EORTC
European Organisation for Research and Treatment of Cancer
The future of cancer therapy

Understanding Statistics in Sarcoma Clinical Trials

9th SPAEN Annual Conference
1st February 2019

Catherine Fortpied, MSc
Lead Statistician @ EORTC HQ, Brussels, Belgium

EORTC
European Organisation for Research and Treatment of Cancer
The future of cancer therapy

Statistics : a black box?

Outline

Three EORTC studies in sarcoma as a basis to present some general statistical concepts & innovative methods:

- ❑ A phase II study in advanced clear-cell sarcoma
- ❑ A phase III study in retroperitoneal sarcoma
- ❑ Trial designs for rare diseases: where can we compromise?

EORTC
European Organisation for Research and Treatment of Cancer
The future of cancer therapy

EORTC 90101 – 'CREATE'

ESMO GOOD SCIENCE BETTER MEDICINE BEST PRACTICE

Annals of Oncology 28: 3000–3008, 2017
doi:10.1093/annonc/mdx527
Published online 18 September 2017

ORIGINAL ARTICLE

Activity and safety of crizotinib in patients with advanced clear-cell sarcoma with MET alterations: European Organization for Research and Treatment of Cancer phase II trial 90101 'CREATE'

P. Schöffski[1,2*], A. Wozniak[2], S. Stacchiotti[3], P. Rutkowski[4,5], J.-Y. Blay[6], L. H. Lindner[7], S. J. Strauss[8], A. Anthoney[9], F. Duffaud[10,11], S. Richter[12,13], V. Grünwald[14], M. G. Leahy[15], P. Reichardt[16], J. Sufliarsky[17], W. T. van der Graaf[18,19], R. Sciot[20], M. Debiec-Rychter[21], T. van Cann[1,2], S. Marréaud[22], M. Lia[22], T. Raveloarivahy[22], L. Collette[22] & S. Bauer[23]

EORTC
European Organisation for Research and Treatment of Cancer
The future of cancer therapy

Study objectives and design

Methods

Study design

This was a multicentre, biomarker-driven, single agent, non-randomized, open-label, two-stage phase II trial, assessing the activity and safety of crizotinib in patients with locally advanced or metastatic CCSA (EORTC 90101,ClinicalTrials.gov:NCT01524926). The patient population was divided into *MET* positive (*MET*+; presence of *EWSR1* gene rearrangement) and *MET* negative (*MET*−; absence of *EWSR1* gene rearrangement) sub-cohorts, which were analysed separately. Investigators were blinded to the centrally assessed MET status.

Ethics approval was obtained by com
ing to national legislation. The stud
with: the Declaration of Helsinki;
participating country/institution; and
Harmonisation-Good Clinical Practice.

Outcomes

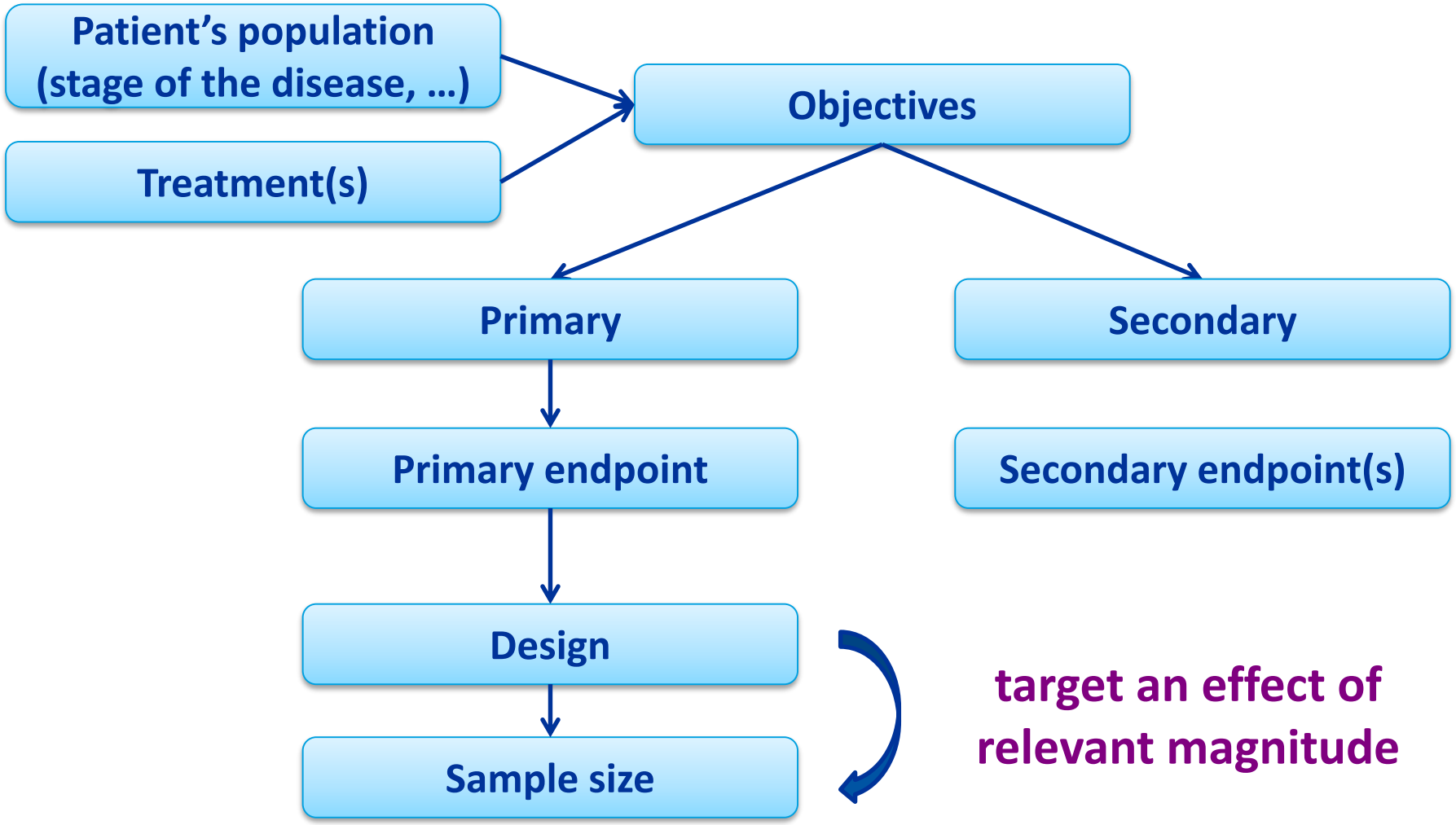
The main objective was to study the activity of crizotinib in *MET*+ CCSA patients. The primary end point was ORR per RECIST 1.1 with response confirmation, assessed by the local investigator. This end point was chosen based on the response pattern seen with crizotinib in the labelled indication of NSCLC and in the absence of reliable reference data on PFS or PFR in CCSA. Secondary end points included: duration of response, DCR, PFS, PFR, OS, OSR, safety and correlative/translational research end points. The DCR was defined as the percentage of patients achieving either a complete (CR) or partial response (PR) or SD.

Statistical analysis

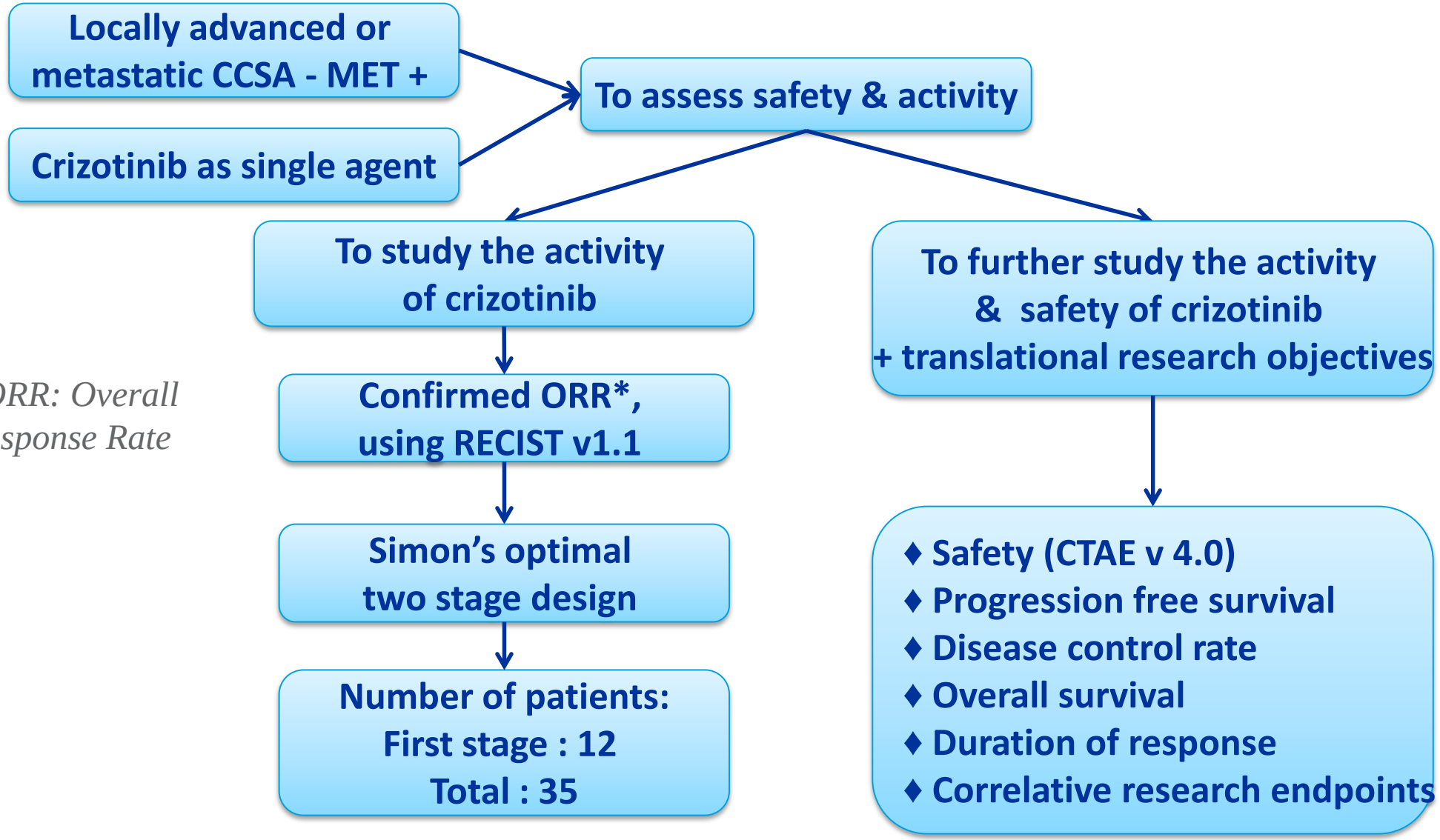
The statistical design was conceptually focused on cases with centrally documented *EWSR1* fusion (*MET*+ sub-cohort). It was decided that showing an ORR >10% (null hypothesis) in CCSA *MET*+ patients, a rare population, resistant to chemotherapy and radiotherapy, would be promising for future research. Therefore a Simon's optimal two-stage design [18] was implemented with the aim of excluding an ORR ≤10% under the alternative assumption that 30% ORR can be achieved with crizotinib. The type I error and power were set at 10%. In stage 1, if at least two out of the first 12 eligible and assessable CCSA *MET*+ patients achieved a confirmed RECIST PR or CR, a maximum of 35 patients were to be enrolled. In stage 2, if less than 6 out of the 35 eligible and assessable patients responded, the treatment was declared ineffective. If at least 6 out of the 35 patients responded, further study of crizotinib in CCSA was

andomized, non-historical and
ll comers' independent of their
of treatment of patients in need
reference data for both subsets.
d ethical due to the lack of vali-
ise.
points details are provided in
available at *Annals of Oncology*
SAS version9.4 (SAS Institute,

From objective(s) to sample size



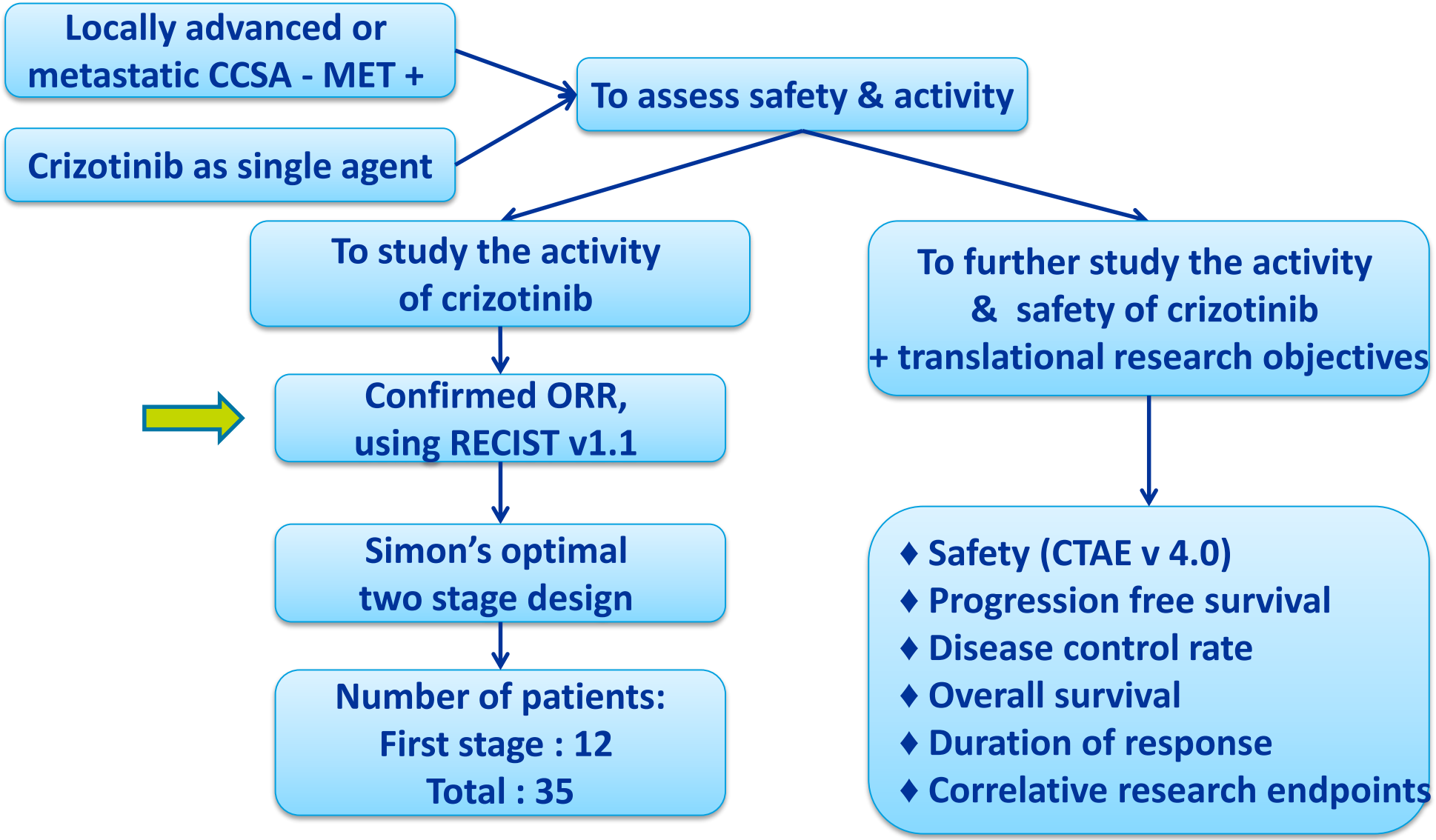
From objective(s) to sample size



**ORR: Overall Response Rate*

EORTC
European Organisation for Research and Treatment of Cancer
The future of cancer therapy

From objective(s) to sample size



Primary endpoint: Confirmed ORR, using RECIST v1.1

The main objective was to study the activity of crizotinib in *MET*+ CCSA patients. The primary end point was ORR per RECIST 1.1 with response confirmation, assessed by the local investigator. This end point was chosen based on the response pattern seen with crizotinib in the labelled indication of NSCLC and in the absence of reliable reference data on PFS

- ORR = Binary Endpoint (yes / no)
- Defined as complete and partial response (RECIST 1.1) observed during the whole treatment period
- Endpoints in phase II trials are most often binary outcome measures of activity and/or toxicity:
    - ORR
    - Clinical benefit = CR+PR+SD
    - Progression-free survival (PFS) at a fixed point in time
    - Toxicity (e.g. grade≥3 tox over treatment period)
    - …

Primary endpoint: Confirmed ORR, using RECIST v1.1

- Why RECIST?
  - Objective criteria which measure the decrease of the tumor burden
  - Standard criteria widely used adopted by academic institutions, cooperative groups, and industry

- Why confirmed response?
  - "In non-randomised trials where response is the primary endpoint, confirmation of PR and CR is required to ensure responses identified are not the result of measurement error."

Eisenhauer et al. New response evaluation criteria in solid tumours: revised RECIST guideline (version 1.1). Eur J Cancer. 2009 Jan;45(2):228-47.

Table 1 – Time point response: patients with target (+/– non-target) disease.

Target lesions	Non-target lesions	New lesions	Overall response
CR	CR	No	CR
CR	Non-CR/non-PD	No	PR
CR	Not evaluated	No	PR
PR	Non-PD or not all evaluated	No	PR
SD	Non-PD or not all evaluated	No	SD
Not all evaluated	Non-PD	No	NE
PD	Any	Yes or No	PD
Any	PD	Yes or No	PD
Any	Any	Yes	PD

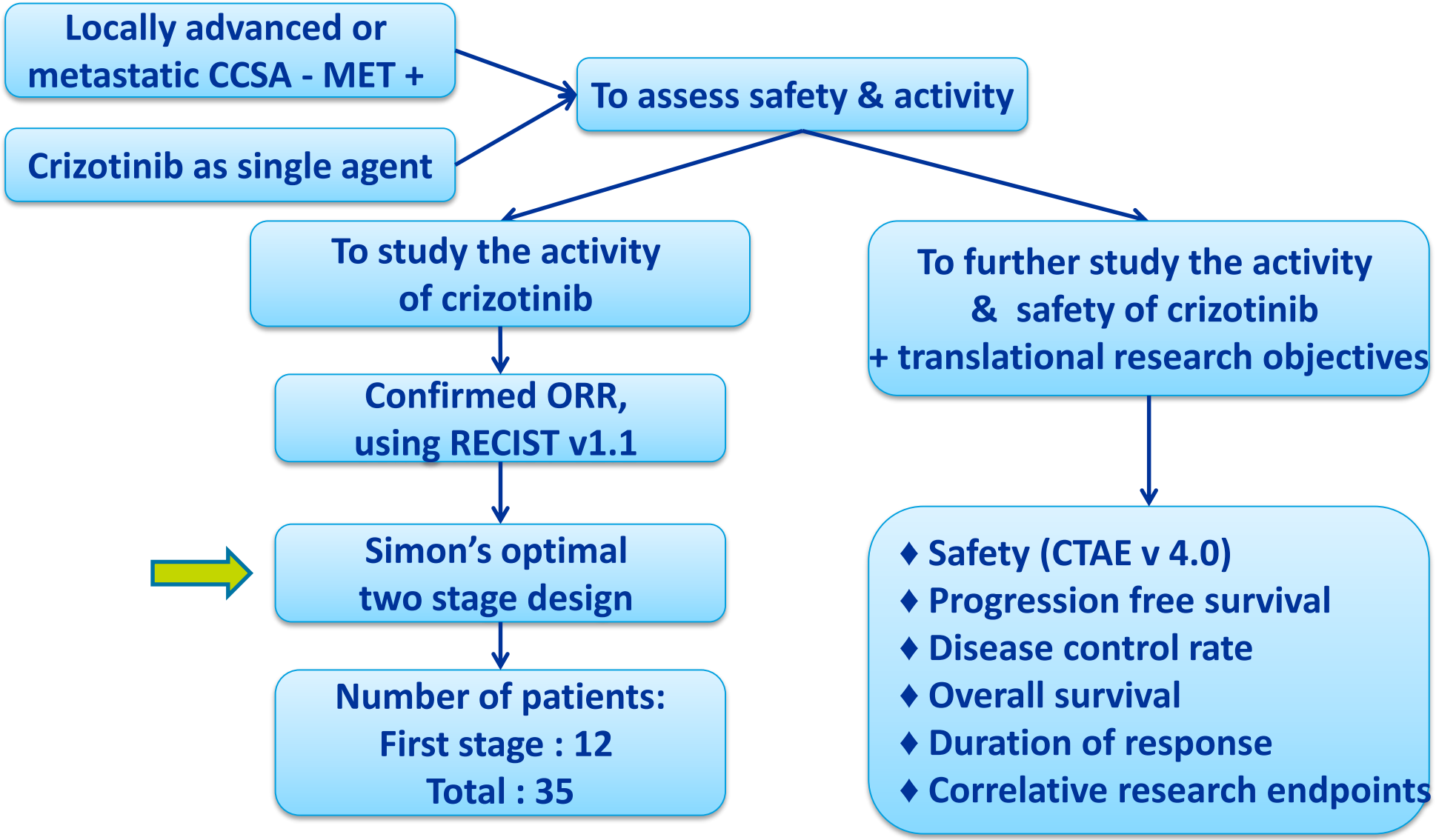
CR = complete response, PR = partial response, SD = stable disease, PD = progressive disease, and NE = inevaluable.

Table 3 – Best overall response when confirmation of CR and PR required.

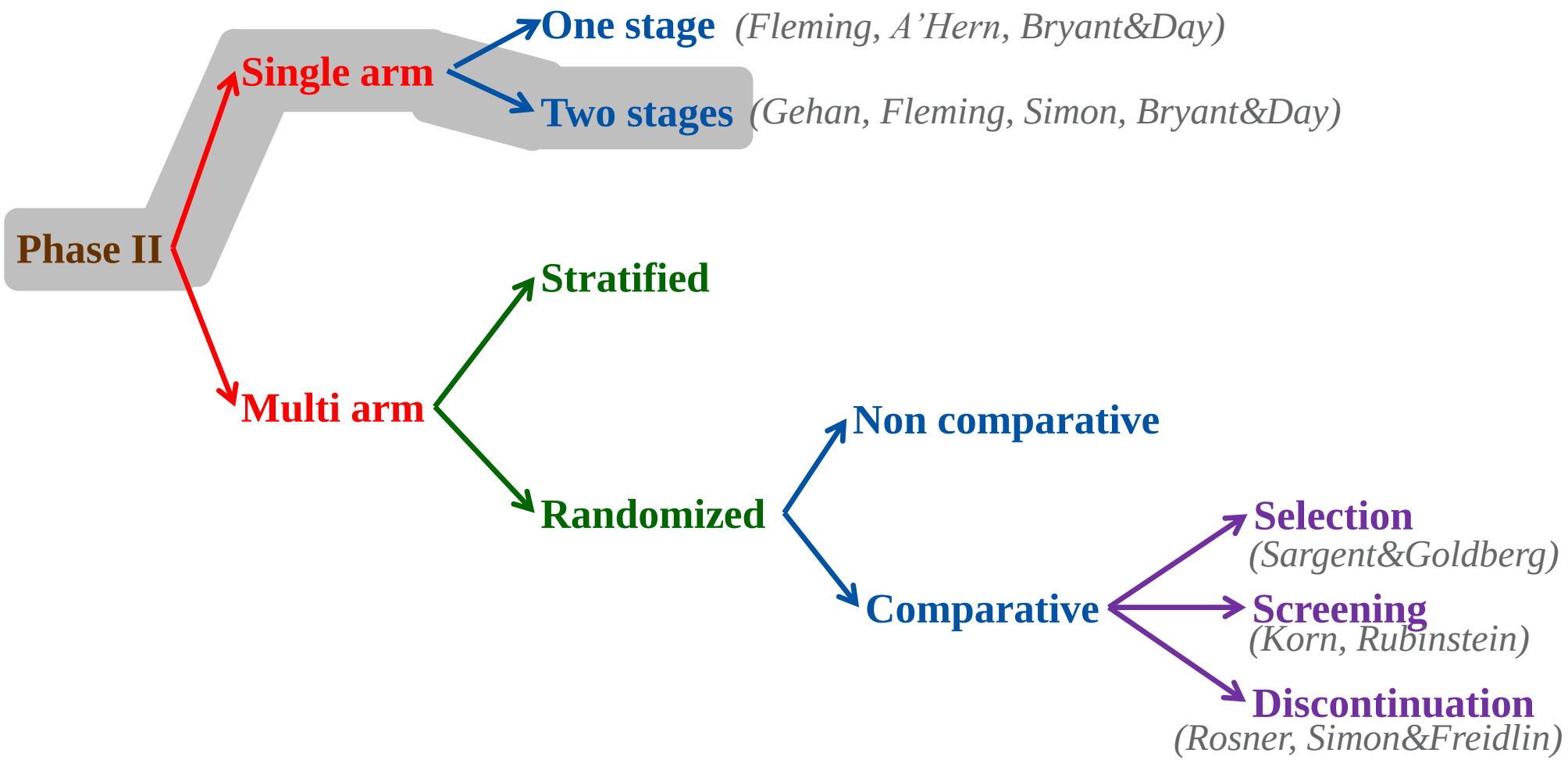
Overall response First time point	Overall response Subsequent time point	BEST overall response
CR	CR	CR
CR	PR	SD, PD or PR[a]
CR	SD	SD provided minimum criteria for SD duration met, otherwise, PD
CR	PD	SD provided minimum criteria for SD duration met, otherwise, PD
CR	NE	SD provided minimum criteria for SD duration met, otherwise NE
PR	CR	PR
PR	PR	PR
PR	SD	SD
PR	PD	SD provided minimum criteria for SD duration met, otherwise, PD
PR	NE	SD provided minimum criteria for SD duration met, otherwise NE
NE	NE	NE

EORTC
European Organisation for Research and Treatment of Cancer
The future of cancer therapy

From objective(s) to sample size



Design


Phase II
Single arm
One stage (Fleming, A'Hern, Bryant&Day)
Two stages (Gehan, Fleming, Simon, Bryant&Day)
Multi arm
Stratified
Randomized
Non comparative
Comparative
Selection (Sargent&Goldberg)
Screening (Korn, Rubinstein)
Discontinuation (Rosner, Simon&Freidlin)

Simon two-stage design

- Two-stage allow early stopping if insufficient activity

- The trial must be temporarily closed to patients entry between the 2 stages, especially if there is:
  - A high recruitment rate
  - Low activity

Simon Optimal & Minimax

Null H0: p0 = 10%

Alternative H1: p1 = 30%

α = 10% 1-β = 90%

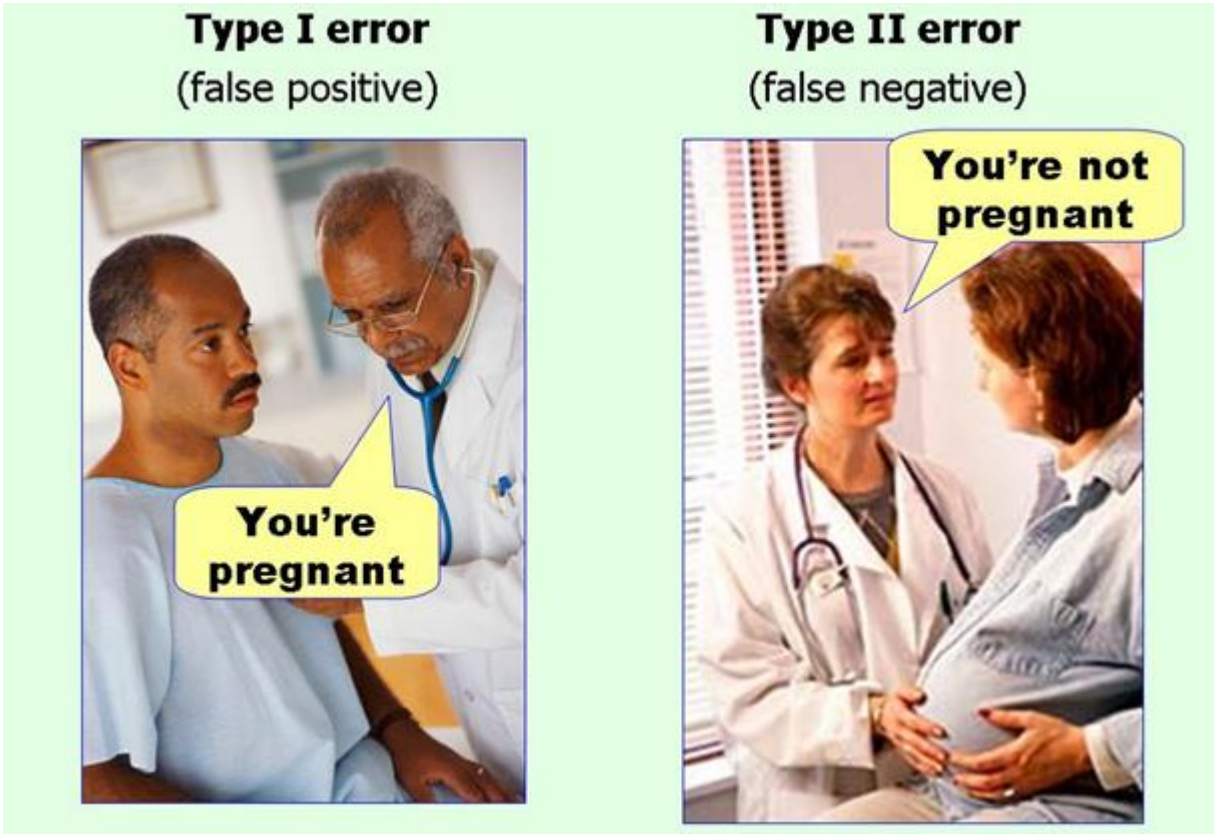
	Simon Optimal	Simon Minimax
N_1	12	16
R_1	2	2
N	35	25
R	6	5
Expected # pts under p0=10%	19.84	20.37
Minimizes…	# pts if trt inactive	maximum # pts

The statistical design was conceptually focused on cases with centrally documented *EWSR1* fusion (*MET*+ sub-cohort). It was decided that showing an ORR >10% (null hypothesis) in CCSA *MET*+ patients, a rare population, resistant to chemotherapy and radiotherapy, would be promising for future research. Therefore a Simon's optimal two-stage design [18] was implemented with the aim of excluding an ORR ≤10% under the alternative assumption that 30% ORR can be achieved with crizotinib. The type I error and power were set at 10%. In stage 1, if at least

Simon R (1989) Optimal Two-stage Designs for Phase II Clinical Trials. Controlled Clinical Trials, 10: 1-10.

Type I and II errors

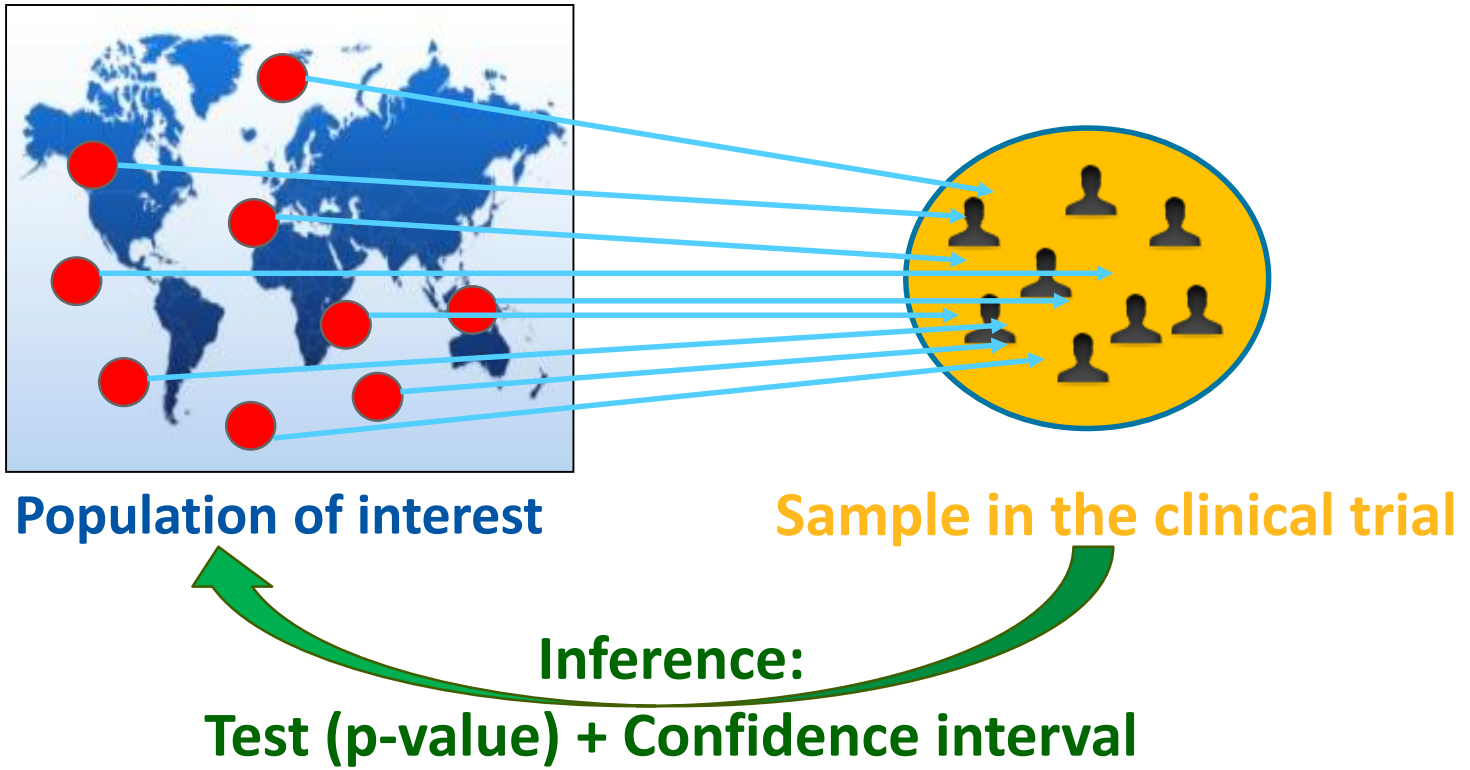
A **type I error** is the incorrect rejection of a true null hypothesis
A **type II error** is the failure to reject the null hypothesis when the alternative is true



Example:
pregnancy test

H_0 not pregnant
H_1 pregnant

How statisticians look at a trial: **sampling**



- What does the **Observed** tells us about the **True** ?
- Is what we have seen in the sampled patients also true in a more **general sense**?

→ STATISTICAL INFERENCE + CLINICAL INTERPRETATION

EORTC
European Organisation for Research and Treatment of Cancer
The future of cancer therapy

Decision making

		Our decision (Trial conclusion)	
		Reject H_0	Do not reject H_0
Reality (Unknown)	H_0 is true	False positive Type I error (α)	True negative (1-α)
	H_1 is true	True positive Power (1-β)	False negative Type II error (β)

The study is designed to control the Type I and II errors to small levels for primary test(s)

Recommended:

- Type I error (α): 5 to 20%
- Type II error (β): 10 to 20% or power (1- β): 80 to 90%

} **Pre-defined!**

EORTC
European Organisation for Research and Treatment of Cancer
The future of cancer therapy

Data analysis and interpretation

Activity of crizotinib in MET+ CCSA

The primary end point was the ORR as assessed by the local investigator, with response confirmation. An objective PR was observed in 1/26 *MET*+ patients [3.8% ORR; 95% confidence interval (CI): 0.1% to 19.6%]. The primary end point of the trial was not met. The duration of response in the responding patient was 851+ days; the patient is still on active treatment at the data cut-off, having received 40+ cycles of treatment. RECIST SD was observed in 17/26 *MET*+ patients (65.4%). Disease progression was seen in eight patients (30.8%). Disease control was achieved in 18/26 *MET*+ patients (DCR; 69.2%, 95% CI: 48.2–85.7). The median PFS was 131 days (95% CI: 49–235; supplementary Figure S5, available at *Annals of Oncology* online). The 3, 6, 12 and 24 month PFR was 53.8% (34.6–73.0), 26.9% (9.8–43.9), 7.7% (1.3–21.7) and 7.7% (1.3–21.7), respectively. The median OS was 277 days (95% CI: 232–442), and the OSR was 36.1% (95% CI: 18.2% to 54.3%) at 1 year and 9.4% (95% CI: 1.7% to 25.3%) at 2 years (supplementary Figure S6, available at *Annals of*

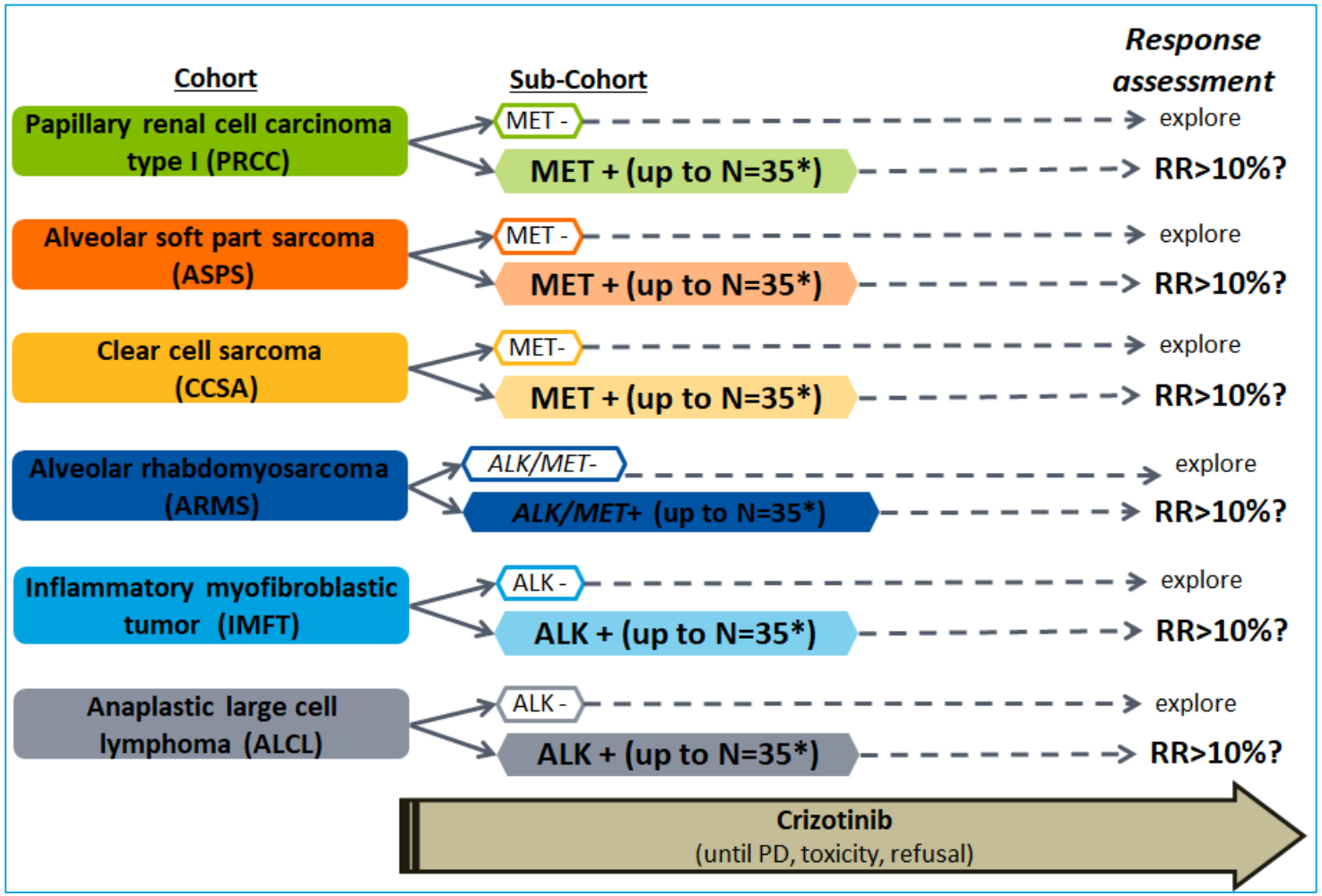
Discussion

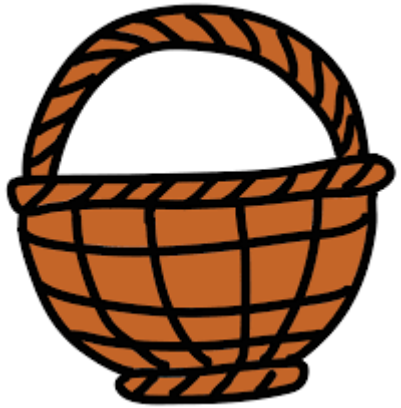
Multiple factors led to over-recruitment of patients beyond stage 1 of Simon's optimal two-stage design. We saw rapid accrual of CCSA cases, with the majority of patients previously untreated, reflecting the high unmet medical need in this orphan

temic agents. Response rate should not be the primary end point for future phase II trials with MET inhibiting agents for CCSA. DCR, PFS and/or PFR would be a more appropriate reflection of the therapeutic effects of treatments in this disease, where progression arrest might be more important than shrinkage of the tumour and its metastasis. Other MET inhibitors, such as small

logy and molecular characterization. Given the inherent limitations of performing larger prospective trials in ultra-rare diseases, innovative trial methodology like the basket approach chosen in EORTC 90101, and new regulatory mechanisms are required to provide patients with orphan malignancies with potentially active drugs such as crizotinib.

This was a study…
…part of EORTC « CREATE » basket trial



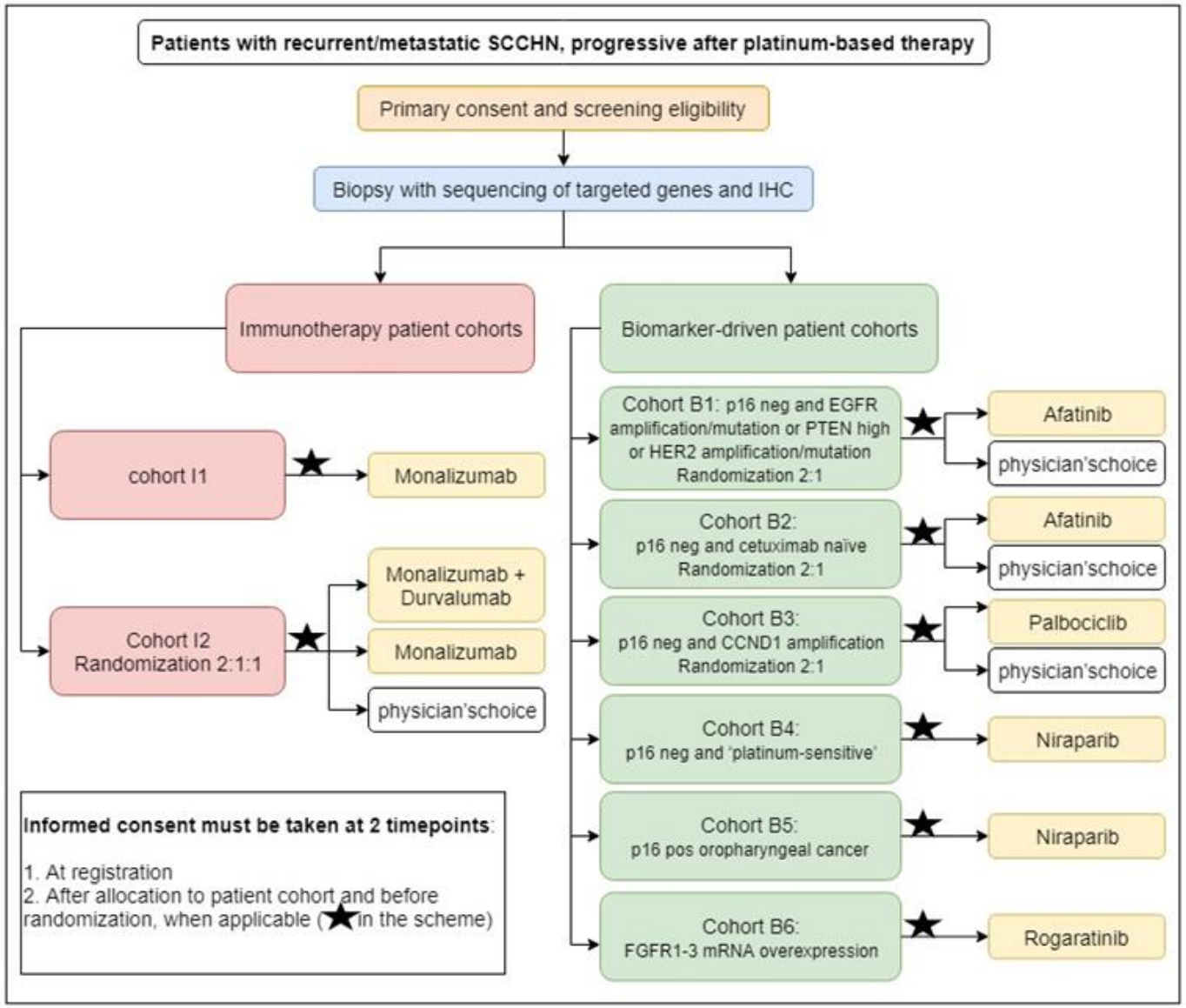


Single treatment

Single biomarker

Different histologies placed in baskets

EORTC « UPSTREAM » : an umbrella trial



Single histology

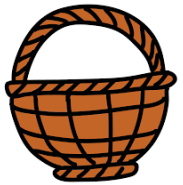
Multiple biomarkers

Each matched to treatments

Basket & Umbrella studies

- **Common** protocol format with parallel studies differentiated by the marker-treatment combination
- **Centralized** screening platform with timely molecular profiling
- Sub studies can be **added or dropped**
- Sub studies may share **common** design features or be designed **differently**

Basket



Single treatment

Single biomarker

Different histologies placed in baskets

Umbrella

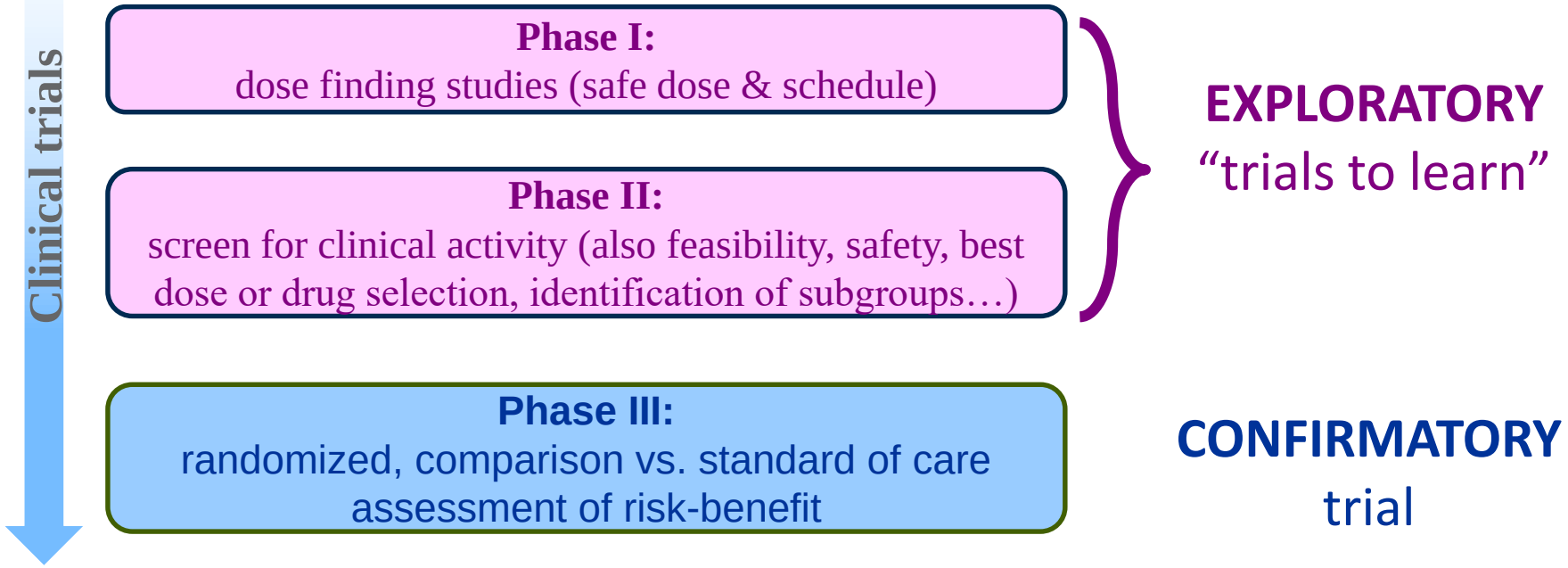
Single histology

Multiple biomarkers

Each matched to treatments

Renfro & Sargent. Ann Oncol. 2017.

From “trials to learn” to “confirmatory trials”



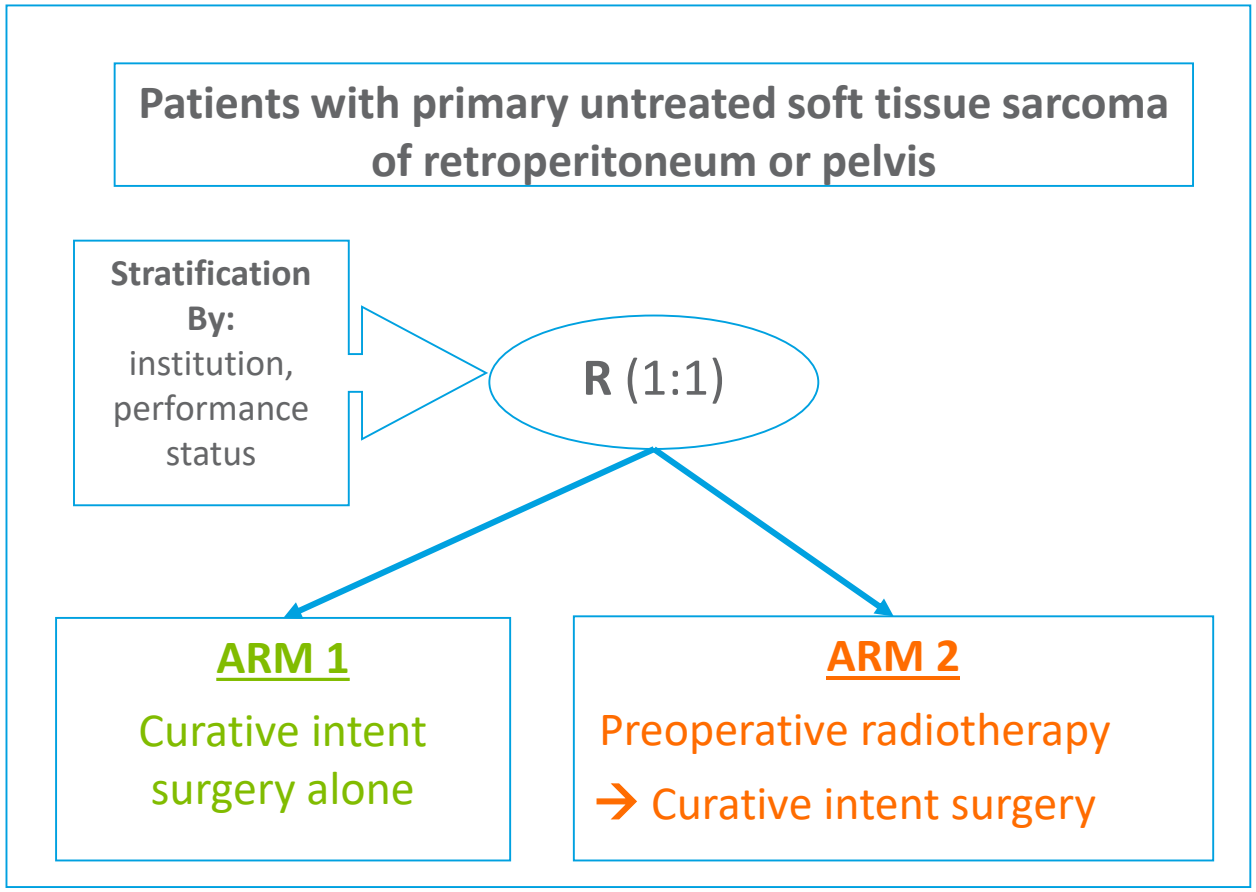
Phase I / II: *phase I continued as phase II*

Phase II / III: *phase II continued as phase III*

Time for first questions ?

EORTC 62092- 22091 – 'STRASS'

A phase III randomized study of preoperative radiotherapy plus surgery versus surgery alone for patients with Retroperitoneal sarcoma (RPS)



FOLLOW UP DISEASE EVALUATION

Arm 1: 14, 24, 36, 48 weeks after randomization and Q6 mo thereafter until recurrence or death.

Arm 2: 24, 36, 48 weeks after randomization and Q6 mo thereafter until recurrence or death.

FOLLOW UP SURVIVAL

Both arms: after recurrence Q6 mo.

Objectives and design

Objective(s)	The main objective is to assess whether preoperative radiotherapy, as an adjunct to curative-intent surgery, improves the prognosis of patients with RPS. Primary objective To assess whether there is a difference in **abdominal recurrence-free survival** between the two arms Secondary objectives • To assess whether there is a difference in metastasis-free survival, abdominal recurrence free interval and overall survival between the two arms • To assess tumor response in patients undergoing preoperative radiotherapy • To assess the toxicity profile of preoperative radiotherapy
Methodology	Superiority phase III trial with stopping rules for the tolerance to protocol treatment
Sample Size & Study Duration	256 patients – Total study duration = 39 + 41 months = 6.7 years

Primary endpoint: Time-to-event endpoint

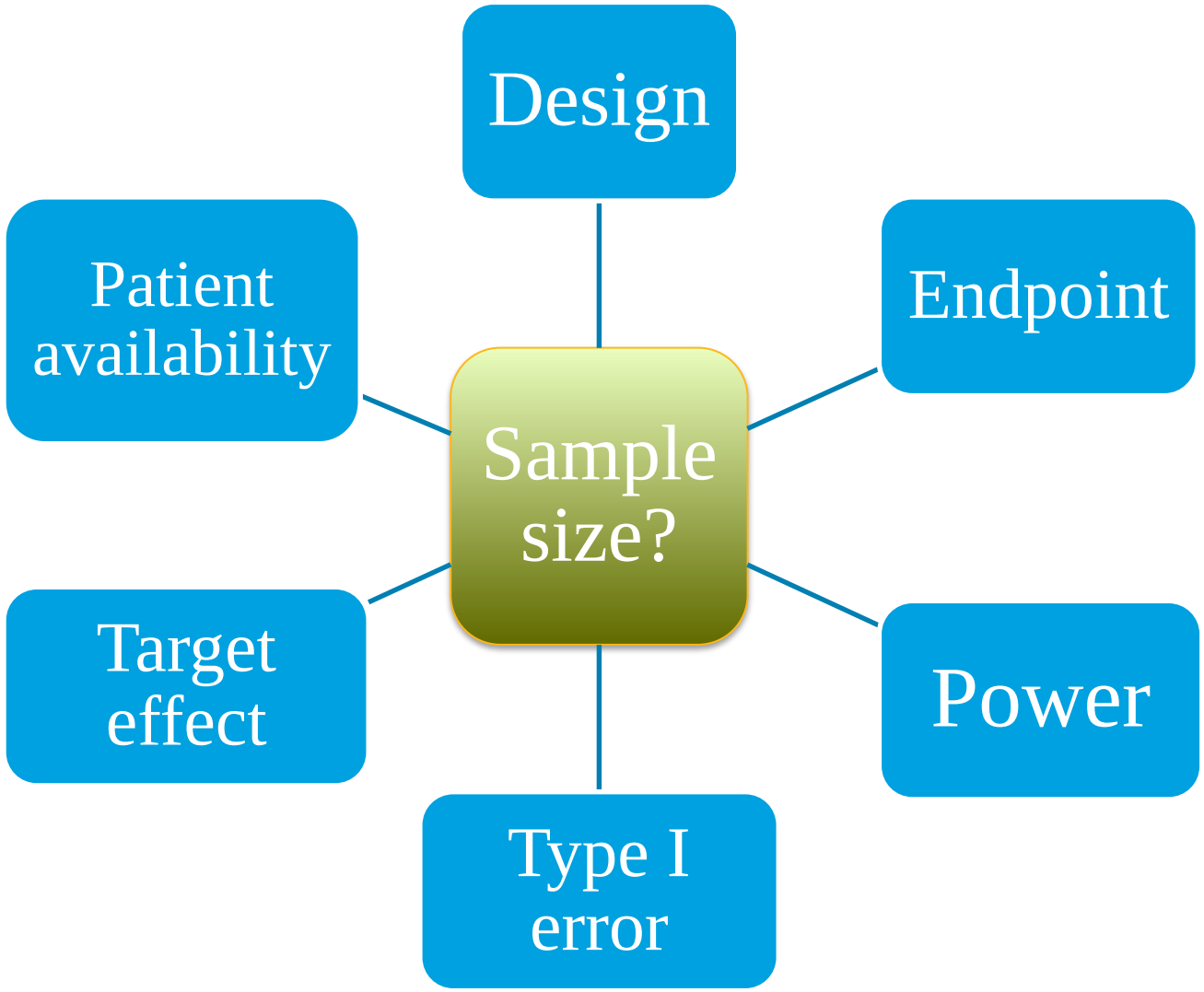
- Abdominal failure free survival: measured from the date of randomization (as reference) to the date of abdominal relapse or death, whichever occurs first.

- Abdominal relapse is defined as:
  - Local relapse (after macroscopically complete resection)
  - Macroscopic residual disease after surgery (R2)
  - Progressive disease during preoperative radiotherapy and/or tumor becoming non-resectable
  - Peritoneal sarcomatosis (presence of peritoneal metastasis)

  Liver metastases will be regarded as distant metastatic events, rather than abdominal recurrence. For patients with distant metastases they will be followed until local failure will be detected.

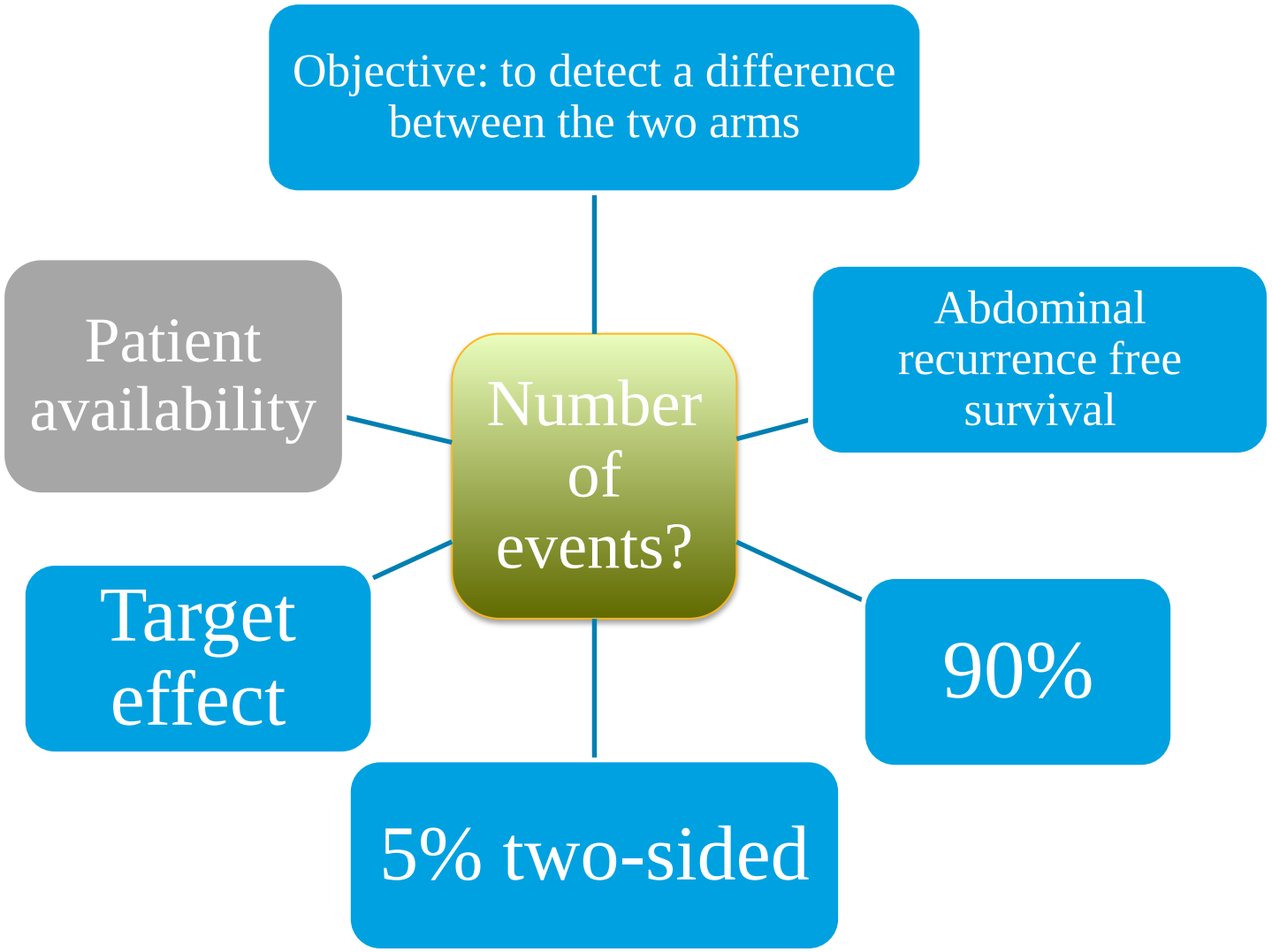
- Patients without one of these events will be censored at the date of last follow-up.

Sample size calculation for time to event endpoints: 2 steps calculation



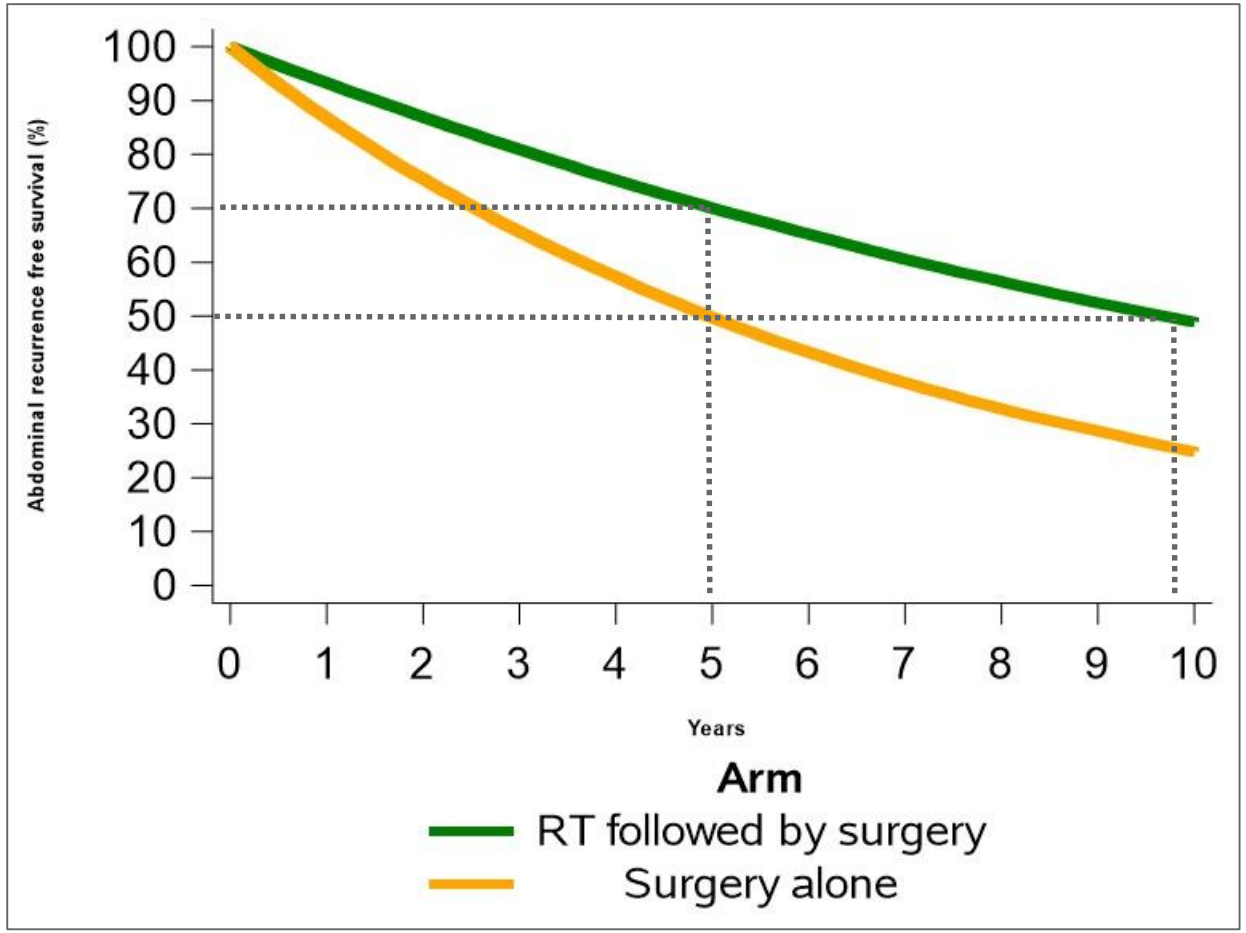
EORTC
European Organisation for Research and Treatment of Cancer
The future of cancer therapy

Step 1: Number of events



Hazard Ratio

Assuming “exponential survival curves”



At 5years : 50% versus 70%

→ Hazard ratio: HR =ln(0.70)/ln(0.5)=0.52

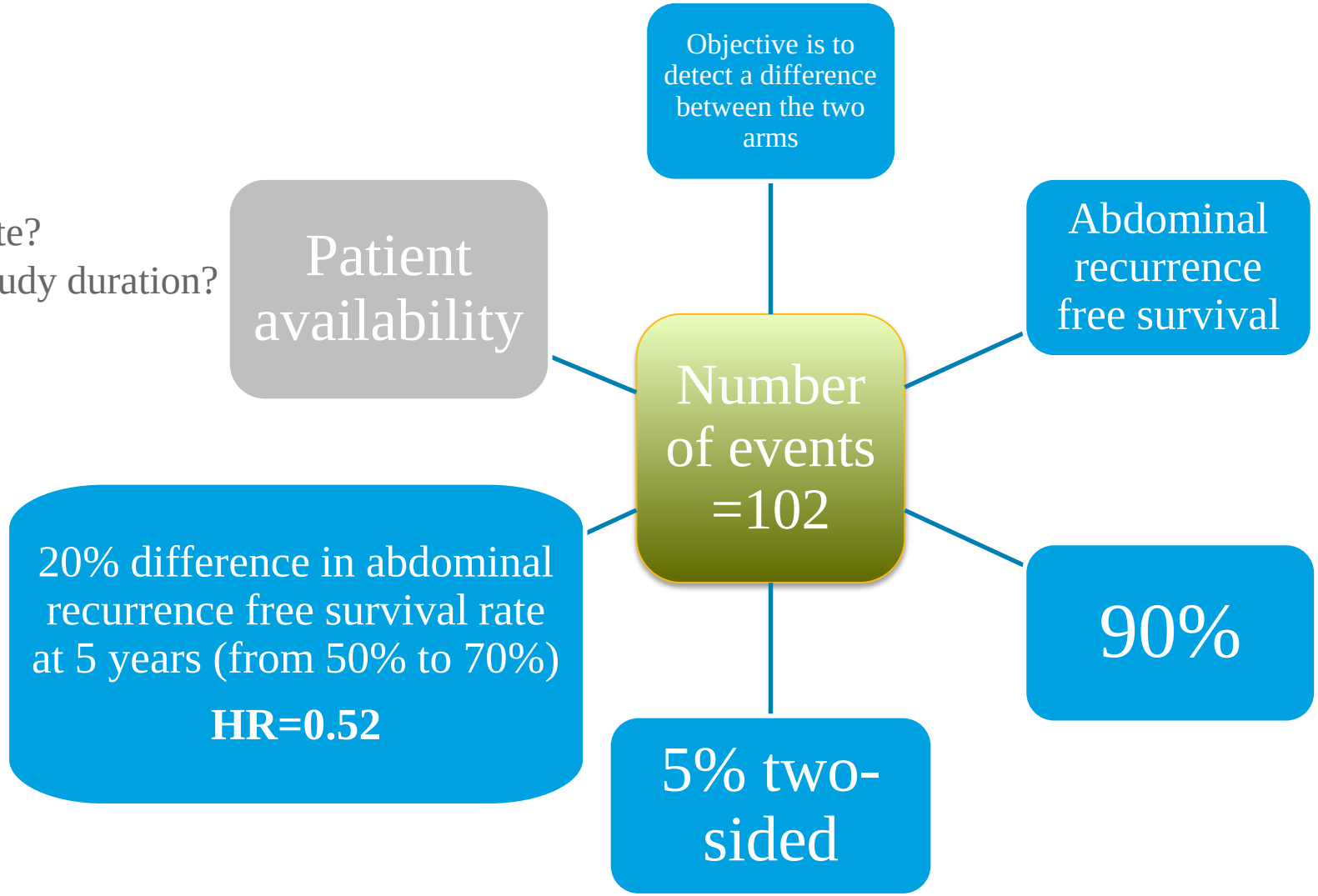
HR<1 : Exp better than std
HR=1: Exp = std
HR>1 : Exp worse than std

Median: 5y versus 9.7y

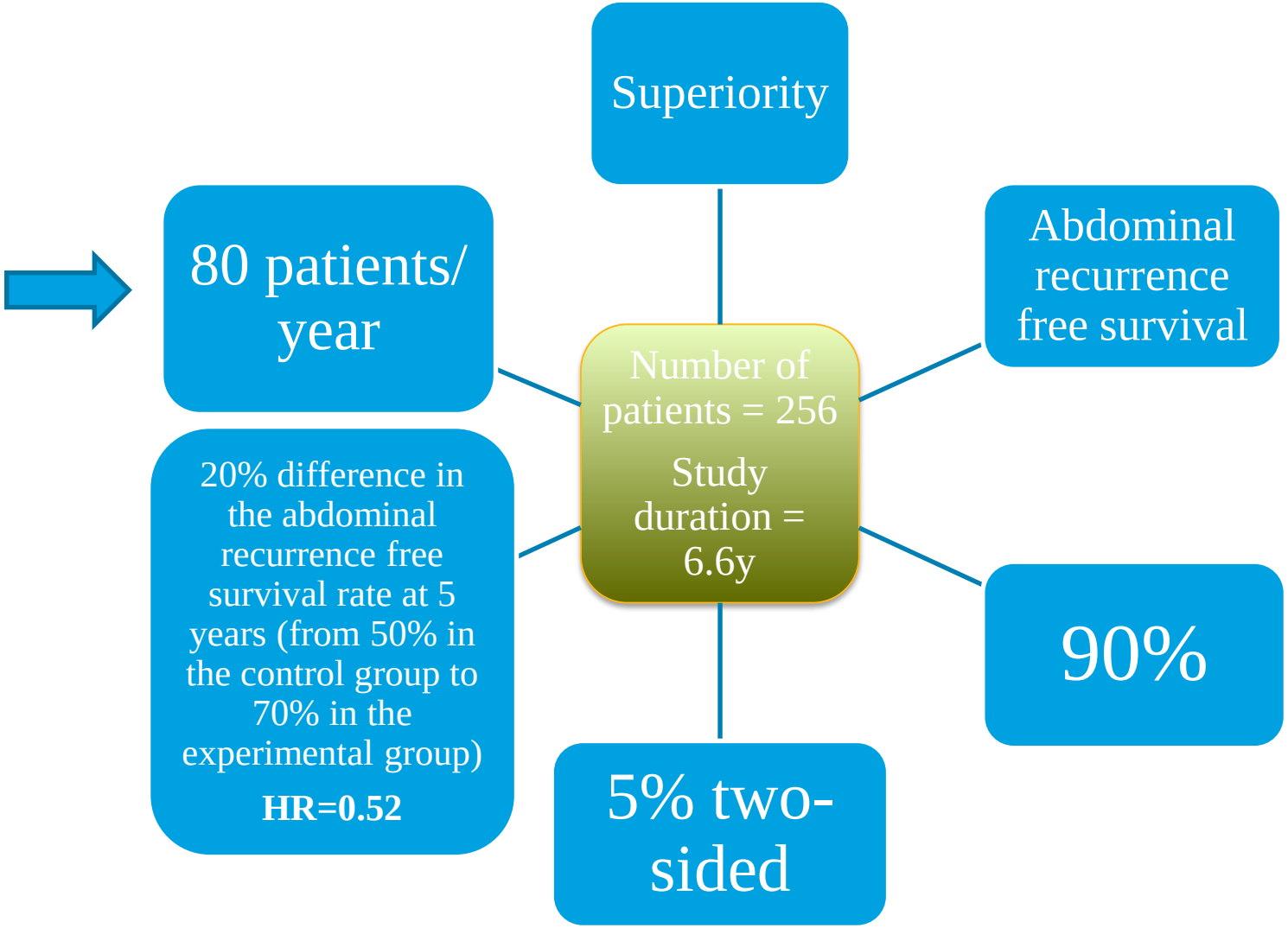
→ HR = 5/9.7=0.52

Step 1: Number of events

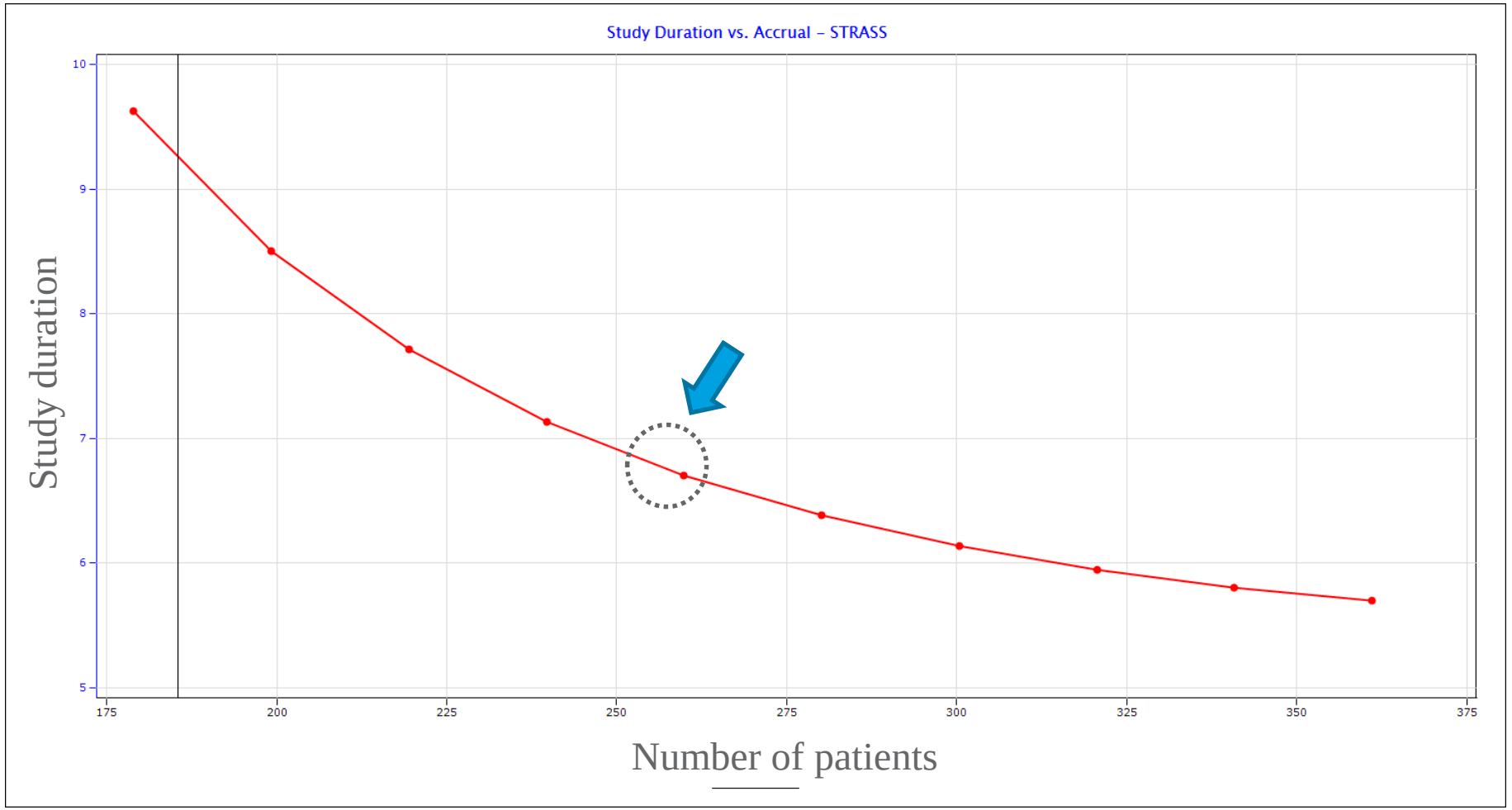
- Accrual rate?
- Feasible study duration?
- (Budget?)



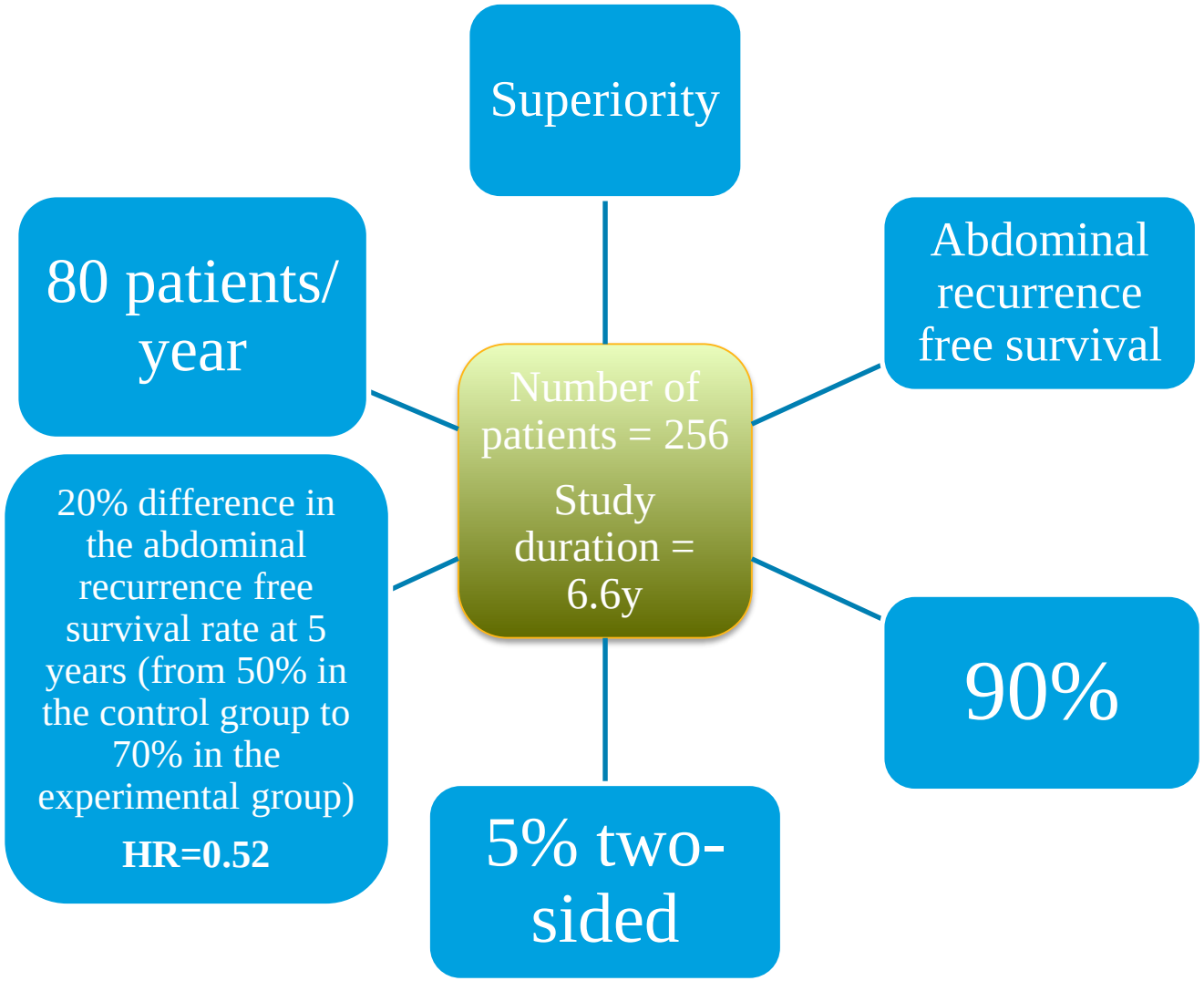
Step 2: Number of patients and study duration



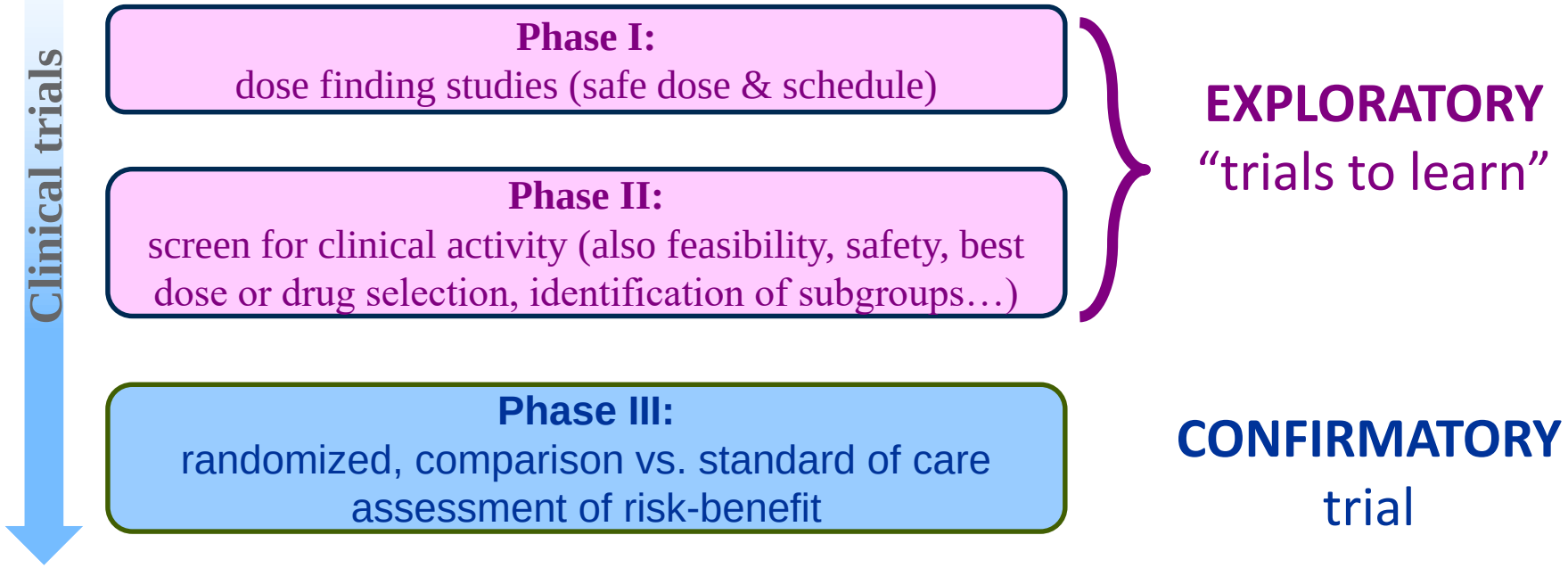
Number of patients versus study duration required to observe 102 events if accrual rate = 80 patients / year



Sample size calculation for time to event endpoints



From “trials to learn” to “confirmatory trials”

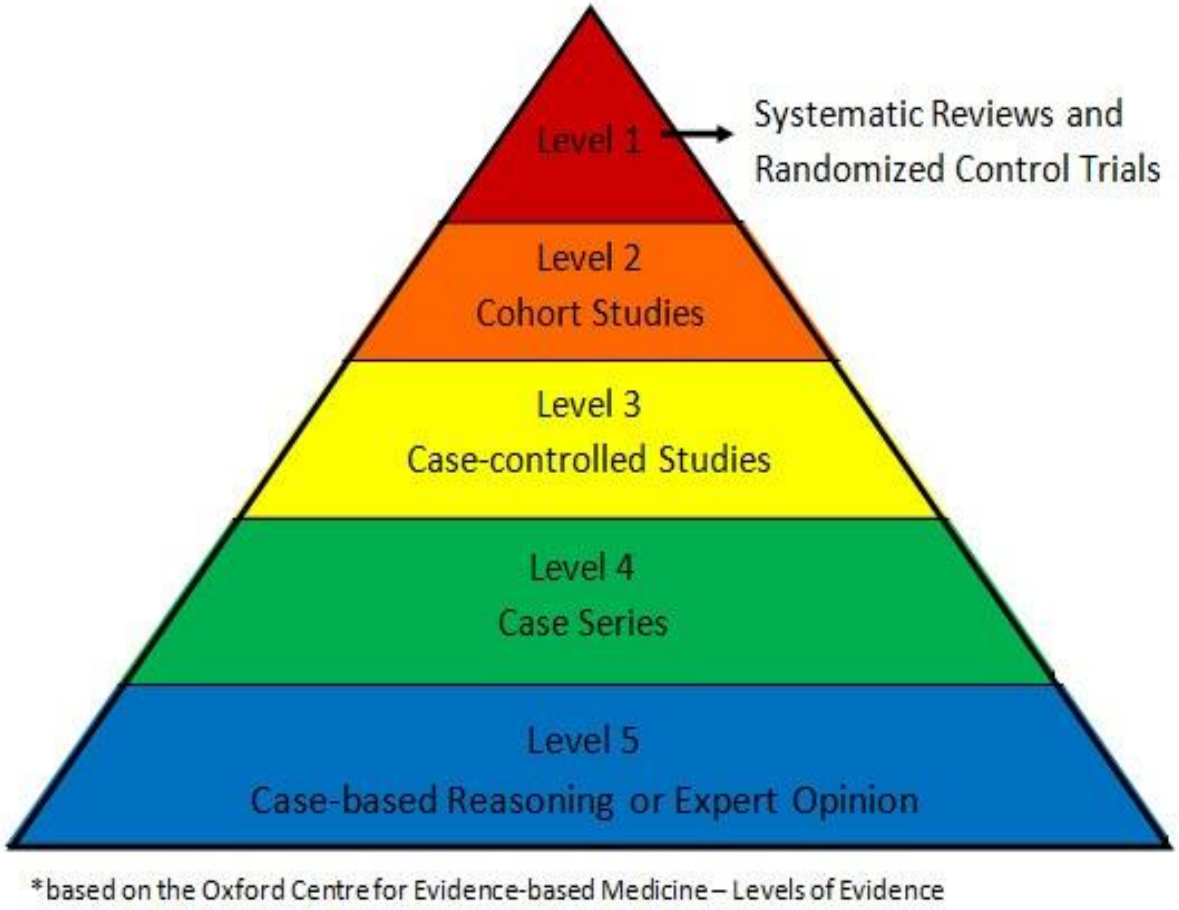


Phase I / II: *phase I continued as phase II*

Phase II / III: *phase II continued as phase III*

Evidence based medicine

Quality of evidence



*based on the Oxford Centre for Evidence-based Medicine – Levels of Evidence

Some questions ?

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…

BDA - Le Deley 19/11/2013 3

Where can we compromise?

- Do we need a control arm?
- Can we allow more uncertainty?
- What about a Bayesian approach?

‘CREATE’ : single arm study

- 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?
- Same imaging / diagnostic tools for assessing outcome?

- Single arm, α=10% for pre-specified p0 (H0) - a 5% absolute error in p0 increases false positive rate to 30%!

‘STRASS’ : a randomized clinical trial (RCT)

- Avoids selection bias
- Ensures that the treatment arms are comparable both in terms of **known** and **unknown** prognostic factors
- Any “time effects” are balanced across arms

But …

- Expensive
- Longer
- Patients reluctant to be allocated to a treatment at random

Equipoise: at the time of the trial, it is not known which treatment is best

Where can we compromise?

- Do we need a control arm?
- Can we allow more uncertainty?
- What about a Bayesian approach?

Allow more uncertainty: relaxing the errors
Back to 'STRASS' study

Remember:

- **Alpha/Type I error rate:** probability of a false positive
- **Power:** probability of finding a meaningful effect

Targeted difference:
HR = 0.52

Number of events

Alpha \ Power	90%	85%	80%	75%
5%	102	88	77	68
10%	84	70	61	53
15%	73	60	51	44
20%	64	53	45	38

→ 120 patients
4.4y study duration

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

Where can we compromise?

- Do we need a control arm?
- Can we allow more uncertainty?
- What about a Bayesian approach?

EORTC
European Organisation for Research and Treatment of Cancer
The future of cancer therapy

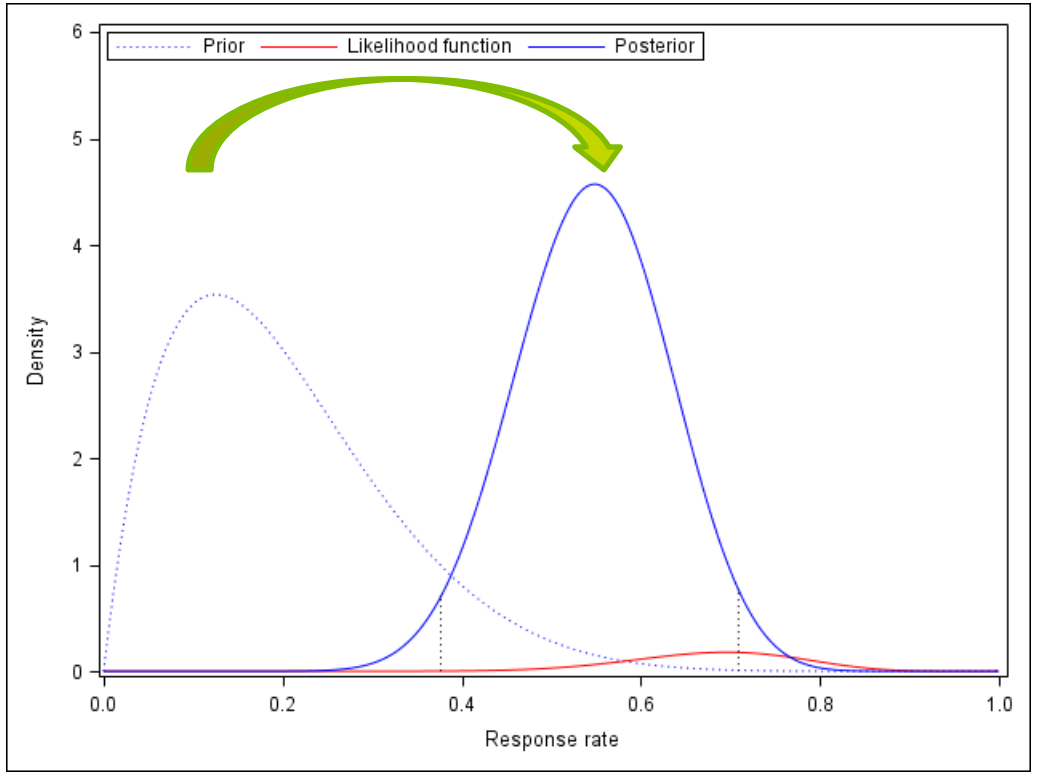
IRCI International Rare Cancers Initiative
Aspiring to improve the lives of patients with rare cancers

Clinical trial designs for rare diseases: studies developed and discussed by the international rare cancers initiative

Table 1
Trial names and registration.

Short name	Full name	Registration number
BALLAD/IRCI 002	A study to evaluate the potential benefit of adjuvant chemotherapy for small bowel adenocarcinoma (SBA)	?
Androgen deprivation therapy in advanced SGCs/IRCI 007	A randomised phase II study to evaluate the efficacy and safety of chemotherapy (CT) versus androgen deprivation therapy (ADT) in patients with recurrent and/or metastatic, androgen receptor (AR) expressing, salivary gland cancer (SGCs)	NCT01969578 Recruiting
HGUS/IRCI 006	A randomised double-blind phase II study evaluating the role of maintenance therapy with cabozantinib in High Grade Undifferentiated Uterine Sarcoma (HGUS) after stabilization or response to doxorubicin +/- ifosfamide following surgery or in metastatic first line treatment	Recruiting
InterAACT/IRCI 003	An International multicentre open label randomised phase II Advanced Anal Cancer Trial comparing cisplatin plus 5-fluorouracil versus carboplatin plus weekly paclitaxel in patients with inoperable locally recurrent or metastatic disease	?
rEECur	Trial of chemotherapy for relapsed and refractory Ewing sarcoma	Recruiting
GOG-0277/IRCI 001	A phase III randomised trial of gemcitabine plus docetaxel followed by doxorubicin versus observation for uterus-limited, high-grade uterine leiomyosarcoma	Active, not recruiting
MEKi ± AKTi in UM/IRCI 005	A randomised two-arm Phase II study of Trametinib alone and in combination with GSK2141795 in patients with advanced uveal melanoma	Completed
InPACT/IRCI 004	International Penile Advanced Cancer Trial	Recruiting

The basics



Prior distribution + Observed data = Posterior distribution

The prior distribution

- Can represent **current knowledge or belief** of the efficacy of the tested regimen.

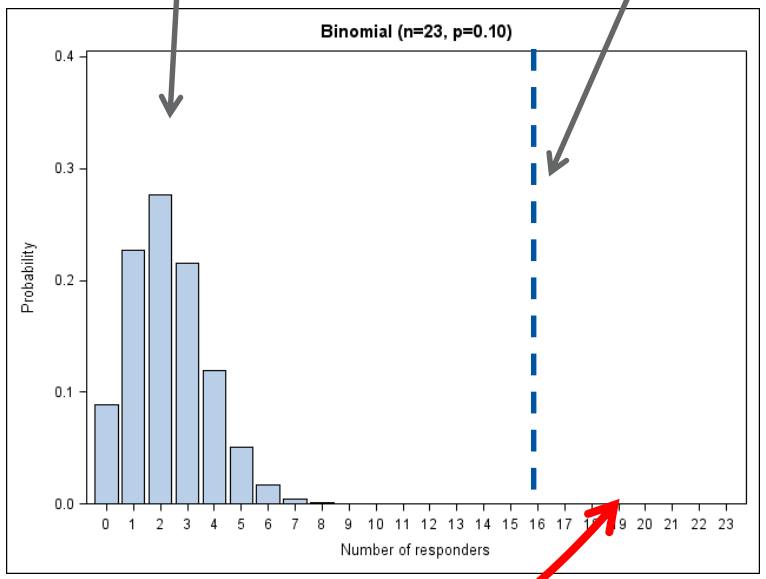
- There is no unique way to choose a prior:
  - Literature review,
  - previous studies and meta-analyses,
  - elicitation of experts opinion.

Classical (**frequentist)** analysis

Test H0: ORR=10% with 23 patients

Distribution of number of responses if p=10%

Observed number of responses = 16/23



p-value <.0001

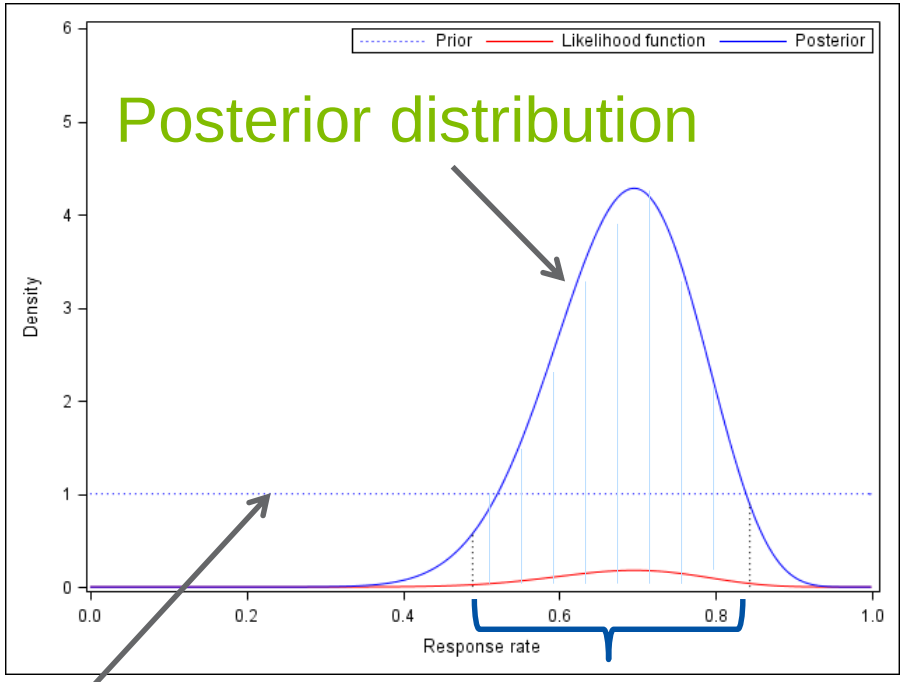
Interpretation

What is the **probability of observing the data** given a pre-specified response rate?

The probability of observing at least 16 responses if H0 true is very low (<0.01%)

→ Reject H0: ORR=10%

Bayesian analysis
Posterior distribution



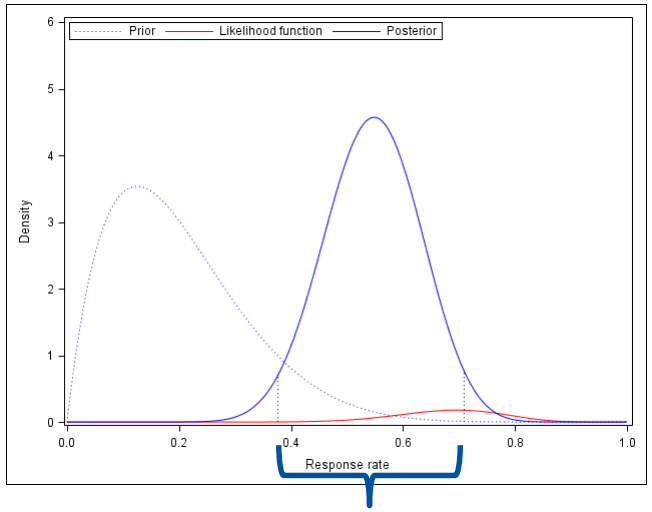
Prior distribution (uninformative)

Interpretation

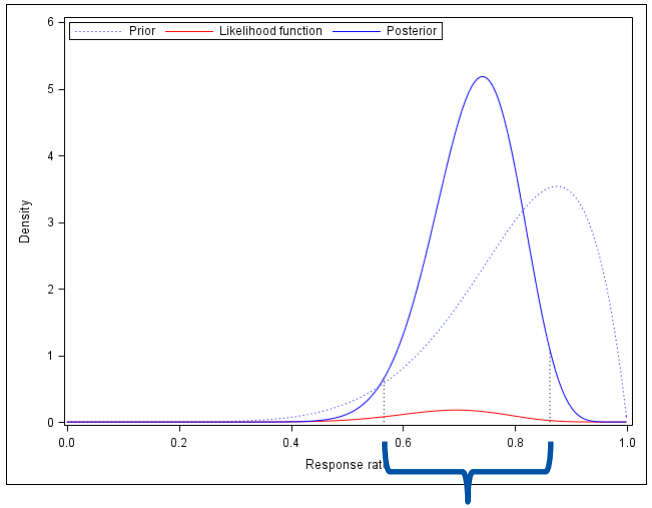
What is the **probability of a given response rate** given the data?

- There is 95% probability that the true response rate lies between 49% and 84%
- The probability that the true response rate is >60% is 81%

Bayesian analysis – Depending on the prior



Pessimistic prior:
There is 95% probability that the true response rate lies between 38% and 71%.



Optimistic prior:
There is 95% probability that the true response rate lies between 57% and 86%.

Bayesian approach

Advantages

- Provides a natural way of updating prior knowledge with newly observed data.
- Provides interpretable answers, such as “the probability that the true response rate is >60% is 81%”

Disadvantages

- It does not tell you how to select a prior.
- Posterior distributions can be heavily influenced by the priors, specially in small trials.

Incorporate Bayesian elements

- *IRCI 002 Small Bowel Adenocarcinoma*
  - Use relaxed type I error
    - Power: 80-90%, Type I error: 20% (one-sided), HR = 0.75
  - If significant, combine with clinician estimates of treatment benefit based on external evidence to provide a more robust estimate
- *IRCI 004 Advanced Penile Cancer*
  - Reverse philosophy: see how many patients could be recruited and then assess whether data has sufficient value
  - Rather than focusing on hypothesis testing, this study would focus on estimation, combining with different prior assumptions (non-informative, sceptic, extreme sceptic, enthusiast)

To conclude on Bayesian designs as compared to classical (frequentist) designs

Methodology	Advantages	Disadvantages
Frequentist	Formal control of type I and type II error	Increased sample size
Bayesian	Inclusion of informative prior information. Particularly useful when a previous trial has data that can be used to inform the prior Smaller sample size Continual assessment is possible	Informative priors reduce sample size, which will increase the trial type I and type II error

"The Bayesian approach provides greater design flexibility, but does not provide additional value over the frequentist approaches when the prior is non-informative."

In rare settings, in which Bayesian is often promoted, often no robust prior information available

EORTC
European Organisation for Research and Treatment of Cancer
The future of cancer therapy

Statistics : a bit less of a black box?

EORTC
European Organisation for Research and Treatment of Cancer
The future of cancer therapy

Thank you !

And thank you to my colleagues statisticians from the EORTC, specially:
Saskia Litiere & Anouk Neven

More questions ?

Some useful reading

- Schöffski et al. Activity and safety of crizotinib in patients with advanced clear-cell sarcoma with MET alterations: European Organization for Research and Treatment of Cancer phase II trial 90101 'CREATE'. Ann Oncol. 2017 Dec 1;28(12):3000-3008.
- Eisenhauer et al. New response evaluation criteria in solid tumours: revised RECIST guideline (version 1.1). Eur J Cancer. 2009 Jan;45(2):228-47.
- Simon R (1989) Optimal Two-stage Designs for Phase II Clinical Trials. Controlled Clinical Trials, 10: 1-10.
- Baey C, Le Deley MC. Effect of a misspecification of response rates on type I and type II errors, in a phase II Simon design. Eur J Cancer. 2011 Jul;47(11):1647-52.
- Renfro, L.A. and D.J. Sargent, Statistical controversies in clinical research: basket trials, umbrella trials, and other master protocols: a review and examples. Ann Oncol, 2017. **28**(1): p. 34-43.
- Mauer et al. Adaptive designs at European Organisation for Research and Treatment of Cancer (EORTC) with a focus on adaptive sample size re-estimation based on interim-effect size. Eur J Cancer. 2012 Jun;48(9):1386-91.
- Bogaerts, Sydes et al. Clinical trial designs for rare diseases: studies developed and discussed by the International Rare Cancers Initiative. Eur J Cancer. 2015 Feb;51(3):271-81.
- Abrahamyan et al. A new toolkit for conducting clinical trials in rare disorders. J Popul Ther Clin Pharmacol. 2014;21(1):e66-78.
- Dutton P, Love SB, Billingham L, Hassan AB. Analysis of phase II methodologies for single-arm clinical trials with multiple endpoints in rare cancers: An example in Ewing's sarcoma. Stat Methods Med Res. 2018 May;27(5):1451-1463.

Backup slides

Where can we compromise?

- Do we need a control arm?
- Can we allow more uncertainty?
- What about a Bayesian approach?
- Can we adapt the design?

EORTC
European Organisation for Research and Treatment of Cancer
The future of cancer therapy

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

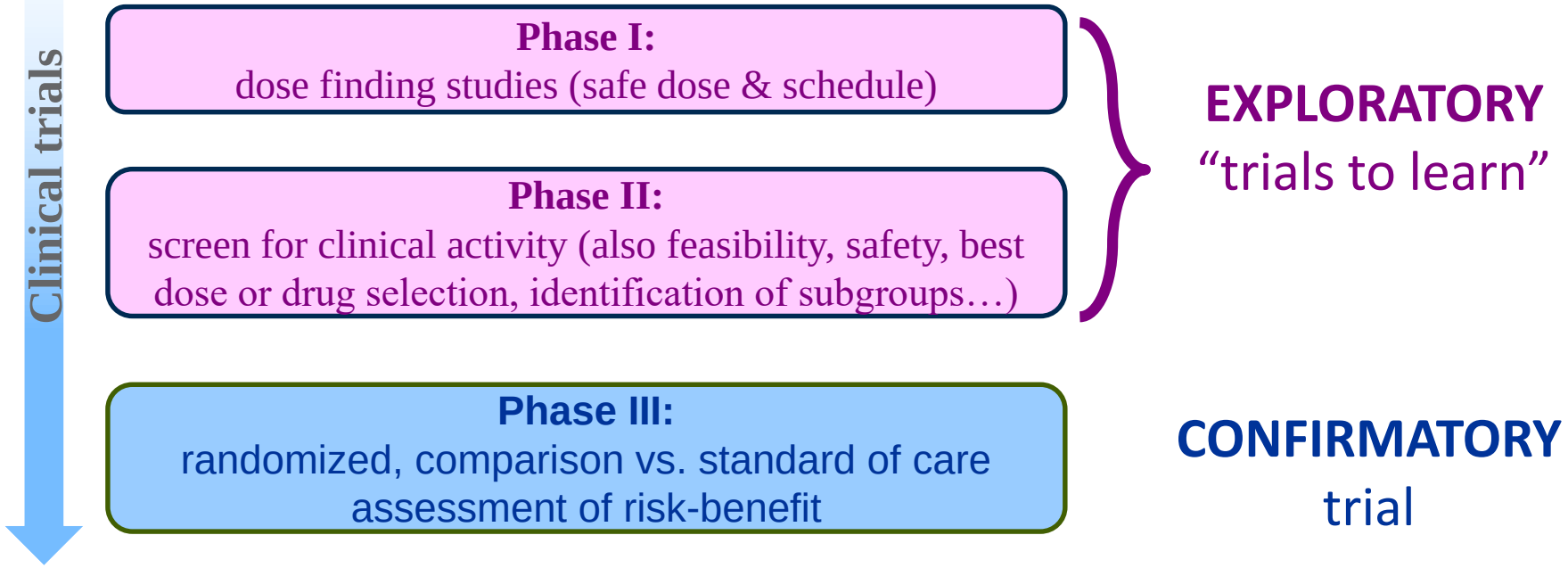
Why do we demand rigorous testing characteristics in a "frequentist" framework, but leave it up to the eye of the "Bayesian" beholder to judge on

- 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

Where can we compromise?

- Do we need a control arm?
- Can we allow more uncertainty?
- What about a Bayesian approach?
- Can we adapt the design?

From “trials to learn” to “confirmatory trials”

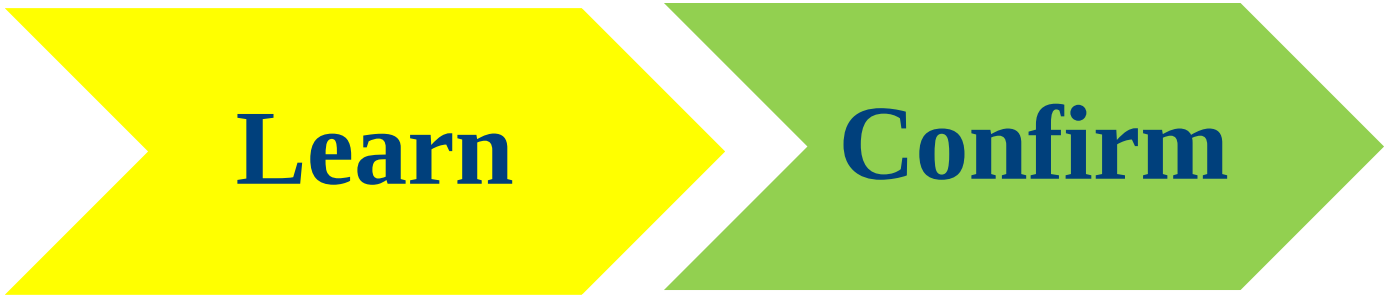


Phase I / II: *phase I continued as phase II*

Phase II / III: *phase II continued as phase III*

Is this possible?

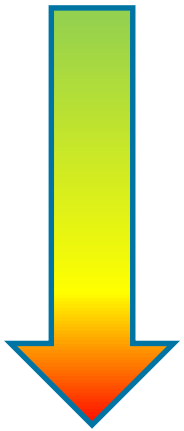
One trial



Change H_0, H_1?
Change design parameters?

Adaptive designs

Well-known



Less understood

- 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 …

Mauer et al. Adaptive designs at European Organisation for Research and Treatment of Cancer (EORTC) with a focus on adaptive sample size re-estimation based on interim-effect size. Eur J Cancer. 2012 Jun;48(9):1386-91.

The challenges

Decision	In truth, the null is... True	In truth, the null is... 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!

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

EORTC
European Organisation for Research and Treatment of Cancer
The future of cancer therapy

A word about non-inferiority trials

EORTC study 1762 – ‘REDUCE’

Reduced dose-density of denosumab for maintenance therapy of unresectable giant cell tumor of bone: a multicenter phase II study

Patients after >=1yr treatment with denosumab (dose of 120mg SC **every 4 weeks**)

Enrollment

Denosumab continued until progression, unacceptable toxicity or patient withdrawal (dose of 120mg SC **every 12 weeks**)

Objectives, design and primary endpoints

Objective(s)	The primary objective of the trial is to evaluate the risk versus benefit of denosumab in maintenance setting in patients requiring long-term use (> 1 year) of denosumab. For that purpose, the treatment schedule with reduced dose density (120mg SC 12-weekly instead of 4-weekly) will be investigated, starting after 1-year (12-15 months) of denosumab full dose, as per current label. The impact on osteonecrosis of the jaw (ONJ) without compromising disease control will be assessed. Secondary objectives are to assess the effect of the reduced dose density schedule on patient clinical and self-reported outcomes, and on the toxicity profile of the treatment.
Methodology	Multi-center, multi-national, open label, single arm phase 2 study of single-agent denosumab. The two co-primary endpoints will be tested hierarchically: 1) PFS at 2 years after enrollment: H0 90% vs H1 >90% 2) ONJ incidence at 2 years after enrollment: H0 6% vs H1 <6%
Sample Size & Study Duration	One hundred (100) patients - Study duration: 5 years = 3y accrual + 2y follow-up

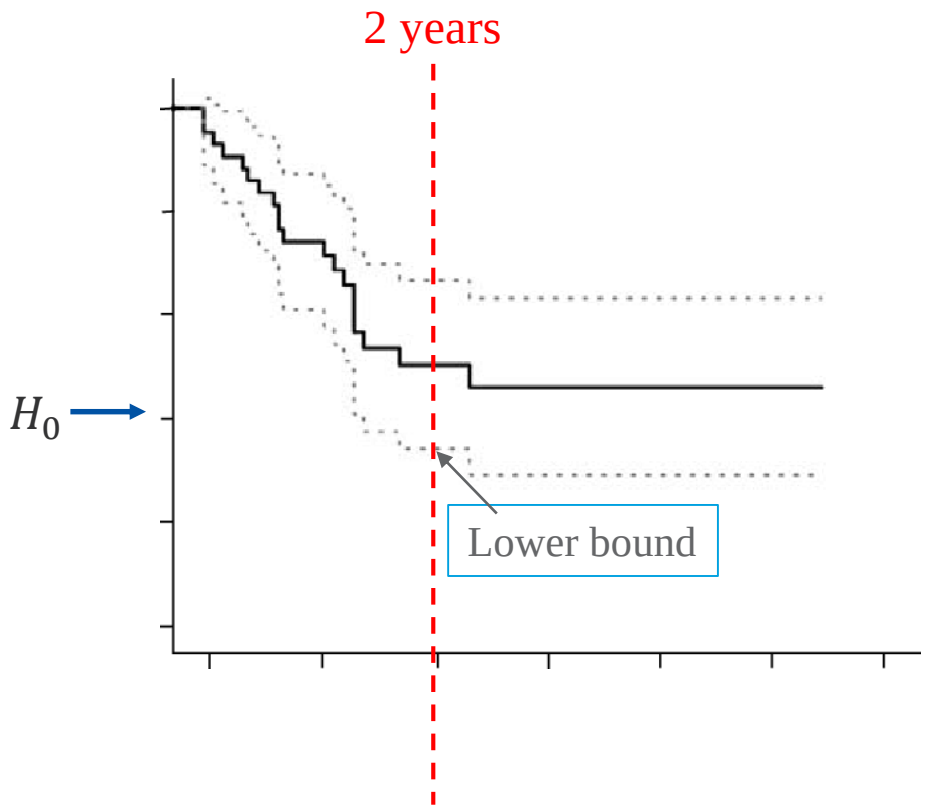
A non-inferiority question

- Background information: Denosumab provides long-term disease control in unresectable or metastatic GCTB with a 3-year PFS estimate of 94.0% and a 5-year PFS of 88% (Palmerini et al, Ann Oncol 2017)

- Lowering the denosumab dose might be an alternative way:
  - To reduce toxicity, i.e. the incidence of ONJ
  - Without compromising disease control (PFS rate) in comparison with historical control

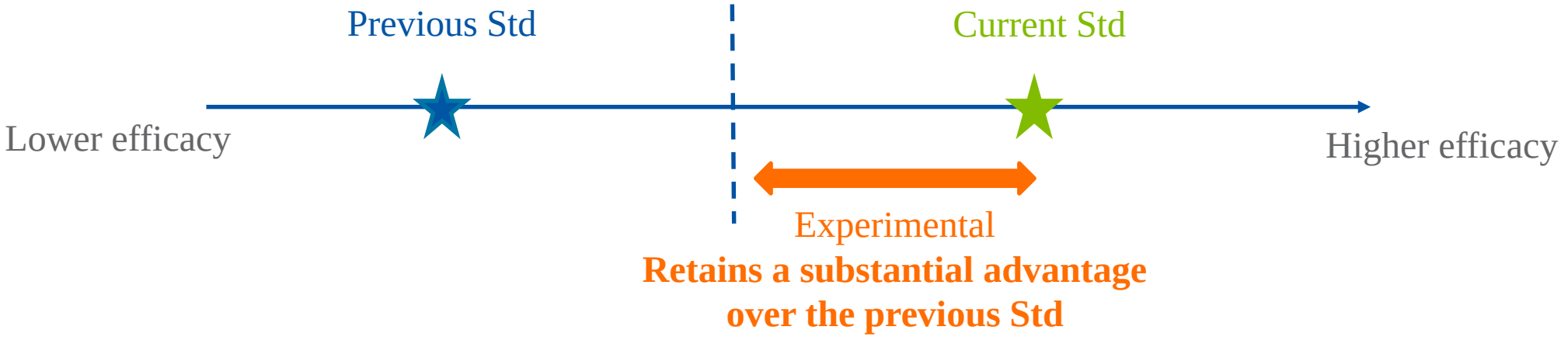
H0: PFS at 2 years = 90%

→ This lower PFS rate is considered to be acceptable, given the expected decreased toxicity



How to choose the non-inferiority margin?

The margin should be less than half the difference that was shown when the standard arm was established



This is why non-inferiority studies require more patients than superiority studies

Co-primary endpoints
Impact on Type I error and power

When statistical significance is required for both endpoints:

- No adjustment of Type I error is necessary

- The overall Type II error is inflated but is lower or equal to the sum of the Type II errors connected to the individual tests. This must be taken into account to power the study adequately.

	PFS at 2 years	ONJ incidence at 2 years	Overall (assuming zero correlation between the 2 endpoints)
Type I error	7.1%	6.9%	7.1% = max(7.1%, 6.9%)
Power	92.2% (under H1: 97%)	91.2% (under H1: 1%)	84.1% = 0.922 x 0.912 (under both H1)

*: assuming 10% drop-out