A framework for the design and analysis of phase III randomised trials in uncommon diseases

Tim Morris, Matthew Sydes, Mahesh Parmar

MRC Clinical Trials Unit at UCL

1 Dec 2015

London Hub

Standard way of designing large trials

Generally accepted by

- funders, regulators, patients, other researchers

Design parameters (type I error, power, targeted effect size,…)

- Lead to large trials
- ‘Doable’ in a reasonable timeframe

What should we do when standard trial design is just too large?

Something has to change in our thinking

Much attention now being paid to very small populations:

- **INSPIRE:** Innovative methodology for small populations research
- **IDEAL:** Integrated design & analysis of small population group trials
- **ASTERIX:** Advances in small trials design for regulatory innovation and excellence

and more…

How can we change our thinking?

- Sacrifice concurrent control?
- Sacrifice randomisation?
- Switch to a Bayesian inferential framework?

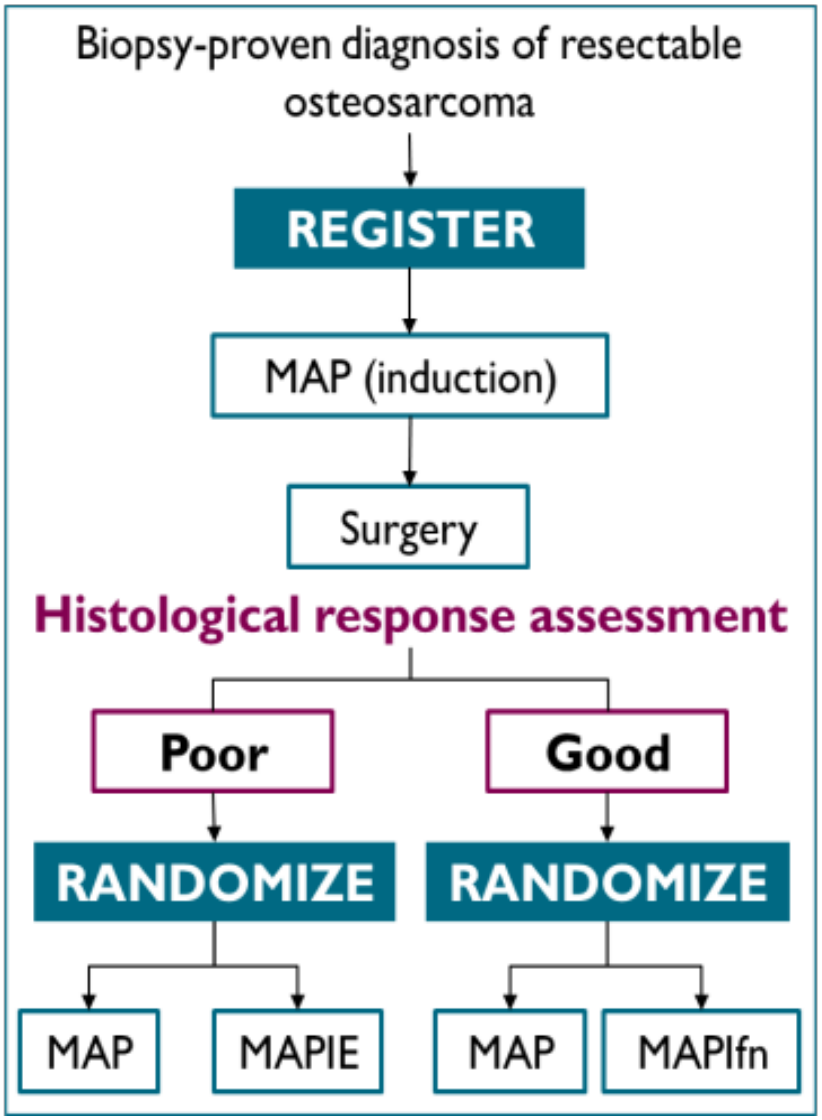
Unfamiliar to funders, regulators, patients and other researchers

How do they compare quality of evidence to a more traditional design?

→ **Reduced acceptance?**

What to do between the extremes?

EURAMOS-1: Bone Tumours



Rare disease

US : <1 case / 1,500 people

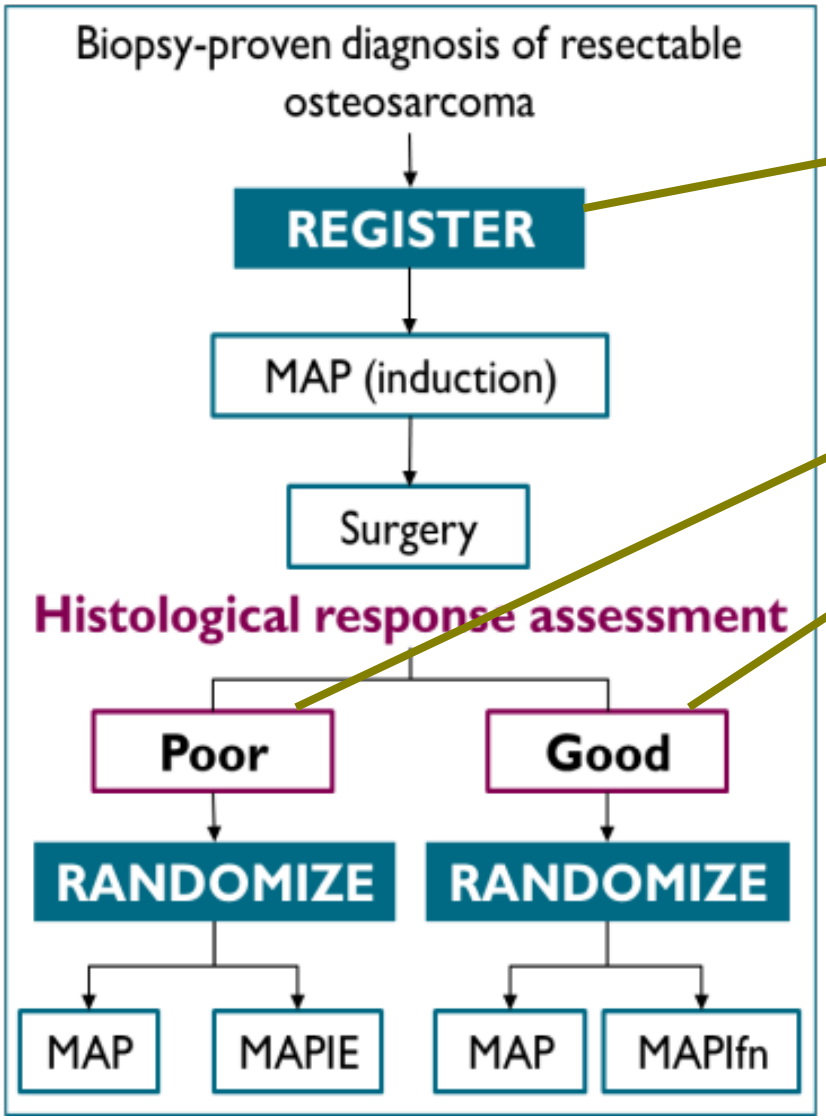
EU : <1 case / 2,000 people

Japan: <1 case / 2,500 people

Osteosarcoma = 200 cases/yr UK

1 case / 300,000 people

EURAMOS-1



Accrual targets

1,400 patients registered

693 poor responders

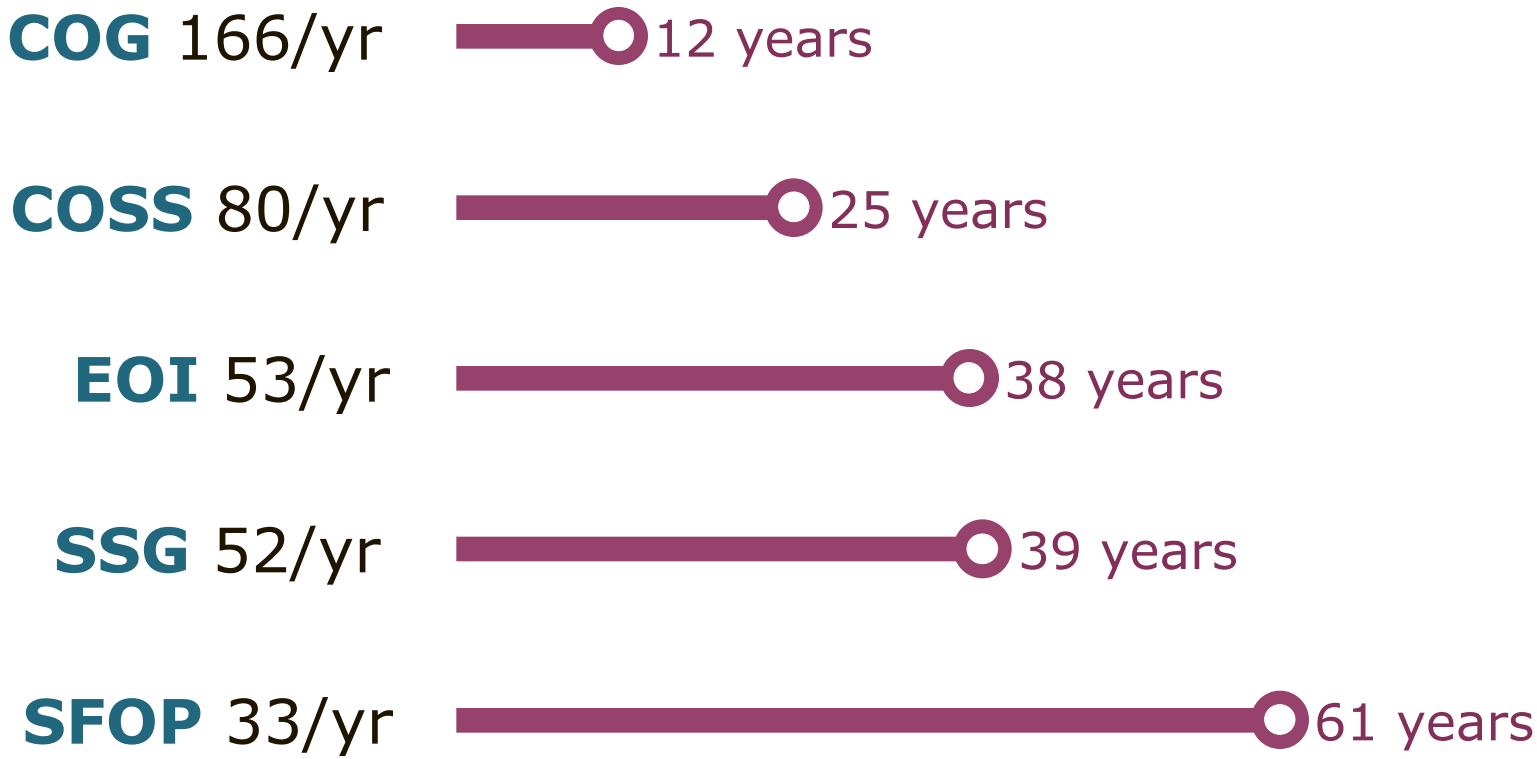
567 good responders

1,400 revised to 2,000 during recruitment

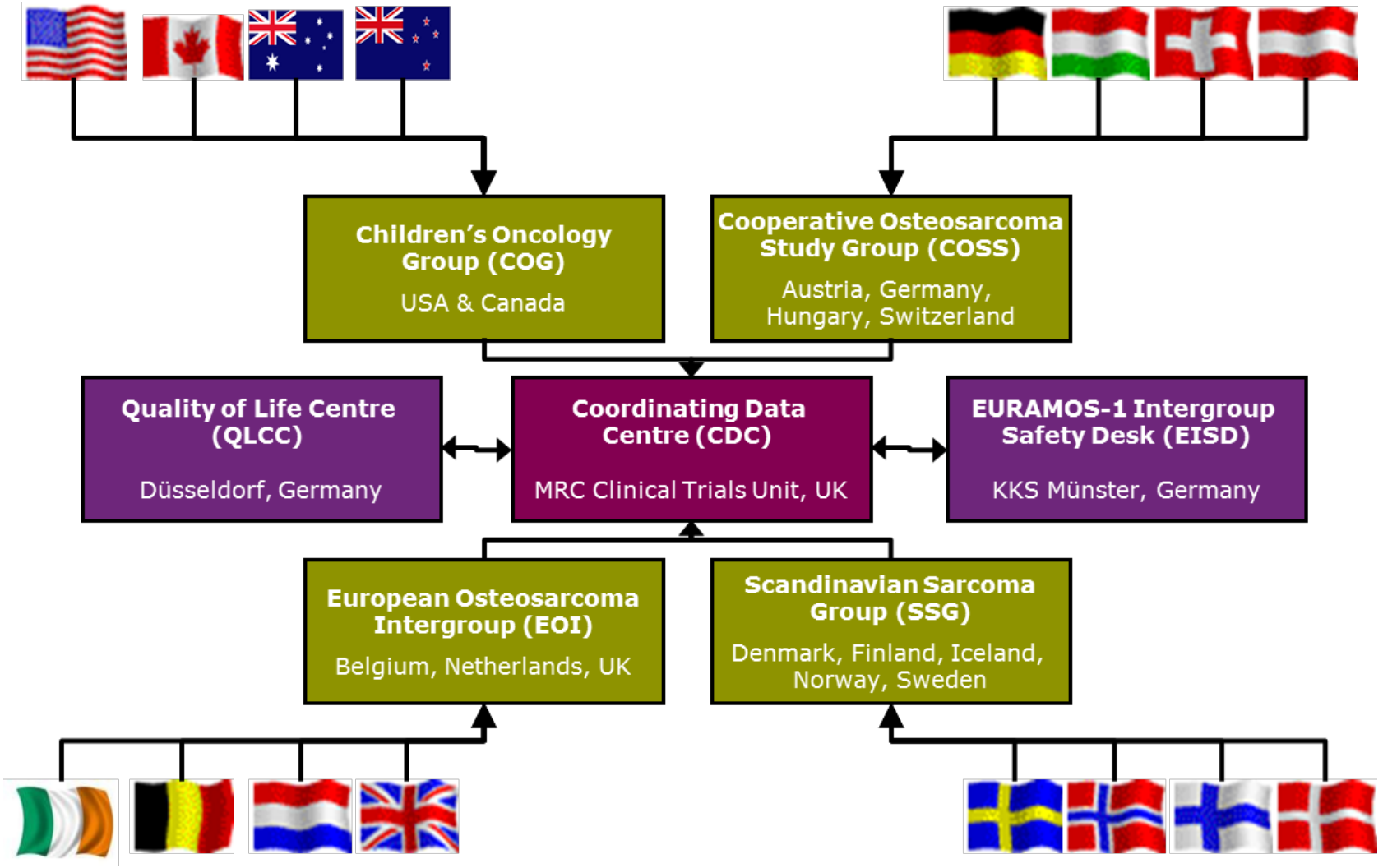
– no regional group can achieve this

Osteosarcoma groups

Time to 2,000 patients



EURAMOS-1 collaboration



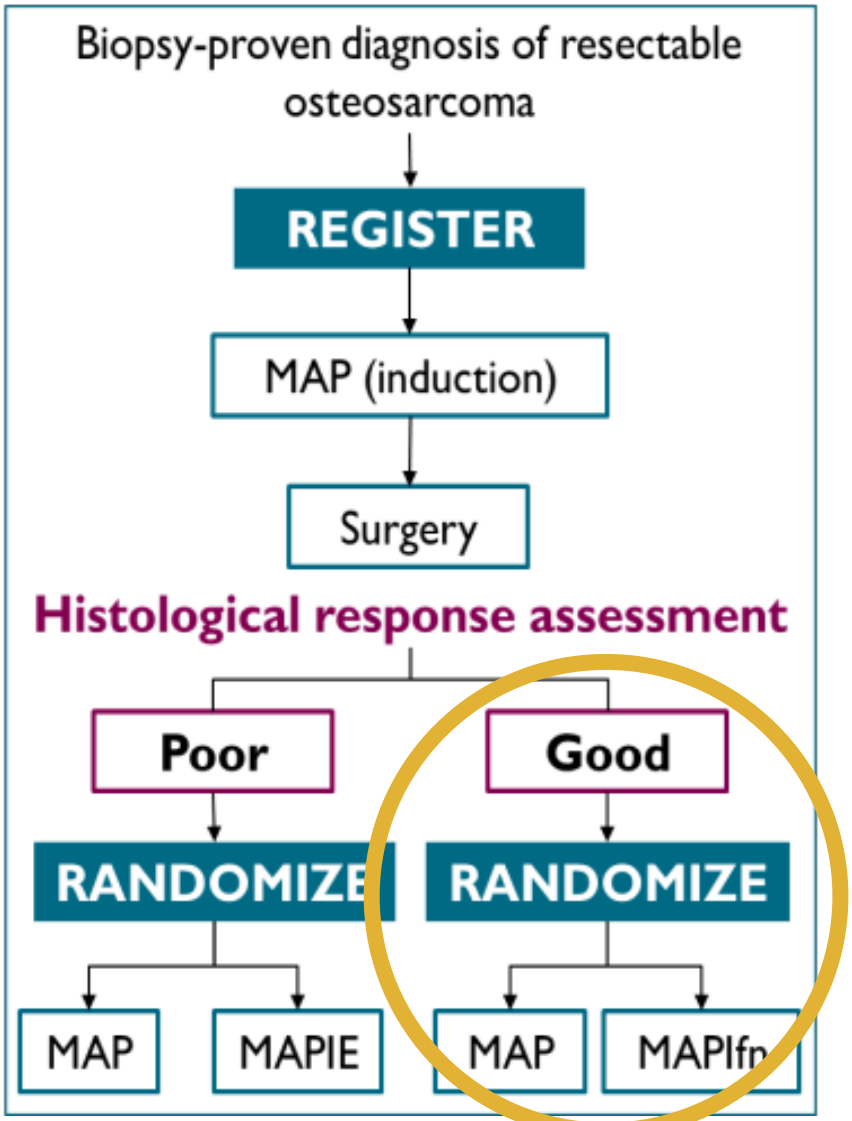
EURAMOS-1

EURAMOS-1 was our best shot at improving treatments for osteosarcoma in 10 years

- Too many patients to switch to designs for really rare diseases
- Too few patients to do a ‘regularly’ designed trial

How could we have designed it differently to maximise randomised information on treatment effect with the numbers we could recruit?

EURAMOS-1 good responders



Primary OM	**Event-free survival**
Ctrl. Events	**70% event-free at 3yr**
Target	**HR=0.63 → 10% abs diff**
Power	**80%**
Alpha	**5%, two-sided**
Events	**147**
Patients	**567**

Framework:
Information-heavy outcomes

- We know we should aim to maximise the information content of the data

Continuous > **time-to-event** > **categorical** > **binary**

- Time-to-event information is defined by no. events. If clinically relevant,

Progression-/disease-free survival > **overall survival**

Framework
Sample size: Power

Context:

- Limited number of trials
- Few chances of improving treatments
- Don’t want to miss something worthwhile

‘Arbitrary’ use of 80 or 90% power in most trials

Framework proposal: **Don’t** compromise on power

Framework
Sample size: Target difference

Best source?

- Other diseases
- History

Some trials target larger difference. However:

- No reason to think treatment effect is larger
- Will miss relevant differences due to low power for realistic difference

Framework proposal: **Do not** compromise (too much) on target difference

Framework
Sample size: Type I error

- Conventions for power and significance levels have come from pragmatic choices
- Scientific parameters determined by what is feasible
- We propose the making these pragmatic choices when patient numbers are more limited

Framework
Sample size: Type I error

5% generally chosen

Can accept larger Type I error rate

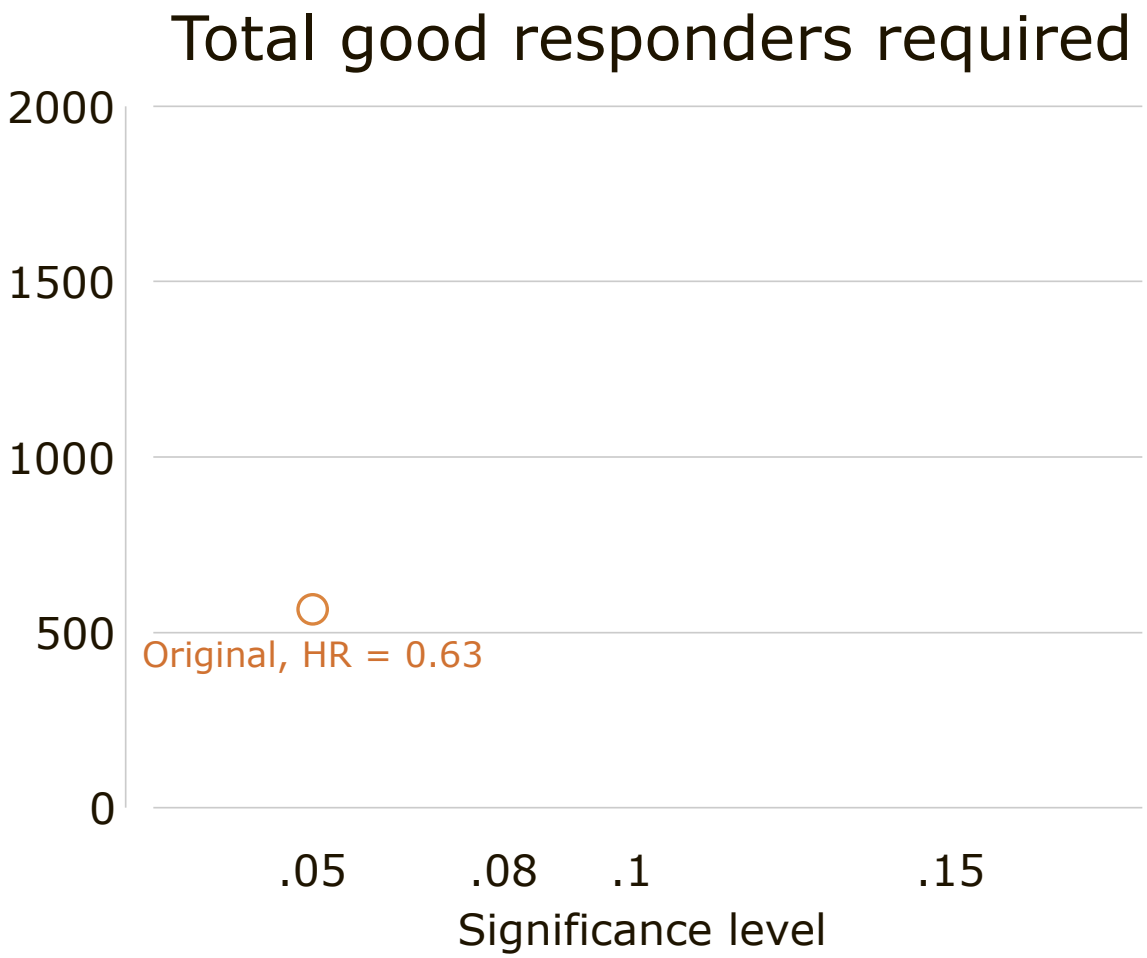
- Very few trials in this 'uncommon disease'
- Many more trials in more 'common diseases'

Are two-sided tests necessary in *superiority* trials?

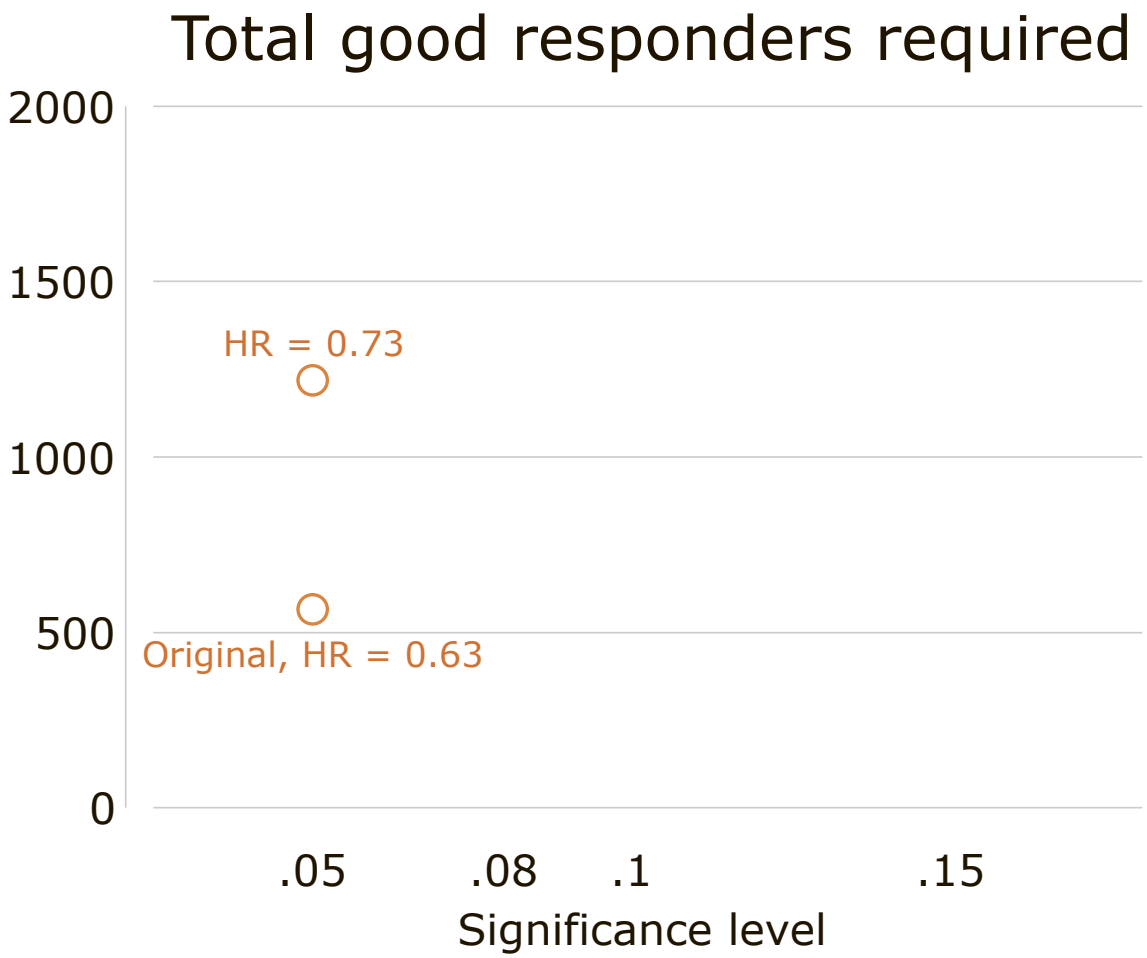
- They are not called *any-difference* trials'!

Framework proposal: **Compromise** on alpha and move to 1-sided tests

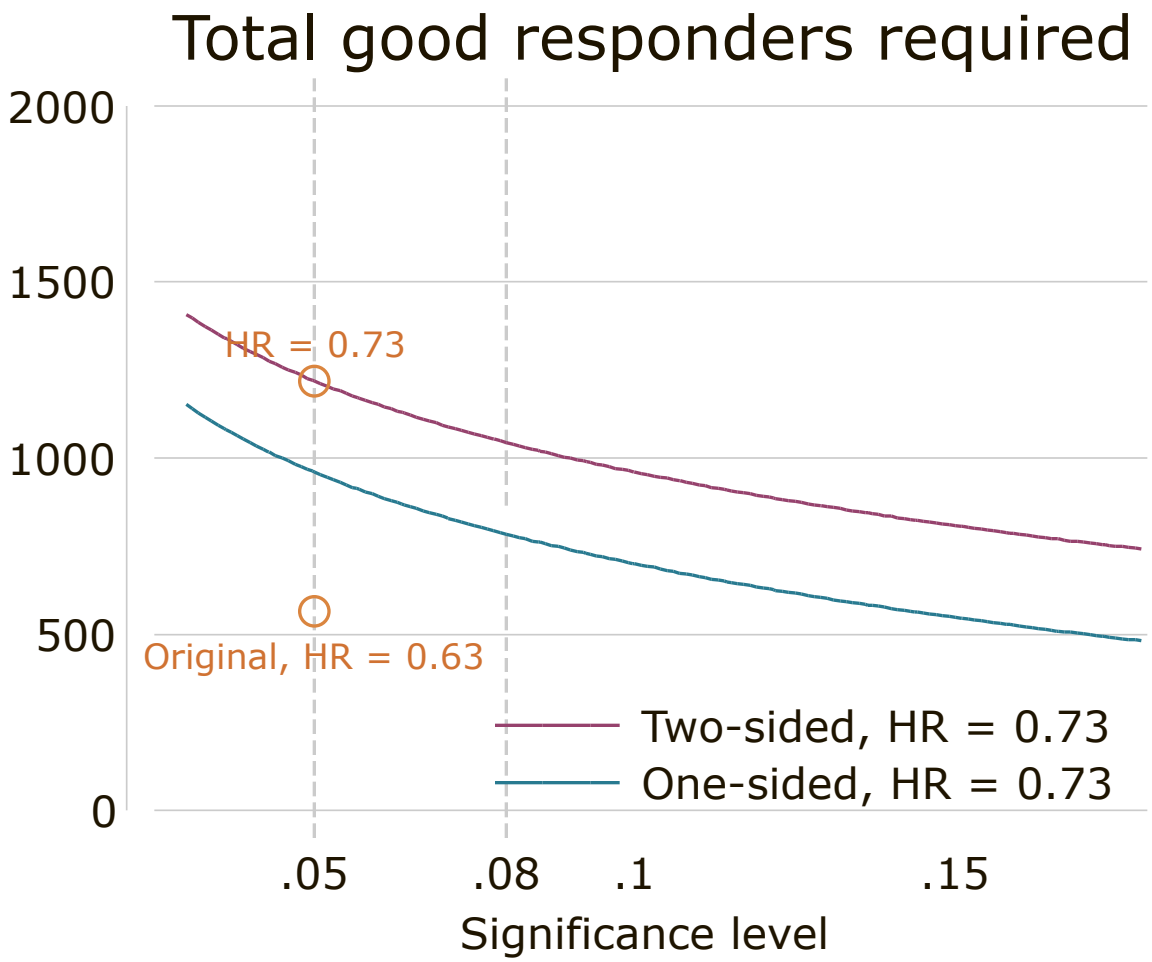
Framework
Sample size: Type I error



Framework
Sample size: Type I error



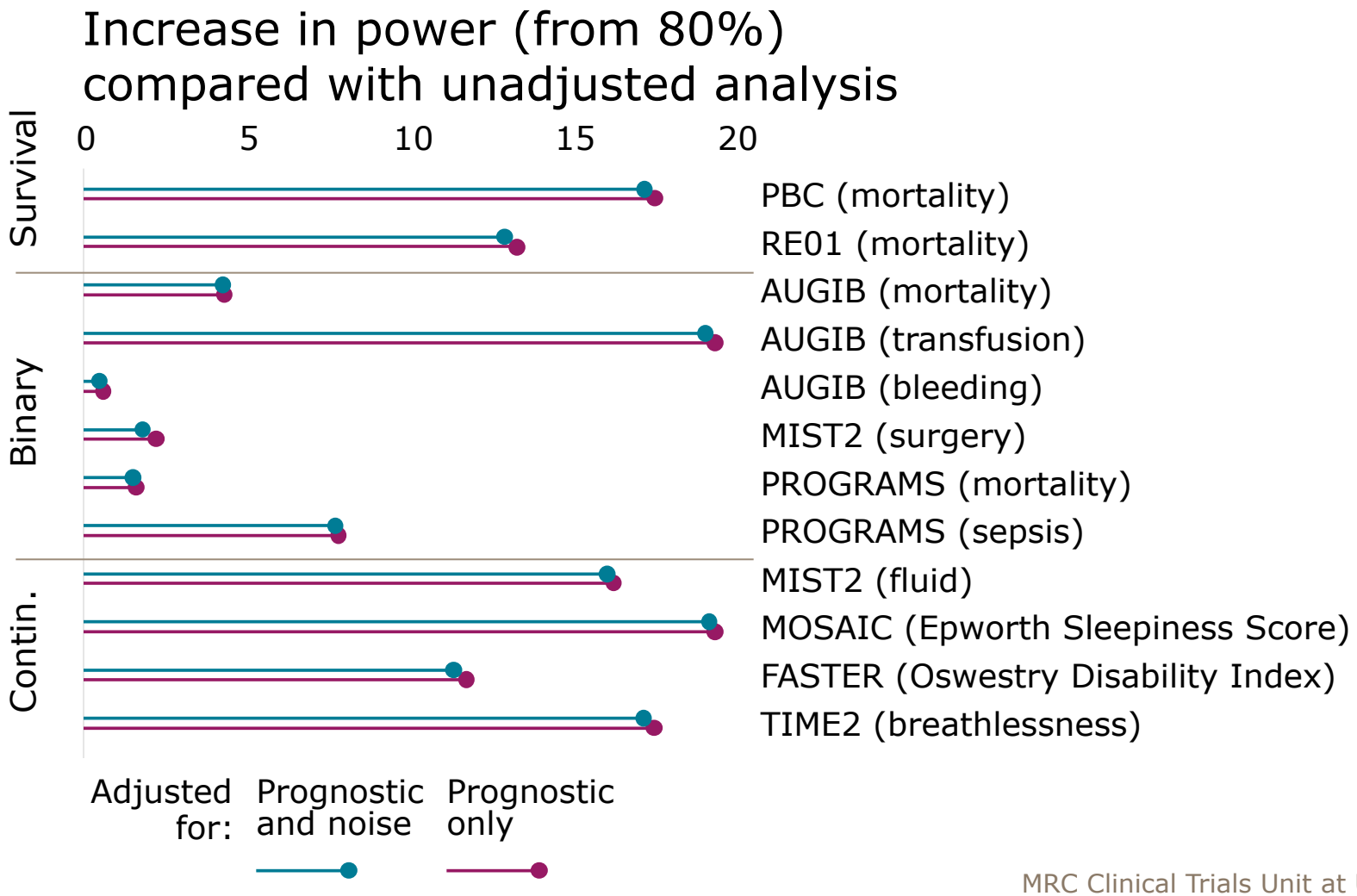
Framework
Sample size: Type I error



With a realistic HR, sample size rockets

We can bring this some way back down by relaxing significance level and using one sided tests

Framework
Including covariates

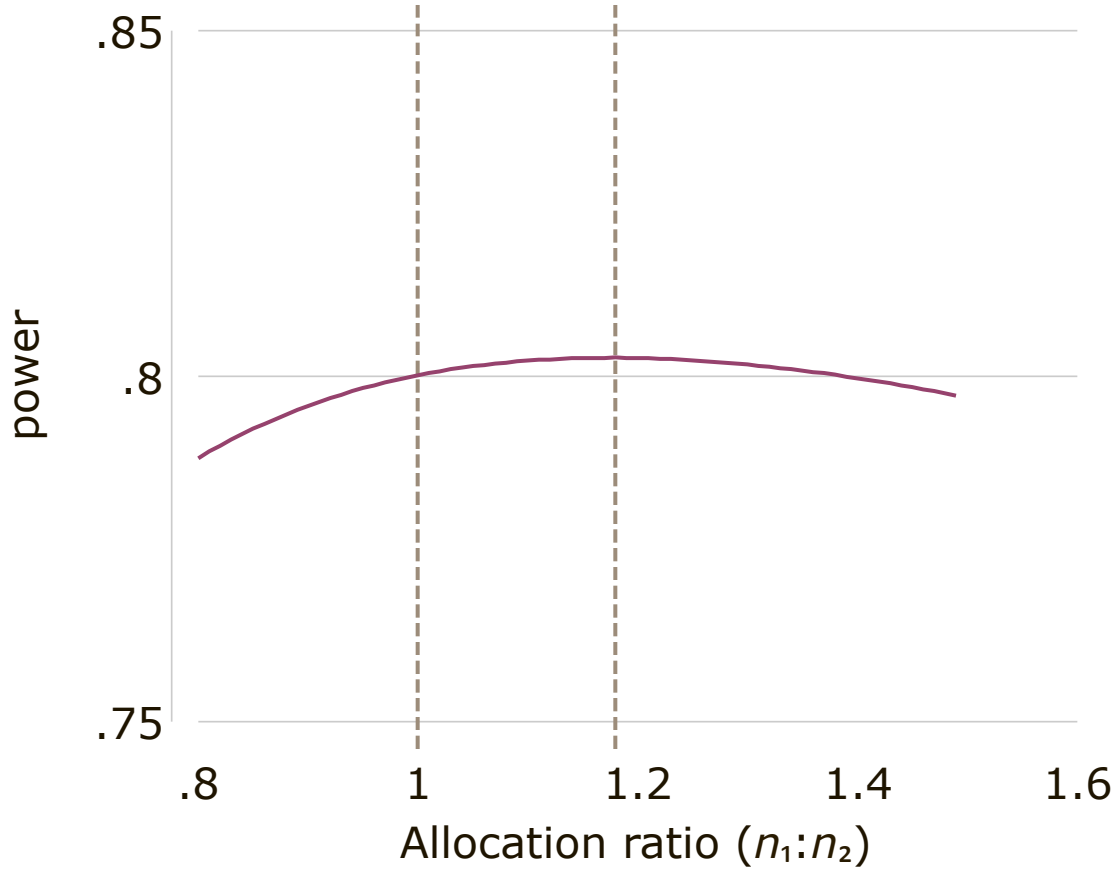


Framework
Including covariates

- Can improve power more than you might expect
- With many covariates and few patients/events, covariate *adjustment* can be tricky
- An attractive recent alternative is to weight on the inverse estimated propensity score (E Williamson, Statist. Med. 2014)

Framework proposal: **Include covariates** you suspect to be prognostic

Framework:
Skewing allocation



Can improve power, but hardly worth the effort

This isn't a reason *not* to skew: if you have another good reason, may lead to slight gain or loss in power

Framework:
Re-randomisation

Can be used when certain conditions are fulfilled:

1. Patients continue to be eligible
2. Patients complete follow up for previous treatment period before being re-randomised
3. Assumption of constant treatment effect across all randomisation periods is reasonable
4. Randomisation must be unrestricted within patient – random whether they switch or stick

Kahan BC, Forbes AB, Doré CJ, Morris TP.

A re-randomisation design for clinical trials.

BMC Medical Research Methodology 2015; **15**(1), 96.

Framework proposal: **Include covariates** you suspect to be prognostic

Framework:
Other aspects

Selection of research treatment

- Do not compare Tweedledum vs. Tweedledee
- Maximise the difference between research and control

Need to carefully think through consequences making a type I error

Using external information

- Particularly important for adverse events?
- Might be important to justify relaxed type I error

Framework applied to EURAMOS-1

Proposals	**EURAMOS-1 re-design**
→ Use info-heavy OMs	**X** EFS already best OM
→ Compromise on alpha	√
→ One-sided tests	√
→ Non-traditional values	√
→ Consider covariates at design	? via simulation?
→ Skew allocation ratios	√ Skew allocation ratios
→ Re-randomise patients	**X** Wouldn't be eligible – progressive condition

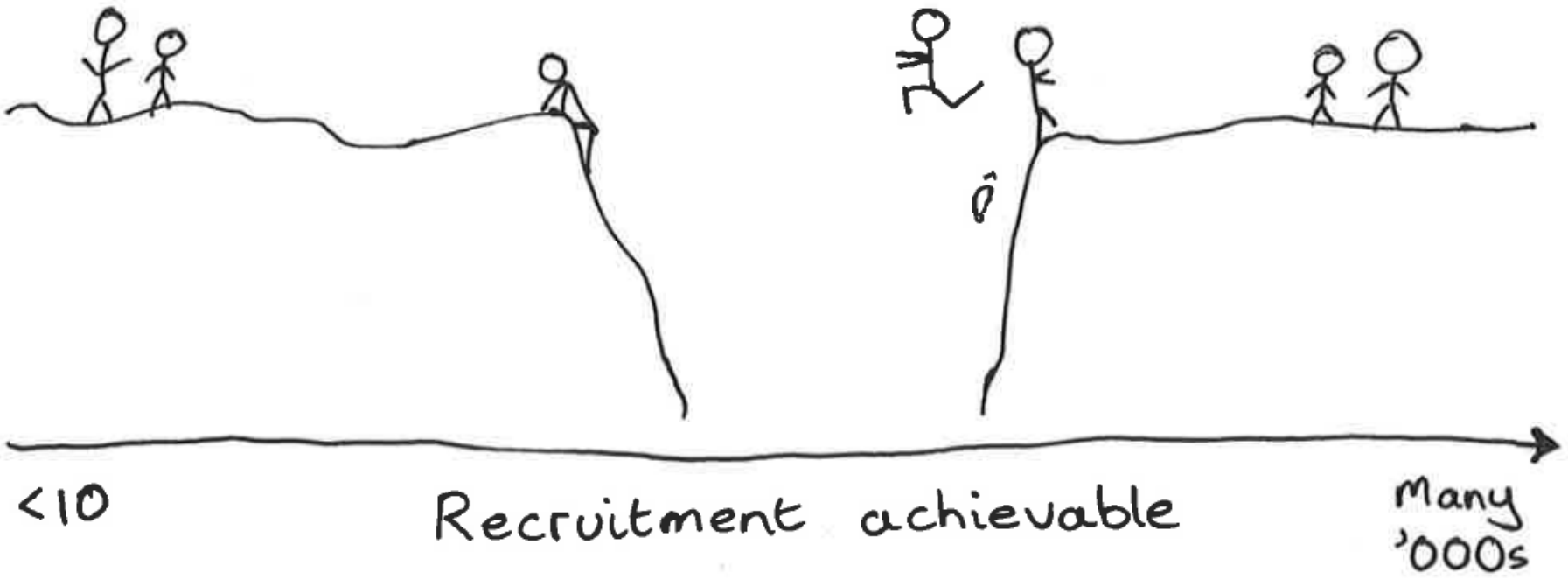
Framework applied to EURAMOS-1

	Original	**Revised #1**
Primary OM	Event-free survival	Event-free survival
Ctrl Events	70% event-free at 3yr	70% event-free at 3yr
Target	HR=0.63 → 10% abs diff	HR=0.73 → 8%
Power	**80%**	**80%**
Alpha	**5%, two-sided**	**15%, one-sided**
Allocation	**1C : 1R**	**4(C) : 5**
Events	**147**	**210**
Patients	**567**	**928**

Framework applied to EURAMOS-1

	Original	**Revised #2**
Primary OM	Event-free survival	Event-free survival
Ctrl Events	70% event-free at 3yr	70% event-free at 3yr
Target	HR=0.63 → 10% abs diff	HR=0.73 → 8%
Power	**80%**	**80%**
Alpha	**5%, two-sided**	**8%, one-sided**
Allocation	**1C : 1R**	**1 : 1**
Events	**147**	**276**
Patients	**567**	**1,208**

What to do between the extremes?



What to do between the extremes?

- Need to bridge the gulf
- ‘All we can do is use the information at hand to make the best decision possible’

  – Wedding Crashers
- ‘A way to do more trials, not a way to do small trials when larger ones are possible’

A framework for the design and analysis of phase III randomised trials in uncommon diseases

Mahesh Parmar, Tim Morris, Matthew Sydes

17 Nov 2015

MRC Hubs for Trials Methodology Research
London Hub

MRC Clinical Trials Unit at UCL

Framework
Sample size: **Type I error**

A thought experiment:

Assume you know nothing about the ritual of using 80 or 90% power and 5% (rather 2.5%) significance.

You are asked to provide acceptable risks of claiming

1. A good treatment doesn't work
2. A useless treatment works

Would you answer (1) **10**% and (2) **2.5**%?

Have you used these in practice?

Framework
Sample size: **Type I error**

No scientific worker has a fixed level of significance at which, from year to year, and in all circumstances, he rejects hypotheses; he rather gives his mind to each particular case in the light of his evidence and his ideas.

– RA Fisher (1956)

1. People do exactly that
2. Scientific workers are not always male

EURAMOS-1 collaboration