IDDI

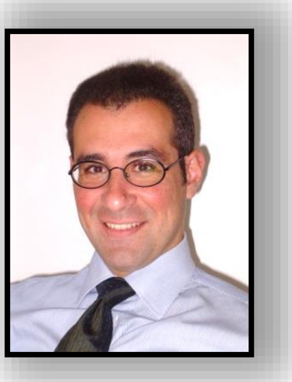
Speakers

A New Method to Assess Benefit/Risk, with Examples in Oncology

Marc Buyse, ScD
Chief Scientific Officer, IDDI
marc.buyse@iddi.com

Everardo D. Saad, MD
Medical Director, IDDI
everardo.saad@iddi.com

Outline



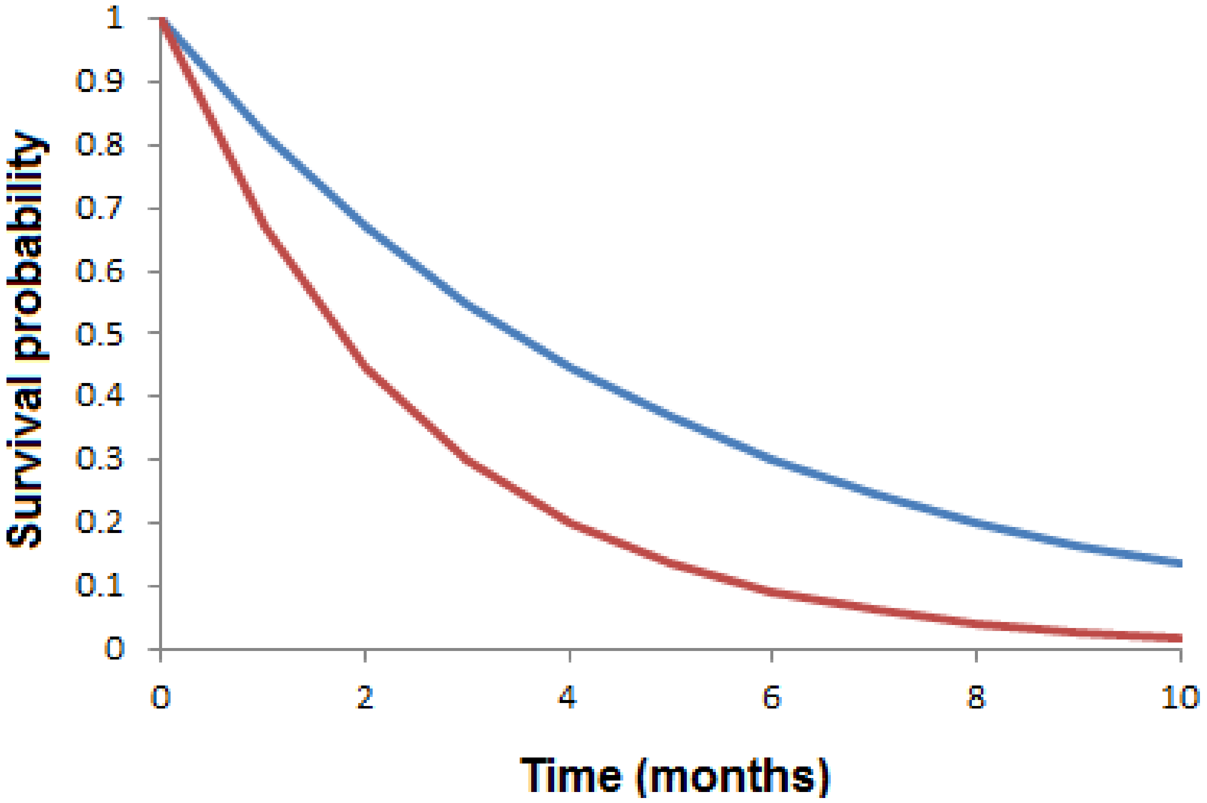
Everardo Saad

- Overview of measures of treatment benefit
- Current attempts to individualize decision-making

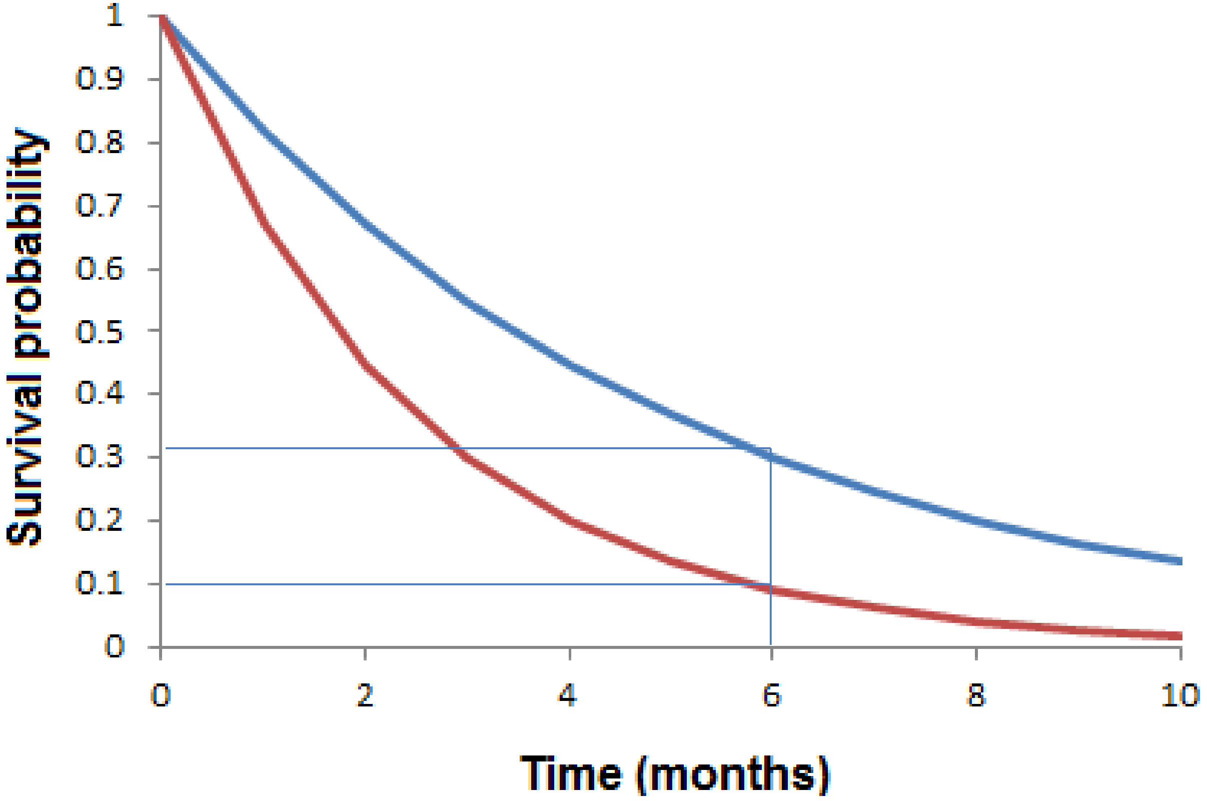
Marc Buyse

- Generalized pairwise comparisons (GPC)
- Prioritising outcomes, with examples in oncology
- GPC in the setting of non-proportional hazards

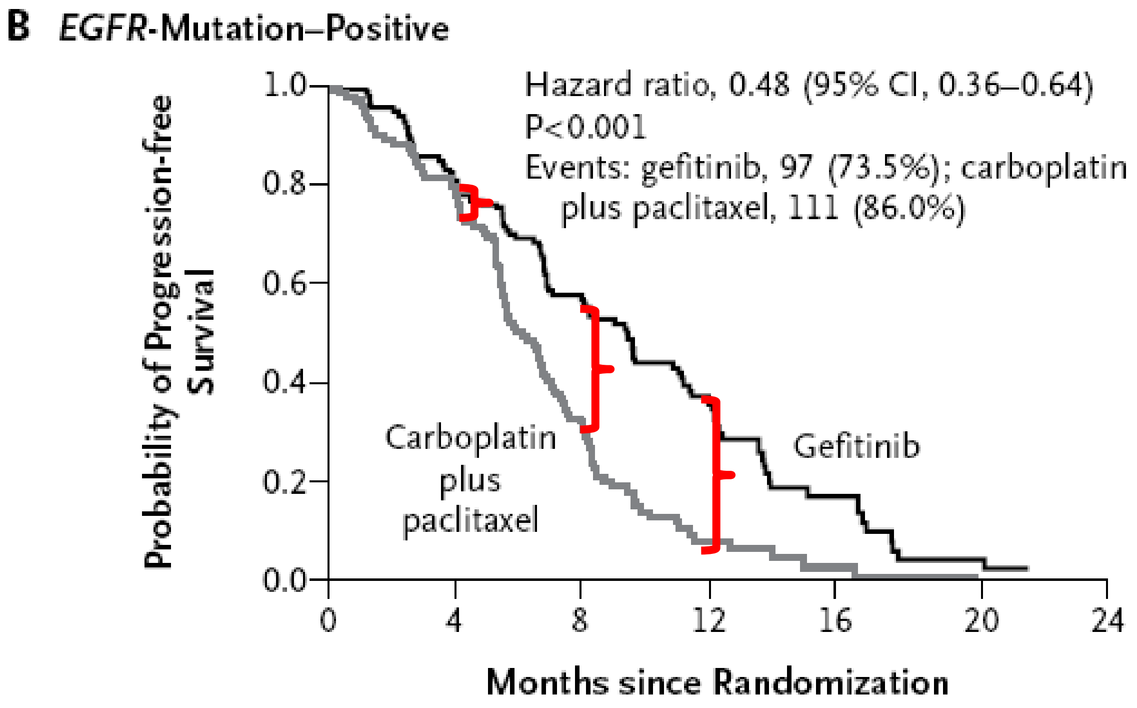
Comparing survival



Survival probability at *t*



Problems

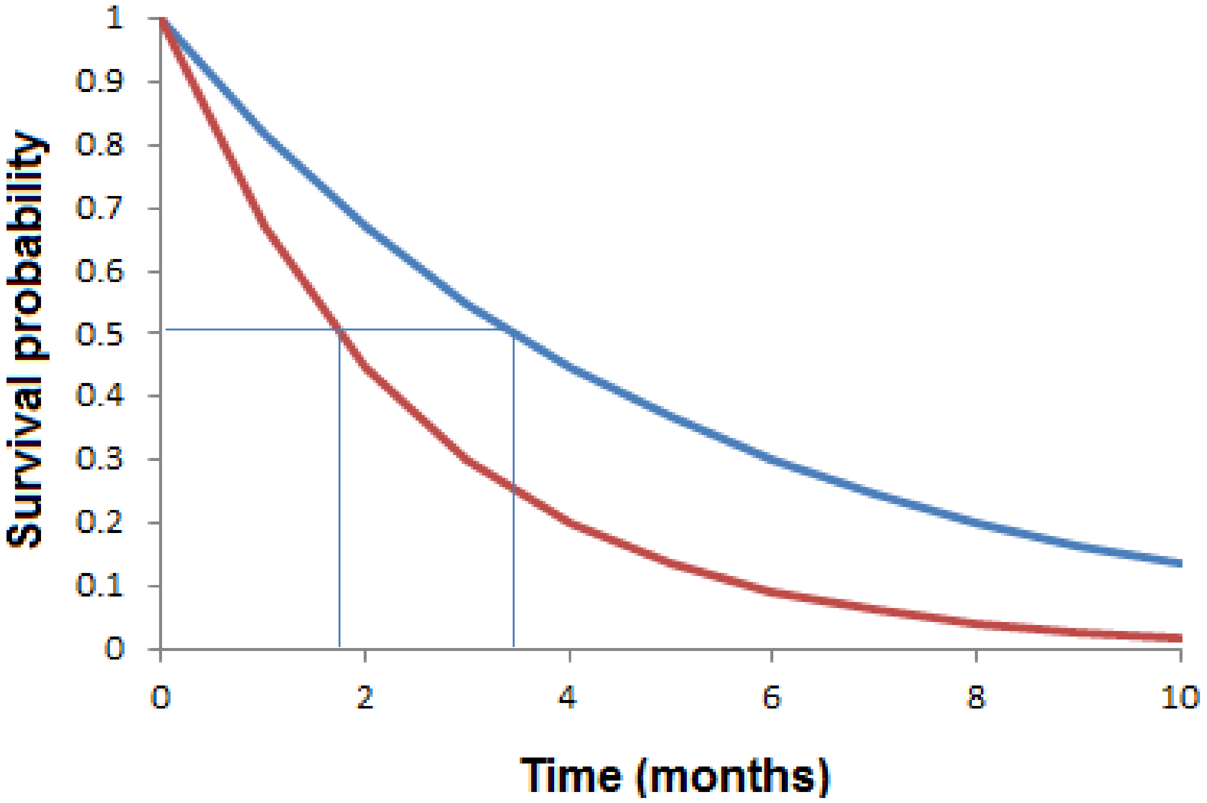


No. at Risk

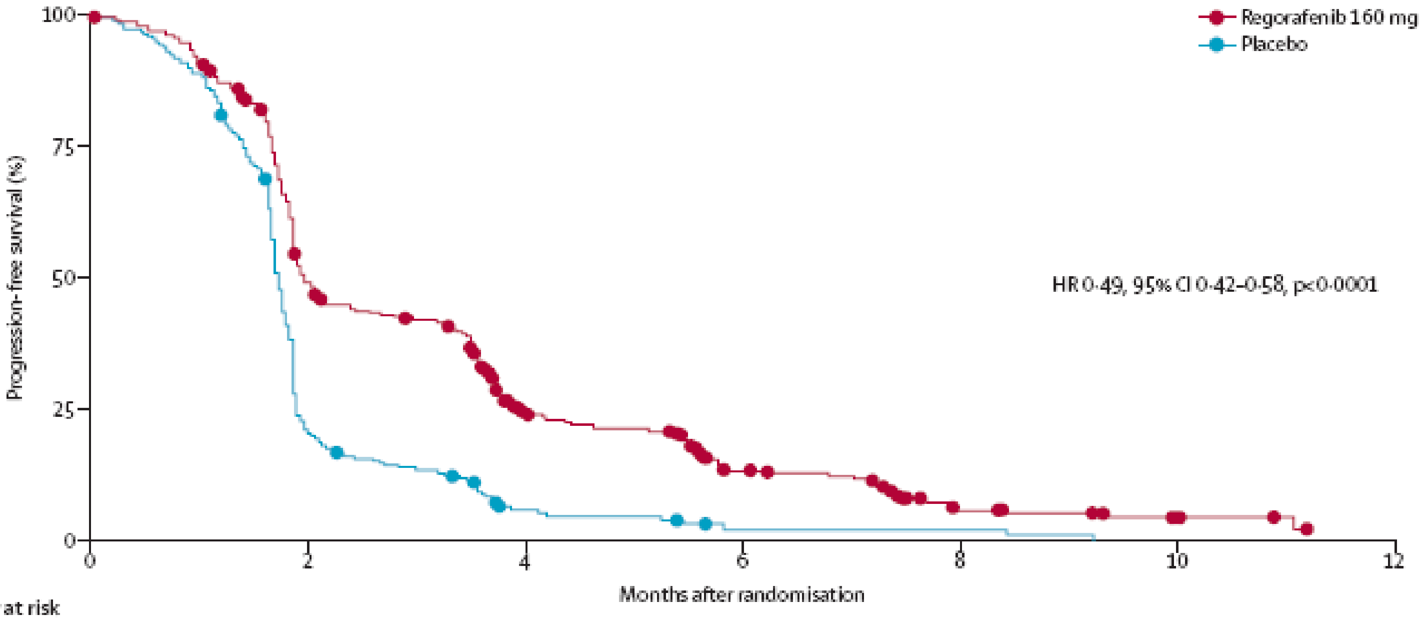
Gefitinib	132	108	71	31	11	3	0
Carboplatin plus paclitaxel	129	103	37	7	2	1	0

Mok et al, N Engl J Med 2009;361:947-57

Difference in medians



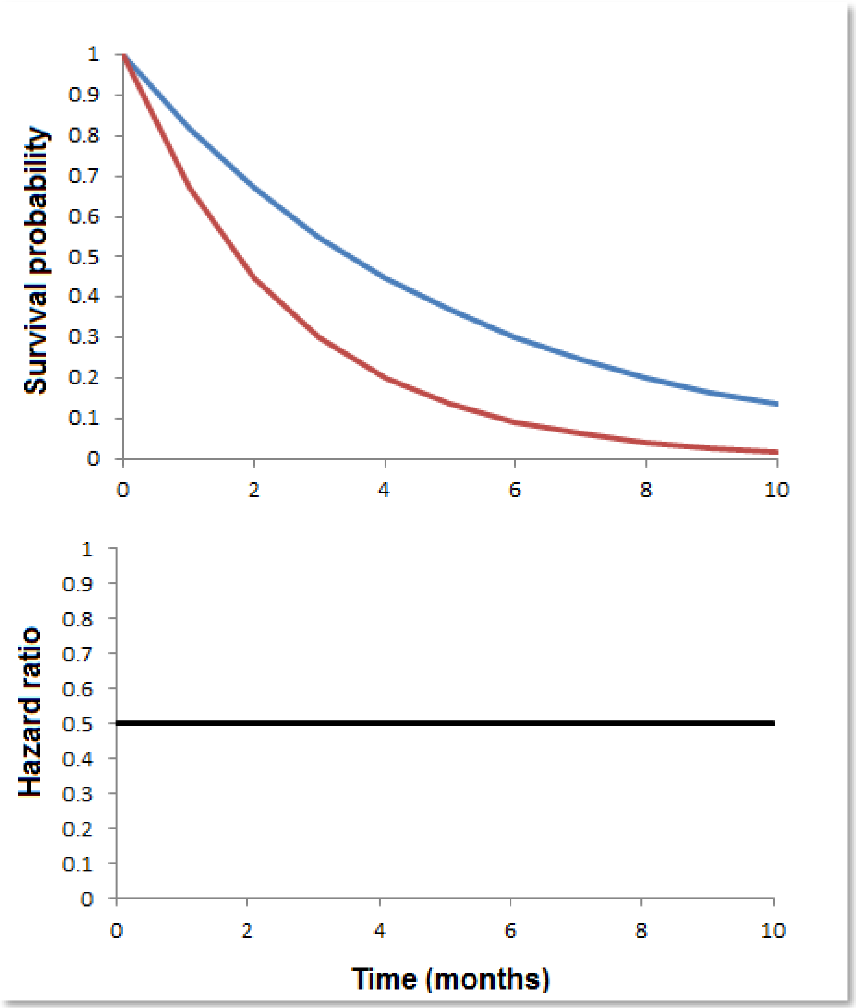
Problems

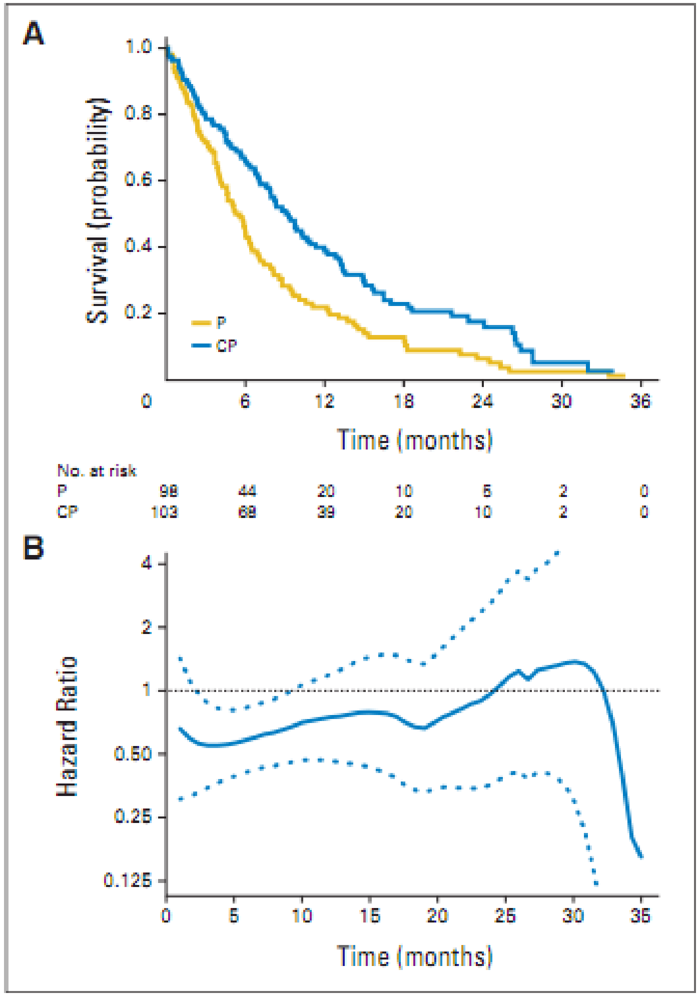


Number at risk					
Regorafenib	238	98	42	12	3
Placebo	51	9	2	2	0

Grothey et al, Lancet 2013; 381:303

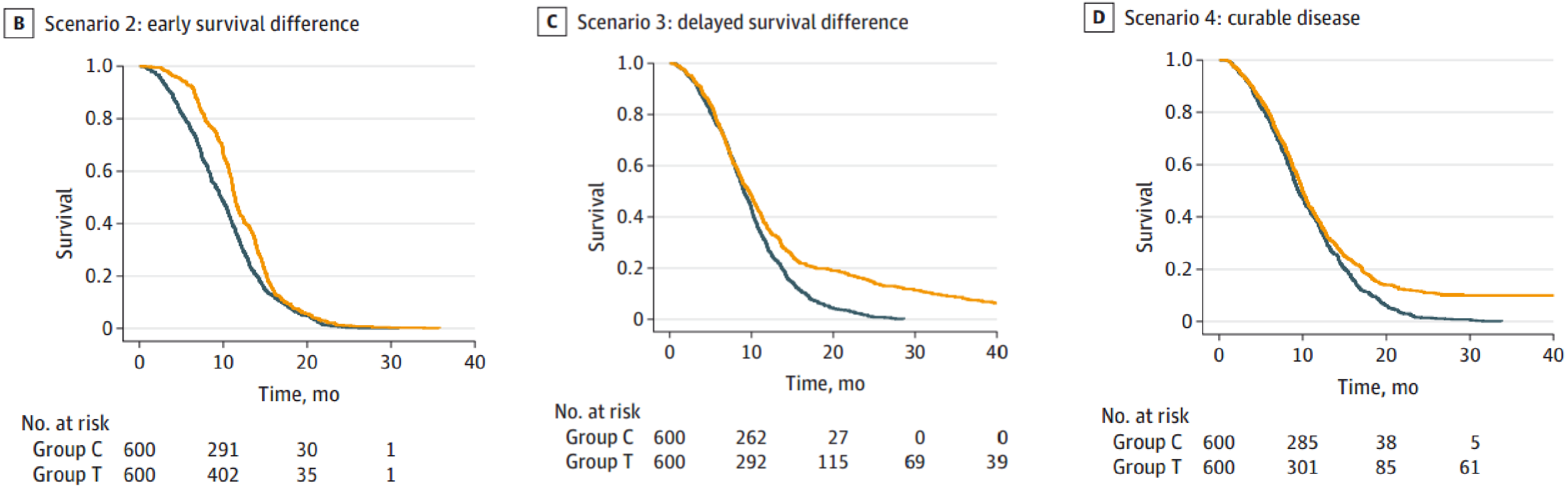
Hazard ratio





Saad et al, J Natl Cancer Inst 2018; Uno et al, J Clin Oncol 2014;32:2380

Non-proportional hazards



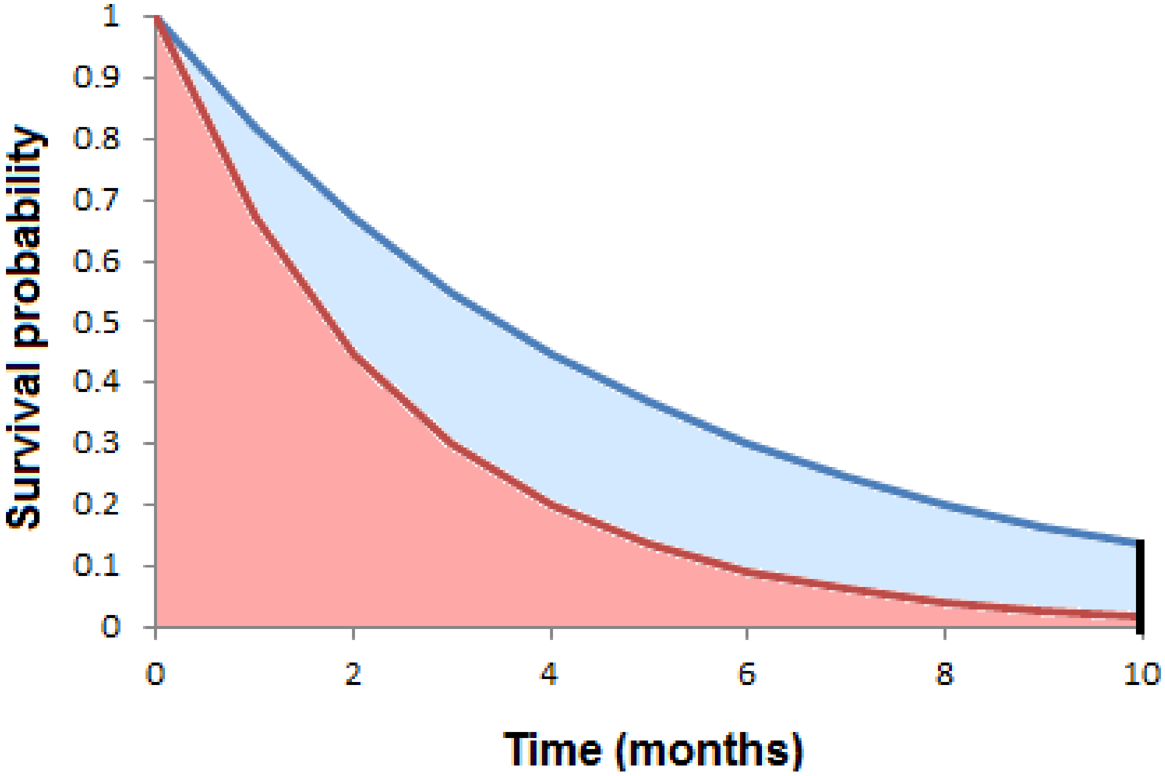
No. at risk				
Group C	600	291	30	1
Group T	600	402	35	1

No. at risk					
Group C	600	262	27	0	0
Group T	600	292	115	69	39

No. at risk				
Group C	600	285	38	5
Group T	600	301	85	61

Péron et al, JAMA Oncol 2016;2:901

Restricted means



Which one should we use?

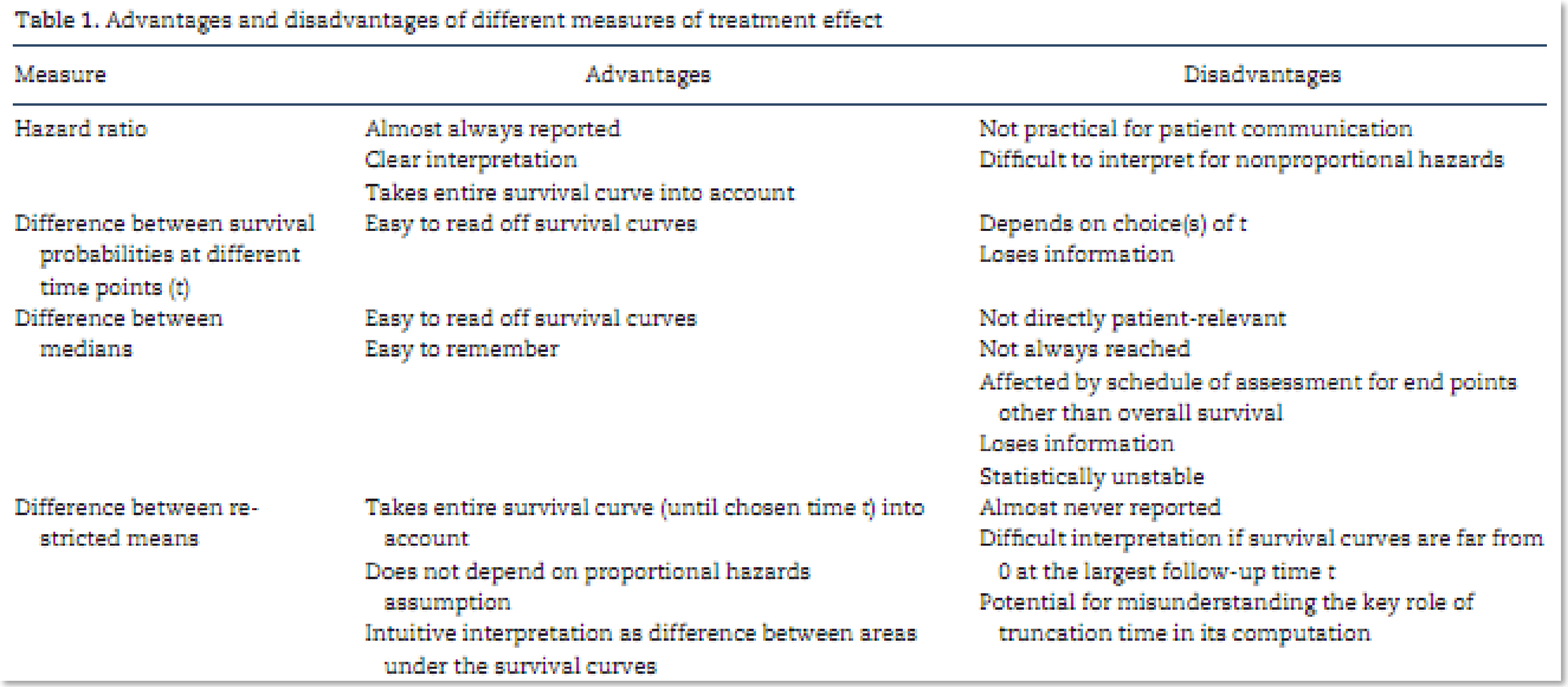
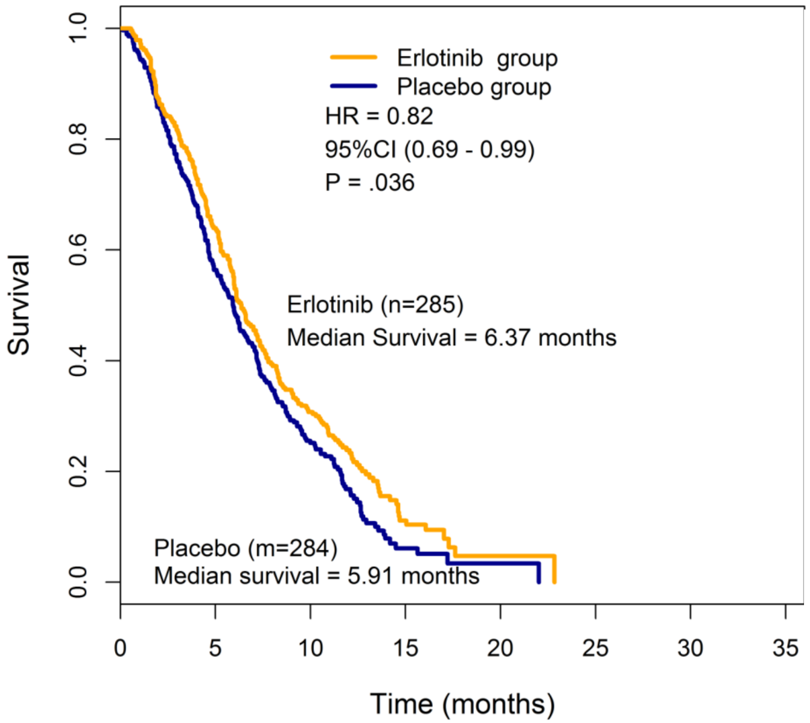


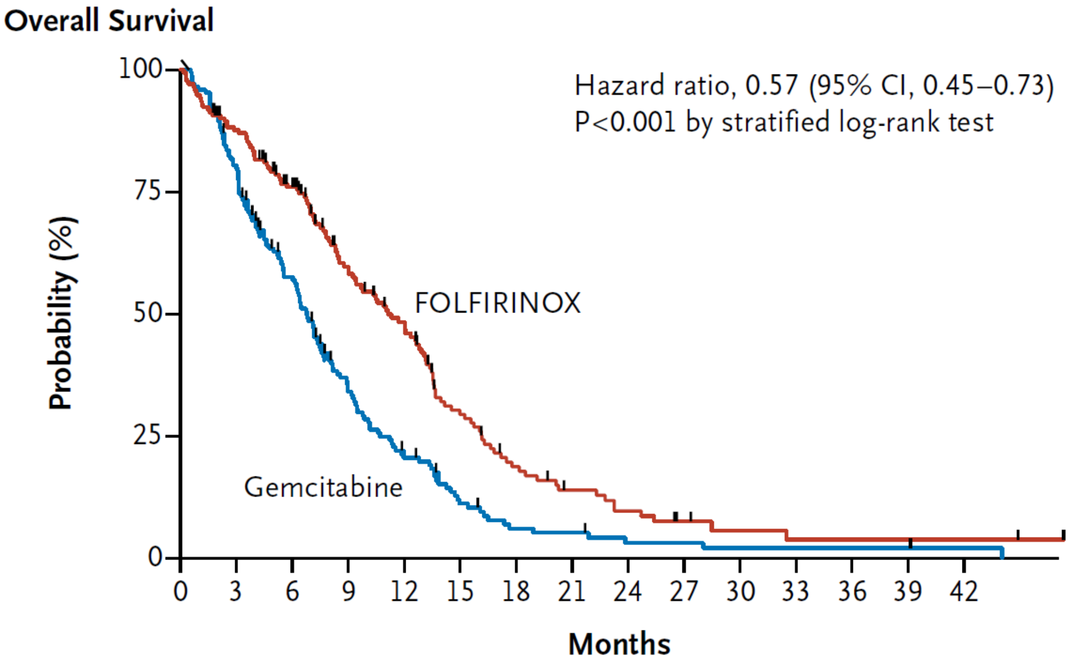
Table 1. Advantages and disadvantages of different measures of treatment effect

Measure	Advantages	Disadvantages
Hazard ratio	Almost always reported Clear interpretation Takes entire survival curve into account	Not practical for patient communication Difficult to interpret for nonproportional hazards
Difference between survival probabilities at different time points (t)	Easy to read off survival curves	Depends on choice(s) of t Loses information
Difference between medians	Easy to read off survival curves Easy to remember	Not directly patient-relevant Not always reached Affected by schedule of assessment for end points other than overall survival Loses information Statistically unstable
Difference between restricted means	Takes entire survival curve (until chosen time t) into account Does not depend on proportional hazards assumption Intuitive interpretation as difference between areas under the survival curves	Almost never reported Difficult interpretation if survival curves are far from 0 at the largest follow-up time t Potential for misunderstanding the key role of truncation time in its computation

Saad et al, J Natl Cancer Inst 2018

Some examples





Moore et al, J Clin Oncol 2007; 25:1960; Conroy et al, N Engl J Med 2011;364:1817

Different views

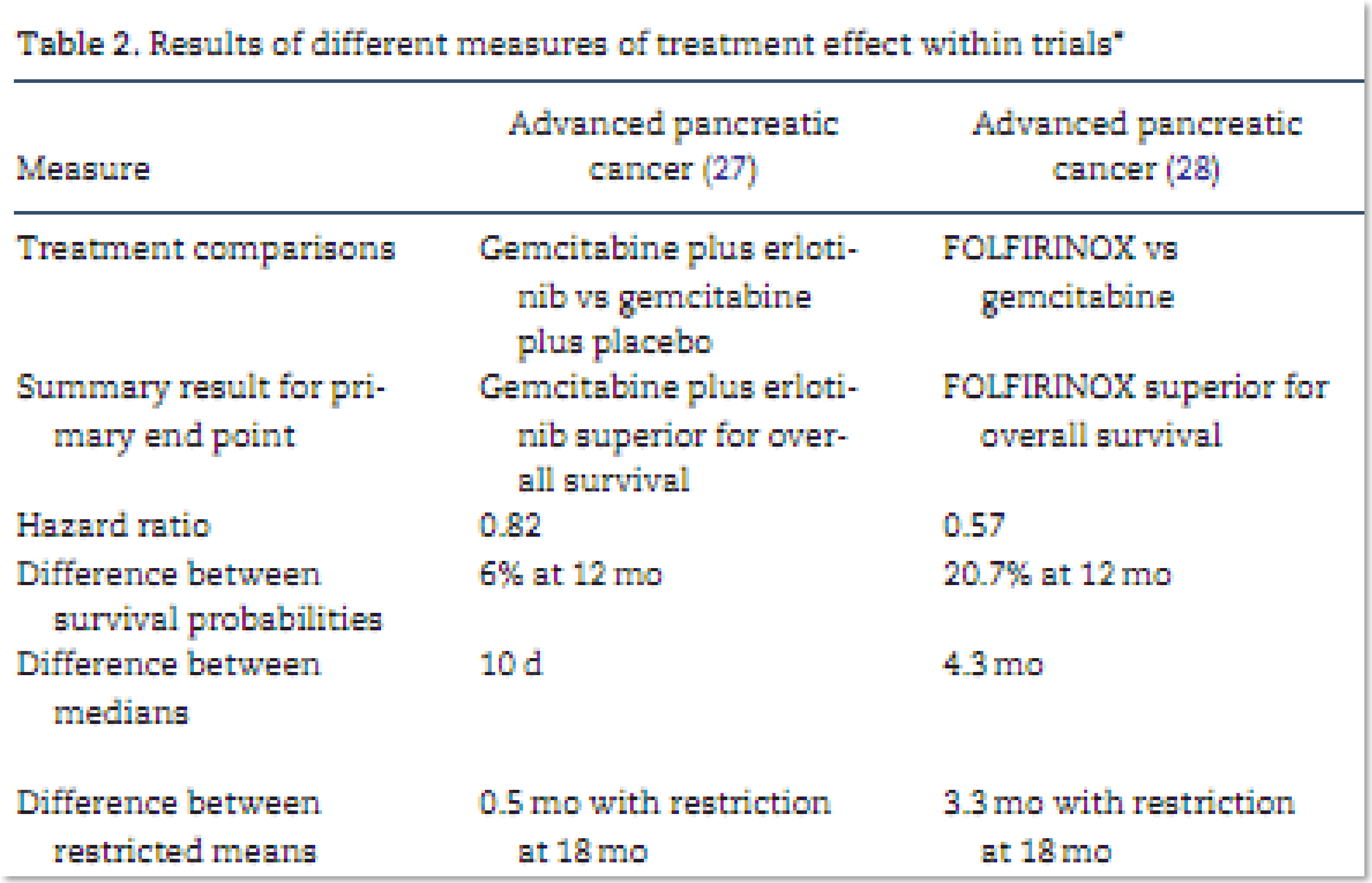


Table 2. Results of different measures of treatment effect within trials*

Measure	Advanced pancreatic cancer (27)	Advanced pancreatic cancer (28)
Treatment comparisons	Gemcitabine plus erlotinib vs gemcitabine plus placebo	FOLFIRINOX vs gemcitabine
Summary result for primary end point	Gemcitabine plus erlotinib superior for overall survival	FOLFIRINOX superior for overall survival
Hazard ratio	0.82	0.57
Difference between survival probabilities	6% at 12 mo	20.7% at 12 mo
Difference between medians	10 d	4.3 mo
Difference between restricted means	0.5 mo with restriction at 18 mo	3.3 mo with restriction at 18 mo

Saad et al, J Natl Cancer Inst 2018

Assessing the worth of treatment

- Formally
  - The primary endpoint, usually related to efficacy, but may be QOL or safety
  - Secondary endpoints
  - Health-economics measures, chiefly cost-effectiveness (QALYs, ICERs)
- Informally
  - Overall assessment of benefit/risk, as done by agencies
  - Issues about value

ASCO and ESMO "scales"

- Focus is on value (benefit/cost)
- Clinical benefit is predefined
  - Assumes a hierarchy within endpoints
  - Ignores potential problems with OS, QOL and surrogates
  - Arbitrary cut-off points of magnitude
- Decisions based on "marginal" results
- From a collective viewpoint, steps in a good direction

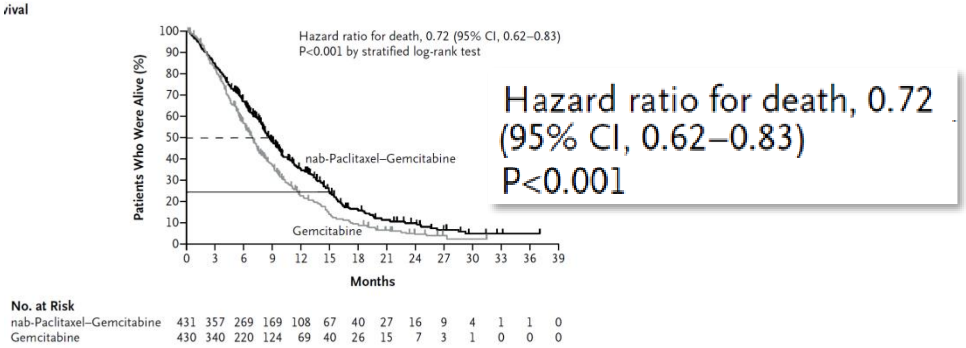
Schnipper et al, J Clin Oncol 2015;33:2563; Cherny et al, Ann Oncol 2015;26:1547

Individualizing treatment choices

- *Evidence-based medicine*
  - RCTs
  - Subgroup analysis, in some cases
- *Precision medicine:* “giving the right treatment to the right patient at the right time”
- *Personalized medicine:* doing this with individualized decisions about the goals of treatment

An unmet need

- Consider the following results



No. at Risk

nab-Paclitaxel–Gemcitabine	431	357	269	169	108	67	40	27	16	9	4	1	1	0
Gemcitabine	430	340	220	124	69	40	26	15	7	3	1	0	0	0

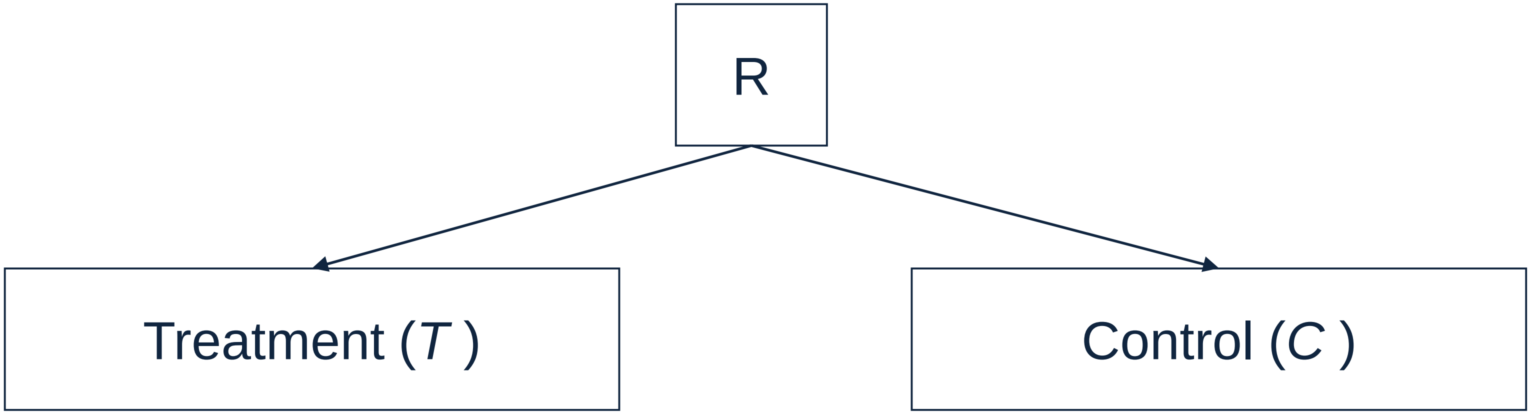
Hazard ratio for death, 0.72
(95% CI, 0.62–0.83)
P<0.001

Worst grade related AE	Monotherapy (n=430)	Combination (n=431)
Grade 3	23%	54%
Grade 4		

- A patient might reason:
  - Taking combination, I'm more likely to live longer
  - 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…

Von Hoff et al, N Engl J Med 2013;369:1691

Setting the stage



R

Treatment (*T*)

Control (*C*)

Let *A* be the result for the primary endpoint in each patient

B, C, D … for secondary endpoints

E, F, G … for untoward effects

Conventional analytic framework

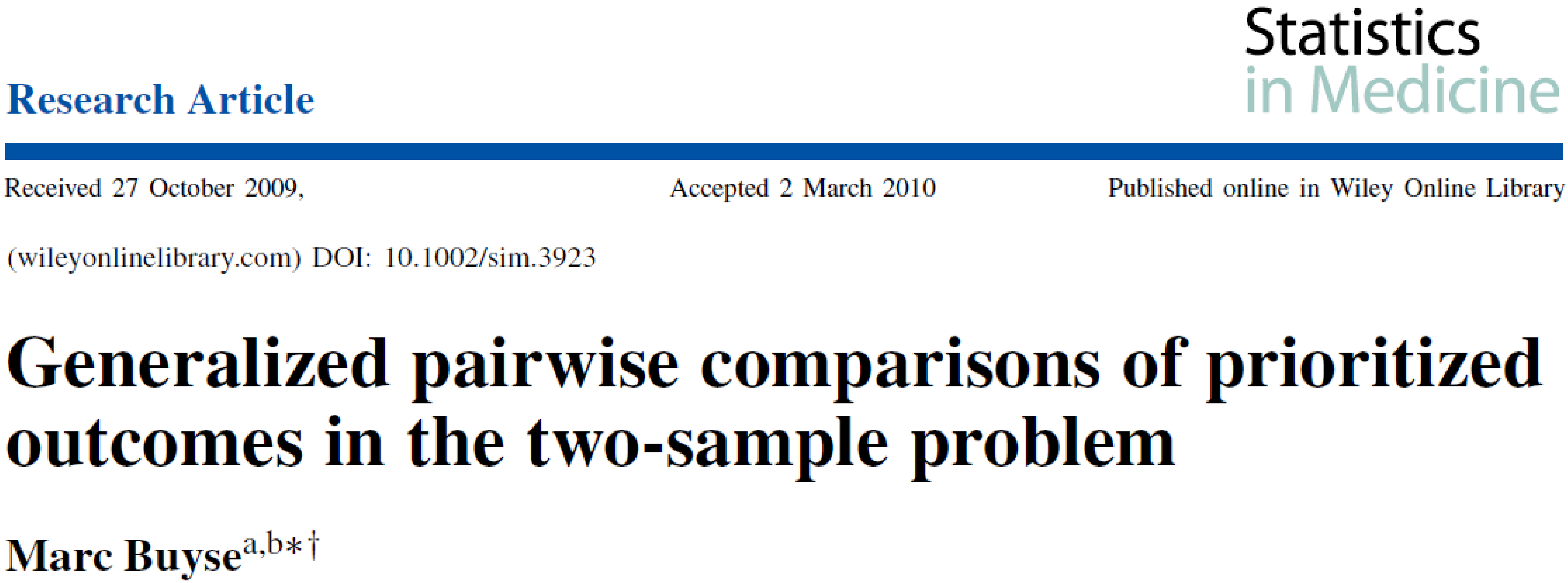
- Compare "average *A*" in each group
- Hope the results for *B, C, D*... agree with those from *A*
- Hope the results for *E, F, G*... are acceptable
- Make recommendations based on these "marginal" results
- Look for predictive factors that tailor recommendations to patient subsets (precision medicine)

Limitations

- General
  - A single 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

Limitations

- General
  - A single 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
- Specific to time-to-event endpoints
  - Non-proportional hazards
  - Composite endpoints consider time to first, not necessarily most relevant, event



Research Article

Statistics in Medicine

Received 27 October 2009, Accepted 2 March 2010 Published online in Wiley Online Library

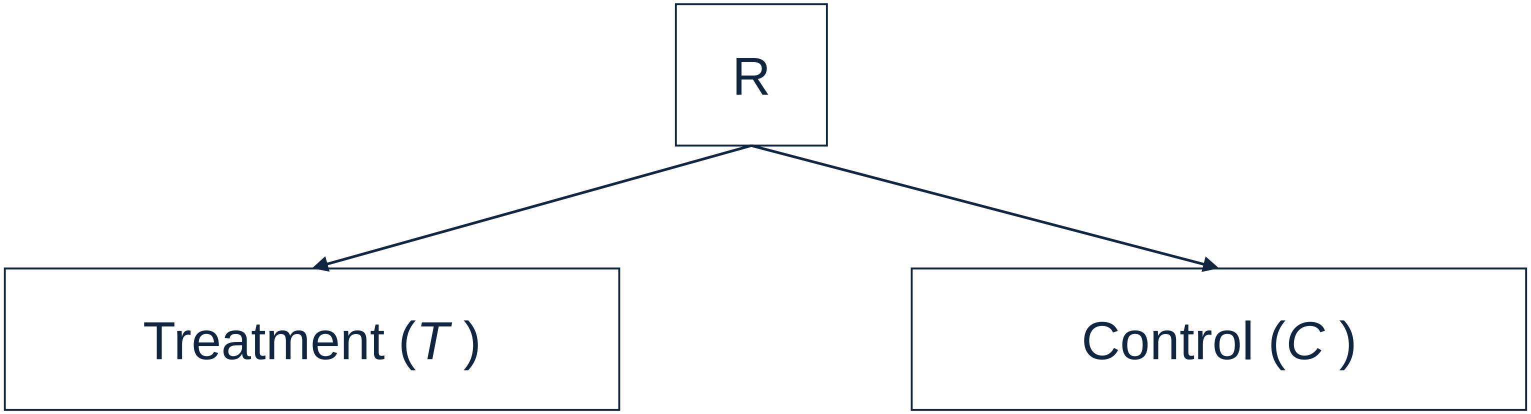
(wileyonlinelibrary.com) DOI: 10.1002/sim.3923

Generalized pairwise comparisons of prioritized outcomes in the two-sample problem

Marc Buyse[a,b*†]

Buyse, Stat Med 2010;29:3245

Randomized trial

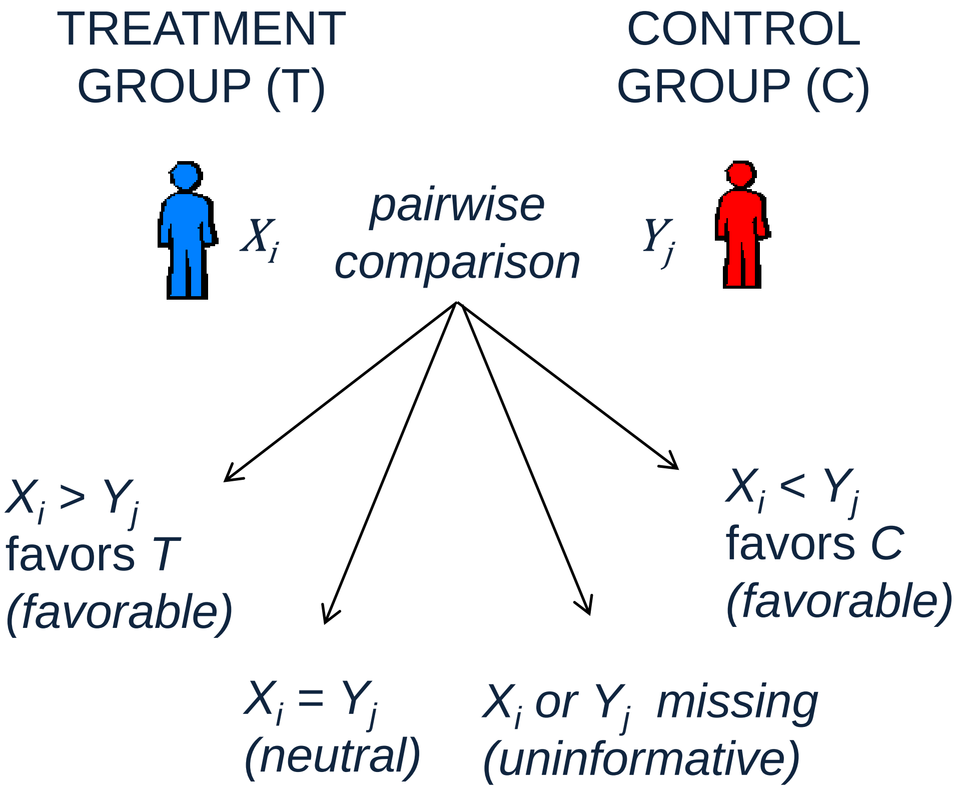


Let X_i be the outcome of i^{th} subject in T ($i = 1, \dots, n$)

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

Buyse, Stat Med 2010;29:3245

Pairwise comparisons



TREATMENT GROUP (T)

CONTROL GROUP (C)

X_i *pairwise comparison* Y_j

$X_i > Y_j$ favors *T* *(favorable)*

$X_i = Y_j$ *(neutral)*

X_i *or* Y_j *missing (uninformative)*

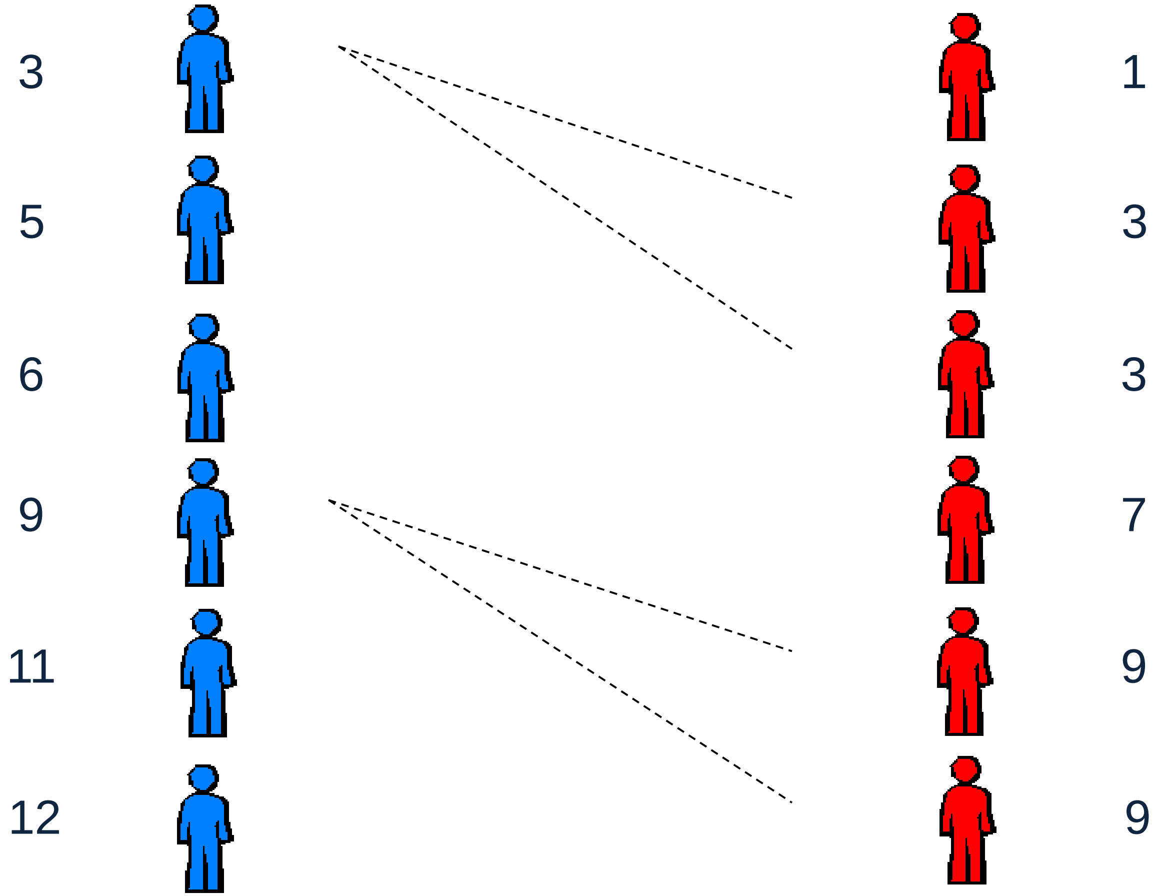
$X_i < Y_j$ favors *C* *(favorable)*

Buyse, Stat Med 2010;29:3245

Illustration of the method

TREATMENT GROUP (T)	CONTROL GROUP (C)
3	1
5	3
6	3
9	7
11	9
12	9

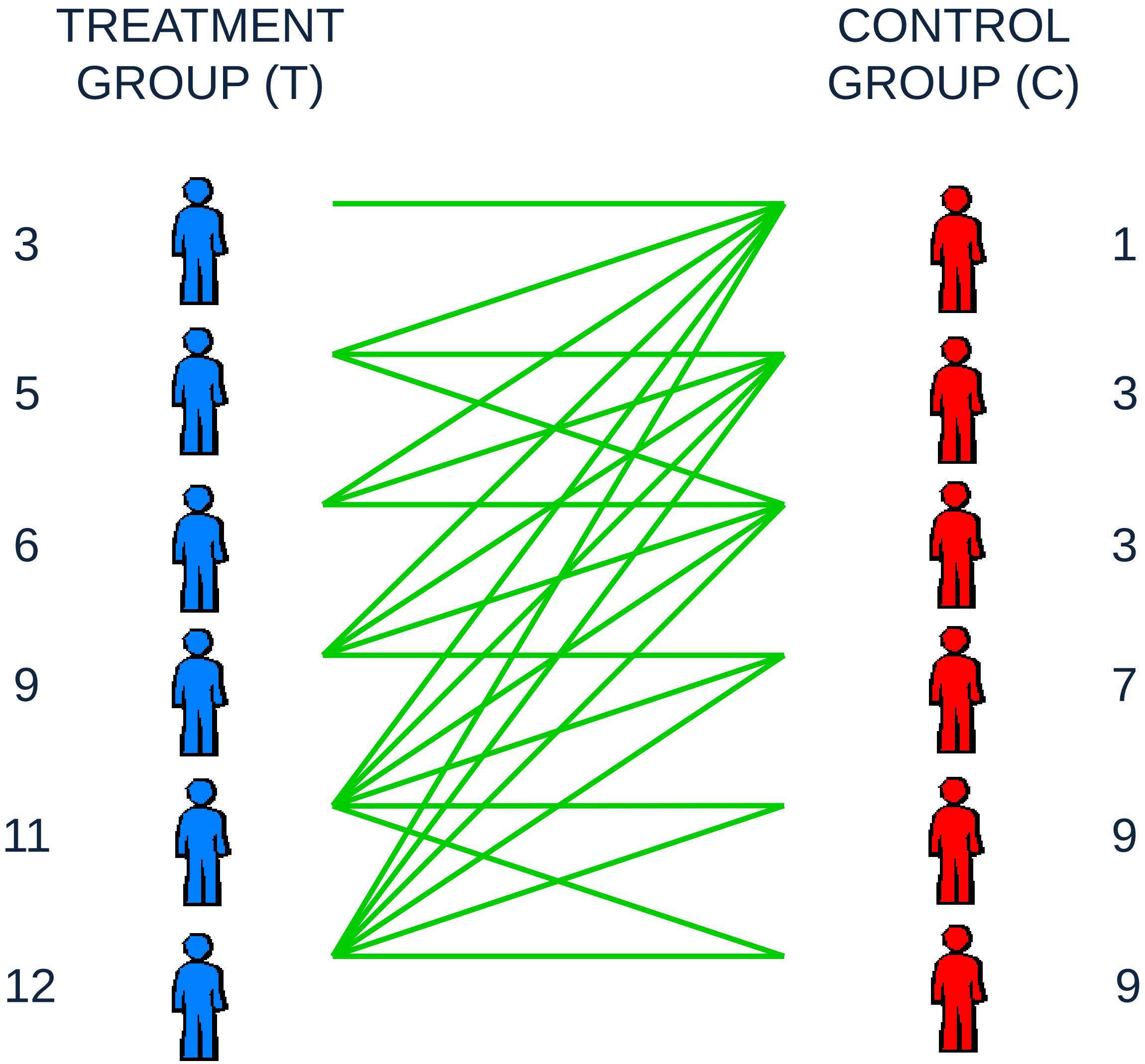
T and C tie



TREATMENT GROUP (T)	CONTROL GROUP (C)
3	1
5	3
6	3
9	7
11	9
12	9

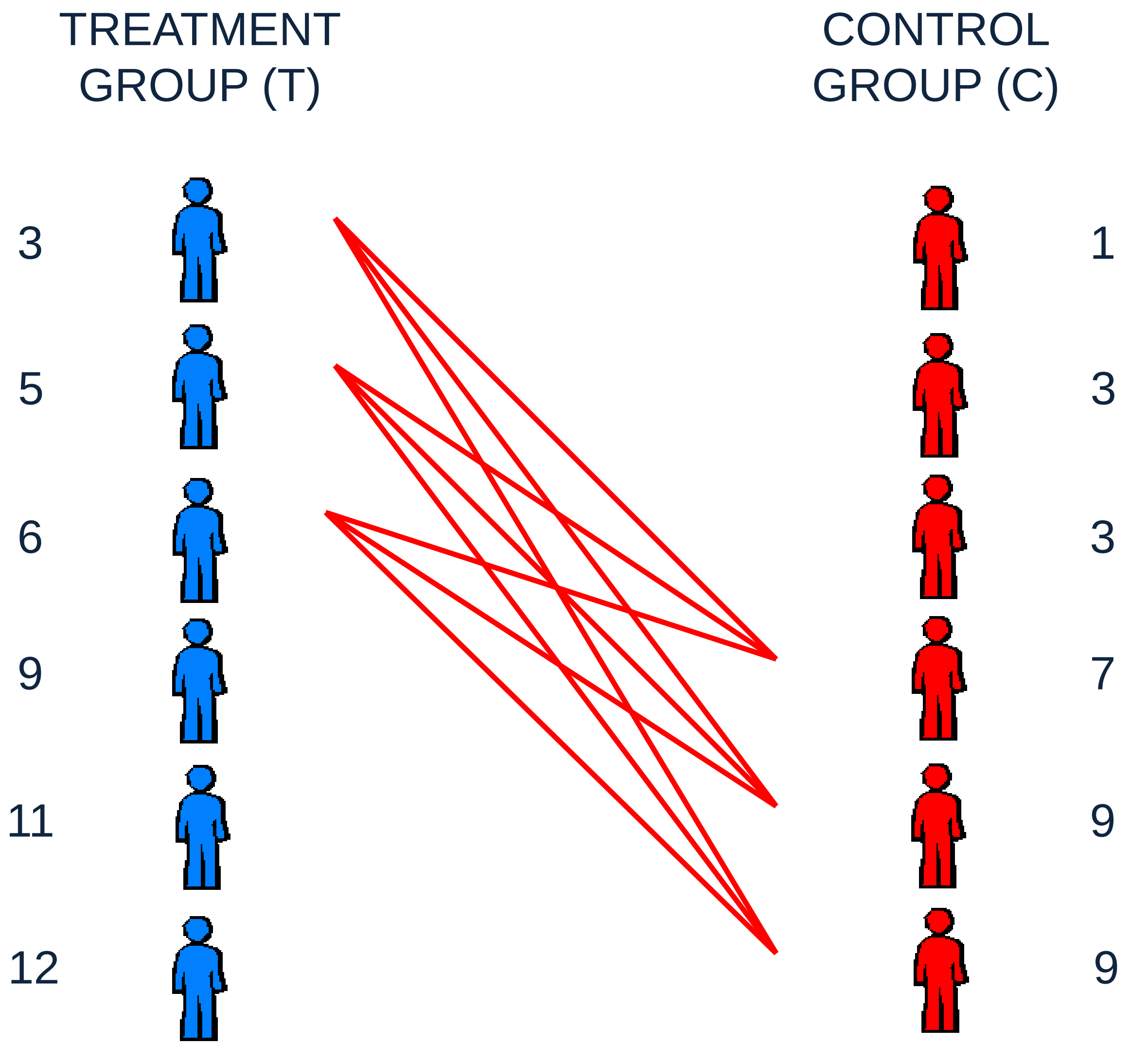
NEUTRAL PAIRS: 4

T is better



FAVORABLE PAIRS: 23

C is better



TREATMENT GROUP (T)

CONTROL GROUP (C)

3
5
6
9
11
12

1
3
3
7
9
9

UNFAVORABLE PAIRS: 9

Who wins?

Neutral	Favorable	Unfavorable	Net benefit
4 / 36 = 0.11	23 / 36 = 0.64	9 / 36 = 0.25	0.64 – 0.25 = 0.39

The probability of a patient having a better outcome

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

The net benefit (or « proportion in favor ») of treatment is 0.39

The net treatment benefit (Δ)

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

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

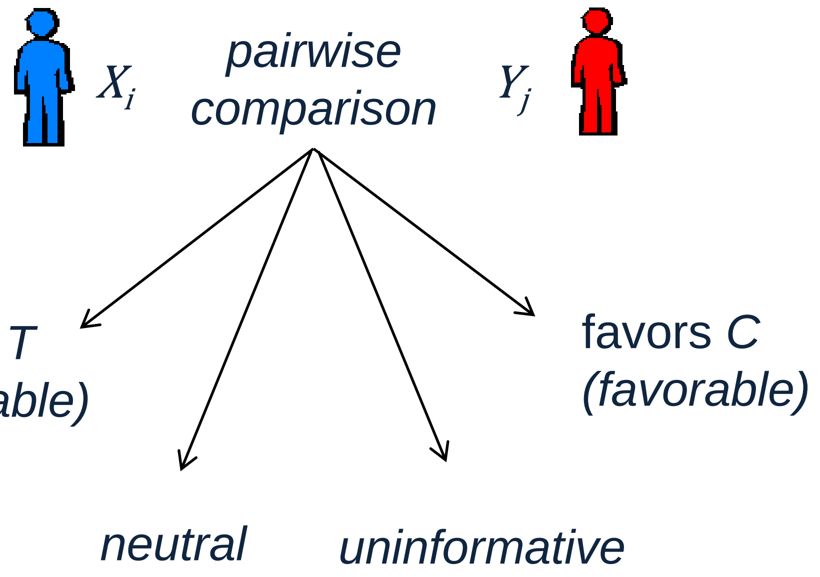
U is the difference between the proportion of favorable pairs and the proportion of unfavorable pairs. It is the « net treatment benefit », denoted Δ.

This measure is analogous to Pocock's « win ratio » (Δ is the « win difference »).

Pocock et al. Eur Heart J 2012; 33: 176

Generalizing the test

Now let X_i and Y_j be observed outcomes for any outcome measure (continuous, time-to-event, binary, categorical, …)



X_i *pairwise comparison* Y_j

favors *T* *(favorable)*

neutral

uninformative

favors *C* *(favorable)*

Generalized pairwise comparisons (GPC)

Buyse, Stat Med 2010;29:3245

Binary outcome measure

Pairwise comparison	Pair is
$X_i = 1, Y_j = 0$	favorable
$X_i = 1, Y_j = 1$ or $X_i = 0, Y_j = 0$	neutral
$X_i = 0, Y_j = 1$	unfavorable
X_i or Y_j missing	uninformative

GPC test is equivalent to χ^2 test

Buyse, Stat Med 2010;29:3245

Continuous outcome measure

Pairwise comparison	Pair is
$X_i - Y_j > \tau$	favorable
$\| X_i - Y_j \| \leq \tau$	neutral
$X_i - Y_j < -\tau$	unfavorable
X_i or Y_j missing	uninformative

$\tau = 0$ is Wilcoxon test

τ can be chosen to reflect clinical relevance

Buyse, Stat Med 2010;29:3245

Time-to-event outcome measure

Pairwise comparison	Pair is
$X_i - Y_j > \tau$	favorable
$\lvert X_i - Y_j \rvert \leq \tau$	neutral
$X_i - Y_j < -\tau$	unfavorable
X_i or Y_j missing	uninformative

$\tau = 0$ is Gehan test (accounting for censoring of *X* or *Y*)

τ can be chosen to reflect clinical relevance

Buyse, Stat Med 2010;29:3245

Prioritizing outcomes

Now let <X_i and X'_i> and <Y_j and Y'_j> be observed results for two outcome measures, X and Y being prioritized over X' and Y'

X_i, X'_i *pairwise comparison* Y_j, Y'_j

X_i / Y_j	X_i' / Y_j'	Pair is
Favorable		favorable
unfavorable		unfavorable
neutral or ?	favorable	favorable
neutral or ?	unfavorable	unfavorable
neutral or ?	neutral	neutral
?	?	?

GPC for prioritized outcomes

Buyse, Stat Med 2010;29:3245

Prioritizing through the use of thresholds of clinical relevance

Survival difference > 12 months	Survival difference ≤ 12 months	Pair is
favorable		favorable
unfavorable		unfavorable
neutral or ?	favorable	favorable
neutral or ?	unfavorable	unfavorable
neutral or ?	neutral	neutral
?	?	?

Buyse, Stat Med 2010;29:3245

Prioritizing through the use of different outcomes

Survival	Serious toxicity (e.g. CTCAE grade 3/4)	Pair is
favorable		favorable
unfavorable		unfavorable
neutral or ?	favorable	favorable
neutral or ?	unfavorable	unfavorable
neutral or ?	neutral	neutral
?	?	?

Buyse, Stat Med 2010;29:3245

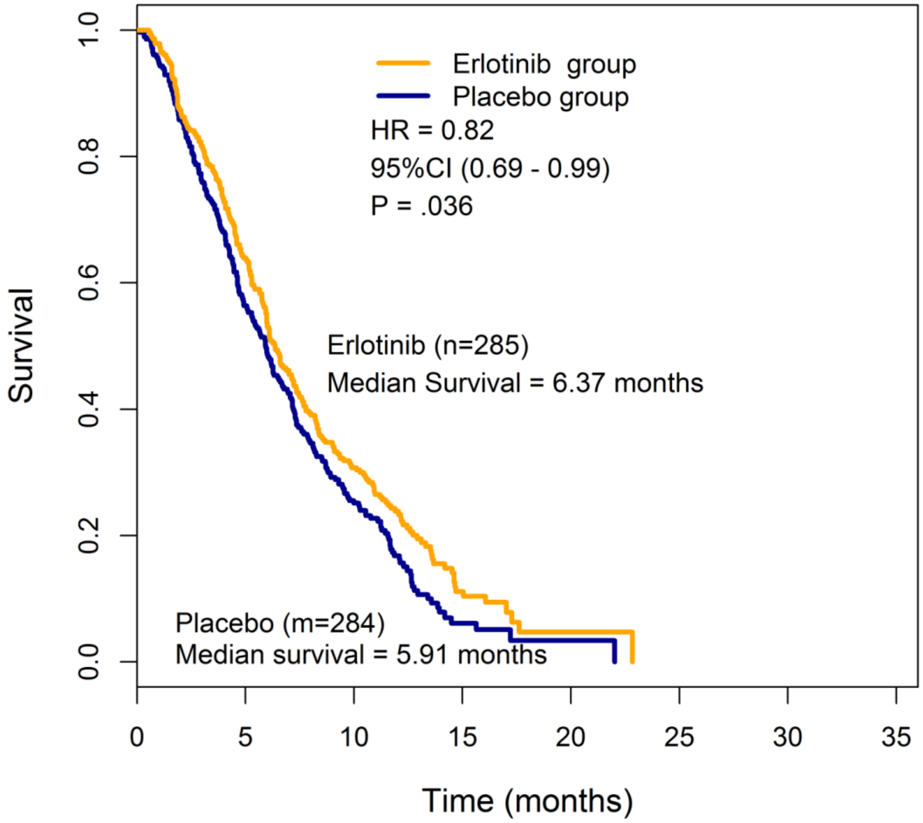
Some examples:

Analyzing benefit/risk in advanced pancreatic cancer

- Re-analysis of individual patient data from three randomized trials:
- Gemcitabine ± erlotinib [1]
- Gemcitabine vs. FOLFIRINOX [2]
- Gemcitabine ± nab-paclitaxel [3]

1. *Moore et al, J Clin Oncol 2007; 25:1960*
2. *Von Hoff et al, N Engl J Med 2013;369:1691*
3. *Conroy et al, N Engl J Med 2011;364:1817*

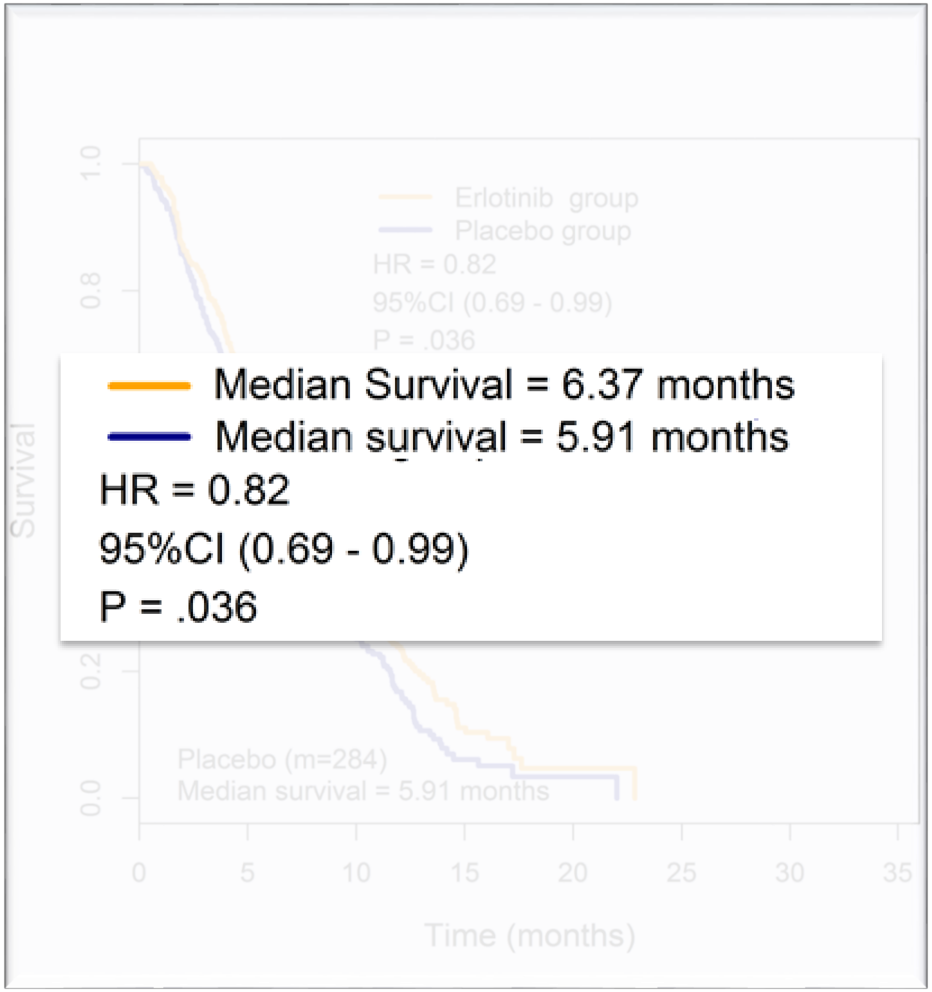
Gemcitabine ± erlotinib



Worst grade related AE	Erlotinib (n=282)	Placebo (n=280)
Grade 1	48 (17%)	69 (24.6%)
Grade 2	118 (41.8%)	89 (31.8%)
Grade 3	72 (25.5%)	47 (16.8%)
Grade 4	11 (3.9%)	6 (2.1%)
Grade 5	4 (1.4%)	3 (1.1%)

Moore et al, J Clin Oncol 2007; 25:1960

Benefit and harm



Worst grade related AE	Erlotinib (n=282)	Placebo (n=280)
Grade 1	48 (17%)	69 (24.6%)
Grade 2	118 (41.8%)	89 (31.8%)
Grade 3	**29%**	**19%**
Grade 4		
Grade 5	4 (1.4%)	3 (1.1%)

Moore et al, J Clin Oncol 2007; 25:1960

Prioritized outcomes: OS and worst toxicity

OS difference > 2 months	Worst toxicity (of any type)	Pair is
favorable	-	favorable
unfavorable	-	unfavorable
neutral or ?	favorable	favorable
neutral or ?	unfavorable	unfavorable
neutral or ?	neutral	neutral
?	?	?

Prioritized outcomes:
OS and worst toxicity

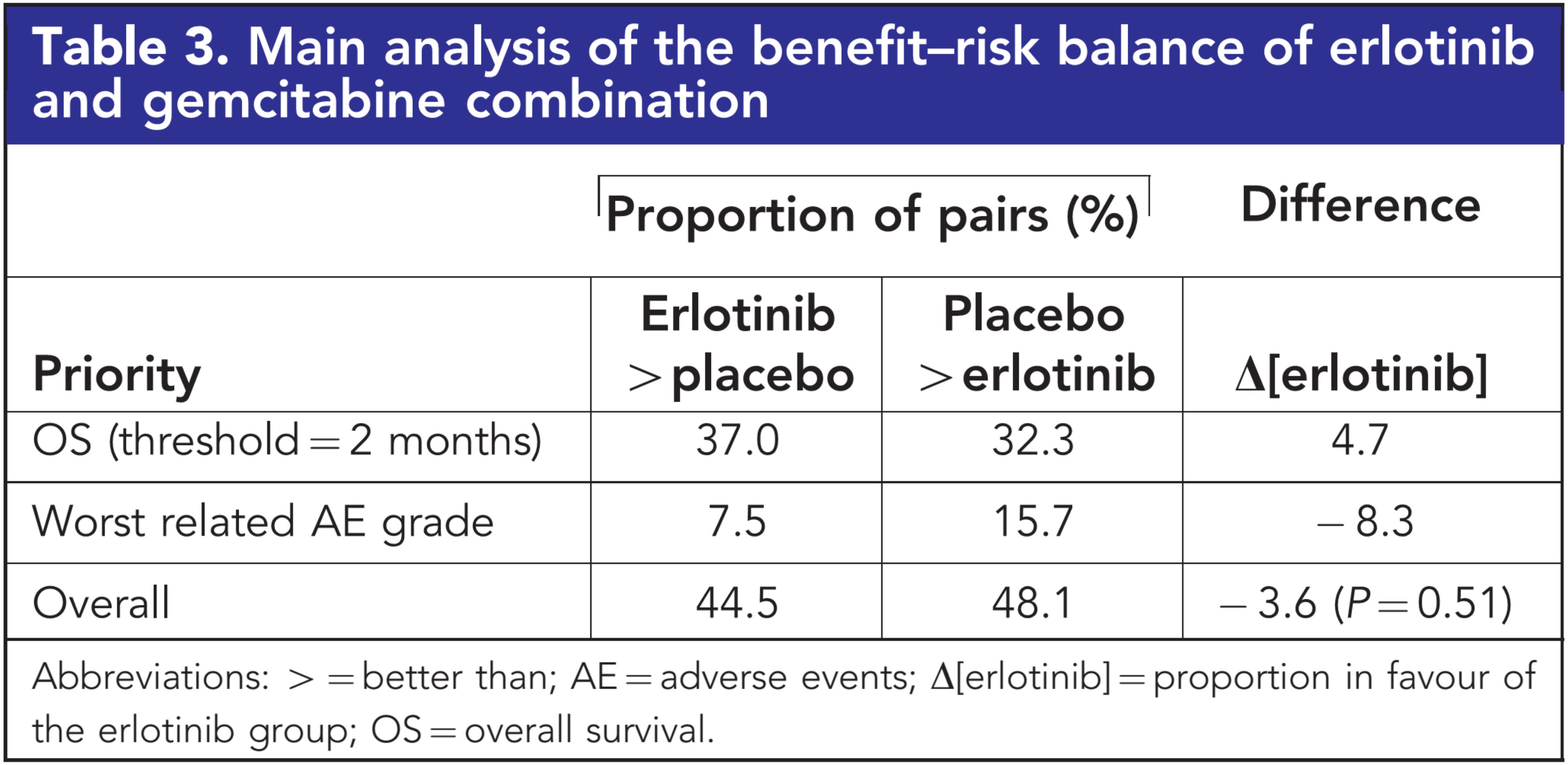


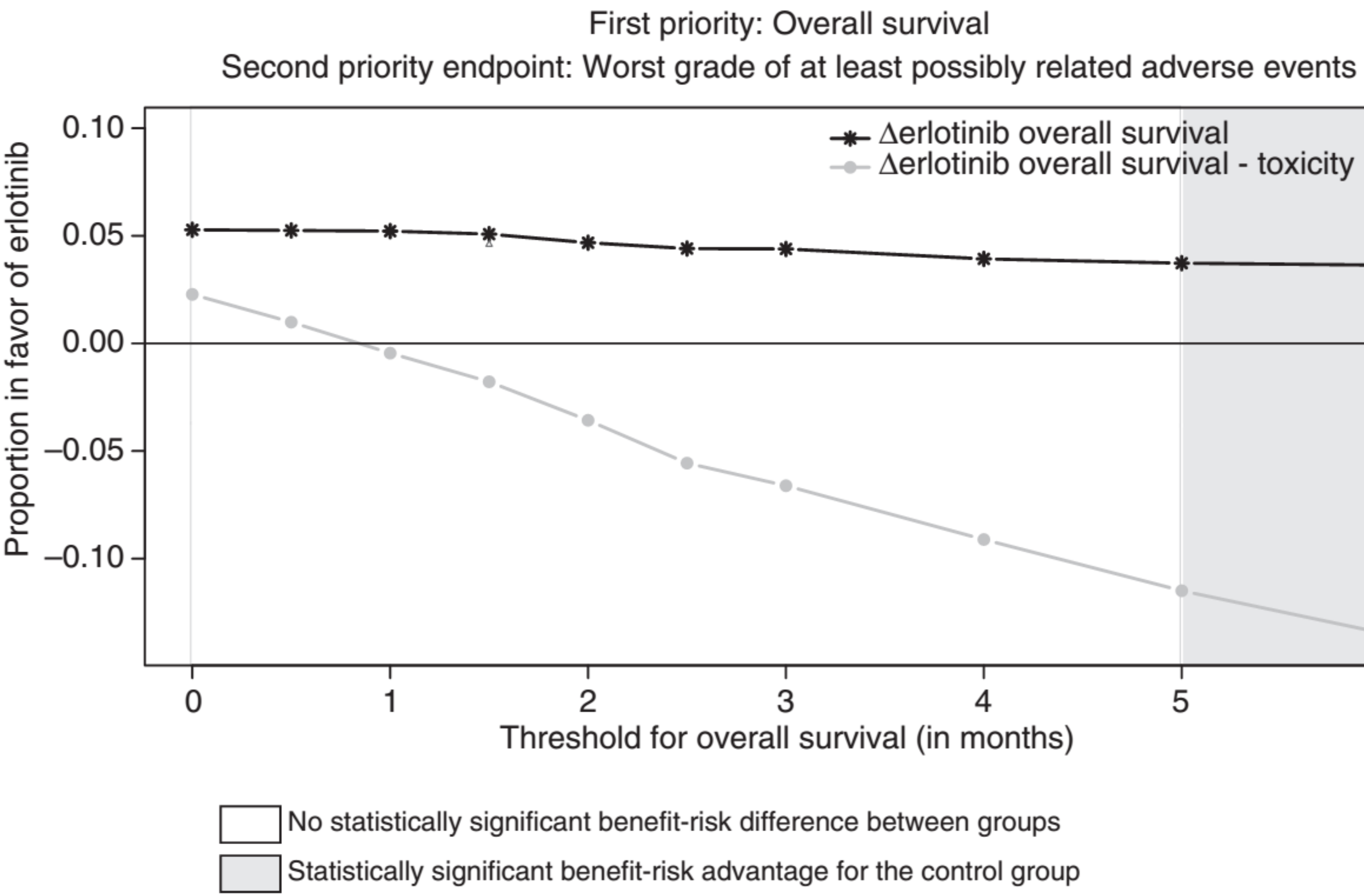
Table 3. Main analysis of the benefit–risk balance of erlotinib and gemcitabine combination

	Proportion of pairs (%)		Difference
Priority	**Erlotinib >placebo**	**Placebo >erlotinib**	**Δ[erlotinib]**
OS (threshold = 2 months)	37.0	32.3	4.7
Worst related AE grade	7.5	15.7	– 8.3
Overall	44.5	48.1	– 3.6 (*P* = 0.51)

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

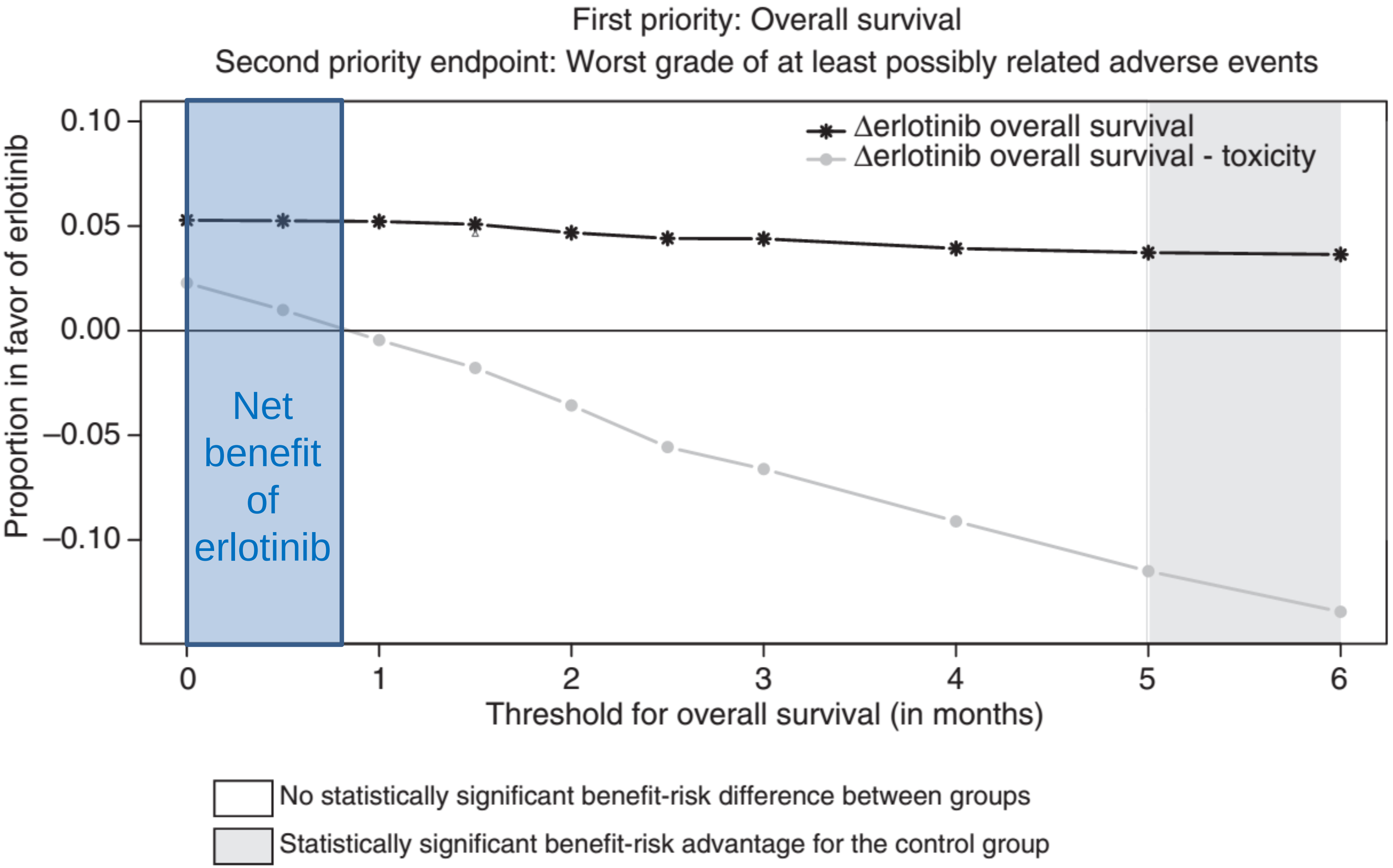
Péron et al, Br J Cancer 2015;112:971

Prioritized outcomes:
OS and worst toxicity



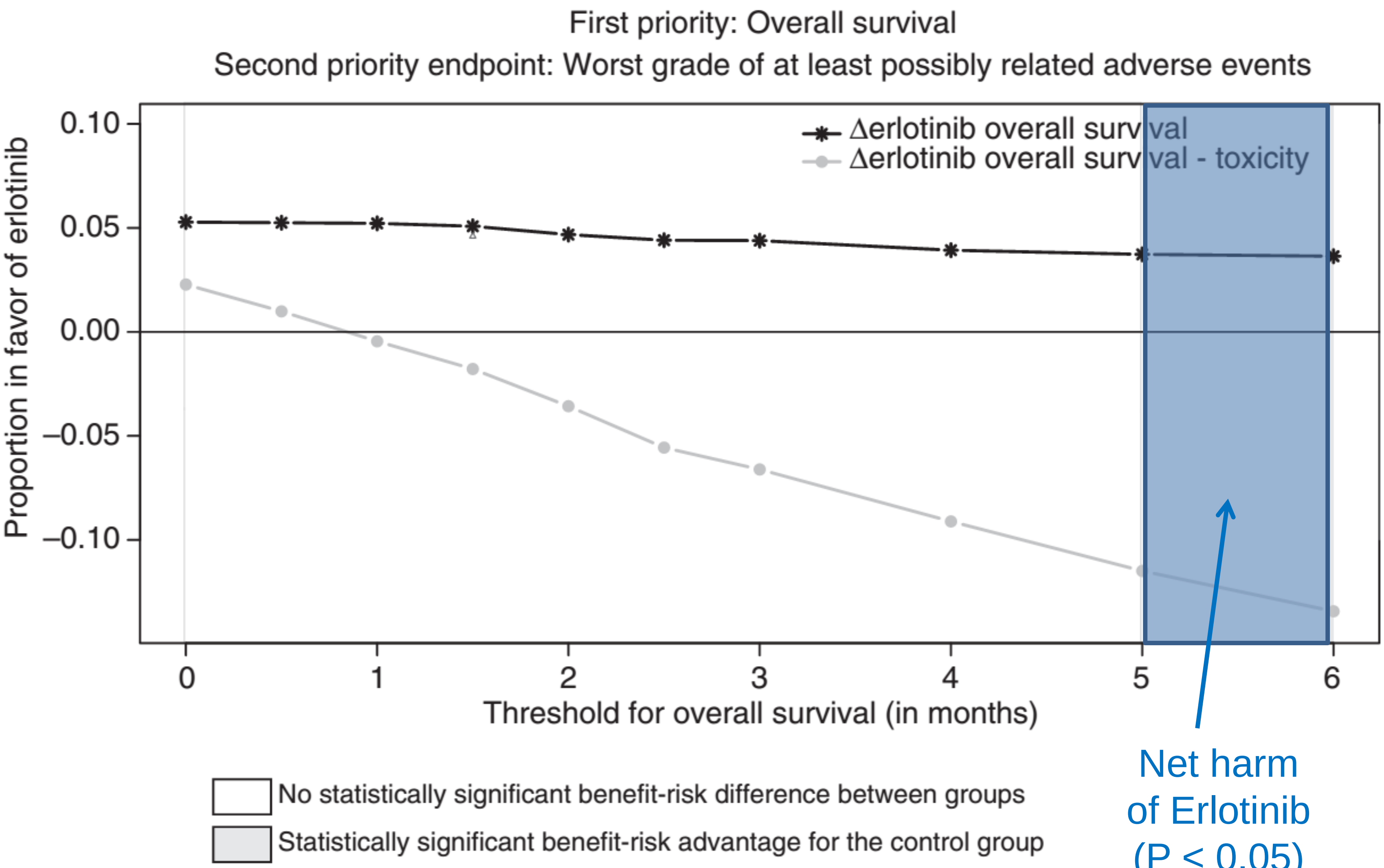
Péron et al, Br J Cancer 2015;112:971

Prioritized outcomes: OS and worst toxicity



Péron et al, Br J Cancer 2015;112:971

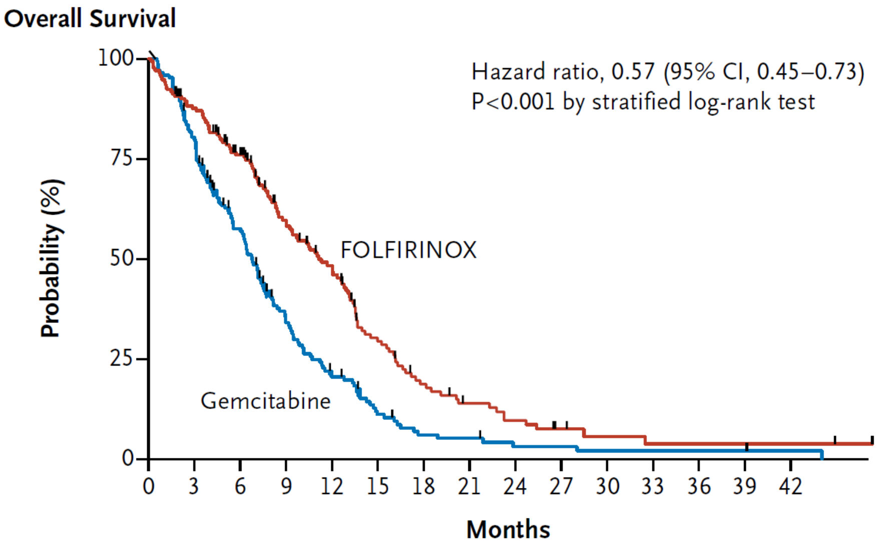
Prioritized outcomes:
OS and worst toxicity



Net harm of Erlotinib ($P < 0.05$)

Péron et al, Br J Cancer 2015;112:971

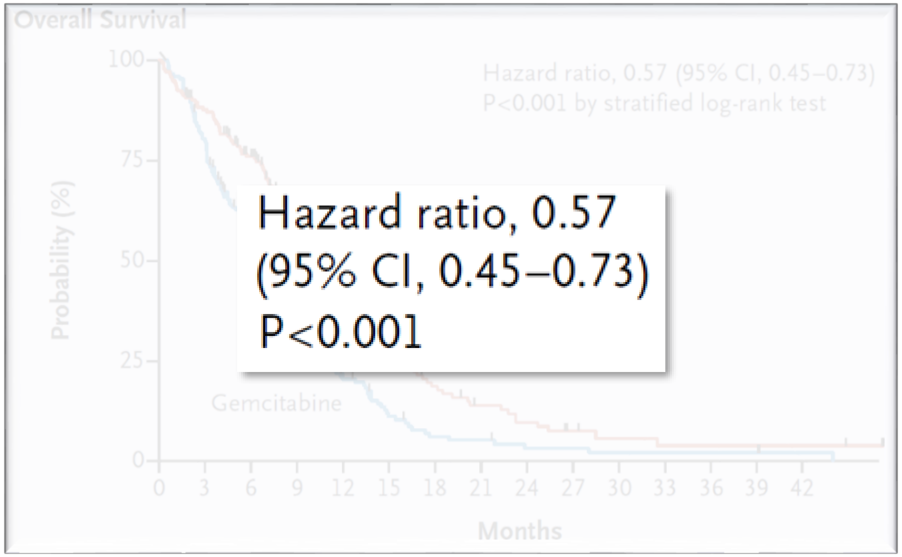
Gemcitabine vs FOLFIRINOX



Worst grade AE	FOLFIRINOX (n=171)	Gemcitabine (n=171)
Grade 0	6 (3.5%)	2 (1.2%)
Grade 1	7 (4.1%)	5 (2.9%)
Grade 2	40 (23.4%)	62 (36.3%)
Grade 3	81 (47.7%)	67 (39.2%)
Grade 4	36 (21.1%)	34 (19.9%)
Grade 5	1 (0.6%)	1 (0.6%)

Conroy et al, N Engl J Med 2011;364:1817

Benefit and harm

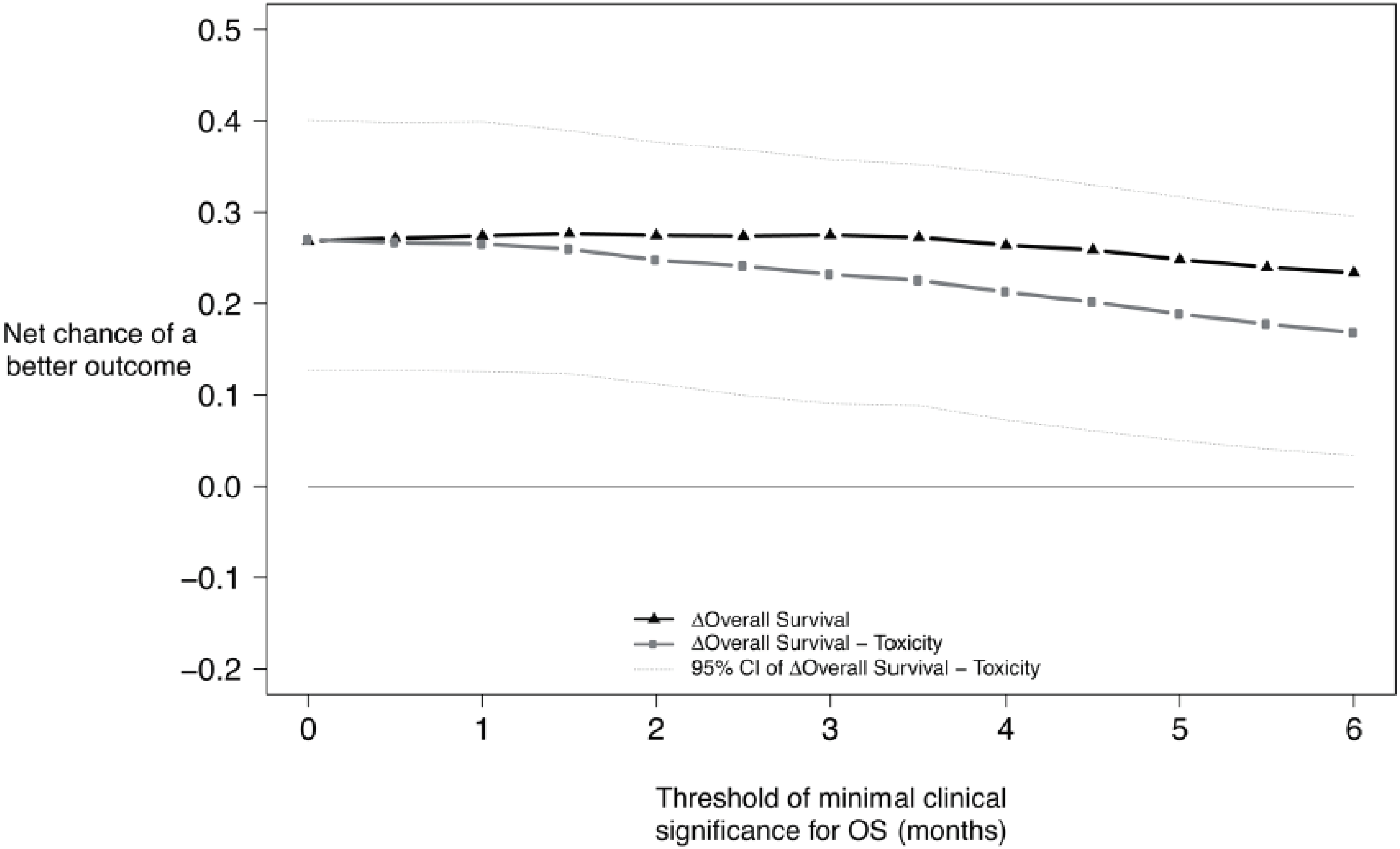


Hazard ratio, 0.57
(95% CI, 0.45–0.73)
P<0.001

Worst grade AE	FOLFIRINOX (n=171)	Gemcitabine (n=171)
Grade 0	6 (3.5%)	2 (1.2%)
Grade 1	7 (4.1%)	5 (2.9%)
Grade 2	40 (23.4%)	62 (36.3%)
Grade 3	**68%**	**59%**
Grade 4		
Grade 5	1 (0.6%)	1 (0.6%)

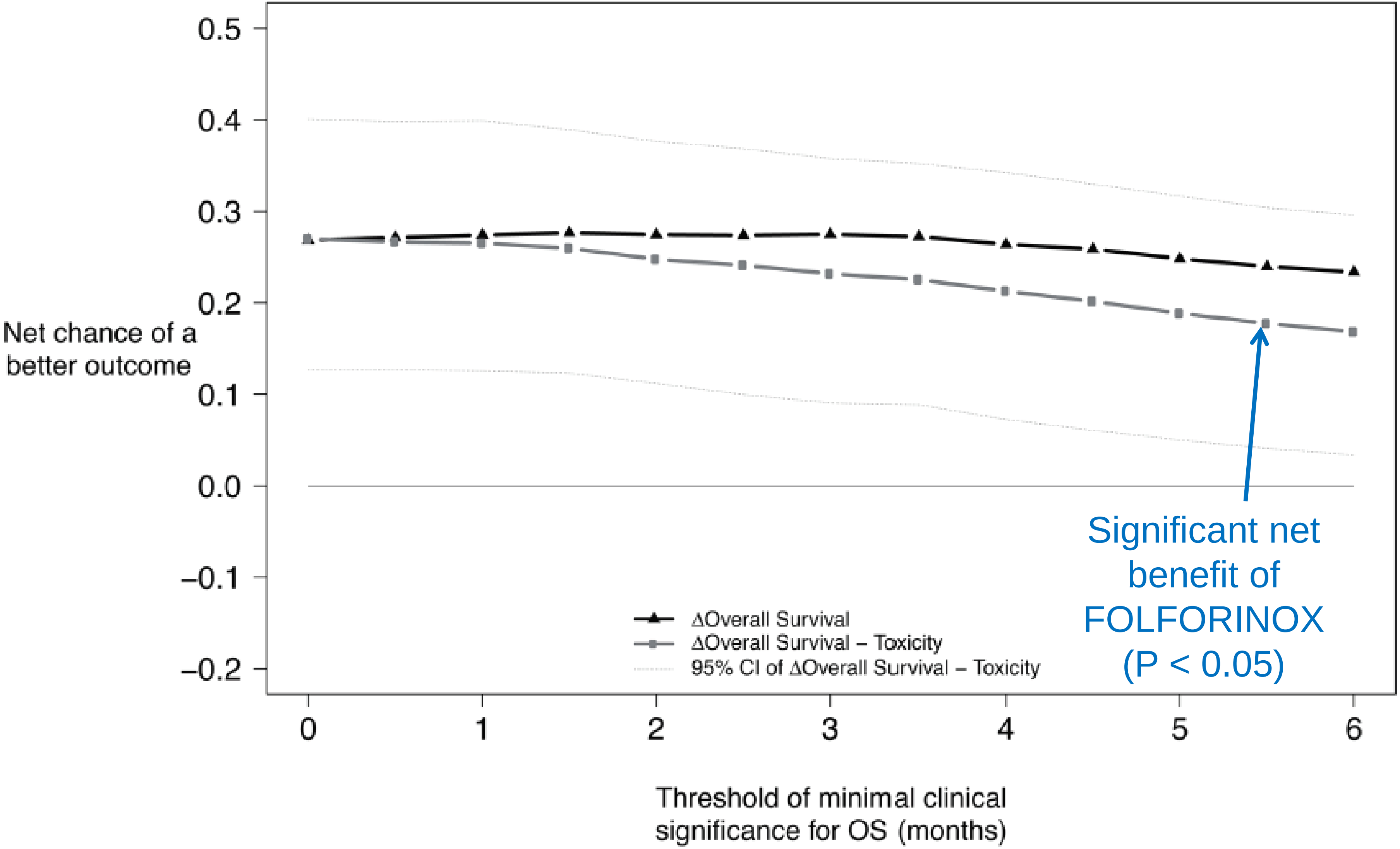
Conroy et al, N Engl J Med 2011;364:1817

Prioritized outcomes: OS and worst toxicity



Péron et al, Oncotarget 2017;7:82953

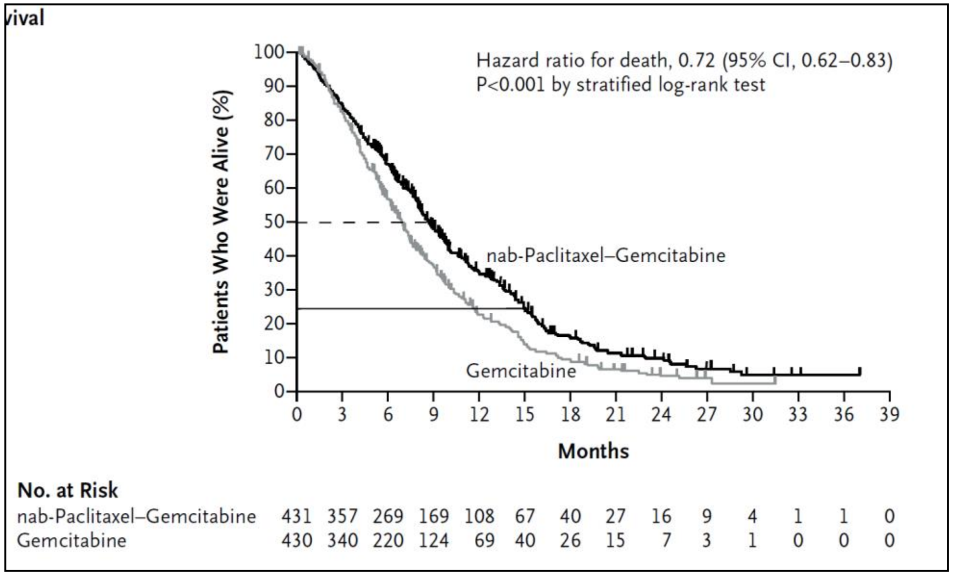
Prioritized outcomes: OS and worst toxicity



Significant net benefit of FOLFORINOX (P < 0.05)

Péron et al, Oncotarget 2016;7:82953

Gemcitabine ± nab-paclitaxel

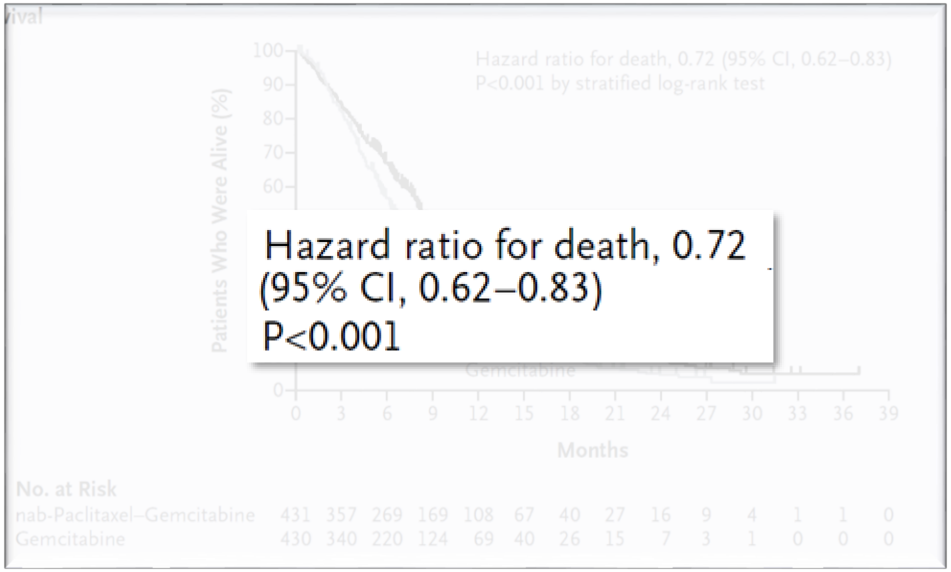


No. at Risk														
nab-Paclitaxel–Gemcitabine	431	357	269	169	108	67	40	27	16	9	4	1	1	0
Gemcitabine	430	340	220	124	69	40	26	15	7	3	1	0	0	0

Worst grade related AE	Monotherapy (n=430)	Combination (n=431)
Grade 0	96 (22.3%)	37 (8.6%)
Grade 1	96 (22.3%)	34 (7.9%)
Grade 2	136 (31.6%)	123 (28.5%)
Grade 3	88 (20.5%)	215 (49.9%)
Grade 4	9 (2.1%)	16 (3.7%)
Grade 5	5 (1.2%)	6 (1.4%)

Von Hoff et al, N Engl J Med 2013;369:1691

Benefit and harm

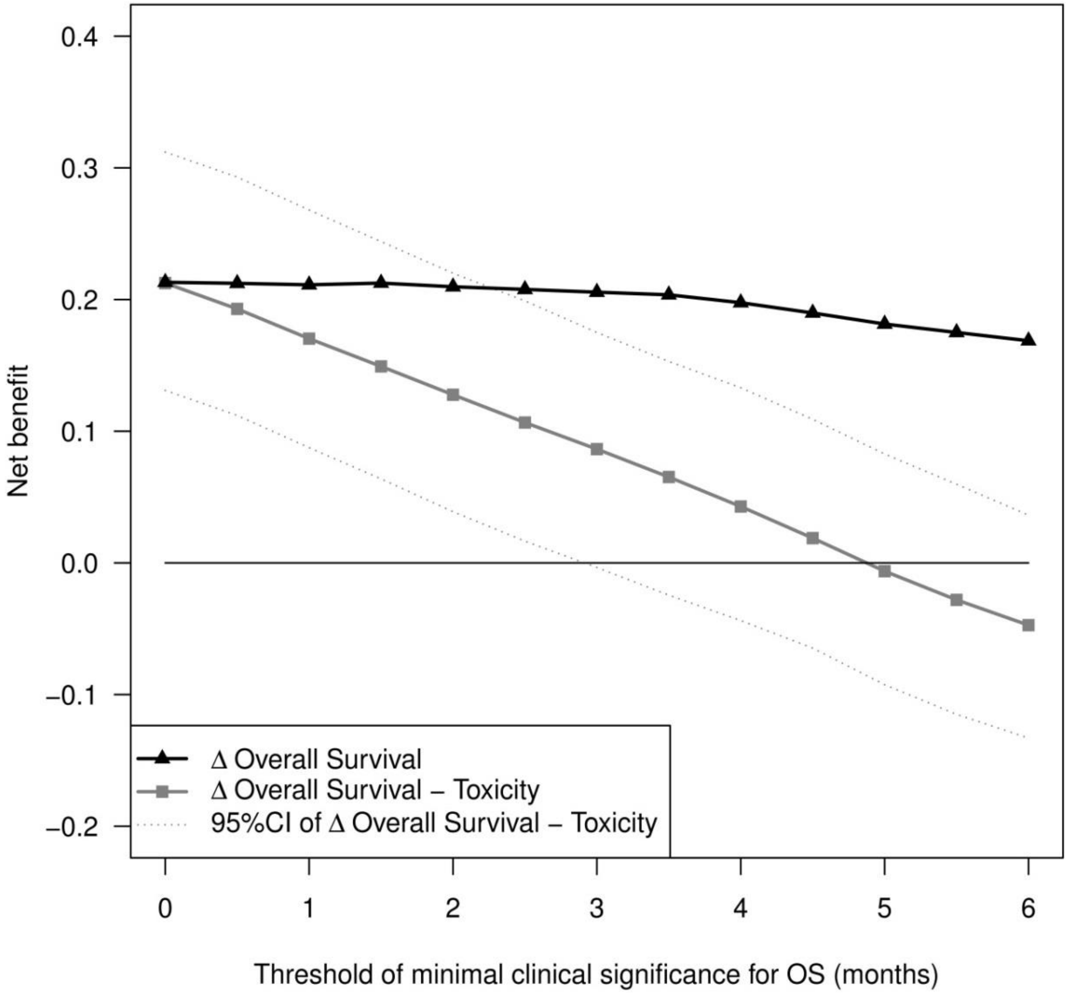


Hazard ratio for death, 0.72
(95% CI, 0.62–0.83)
P<0.001

Worst grade related AE	Monotherapy (n=430)	Combination (n=431)
Grade 0	96 (22.3%)	37 (8.6%)
Grade 1	96 (22.3%)	34 (7.9%)
Grade 2	136 (31.6%)	123 (28.5%)
Grade 3	**23%**	**54%**
Grade 4		
Grade 5	5 (1.2%)	6 (1.4%)

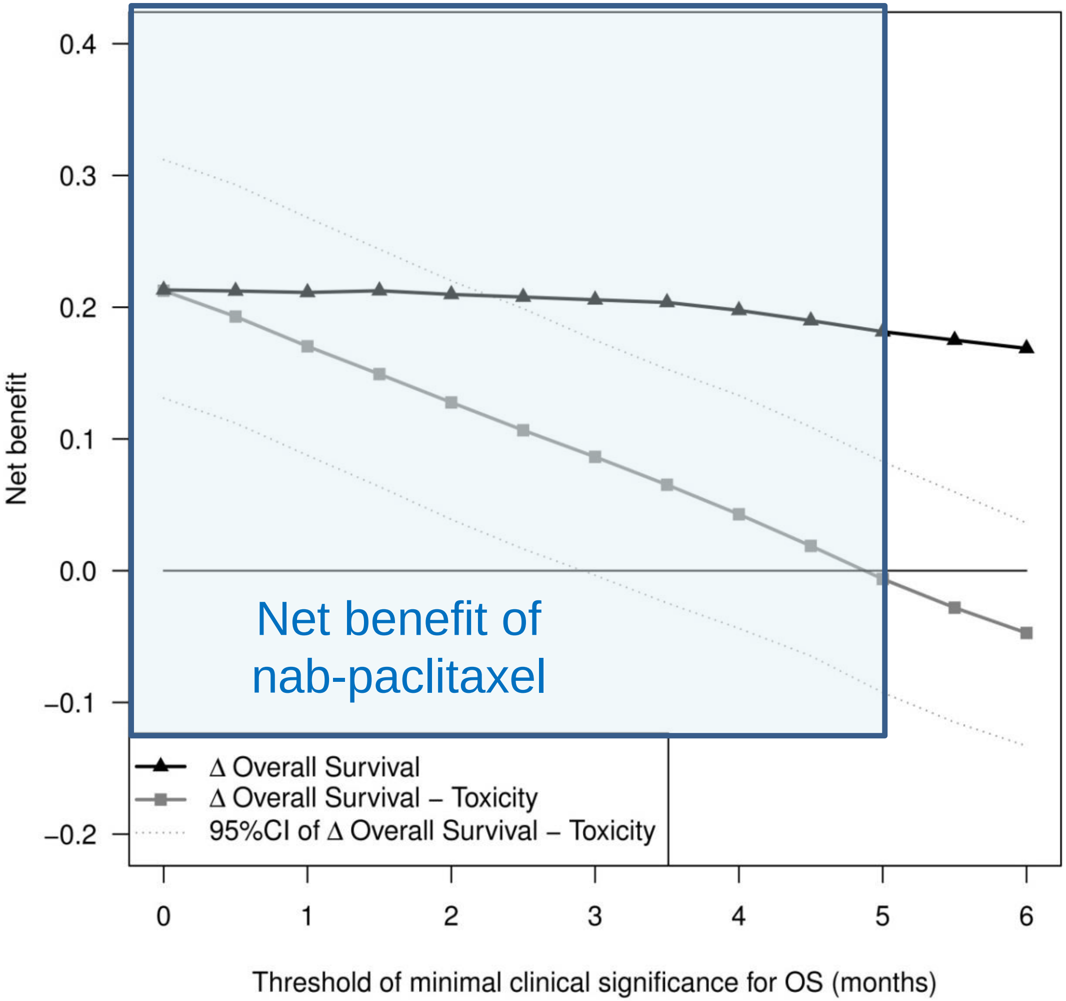
Von Hoff et al, N Engl J Med 2013;369:1691

Prioritized outcomes: OS and worst toxicity



Péron et al,

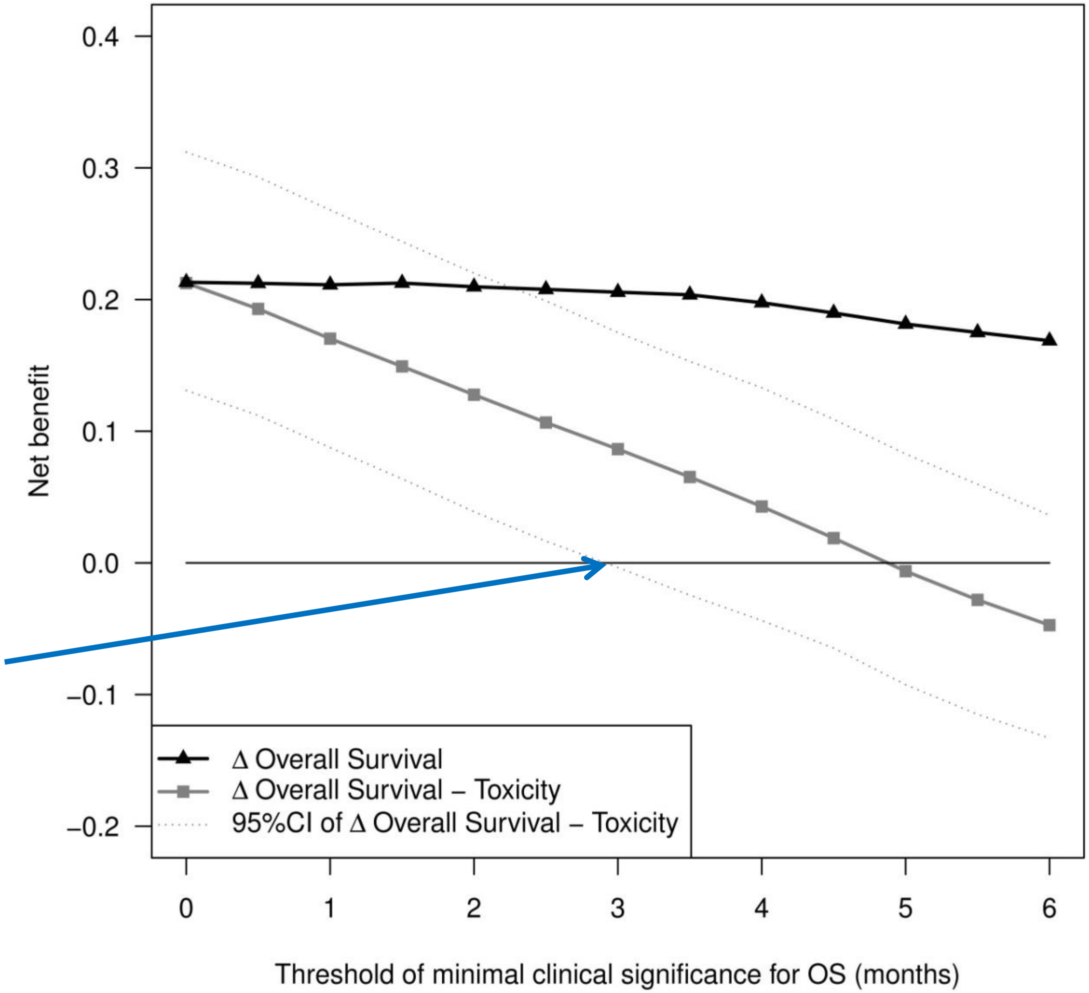
Prioritized outcomes: OS and worst toxicity



Net benefit of nab-paclitaxel

Péron et al,

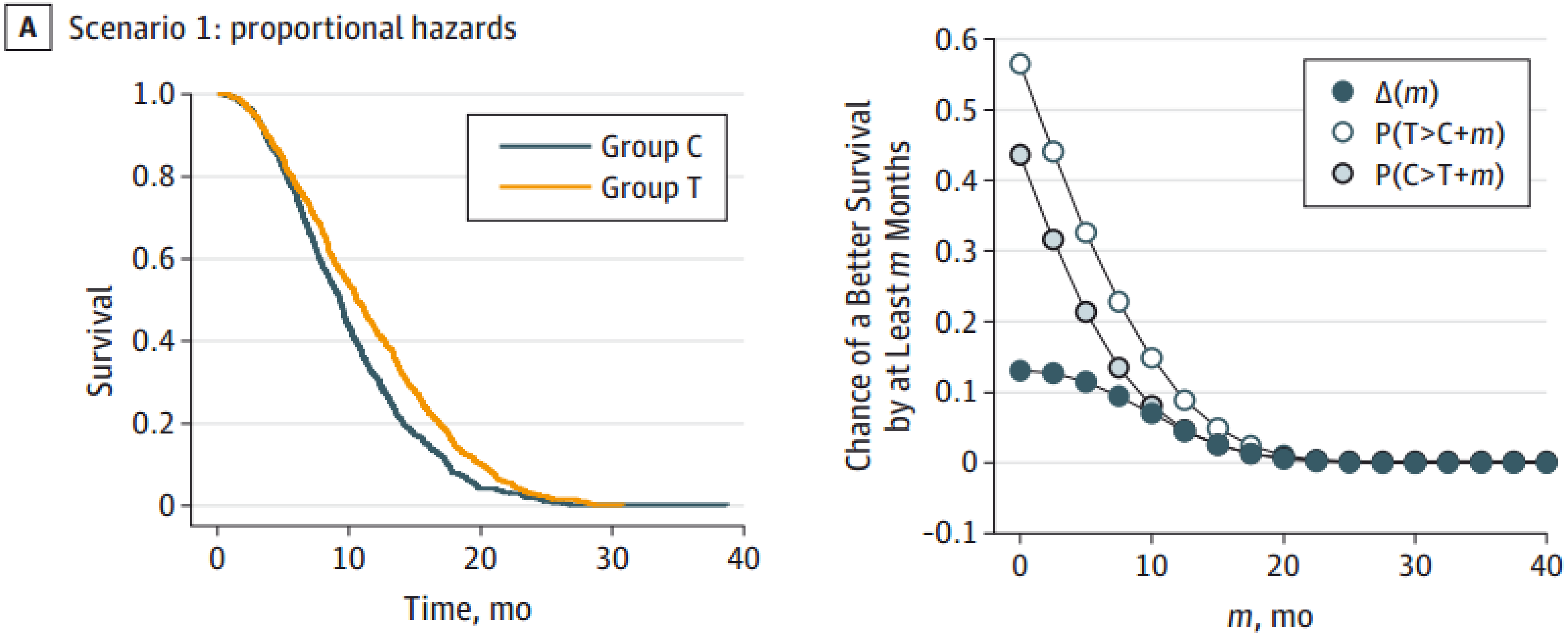
Prioritized outcomes: OS and worst toxicity



Significant net benefit of nab-paclitaxel (P < 0.05)

Péron et al,

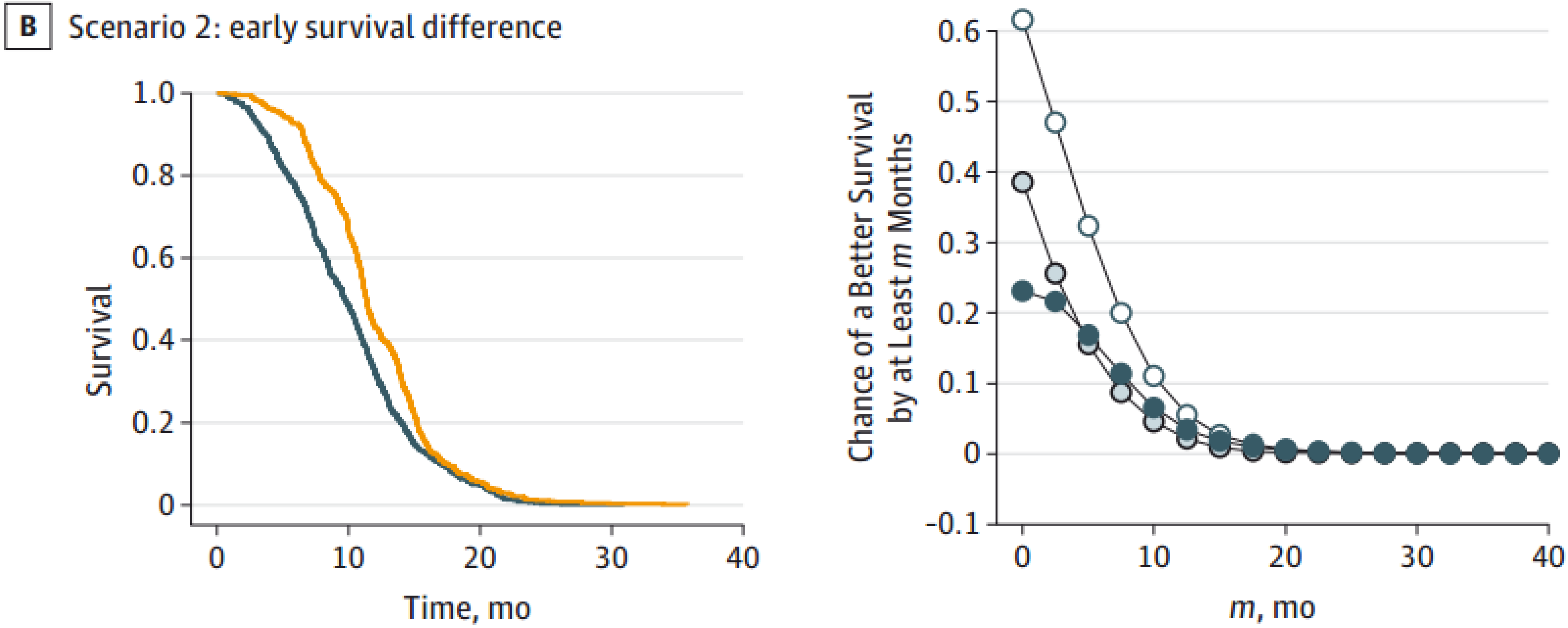
Net benefit – proportional hazards



No. at risk				
Group C	600	263	26	1
Group T	600	324	62	1

Péron et al, JAMA Oncol 2016;2:901

Net benefit – early difference



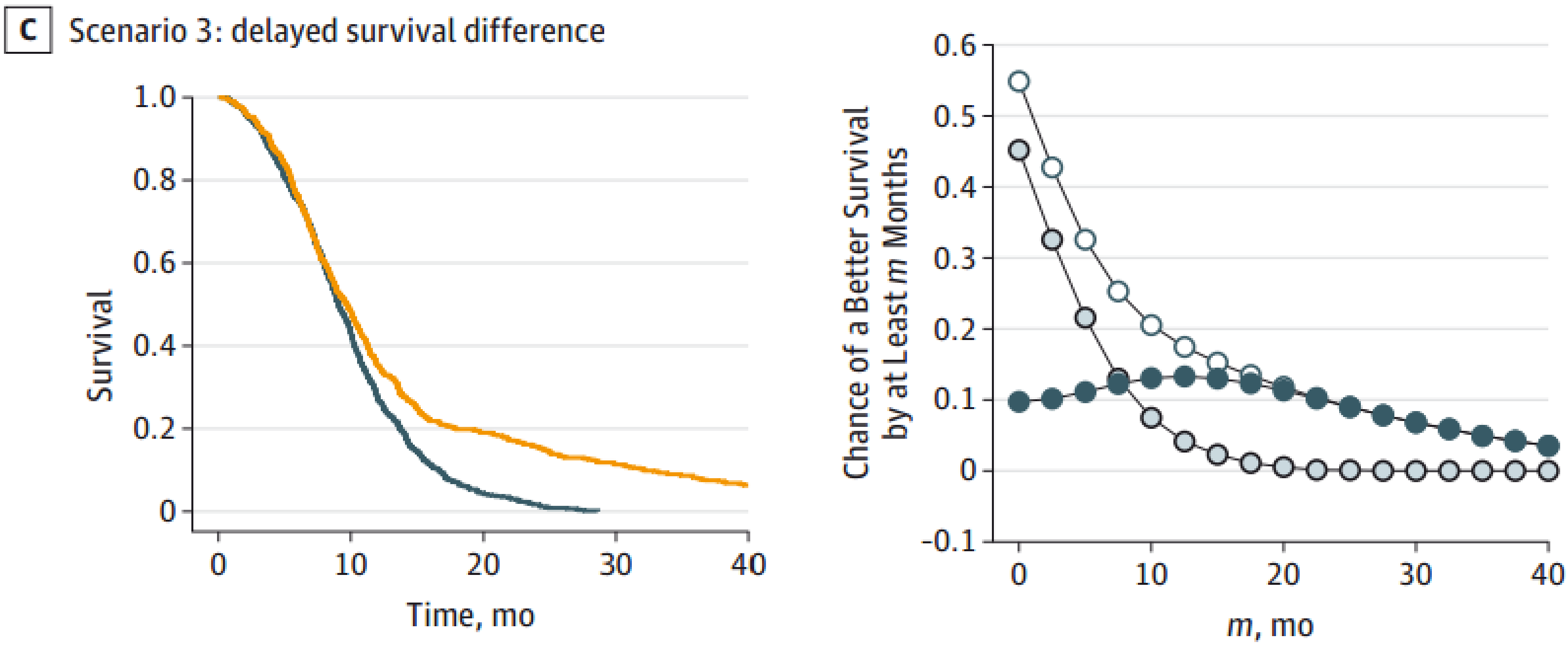
No. at risk

Group C	600	291	30	1
Group T	600	402	35	1

Example: cytotoxics

Péron et al, JAMA Oncol 2016;2:901

Net benefit – delayed difference

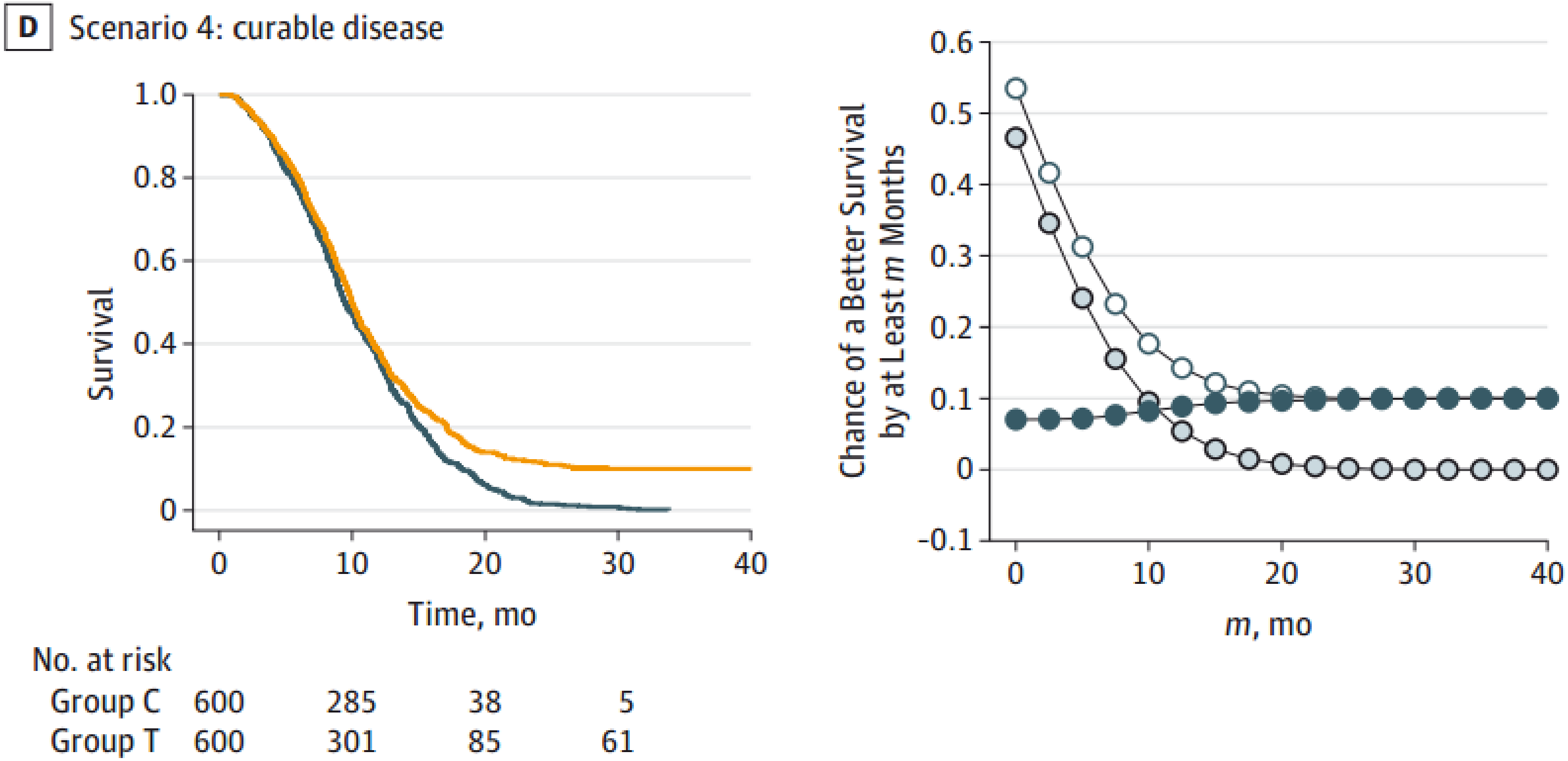


No. at risk					
Group C	600	262	27	0	0
Group T	600	292	115	69	39

Example: immunotherapy for advanced solid tumors

Péron et al, JAMA Oncol 2016;2:901

Net benefit – cure rate

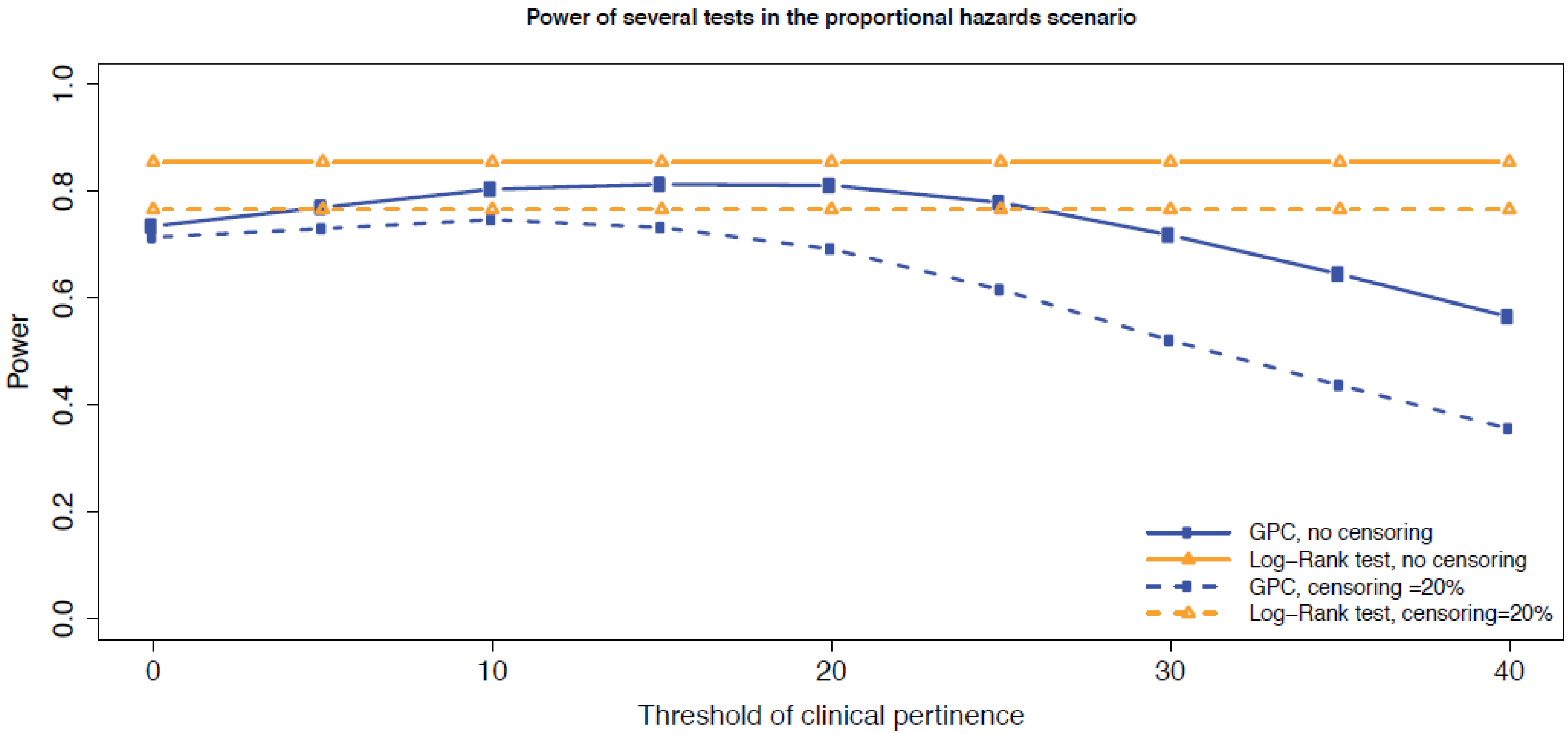


No. at risk				
Group C	600	285	38	5
Group T	600	301	85	61

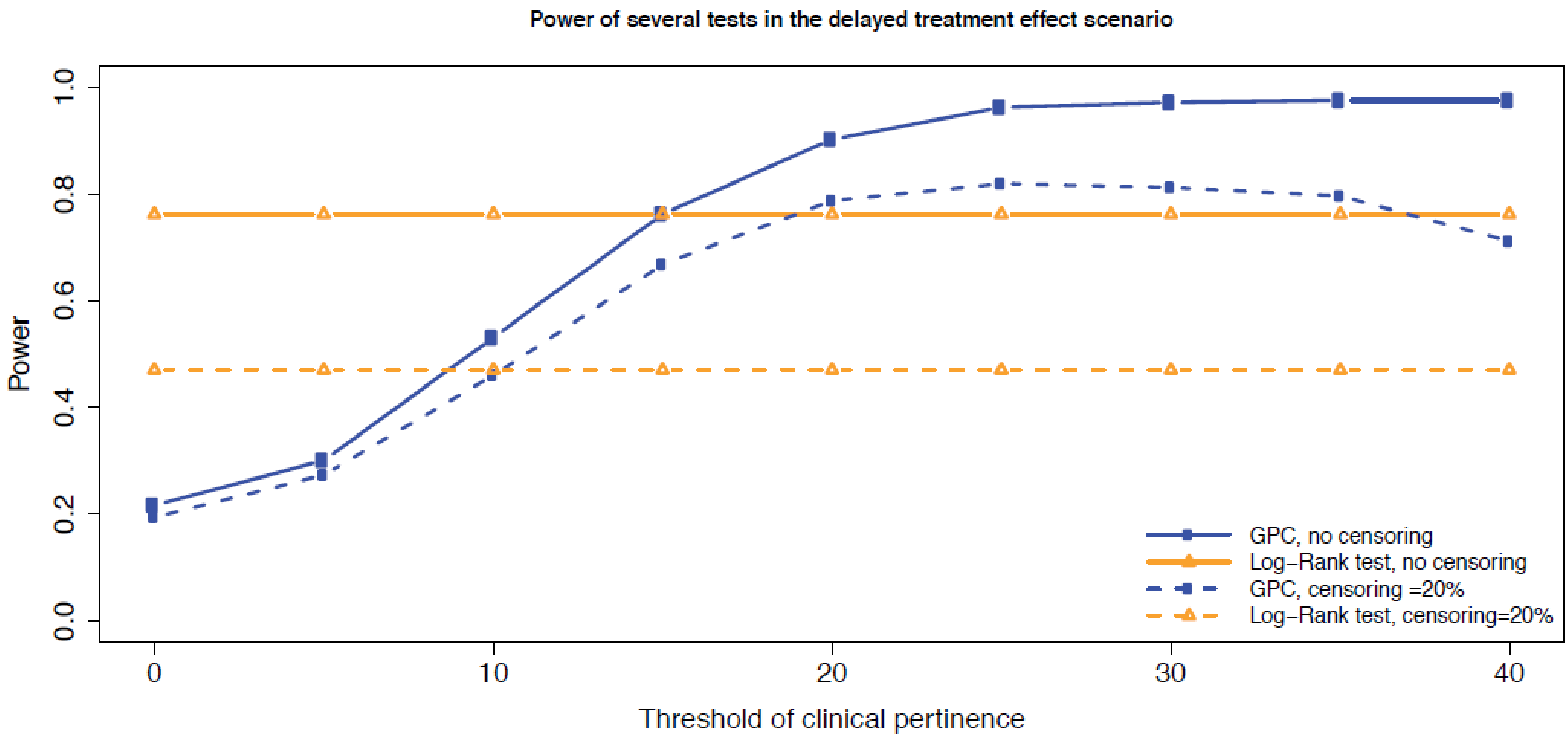
Example: allografts in childhood tumors

Péron et al, JAMA Oncol 2016;2:901

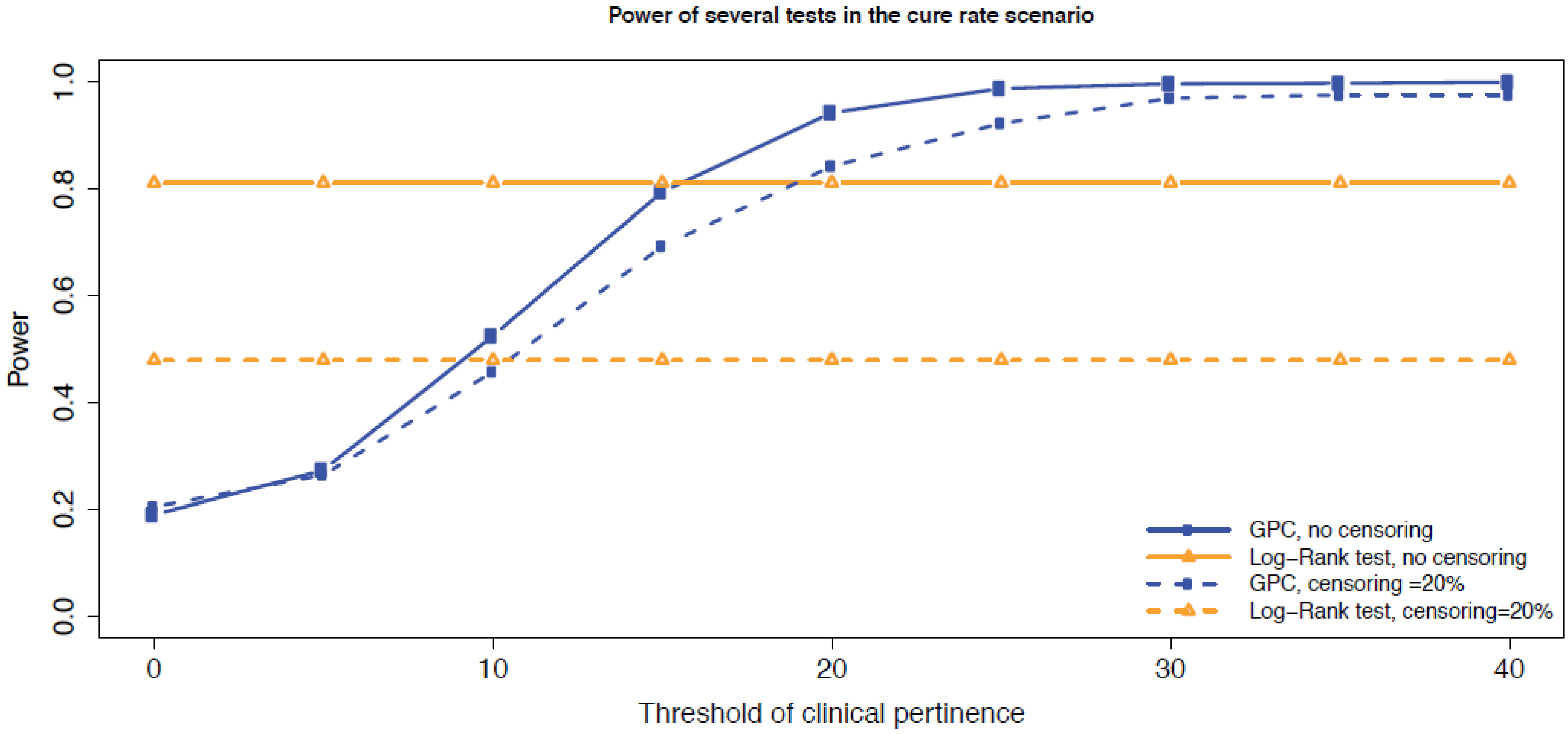
Power – proportional hazards



Power – delayed difference



Power – cure rate



Closing remarks

- Assessing benefit/risk in an individualized manner is key to personalized medicine
  - Marginal (one outcome at a time) benefit/risk analyses ignore the correlation between the outcomes
  - GPCs account naturally for the correlation, but require prioritization of the outcomes

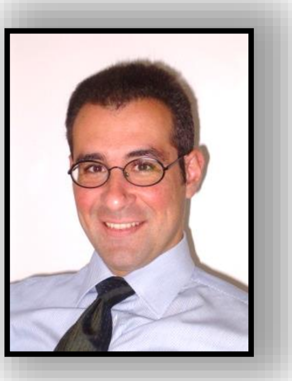
Evans and Follmann, Using outcomes to analyze patients rather than patients to analyze outcomes: A step toward pragmatism in benefit:risk evaluation. Stats Biopharml Res 2016;8:386.

Closing remarks

- GPCs are attractive
  - In terms of patient centricity:
    - They lead to the “net benefit”, a patient-relevant measure
    - They use prioritized outcomes (according to patient preferences)
  - In statistical terms:
    - They are equivalent to standard non-parametric tests in simple cases
    - They may have better power than the logrank test (for delayed treatment benefits)
    - They allow for testing of clinically relevant differences

Thank You!

Marc Buyse, ScD
Chief Scientific Officer, IDDI
marc.buyse@iddi.com



Everardo D. Saad, MD
Medical Director, IDDI
everardo.saad@iddi.com