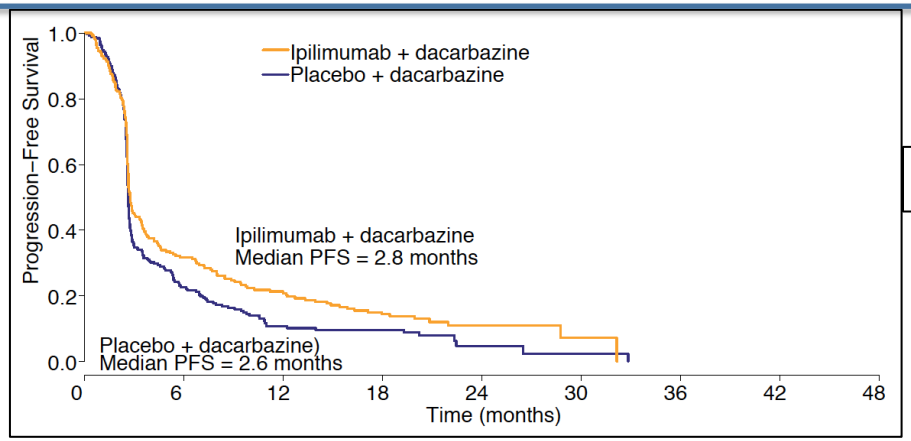
Pcb	252	52	20	13	2	1	0	0	0
Ipi	250	70	40	25	6	2	0	0	0

PFS results in the CA184-024 trial

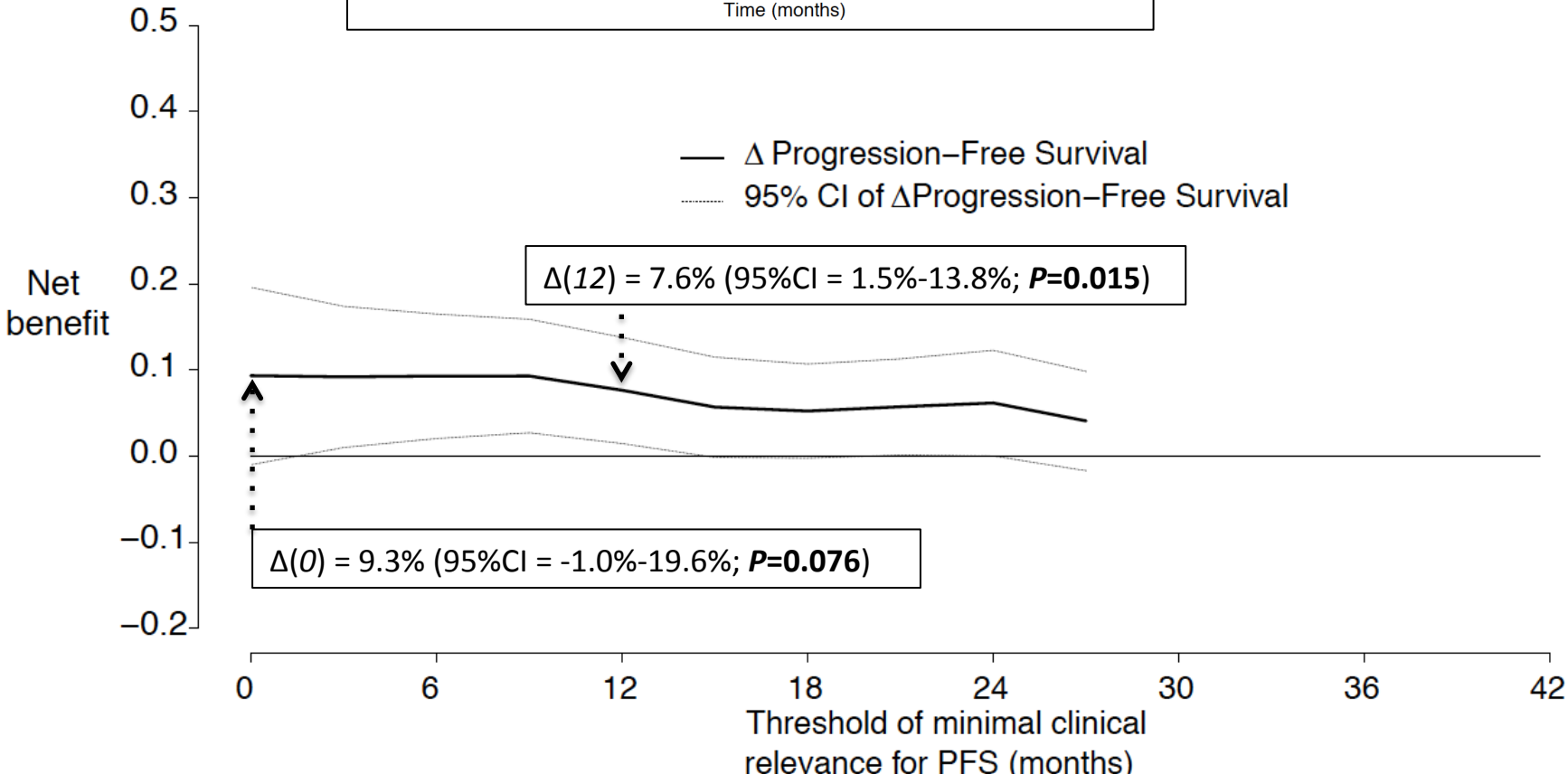




PFS results in the CA184-024 trial



Log rank P = 0.022



Software implementation

A package 

- Available on CRAN (“BuyseTest”)
- Available on github (“https://github.com/bozenne/BuyseTest”)

Conclusions

The net benefit

- Is equivalent to standard non-parametric tests in simple cases
- Is **meaningful and patient-relevant**
- Can focus on **long-term survival differences**
- Allows **multicriteria analysis**
- May have better power than the logrank test (e.g. for delayed treatment effect)
- Is OK when hazards are **not proportionals**
- Is **available**

Thank you

References

Buyse M. Reformulating the hazard ratio to enhance communication with clinical investigators. *Clin Trials* 5: 641-2, 2008.

Buyse M. Generalized pairwise comparisons for prioritized outcomes in the two-sample problem. *Statist Med* 29: 3245-57, 2010.

Péron J, Buyse M, Ozenne B, Roche L, Roy P. An extension of generalized pairwise comparisons for prioritized outcomes in the presence of censoring. *Statist Meth Med Res* DOI: 10.1177/0962280216658320, 2017.

Péron J, Roy P, Ding K, Parulekar W, Roche L, Buyse M. Benefit-risk assessment of adding erlotinib to gemcitabine for the treatment of advanced pancreatic cancer. *Brit J Cancer* 112: 971-976, 2015.

Péron J, Roy P, Ozenne B, Roche L, Buyse M. The net chance of a longer survival as a patient-oriented measure of benefit in randomized clinical trials. *JAMA Oncology* DOI: 10.1001/jamaoncol.2015. 6359, 2016.

Methods – Definition of priority

First priority outcome	Second priority outcome	Pair rating
Favorable	NA	Favorable
Unfavorable	NA	Unfavorable
Neutral/Uninf	Favorable	Favorable
Neutral/Uninf	Unfavorable	Unfavorable
Neutral/Uninf	Neutral/Uninf	Neutral/Uninf

Buyse. stat in med 2010

Methods – Definition of priority

First priority outcome	Second priority outcome	Pair rating
Favorable	NA	Favorable
Unfavorable	NA	Unfavorable
Neutral/Uninf	Favorable	Favorable
Neutral/Uninf	Unfavorable	Unfavorable
Neutral/Uninf	Neutral/Uninf	Neutral/Uninf

Buyse. stat in med 2010

Methods – Definition of priority

First priority outcome	Second priority outcome	Pair rating
Favorable	NA	Favorable
Unfavorable	NA	Unfavorable
Neutral/Uninf	Favorable	Favorable
Neutral/Uninf	Unfavorable	Unfavorable
Neutral/Uninf	Neutral/Uninf	Neutral/Uninf

Buyse. stat in med 2010

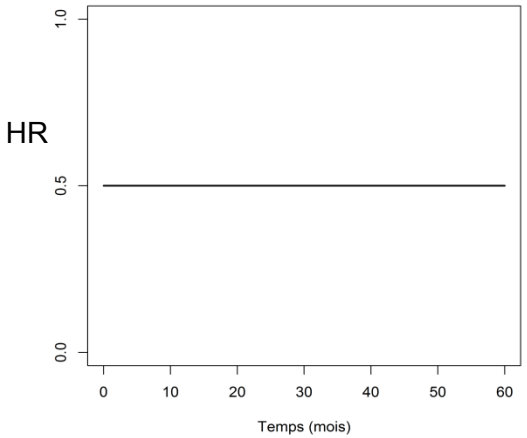
Simulation study - Design

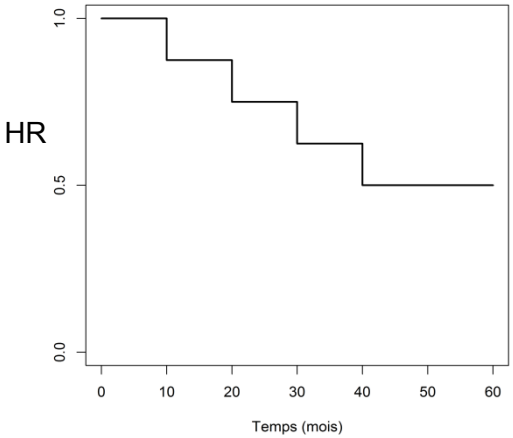
- **Objective:** To compare the standard and the extended procedures of generalized pairwise comparison
- Simulation of M = 1000 datasets of with N = 200 patients
  - One time-to-event outcome
  - Threshold τ = 0 months

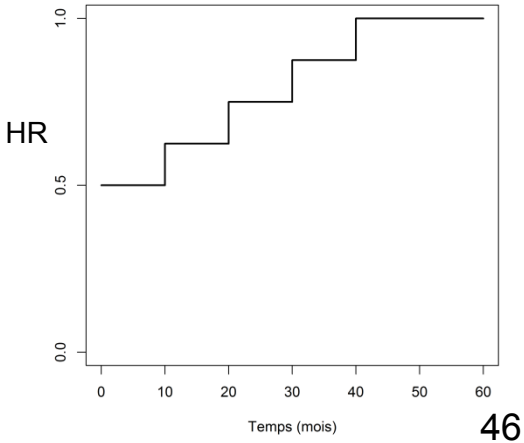
Simulation study - Design

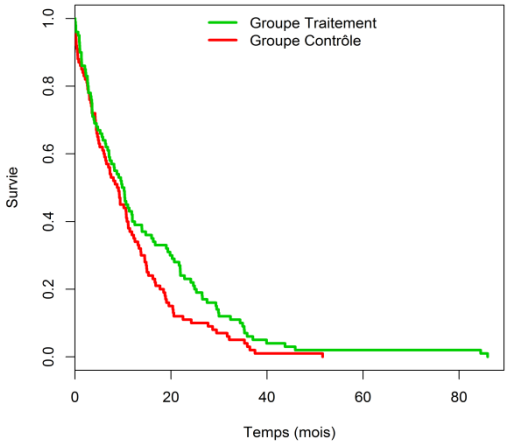
- Survival time: exponential distributions

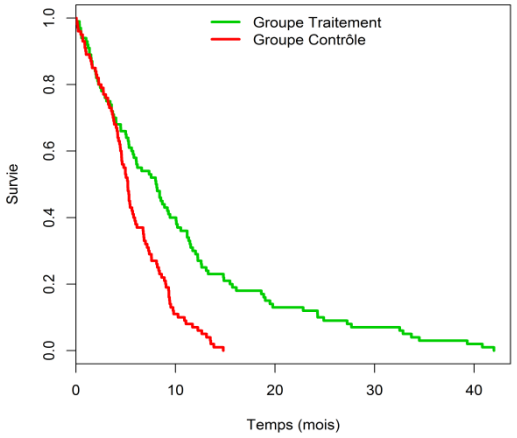
Scenario 1 : Proportional hazards	Scenario 2 : Late treatment effect	Scenario 3 : early treatment effect

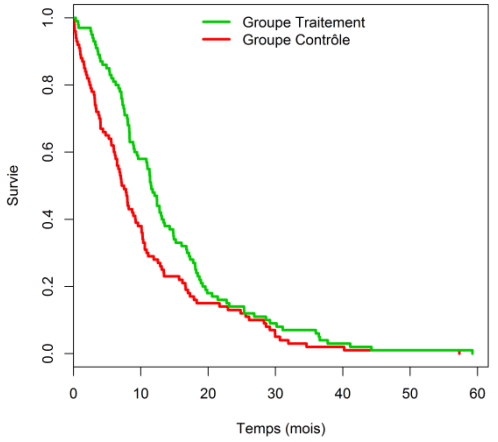












Simulation study - Design

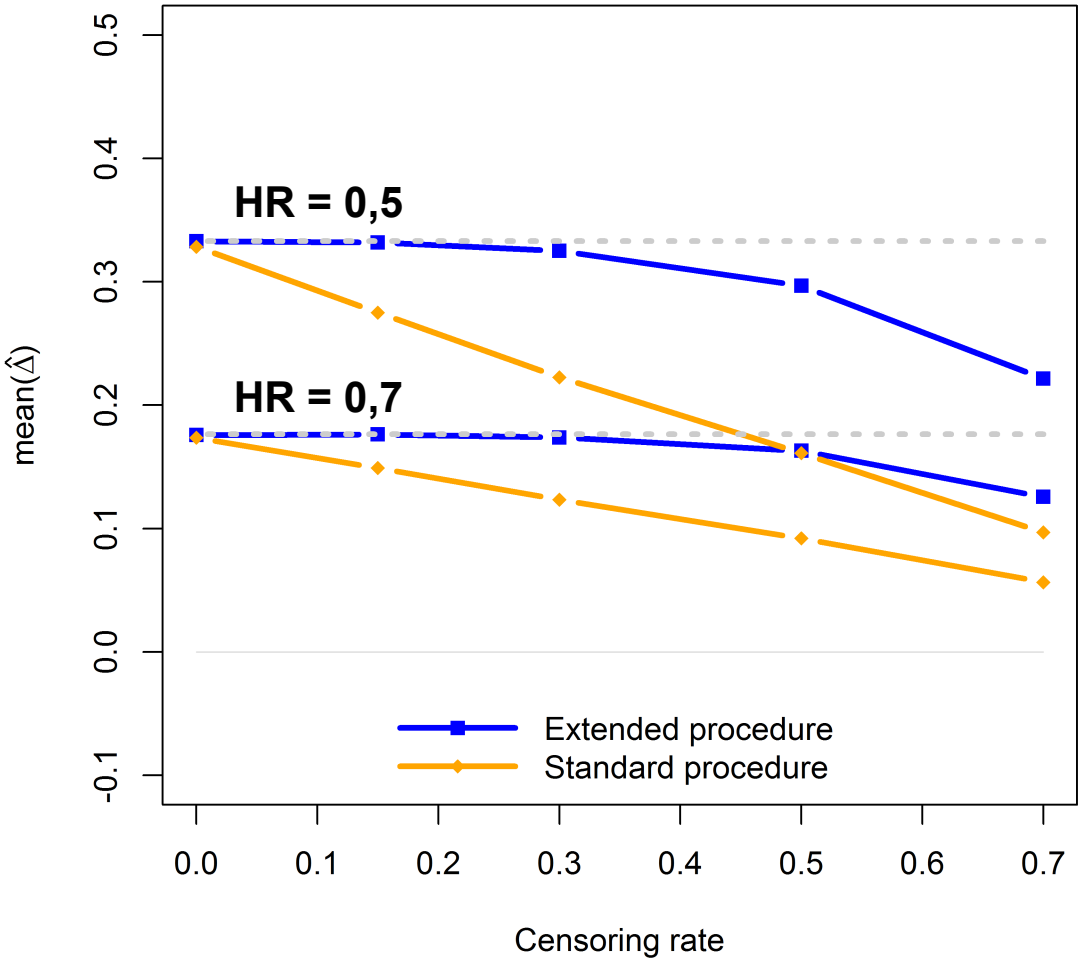
- Several treatment effect size
  - Hazard ratio $\{0{,}5; 0{,}7; 1\}$
- Administrative censoring proportion
  - Uniform distribution
  - Between 0% and 70%

Simulation study - Design

- For each simulated dataset
  - Estimation of the net chance of a better outcome (standard and extende procedure)
  - Test of the null hypothesis (Permutation test, Log-Rank test)

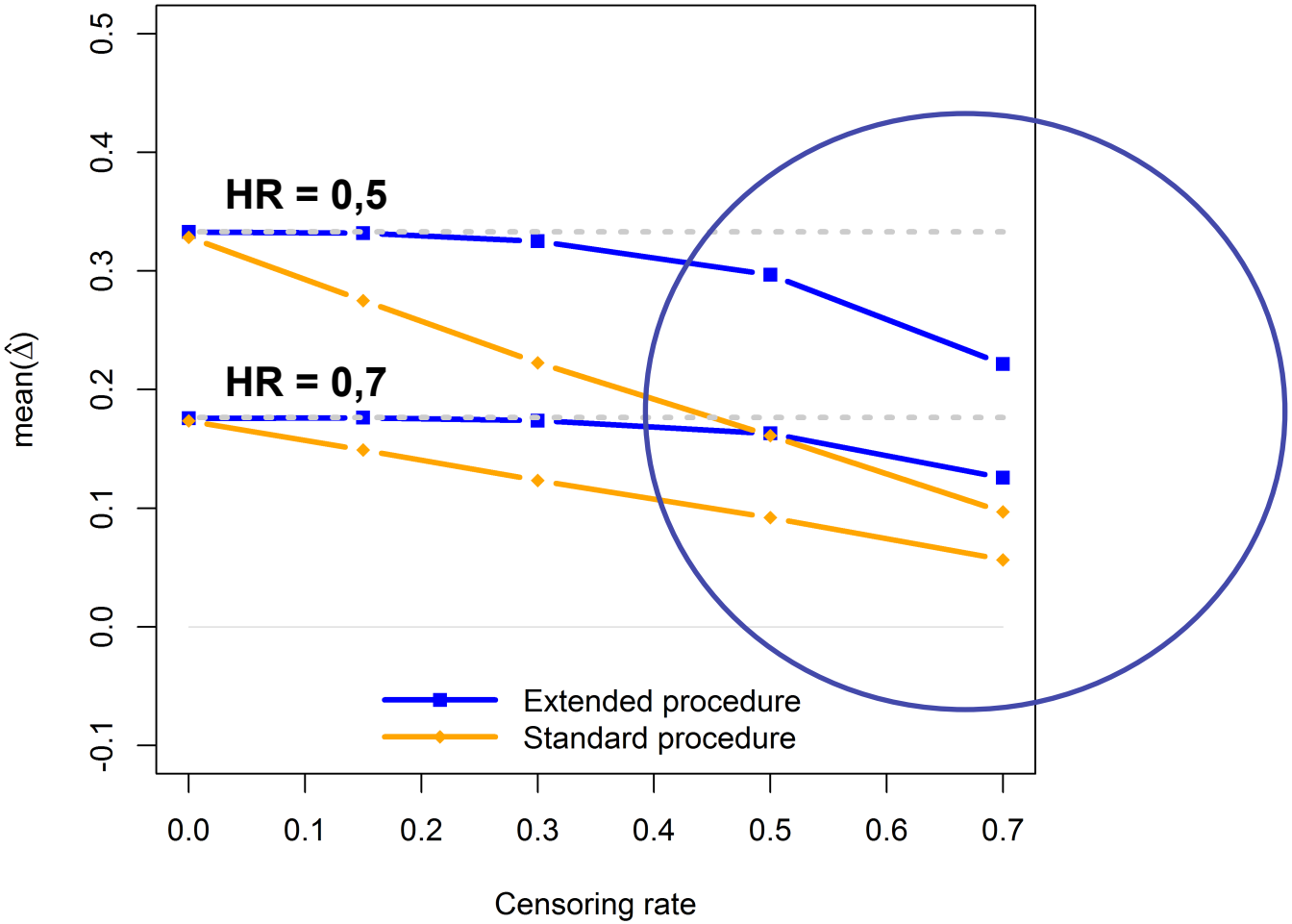
- Endpoints
  - Bias
  - Power
  - Type 1 error

Scenario 1 – Proportional hazards



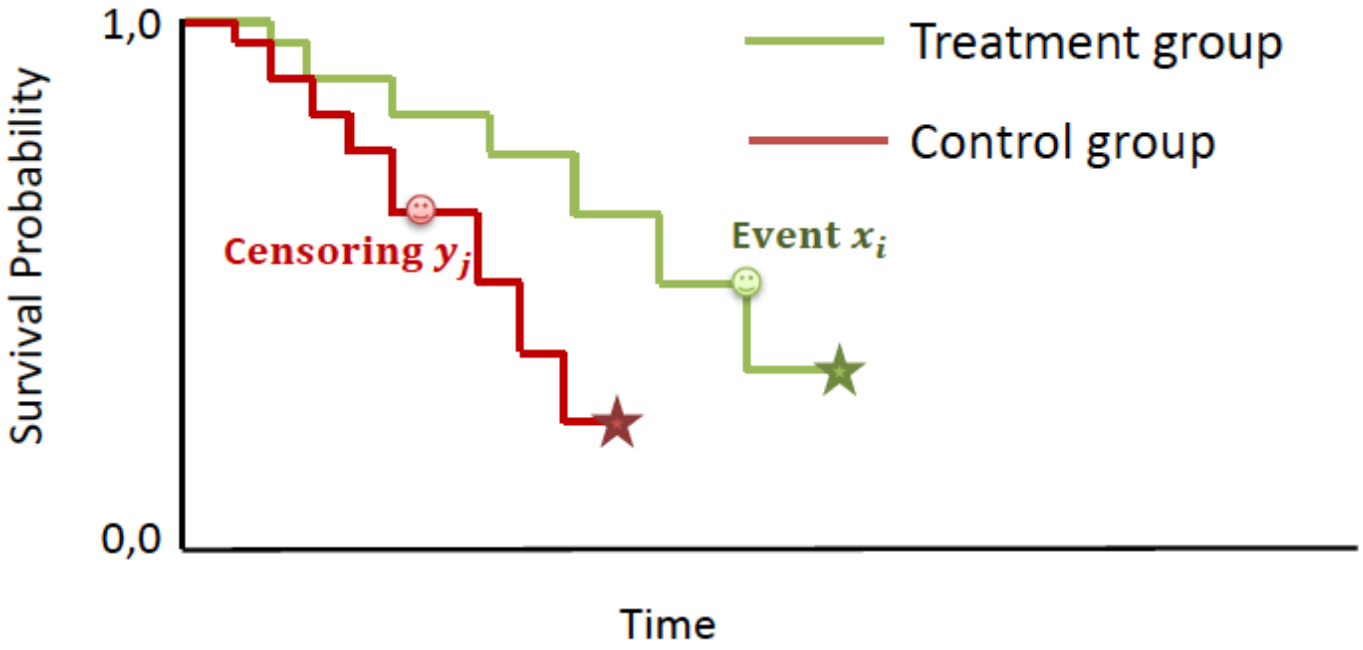
Péron, et al. *SMMR 2016*

Scenario 1 – Proportional hazards



Péron, et al. *SMMR 2016*

An explanation for this bias?



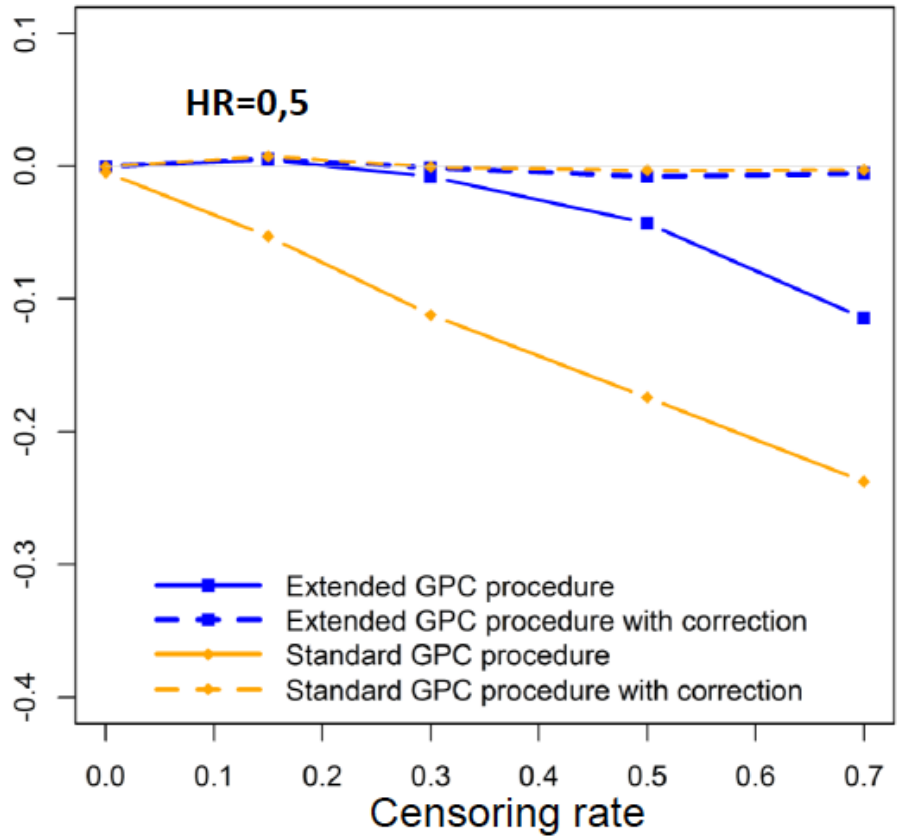
Standard procedure: **Uninformative** $\rightarrow p_{ij} = 0$

Extended procedure: $\begin{cases} \mathbb{P}\left[x_i^0 > y_j^0 + \tau\right] = 1 - \frac{\hat{S}_C(x_i - \tau)}{\hat{S}_C(y_j)} \\ \mathbb{P}\left[y_j^0 > x_i^0 + \tau\right] = \frac{\hat{S}_C(x_i + \tau)}{\hat{S}_C(y_j)} \end{cases}$ **Uninformative also** $\rightarrow p_{ij} = 0$

A correction for this bias

f_m : Proportion of informative pairs for the trial $m (m = 1, \ldots, M)$

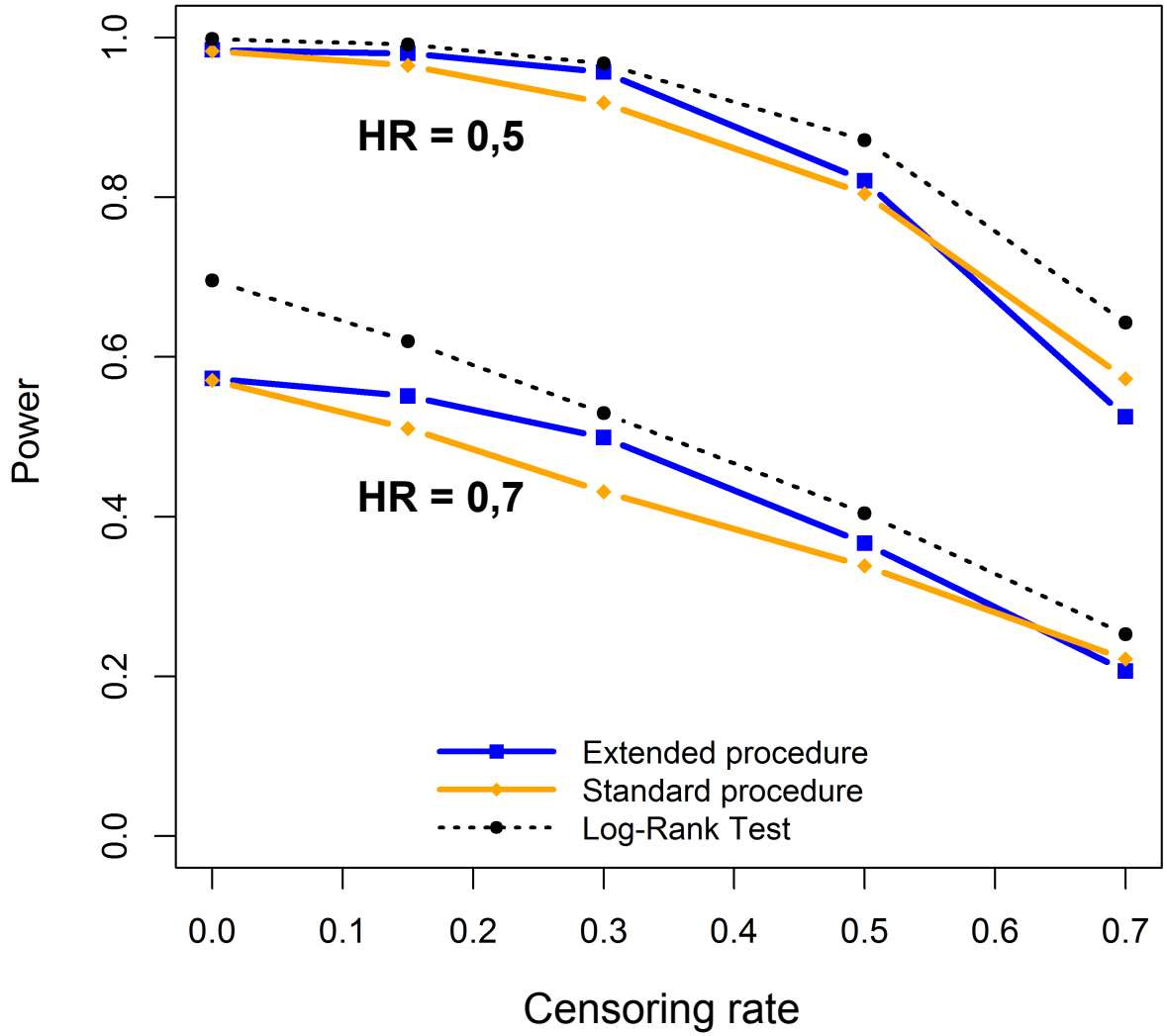
$\widehat{\Delta}corr_m = \frac{\widehat{\Delta}_m}{f_m}$: corrected net chance of a better outcome



$$Bias = \mathbb{E}\left(\widehat{\Delta} - \Delta\right)$$

Péron, et al. *SMMR 2016*

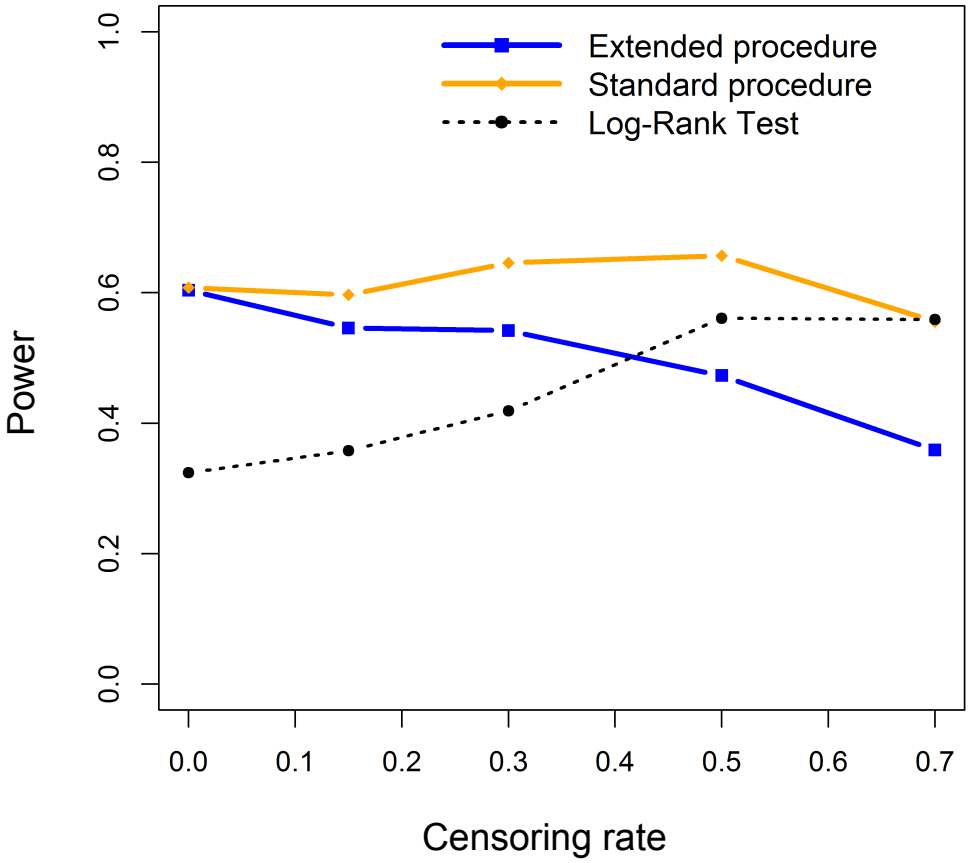
Scenario 1 – Proportional hazards



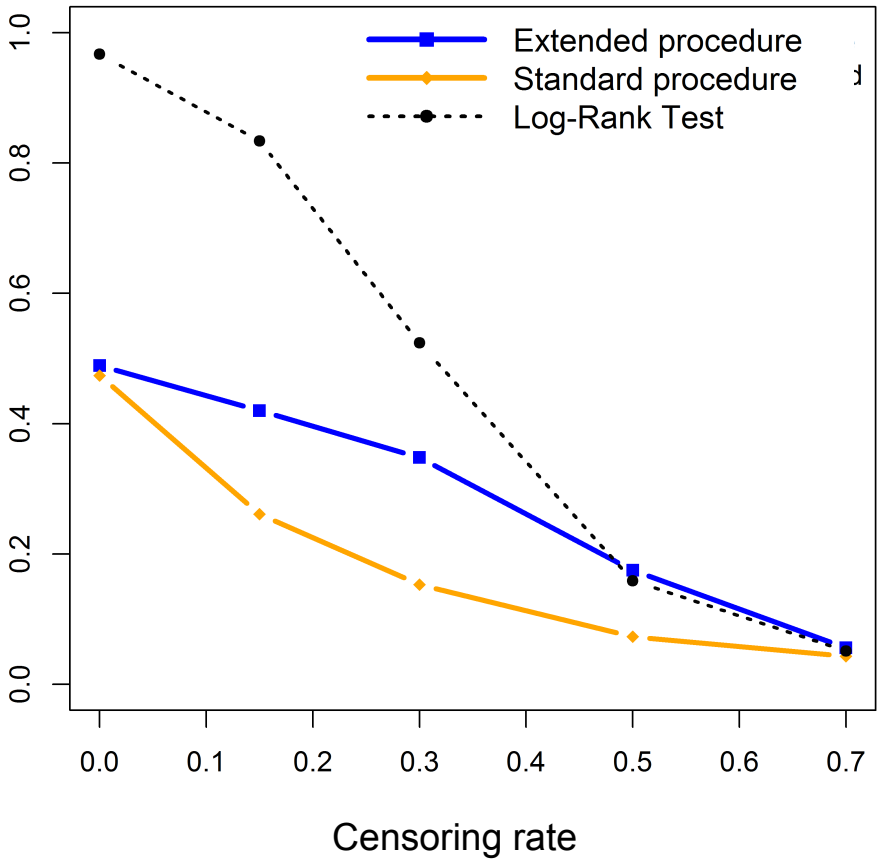
Péron, et al. *SMMR 2016*

Scenario 2 et 3 – Non Proportional hazards

Early treatment effect



Late treatment effect



Type 1 error rate ≈ 5%

Péron, et al. *SMMR 2016*