Study designs in rare cancer

Saskia Litière

EORTC Biostatistician

RCE Patient advocates course 20/07/2015

EORTC The future of cancer therapy

EORTC The future of cancer therapy

The randomized comparative trial (RCT)

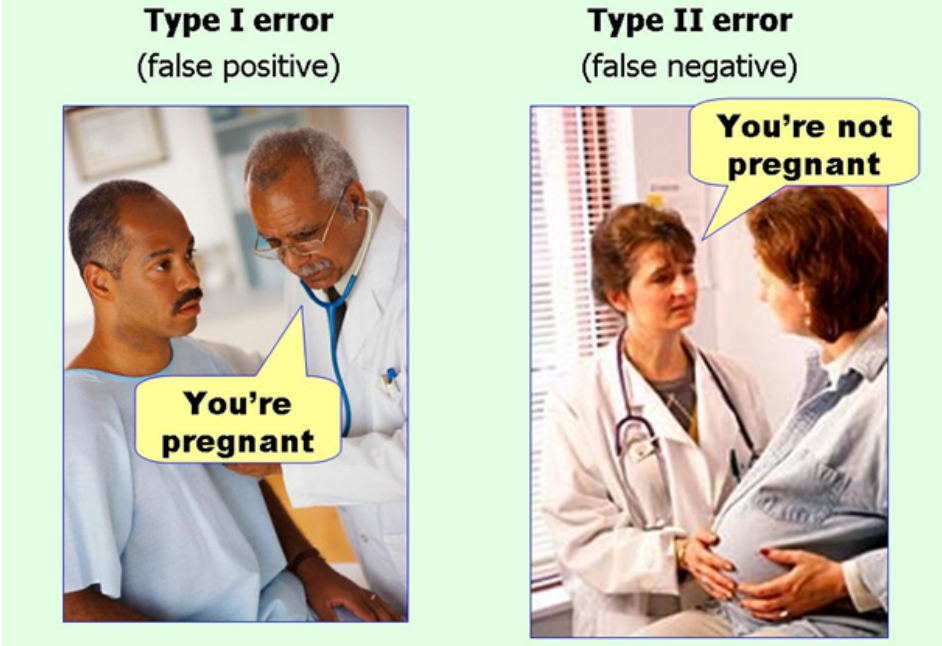
RCTs and systematic reviews of RCTs remain the gold standard for assessing benefit of treatment

- generate unbiased evidence
- size is determined to be most efficient to detect a targeted treatment effect that is considered clinically relevant, given some pre-specified design characteristics

As with any assay or test looking for a yes/no answer, we can make a mistake

- Type I error (α): 5-20%
- Type II error (β): 10-20%

  or power (1- β): 90-80%



When facing a rare cancer …

Preambule

You would like the statisticians to help you
in making the right decisions
in the situations of great uncertainty.

There is no magic solution
but we should find some (smart) compromises…

http://cddf.org/files/2014/07/LEDELEY-Innovative-designs-for-efficacy-trials-in-rare-conditions.pdf

Where can we compromise?

- Do we need a control arm?
- Can we allow more uncertainty?
- Can we use a different endpoint -> next talk
- Which patients do we select?
- Can we adapt the design?
- What about a Bayesian approach?

Single-arm/non-comparative approaches

- Observing some response is an improvement in itself
- Stopping progression is an improvement in itself
- More ethical if control (standard?) treatment is
  - believed to have little effect
  - and/or very toxic
- If robust historical data is available …

But …

- Still applicable?
- Same patient population?
  - Same characteristics?
  - Same staging system?
  - Biomarker driven?
- Single arm, α=10% for pre-specified p0 (H0)
  A 5% absolute error in p0 increases false positive rate to 30%!

Baey & Le Deley EJC 2011

Allow more uncertainty: relaxing the errors

- **Type I error rate (alpha):** probability of a false positive
- **Power:** probability of finding a meaningful effect

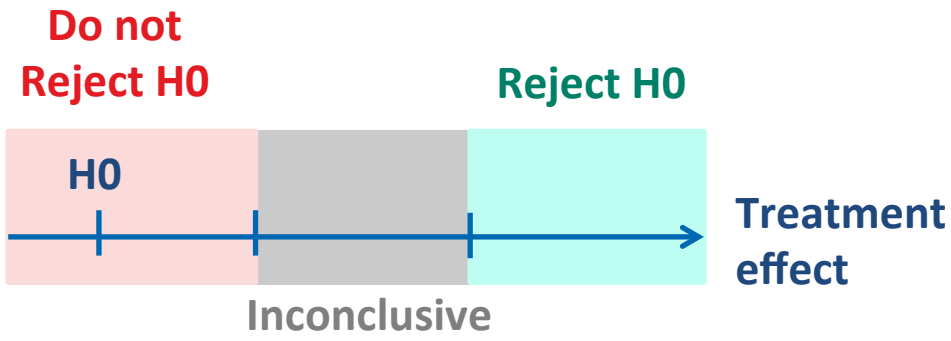
Randomized study
Outcome: time to event
Targeted difference:
HR = 0.65

Alpha \ Power	90%	85%	80%	75%
5%	**227**	194	170	150
10%	185	155	134	116
15%	160	133	113	97
20%	142	116	98	**83**

Number of events obtained from EAST 6.0

Need to be careful with the consequences of relaxing the errors, given that it is unlikely that another trial will be conducted to confirm the results

Allow more uncertainty: the grey zone



Factors other than the success rate (such as toxicity, cost, or QoL) may dominate the decision regarding whether the agent should undergo further testing.

A Three-Outcome Design for Phase II Clinical Trials

Controlled Clinical Trials 22:117–125 (2001)

Daniel J. Sargent, PhD, Victor Chan, MS, and Richard M. Goldberg, MD

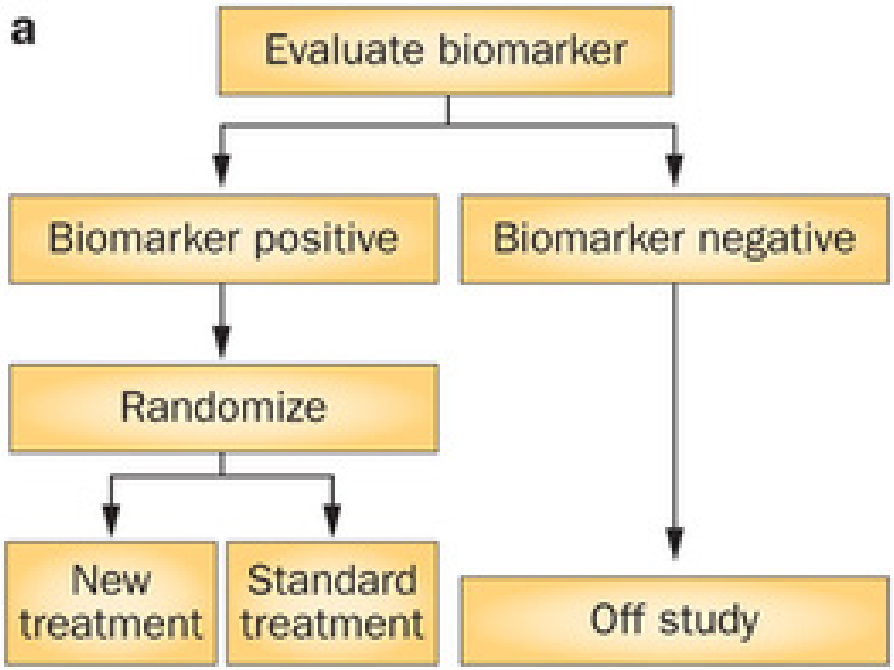
Section of Biostatistics (D.J.S.) and Division of Medical Oncology (R.M.G.), Mayo Clinic, Rochester, Minnesota, and Department of Statistics, Iowa State University, Ames, Iowa (V.C.)

Table 1 Comparison of Sample Sizes

				Standard Design		Three-Outcome Design		
p_0	p_a	$\leq\alpha$	$\leq\beta$	r	N	r	s	N
0.05	0.20	0.10	0.10	3	32	2	4	27
		0.05	0.10	4	38	2	4	27
0.10	0.25	0.10	0.10	6	40	4	6	30
		0.05	0.10	8	49	6	8	40
0.20	0.35	0.10	0.10	15	58	10	12	41
		0.05	0.10	19	70	13	16	53
0.30	0.45	0.10	0.10	24	66	18	20	51
		0.05	0.10	32	85	22	25	62
0.40	0.55	0.10	0.10	34	73	25	28	56
		0.05	0.10	43	90	31	35	69
0.50	0.65	0.10	0.10	41	72	32	34	57
		0.05	0.10	51	88	36	40	66

Each design has the smallest sample size N among all the designs of the same type with actual α and β no larger than the given α and β. For each three-outcome design, actual η and π must be ≥ 0.8.

Select patient population?



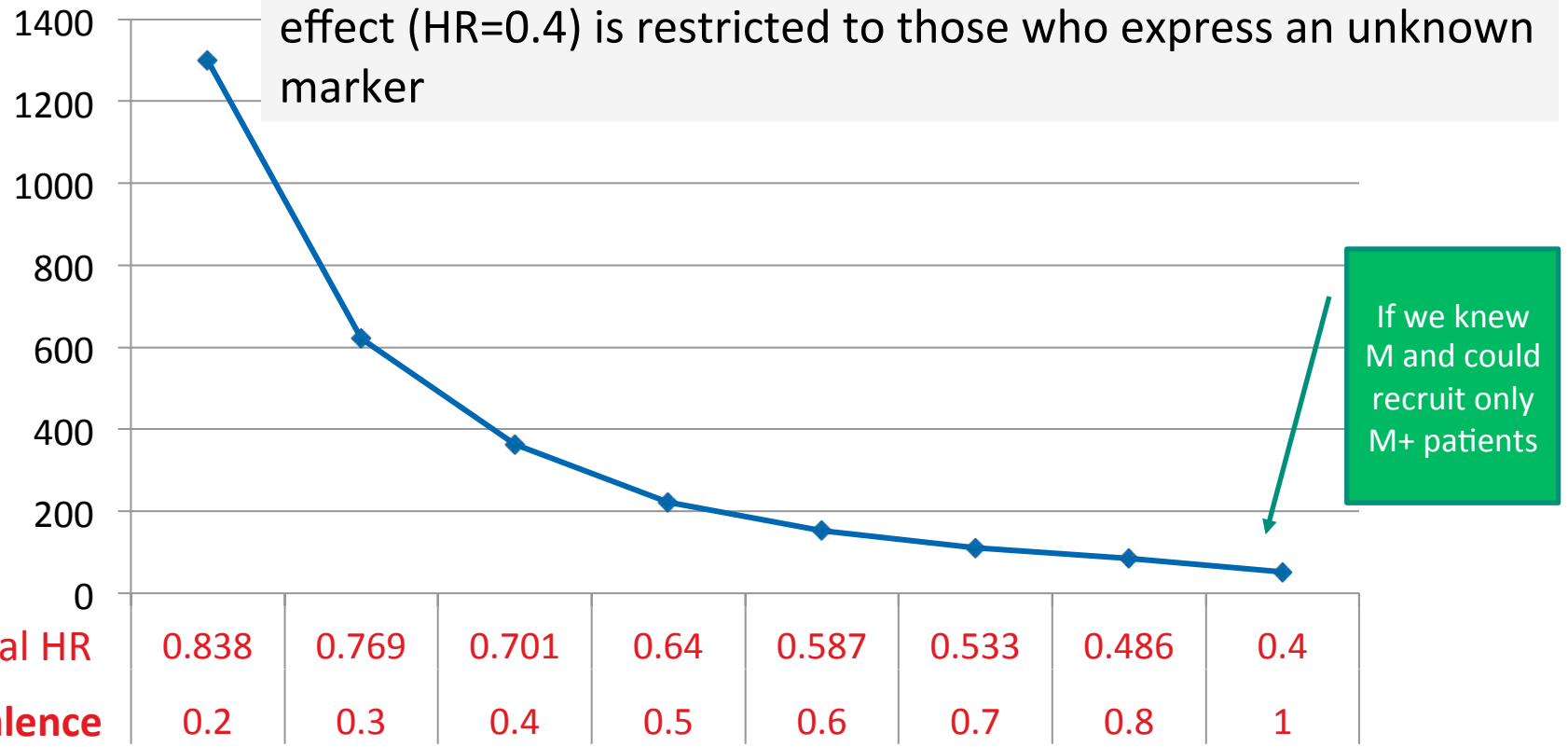
Freidlin and Korn,
Nature Reviews Clinical Oncology 2014

Identify a small selected population where large benefits are expected

To be preferred to a large study on unselected population where moderate or small benefit is expected

Select patient population?

Sample size of an unselected phase III when the treatment effect (HR=0.4) is restricted to those who express an unknown marker

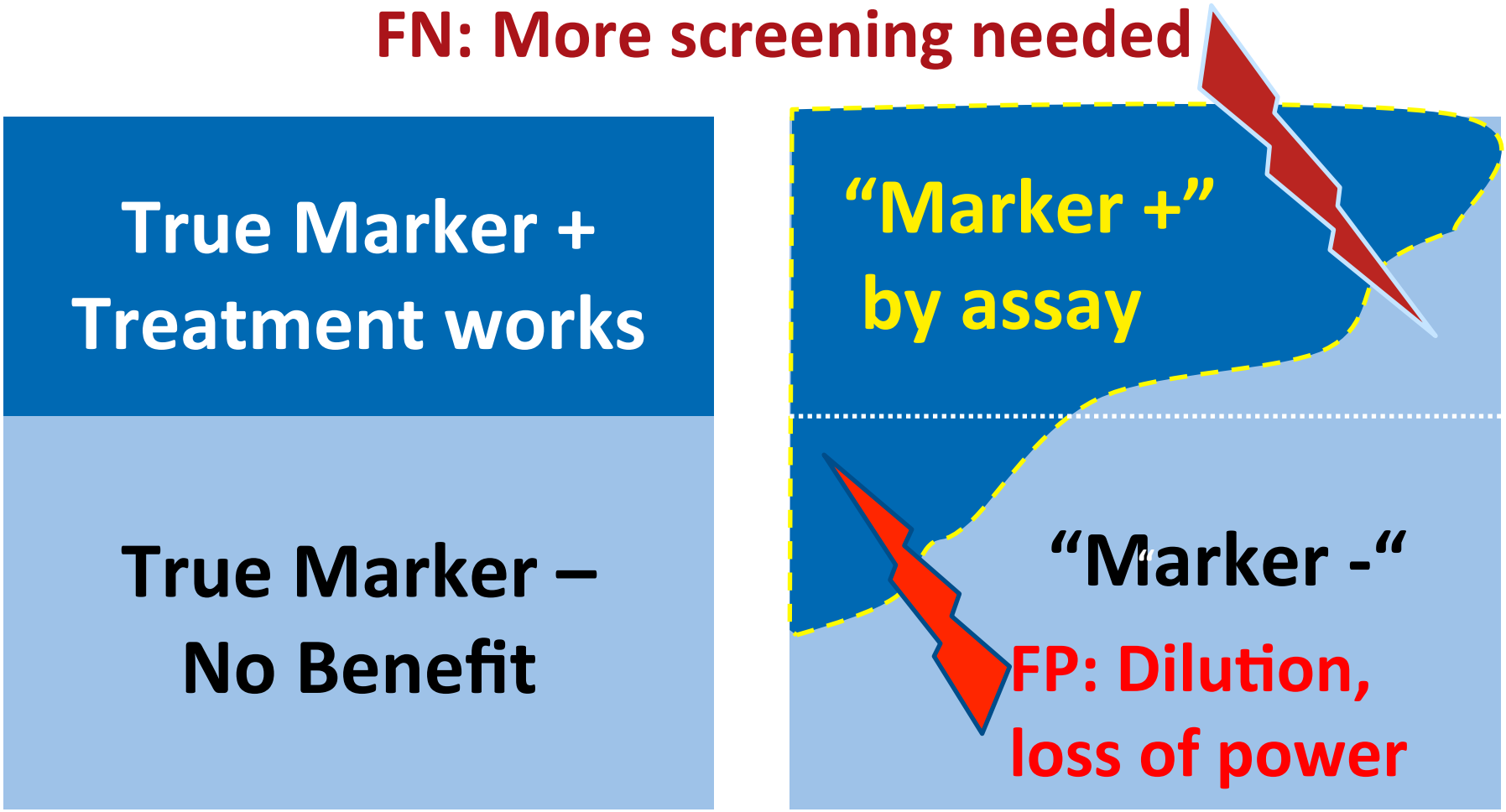


Global HR	0.838	0.769	0.701	0.64	0.587	0.533	0.486	0.4
Prevalence	0.2	0.3	0.4	0.5	0.6	0.7	0.8	1

Adapted from S. Sleijfer et al., JCO 2013

EORTC The future of cancer therapy

A word of caution: how good is the assay? …



Adaptive designs

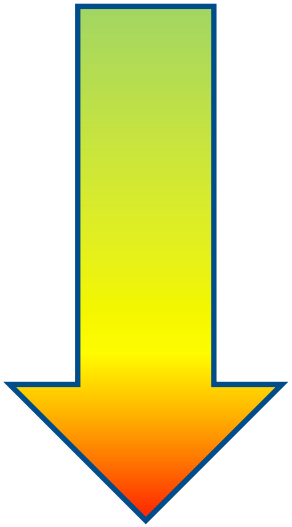
- Early stopping for futility and/or efficacy

- Drop treatment arm(s) or pick the winner designs
- Biomarker adaptive designs

- Sample size re-estimation
- Adaptive randomization…

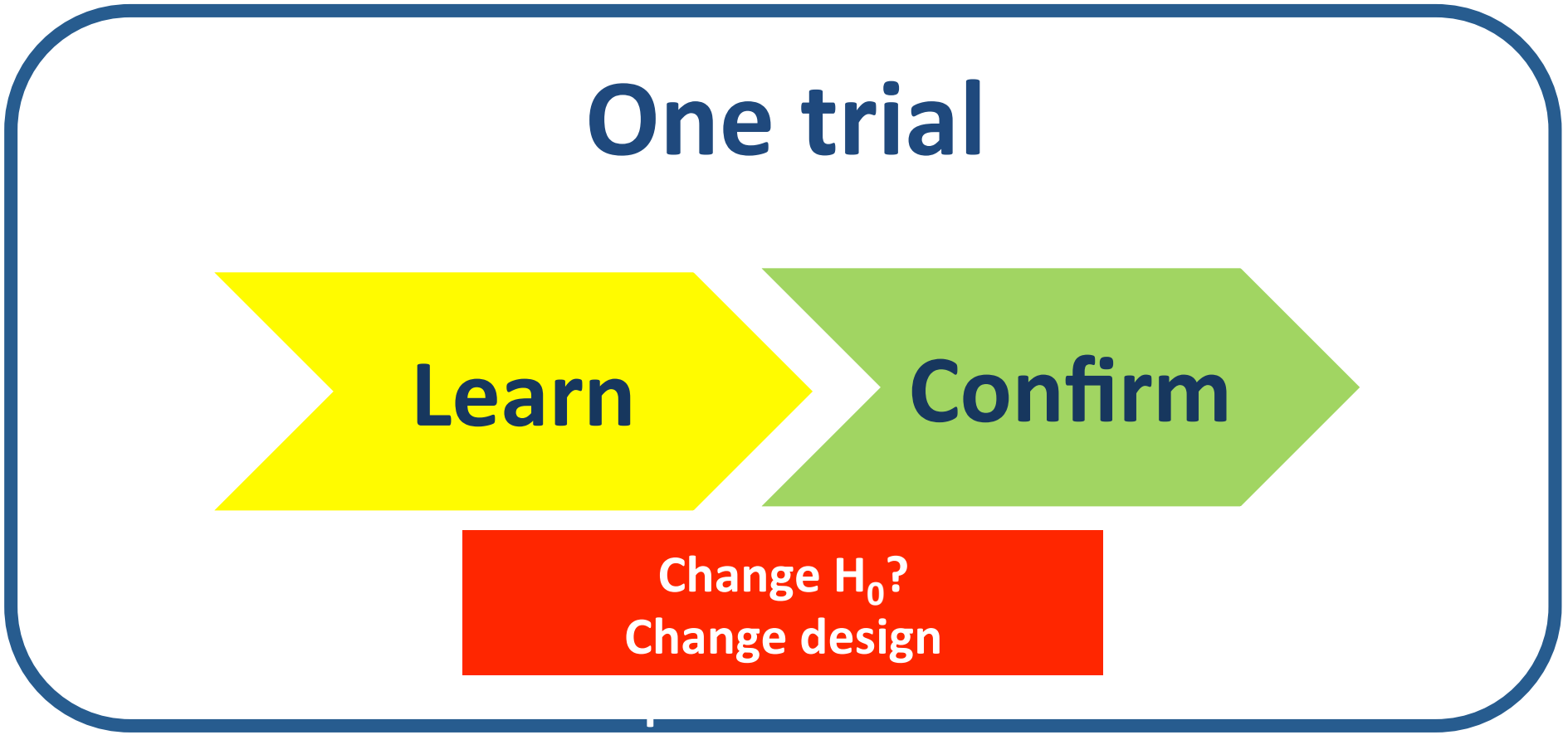
To name but a few …

Well-known



Less understood

Most of them come down to …



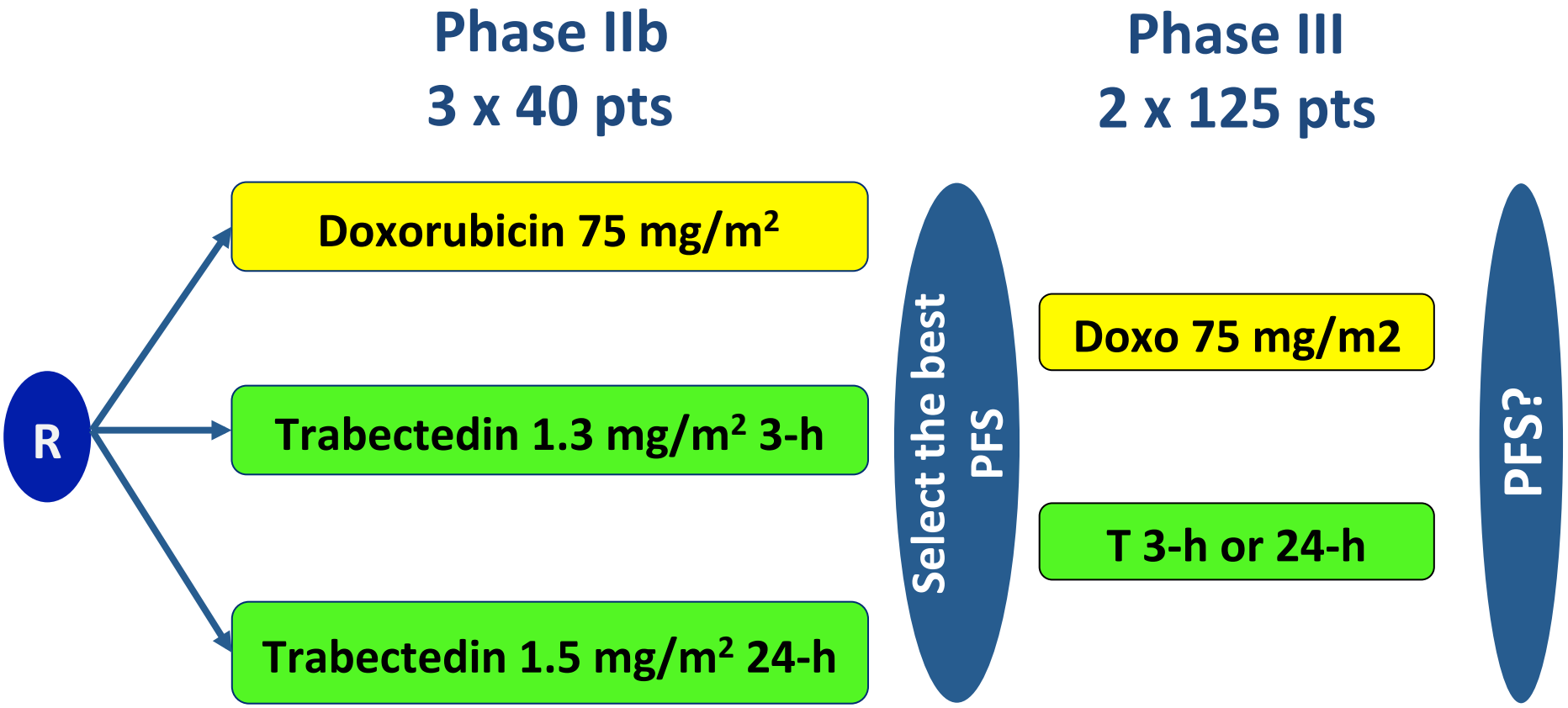
The challenges

Decision	In truth, the null is... True	False
Accept Null H_0	Correct!	Type II
Accept Alt. H_a	Type I	Correct!

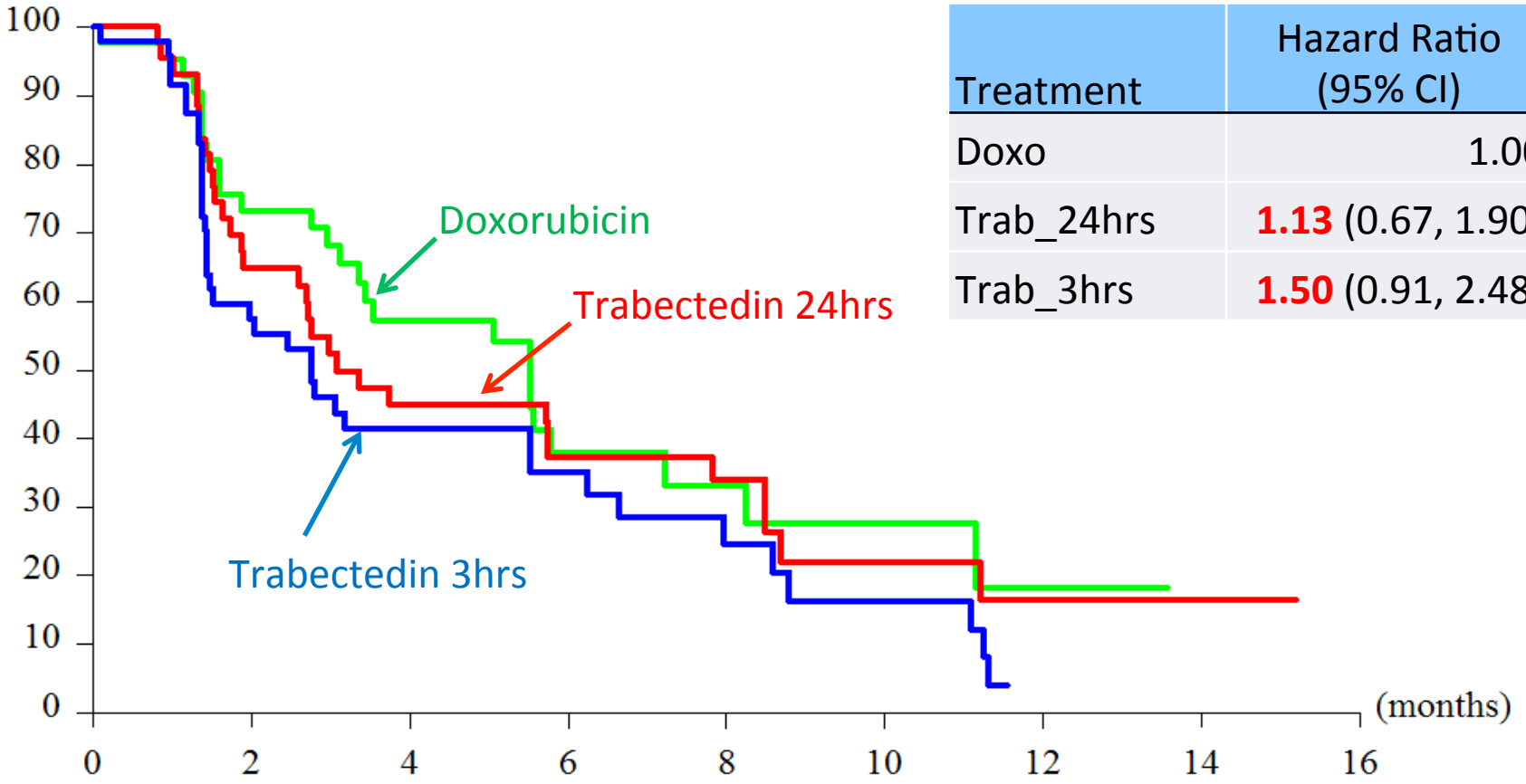
- To control the operating characteristics
- To control the bias due to the adaptation
  - Statistical
  - Operational

- To guarantee that the results can be interpreted and explained!

Example: **TRUSTS (EORTC 62091)** in advanced or metastatic untreated soft tissue sarcomas



Progression free survival



Treatment	Hazard Ratio (95% CI)
Doxo	1.00
Trab_24hrs	**1.13** (0.67, 1.90)
Trab_3hrs	**1.50** (0.91, 2.48)

O	N	Number of patients at risk :						
26	43	30	20	10	6	5	1	0
31	43	27	18	13	10	5	3	1
37	47	26	16	11	6	4	0	0

Treatment
Doxo
Trab_24hrs
Trab_3hrs

Adaptive designs in rare disease

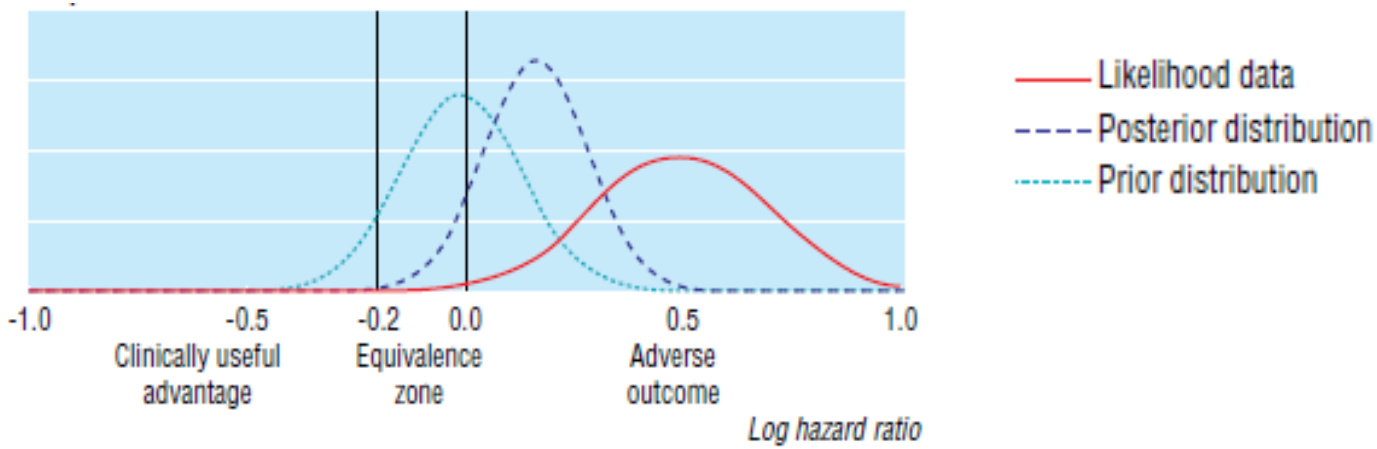
- Adaptive designs provide an appealing alternative because
  - shorten development process
  - ineffective treatments can be identified earlier on
  - more efficient use of limited patient numbers

- Being adaptive comes at a cost
  - Complex design with statistics that can become difficult to explain
  - Logistically challenging
  - Difficult in studies with long-term endpoints

Adaptive designs cannot make a drug more effective

EORTC The future of cancer therapy

A Bayesian approach



- Formally incorporating historical data into the design and analysis
  - Formulate prior distribution from existing evidence
    - Literature review
    - Score according to pertinence, validity and precision (Tan et al, BMJ, 2003)
  - In a small trial, the choice of the prior may carry heavy weight
  - In rare cancers, only weak prior evidence might be available
  - May require a lot of data to overcome a "bad" prior …

EORTC The future of cancer therapy

An example

A NEW TOOLKIT FOR CONDUCTING CLINICAL TRIALS IN RARE DISORDERS

Lusine Abrahamyan[1], Ivan R Diamond[2], Sindhu R Johnson[3], Brian M Feldman[4]

J Popul Ther Clin Pharmacol Vol 21(1):e66-e78; February 23, 2014

This can be illustrated with data from a hypothetical placebo controlled trial with 30 participants (15 in each group) of a drug for symptom management of a rheumatologic disorder. The outcome of this study is a quality of life instrument. A difference of 2-points is believed to be clinically important. The mean value in the experimental arm is 13.6 units (standard deviation: 4.97) and in the control arm 10.5 units (standard deviation: 4.36). A traditional frequentist analysis with a t-test provides insufficient evidence to disprove the null hypothesis – t-statistic 1.797, df = 28, p = 0.083, mean difference 3.06: 95% confidence interval: -0.42 to 6.56. For this small study, the power is low and the results may be false negative. A Bayesian analysis (using an uninformative prior[1]) provides very similar estimates for the difference in means between the groups (median value: 3.05, 95% credible interval -0.58 to 6.61). However, unlike the frequentist analysis the Bayesian analysis allows us to determine that the probability of a clinically important difference of 2 points is 74% – which can be thought of as 3:1 odds favoring the experimental treatment. For an inexpensive and safe treatment, this may be enough evidence to support treatment. There is a 96% probability that the experimental treatment is at least a little bit better. Therefore, this small study, that would have likely been regarded as "negative" with the frequentist approach, may provide useful information when analyzed by the Bayesian approach. A Bayesian reanalysis of a

The probability of a clinically important difference of at least 2 points is 74% - is this enough if ?

26% chance that it will not make an important difference

Many treatment options	Expensive	Toxic
YES	NO	NO
YES	YES	NO
YES	NO	YES
NO	NO	NO
NO	YES	YES

A word of caution

"Frequentist" framework

- 0.083 (or 0.042 if you are looking one-sided) is the likelihood of your data if there were no difference
- Note that 0.05 is a pre-specified (arbitrary!) cut-off requiring strong evidence to support a decision

Why do we demand rigorous testing characteristics in a "frequentist" framework, but leave it up to the eye of the "Bayesian" beholder to judge on

- 74% probability of having a clinically important difference of at least 2 points
- 96% probability of the experimental treatment being at least a bit better

FDA guidance for the use of Bayesian statistics in medical device clinical trials (2010)

"Because of the inherent flexibility in the design of a Bayesian clinical trial, a thorough evaluation of the operating characteristics should be part of the trial planning"

⇨ type I error rate, type II error rate, power, sample size distribution, prior probability and probability of stopping at each interim look

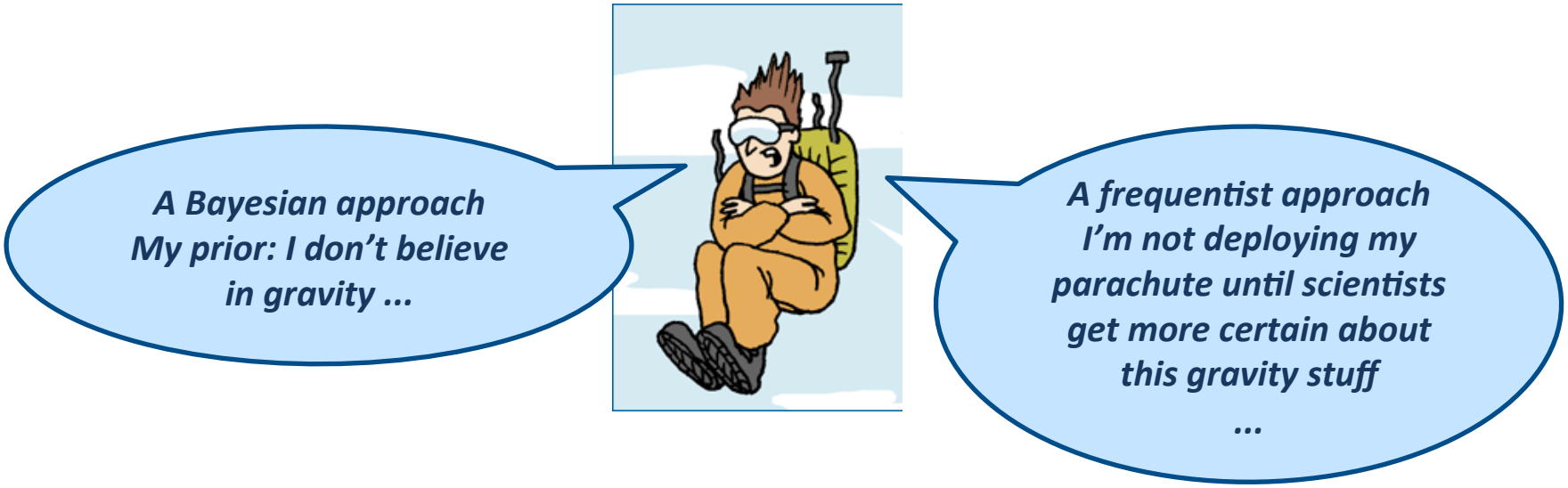
You might be surprised at how much this differs from the frequentist approach … **or not**?

Conclusion

- The gold standard for empirical evidence remains an adequately powered comparative trial.

- In rare cancers however we need to look for
  - flexible research strategies
  - more efficient use of resources

  However, we may need to accept that there is a price to pay
  - More uncertainty
  - Complexity
  - Interpretation of results



Thank you

Special thanks to my EORTC stat colleagues, in particular Laurence Collette, Jan Bogaerts, Catherine Fortpied and Leen Slaets

EORTC The future of cancer therapy

EORTC The future of cancer therapy