Phase 2 studies

Professor Bruce Morland
Birmingham Children's Hospital

It’s all about response!

What, how, when, how big

Critical design issues

- What are you going to measure as a measure of efficacy?
  - Tumour shrinkage
  - Biomarker
  - Survival
    - PFS, EFS, OS
- How will you measure it accurately?
  - Must be a validated tool
  - Is it reproducible?

Critical design issues

- When are you going to measure response?
  - After x number of “cycles”
    - How do you define a “cycle”?
  - Maximum response
    - Within a time limit?
- How “big” does response needed to be?
  - Typically a combination of CR and PR
  - But could be prolonged SD be good?

Endpoints

- Be very careful about selection of endpoints
- What convinces you/others/regulatory bodies this drug is active?
- Comparison with historical data essential unless doing a randomised study

Possible endpoints

- Radiological response
  - Most common
  - What modality
  - What is “measurable” (e.g. mIBG scans)
  - What “response” are you expecting?
    - Is prolonged stabilisation of disease enough?
- Survival
  - EFS
  - PFS
  - OS

Possible endpoints

- Biological
  - Histological (bone marrow)
  - MRD in leukaemia
  - AFP in HB/GCT
  - Circulating VEGF (antiangiogenics)

  Validated

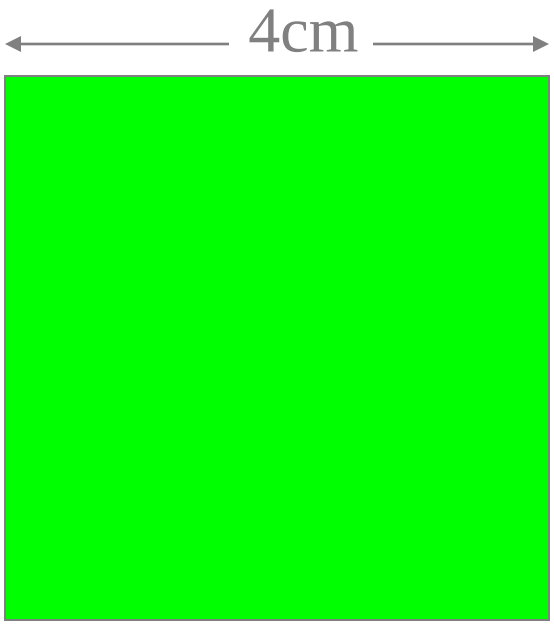
  Biologically meaningful

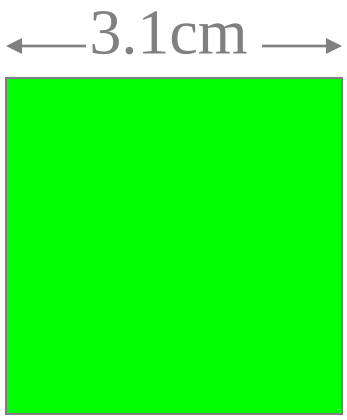
  Directly correspond to patient benefit

Radiological response to treatment, it used to be easy!

- Complete Response
  - No measurable disease
- Partial response
  - 50% or greater reduction
- Stable disease
  - <25% decrease and <25% increase in size
- Progressive disease
  - >25% increase in size

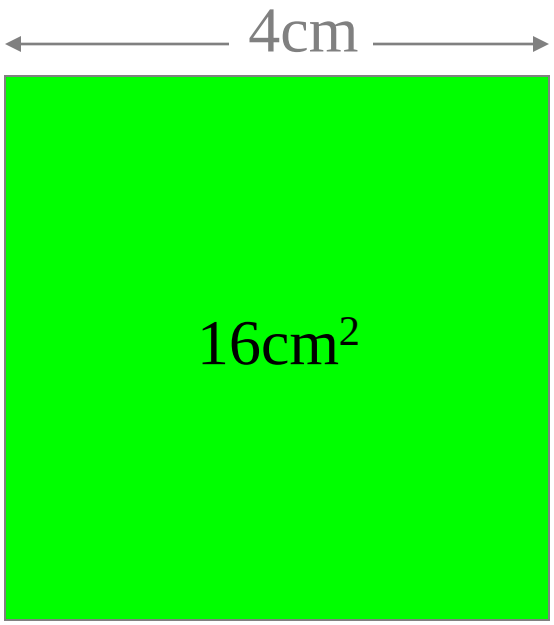
How big is this tumour?
Has it responded?

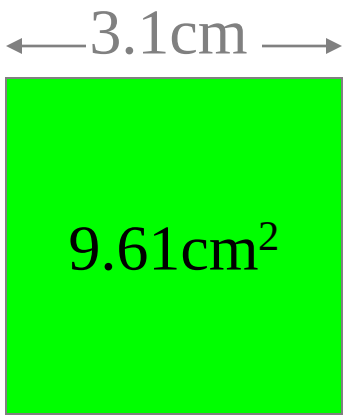




23%

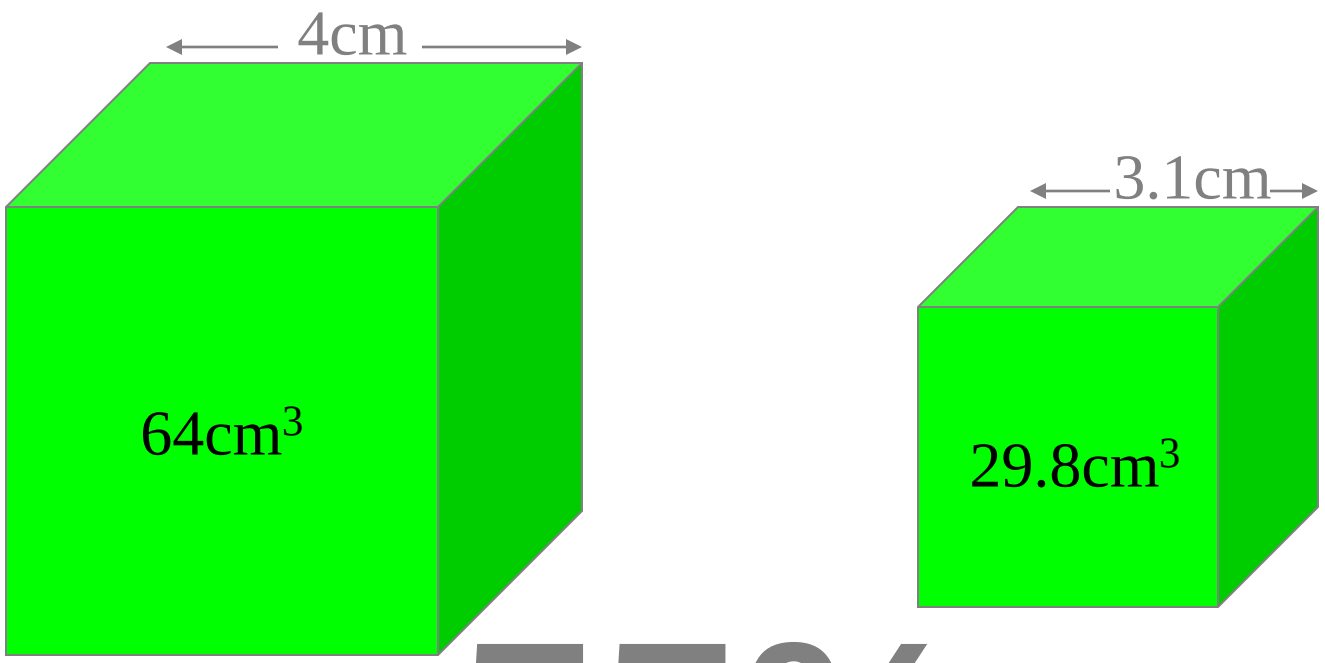
How big is this tumour?
Has it responded?





40%

How big is this tumour?
Has it responded?



55%

Things that might help

Tumour aren't perfect squares or cubes!

Tumour volume = a x b x c x F,

where a, b, and c represent the maximum tumour dimensions in three planes,

with F = 0.52 for spherical tumours,

or F = 0.785 for cylindrical tumours

“RECIST”

- Response evaluation criteria in solid tumours
- Defines tumour response by single tumour dimension
- Not evaluated in paediatric population


EUROPEAN JOURNAL OF CANCER 45 (2009) 228–247

ELSEVIER

available at www.sciencedirect.com

ScienceDirect

journal homepage: www.ejconline.com

EJC

New response evaluation criteria in solid tumours: Revised RECIST guideline (version 1.1)

E.A. Eisenhauer[a,*], P. Therasse[b], J. Bogaerts[c], L.H. Schwartz[d], D. Sargent[e], R. Ford[f], J. Dancey[g], S. Arbuck[h], S. Gwyther[i], M. Mooney[g], L. Rubinstein[g], L. Shankar[g], L. Dodd[g], R. Kaplan[j], D. Lacombe[c], J. Verweij[k]

RECIST concepts

- Measurable lesions
  - Accurately measured in at least one dimension >10mm CT (>20mm CXR)
- Non-measurable lesions
  - Ascites, pleural effusion, leptomeningeal infiltration

RECIST concepts

- Target lesions
  - Where there is 1 lesion…..easy
  - Where there are more than 1 lesion
    - Measure up to 5
    - Max 2 lesions per organ involved
    - Measure biggest lesions
    - Pick lesions than will be reproducible in future scans
    - Sum largest diameter of the lesions

RECIST evaluation - target lesions

- CR – disappearance of all target lesions
- PR – at least 30% decrease in the sum of maximum diameter of target lesions
- PD – at least 20% increase in the sum of maximum diameter of target lesions
- SD – Neither PD or PR

RECIST evaluation – non-target lesions and markers

- CR – disappearance of all non-target lesions and normalisation of markers
- SD – persistence of one or more non-target lesions and/or maintenance of abnormal markers above normal
- PD – Appearance of new lesions and/or unequivocal progression of markers

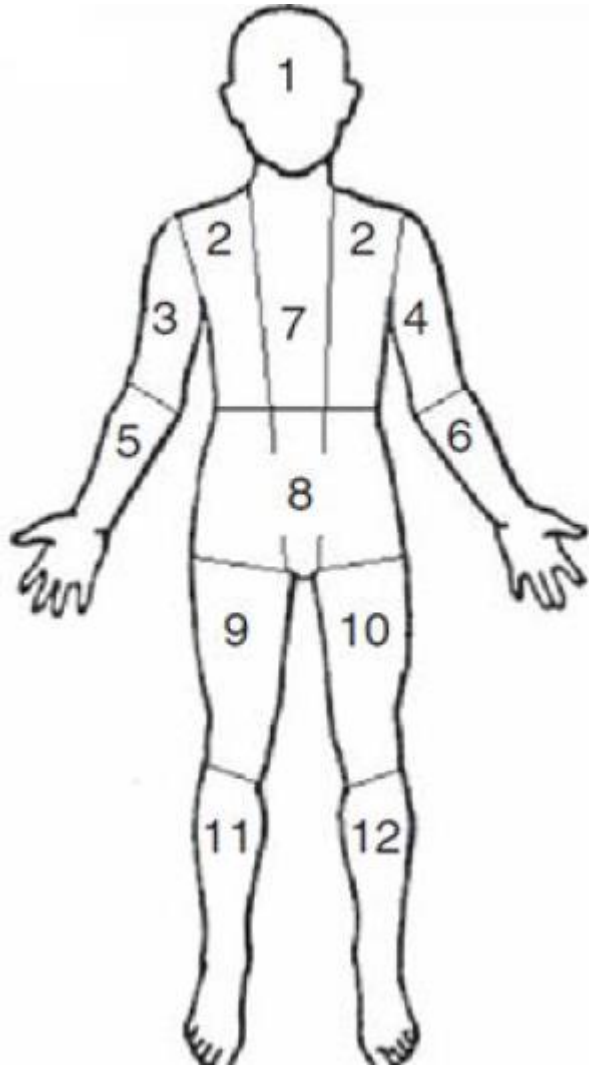
RECIST – overall evaluation

Target lesion	Non-target lesion	New lesion	Overall response
CR	CR	No	CR
CR	Non-CR/Non-PR	No	PR
PR	Non-PD	No	PR
SD	Non-PD	No	SD
PD	Any	Yes or No	PD
Any	PD	Yes or No	PD
Any	Any	Yes	PD

Neuroblastoma

- Need to allow for combinations of
  - Solid tumours (RECIST)
  - mIBG responses (SIOPEN/Curie score)
  - Bone marrow responses
  - Catecholamine changes

SIOPEN mIBG scores



0. No lesion
1. 1 lesion
2. 2 lesions
3. 3 lesions
4. >3 lesions or diffuse disease <50% area
5. Diffuse disease 50-95% area
6. Complete area involved

RESPONSE	ANATOMICAL IMAGING + MIBG (FDG-PET)
CR	• Complete resolution of non-primary measurable lesions • MIBG non-avid (no increased FDG-PET, uptake for MIBG non-avid tumours) of non-primary lesions
PR	• ≥30% decrease in size/sum of non-primary measurable disease (RECIST), OR • ≥50% reduction in MIBG score (relative MIBG score >0 to ≤ to 0.5)
MR	• CR or PR for one compartment (bone or soft-tissue) with at least SD in the other as long as no PD
PD	• Any new lesion by CT/MRI or MIBG • >20% increase in size AND a minimum absolute increase of 5mm in longest dimension in existing lesions • Relative MIBG (FDG-PET for MIBG non-avid tumours) score ≥1.2
SD	• Neither sufficient shrinkage for MR or PR nor sufficient increase for PD

CR: Complete Response
PR: Partial Response
MR: Minor Response
PD: Progressive Disease
SD: Stable Disease

INRG phase 2 task force/INRC working group metastatic response criteria (soft tissue/bone)

INRC response definition

CT/MRI Lesions	MIBG Lesions	Bone Marrow	Catechols	Overall Response
PD	Any	Any	Any	PD
Any	PD	Any	Any	PD
Any	Any	PD	Any	PD
CR	CR	CR	Normal	CR
VGPR	CR in bone lesions ; May have SD/CR in soft tissue sites corresponding to lesions on CT/MRI	CR	Normal	VGPR
PR	PR/CR in bone lesions; May have SD/CR in soft tissue sites corresponding to lesions on CT/MRI	CR	Any	PR
SD	SD	SD	Any	SD
SD/PR/VGPR/CR	SD	SD/CR	Any	SD
SD/PR/VGPR/CR	SD/PR/CR	SD	Any	SD

Revised response criteria for malignant lymphoma

Response	Definition	Nodal Masses	Spleen, Liver	Bone Marrow
CR	Disappearance of all evidence of disease	a) FDG-avid or PET positive prior to therapy; mass of any size permitted if PET negative b) Variably FDG-avid or PET negative; regression to normal size on CT	No palpable, nodules; nodules disappeared	Infiltrate cleared on repeat biopsy; if indeterminate on morphology and immunohistochemistry; bone marrow should be negative
PR	Regression of measurable disease and no new sites	A ≥ 50% decrease in SPD of up to six largest dominant masses; no increase in size of other nodes a) FDG-avid or PET positive prior to therapy; one or more PET positive at previously involved site b) Variably FDG-avid or PET negative; regression on CT	A ≥ 50% decrease in SPD of nodules (for single nodule in the greatest transverse diameter); no increase in size of liver or spleen	Irrelevant if positive prior to therapy; cell type should be specified
SD	Failure to attain CR/PR or PD	a) FDG-avid or PET positive prior to therapy; PET positive at prior sites of disease and no new sites on CT or PET b) Variably FDG-avid or PET negative; no change in size of previous lesions on CT		
Relapsed disease or PD	Any new lesion or increase by ≥ 50% of previously involved sites from nadir	Appearance of a new lesion(s) > 1.5 cm in any axis, ≥ 50% increase in SPD of more than one node, or ≥ 50% increase in longest diameter of previously identified node > 1 cm in short axis Lesions PET positive if FDG-avid lymphoma or PET positive prior to therapy	50% increase from nadir in the SPD of any previous lesions	New or recurrent involvement

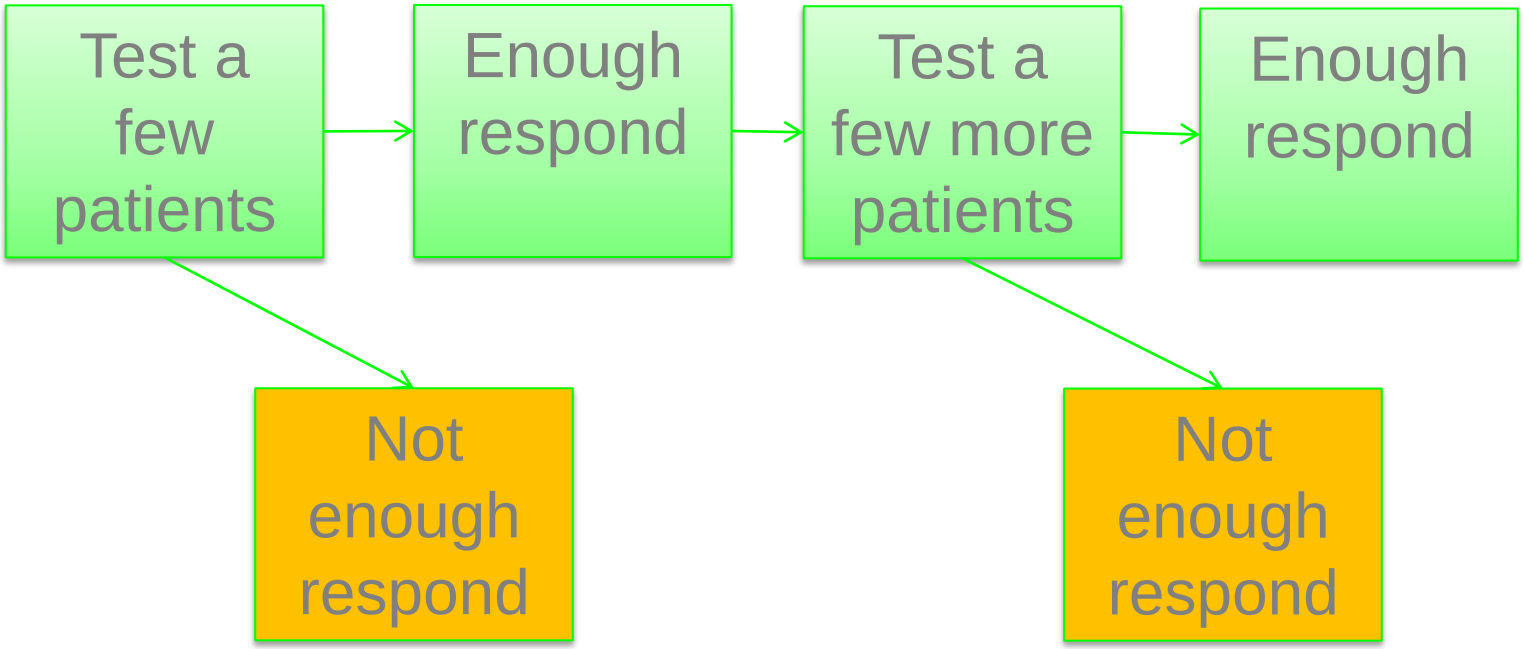
Survival as an endpoint

- Overall survival
  - You're only dead once!
  - But a lot may have happened to you between the trial and your death….
  - You may have been knocked over by a bus!
- Progression free survival
  - How sure are you to measure/describe progression?
  - Pseudo progression in CNS tumours is well described
  - Dependent on how frequently you measure to pick up progression e.g. you may be doing scans only 3 monthly
- Event-free survival
  - What's an event?
  - Need define carefully

“Biomarkers” as endpoints

- AFP is a good tumour marker, correlates with disease burden and response and ?prognosis
  - But what’s a good response? 10%, 50%, 90%?
- Your favourite TKI may completely obliterate the presence of a targeted biomarker
  - But need to validate this with meaningful clinical outcome (tumour shrinkage/survival)

Classical phase II trials (Simon, Fleming…)



The “null hypothesis”

- It says

  “This treatment doesn’t work”

Classical phase II trials (Simon, Fleming…)

- Single arm, multistage design (2-stage Simon's plan)
- Endpoint: efficacy, considering response *versus* failure
- When designing a phase II trial, we need to define
  - Four parameters: p0, p1, alpha, beta
  - The maximal number of patients
- But parameters are dependent on the tumour type being investigated

Classical phase II trials (Simon, Fleming…)

- p0: response rate p0 of “inefficacy”.
  If the new treatment provides a response rate significantly higher than this p0, *i.e.* if we can reject the null hypothesis H0, => we will conclude to the efficacy of the new treatment.
- The risk alpha of wrongly concluding to efficacy whereas the new treatment is in fact ineffective; a “false positive” result
- The minimal response rate p1 for which we want to be able to conclude to the efficacy of the new treatment with a specified power. Power = 1 – beta
- The risk beta = probability of not concluding to the efficacy whereas the new treatment is effective; a “false negative result”

An example

- We think a drug is ineffective if the response rate is less than 20% (p0)
- We think a drug is effective if the response rate is more than 40% (p1)
- We want to keep the risk of using a drug that is ineffective, as low as possible, say 10% (alpha), i.e. a low false positive rate
- We want to maximise the chance of not missing an effective drug by declaring it ineffective, say 10% (beta) i.e. a low false negative rate

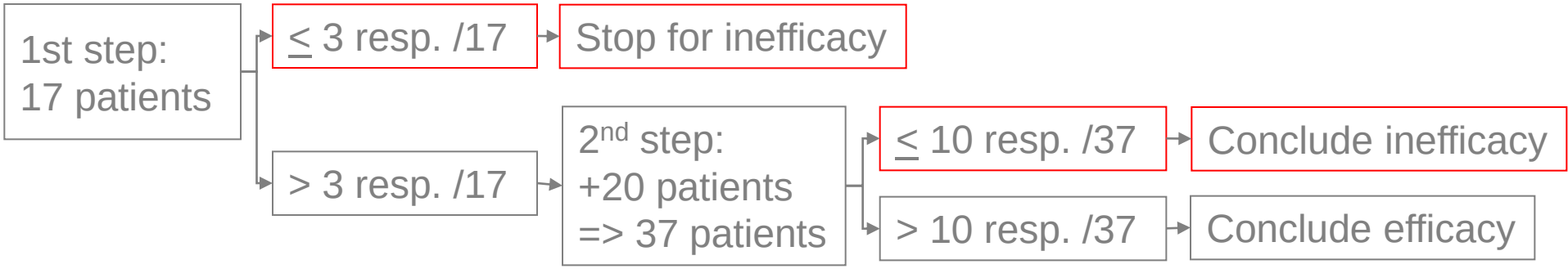
Classical phase II trials

Some examples of Simon's Optimal designs

p0	0.20	0.20	0.25
p1	0.40	0.40	0.45
Alpha	0.10	0.05	0.10
Beta	0.10	0.20	0.10
n1	17	13	14
k1	3	3	3
N	37	43	44
K	10	12	14

What you define

The subsequent decision rule



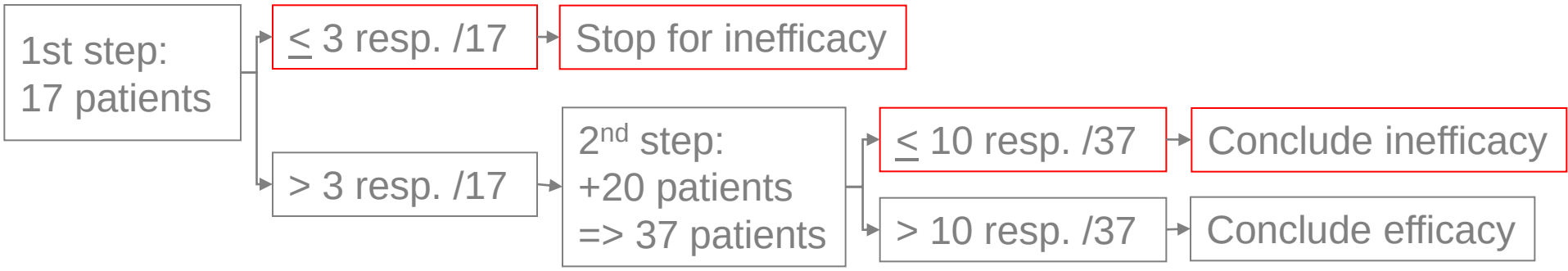
Classical phase II trials

- Usual statistical approach with usual tests (frequentist approach)
- No control within the trial
  but historically controlled (choice of p0, p1)
- At the end of a phase II trial,
  considering the specified hypotheses
  and the observed data,
  there is one (and only one) binary conclusion:
  we will conclude to efficacy or to inefficacy

- What are the consequences of a misspecification of p0 or p1 on the risk of wrongly concluding to efficacy or inefficacy?

Limits of classical phase II trials

- Imagine you choose a Simon's optimum design with p0=0.20, p1=0.40, alpha=0.10 and beta=0.10
  => we define and apply the following rule



If in fact

- p1=0.35 => beta=0.236 => power =76%
- p1=0.45 => beta=0.032 => power =97%
- If the drug should be considered as ineffective for p0=0.25 (instead of 0.20),
  the calculated alpha risk is 0.28 (much greater than 0.10 warranted by the plan for p0=0.20)

Randomised phase II studies

- Control vs experimental
  - Allows some security around defining p0 and p1
- Experimental vs experimental
  - “pick a winner”

Randomised phase II trials

- Reduces selection bias
- May identify a significantly different response rate between therapies
- Does not obviate the need for subsequent phase III with clinically relevant end points e.g. survival, QoL
- Gives confidence that the regimen selected for phase III is the more active
- Powered to select a substantial benefit therefore may not detect a moderate activity which may be clinically relevant

UNIVERSITY OF BIRMINGHAM

CANCER RESEARCH UK

@CRCTU Cancer Research UK Clinical Trials Unit

A randomised phase IIb trial of BEvACizumab added to Temozolomide ± IrinOtecan for children with refractory/relapsed Neuroblastoma -

BEACON-Neuroblastoma

	Day 1	Day 2-5	Day 15	Day 22	Day 29
Drug schedule on **temozolomide-**based arms *Treatment duration: 6 cycles = 24 weeks, response assessed every 2 cycles*					
T	Temozolomide 200 mg/m^2 po	Temozolomide 200 mg/m^2 po			Day 1 of next cycle starts
BT	Bevacizumab 10 mg/kg iv Temozolomide 200 mg/m^2 po	Temozolomide 200 mg/m^2 po	Bevacizumab 10 mg/kg iv		Day 1 of next cycle starts
Drug schedule on **irinotecan/temozolomide**-based arms *Treatment duration: 6 cycles = 18 weeks, response assessed every 2 cycles*					
IT	Irinotecan 50 mg/m^2 iv Temozolomide 100 mg/m^2 po	Irinotecan 50 mg/m^2 iv Temozolomide 100 mg/m^2 po		Day 1 of next cycle starts	
BIT	Bevacizumab 15 mg/kg iv Irinotecan 50 mg/m^2 iv Temozolomide 100 mg/m^2 po	Irinotecan 50 mg/m^2 iv Temozolomide 100 mg/m^2 po		Day 1 of next cycle starts	
Drug schedule on **topotecan/temozolomide**-based arms *Treatment duration: 6 cycles = 24 weeks, response assessed every 2 cycles*					
TTo	Temozolomide 150 mg/m^2 po Topotecan 0.75 mg/m^2/day iv	Temozolomide 150 mg/m^2 po Topotecan 0.75 mg/m^2/day iv			Day 1 of next cycle starts
BTTo	Bevacizumab 10 mg/kg iv Temozolomide 150 mg/m^2 po Topotecan 0.75 mg/m^2/day iv	Temozolomide 150 mg/m^2 po Topotecan 0.75 mg/m^2/day iv	Bevacizumab 10 mg/kg iv		Day 1 of next cycle starts

Trial questions

- “Main randomisation” Does the addition of Bevacizumab add anything?
  - (T, IT, TT) v (BT, BIT, BTT)
- “Randomisations for free” Does the addition of irinotecan or topotecan add anything to a backbone of temozolomide
  - (T, BT) v (IT, BIT)
  - (T,BT) v (TT, BTT)

Bevacizumab randomisation

- Standard 2 stage design
- p0=25% p1=40% α=0.2 ß=0.2
- Enter 84pts if ≥2pts respond continue
- Enter 106pts if ≥4pts respond declare active

Irinotecan randomization

- A probability based Bayesian design
- 60pt each arm

Control Rate	Difference	Experimental Rate	# Events	RR	Pr (true RR>1.0| data) (%)	Pr (true RR>1.2| data) (%)	Pr (true RR>1.4| data) (%)	Pr (true RR>1.6| data) (%)
0.25	0	0.25	30	1.0	50	28	14	7
0.25	0.05	0.3	33	1.2	73	50	30	17
0.25	0.1	0.35	36	1.4	88	71	50	32
0.25	0.15	0.4	39	1.6	96	85	69	50

Topotecan randomization

- 40pts each arm

Control Rate	Difference	Experimental Rate	# Events	RR	Pr (true RR>1.0| data) (%)	Pr (true RR>1.2| data) (%)	Pr (true RR>1.4| data) (%)	Pr (true RR>1.6| data) (%)
0.25	0	0.25	20	1.0	50	32	19	11
0.25	0.05	0.3	22	1.2	69	50	34	22
0.25	0.1	0.35	24	1.4	83	67	50	35
0.25	0.15	0.4	26	1.6	92	80	65	50

Multi-arm multi-stage studies MAMS

- 'New' is more often not better than standard
- Academia
  - 624 NCI sponsored phase III trials (Arch Int Med 2008)
  - ~30% of trials 'statistically significant'
  - ~40% of trials 'new' therapy preferred
- Industry
  - Agents successful at phase I: only 10-20% receive a marketing authorisation
  - Success rate of phase III trials ~30-40%

Principles

- Need better mechanism for phase III choice
  - Than single arm phase II trial
- Test many new promising treatments
  - In the same timescale
- Potential to discontinue unpromising arms
  - Quickly and reliably
- Start to randomise as quickly as possible
- Maximise potential for a 'positive trial'

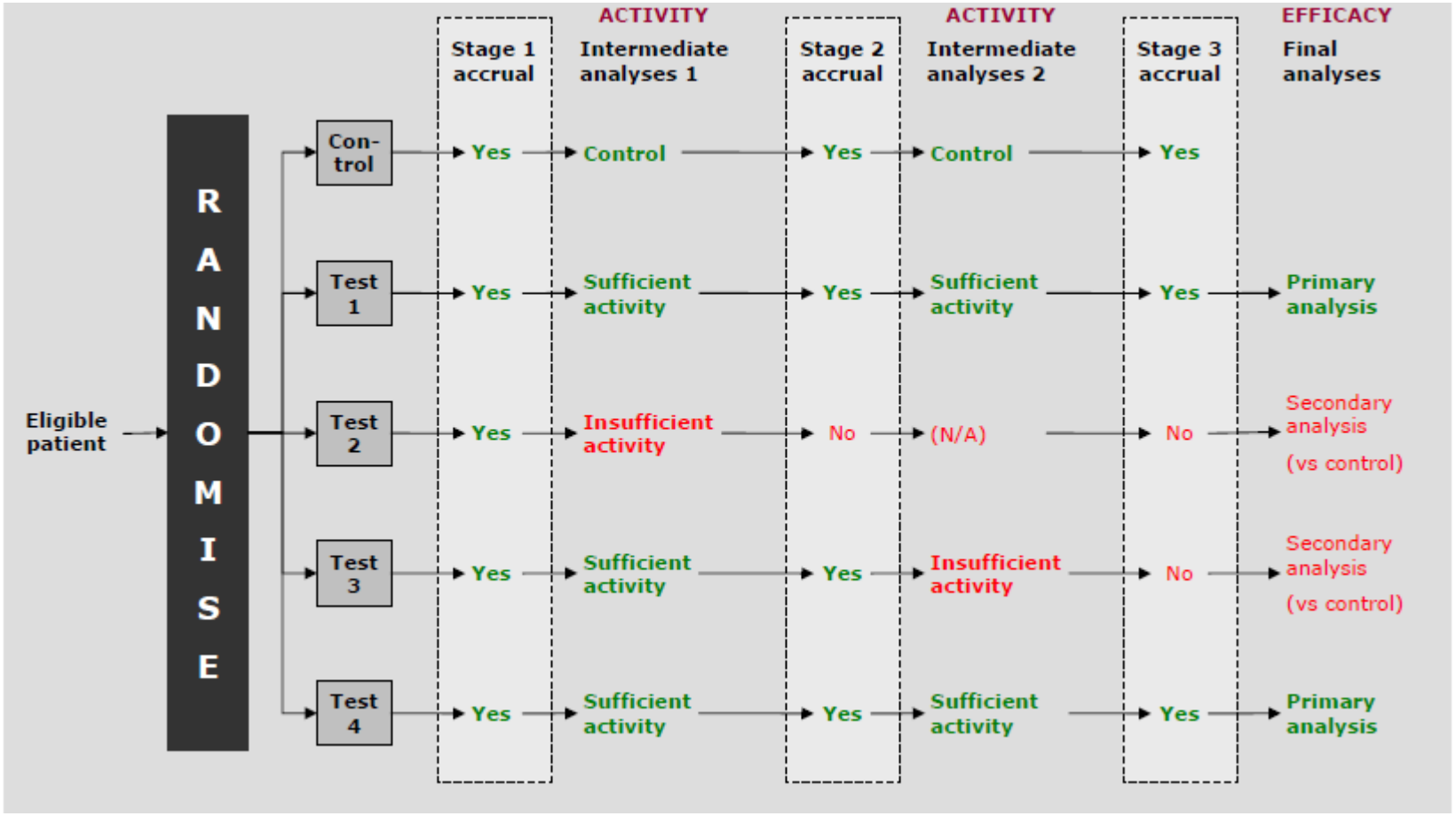
Activity (phase II stages)

- Ask the question:
  - Are there reasons why we should continue investigating a treatment?
  - Need to see sufficiently encouraging activity to continue assessment
- Testing for a lack-of-activity
  - Emphasis not testing for activity but for lack-of-sufficient-benefit
  - Focus away from insufficiently active regimens

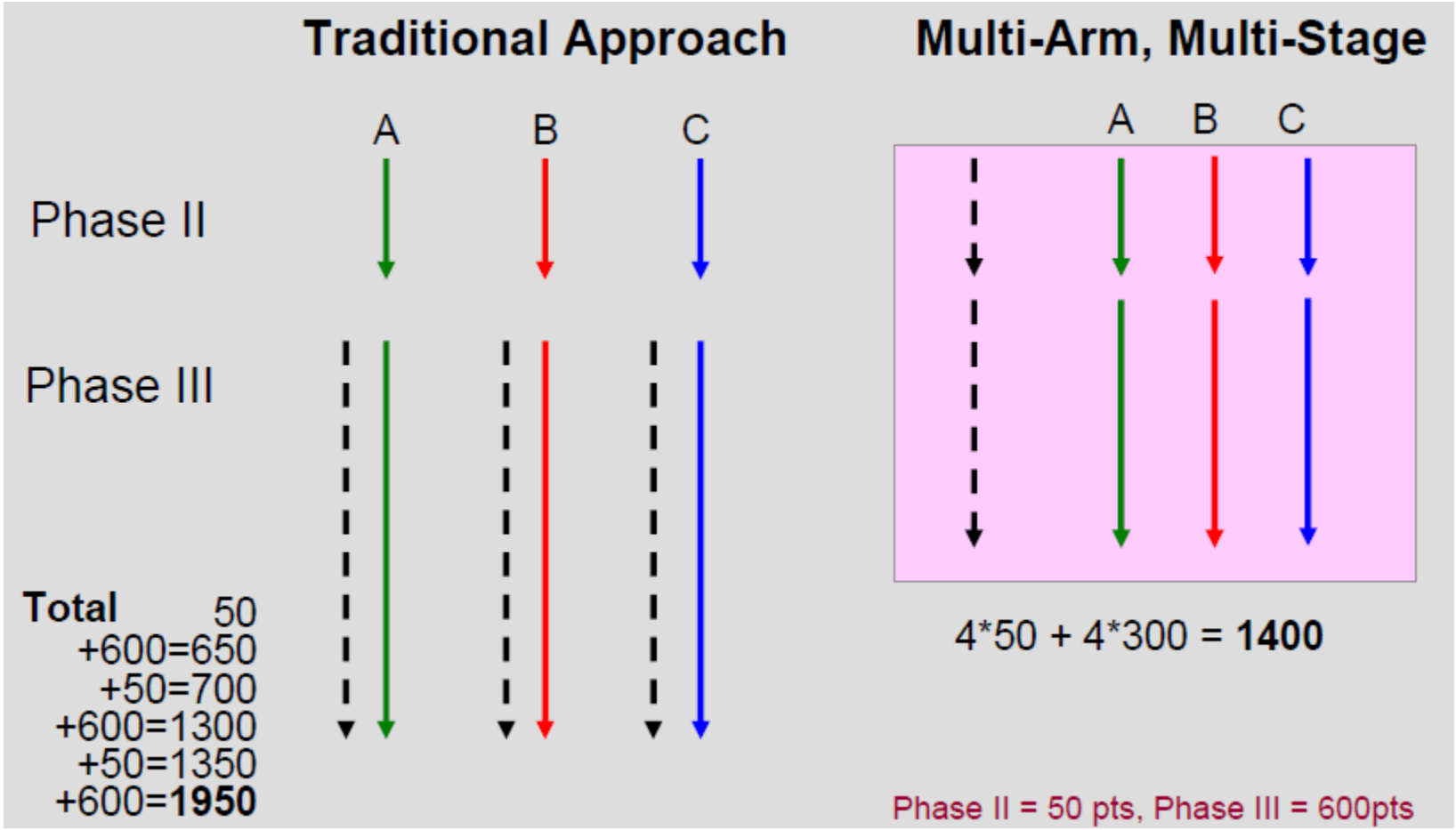
Activity (phase II stages)

- In Activity Stages use earlier outcomes
  - Even if interested in longer-term outcomes
  - Focus on Event-Free Survival (EFS)
- More EFS events than deaths
  - Therefore, more power for EFS than survival
- Design assumes:
  - To see an effect on OS you have to see an effect on FFS
  - Just because you see an effect on FFS does not mean that you will see an effect on OS

MAMS trial


ACTIVITY
ACTIVITY
EFFICACY
Stage 1 accrual
Intermediate analyses 1
Stage 2 accrual
Intermediate analyses 2
Stage 3 accrual
Final analyses
Eligible patient
RANDOMISE
Con-trol
Test 1
Test 2
Test 3
Test 4
Yes
Control
Yes
Control
Yes
Yes
Sufficient activity
Yes
Sufficient activity
Yes
Primary analysis
Yes
Insufficient activity
No
(N/A)
No
Secondary analysis
(vs control)
Yes
Sufficient activity
Yes
Insufficient activity
No
Secondary analysis
(vs control)
Yes
Sufficient activity
Yes
Sufficient activity
Yes
Primary analysis

Traditional v MAMS



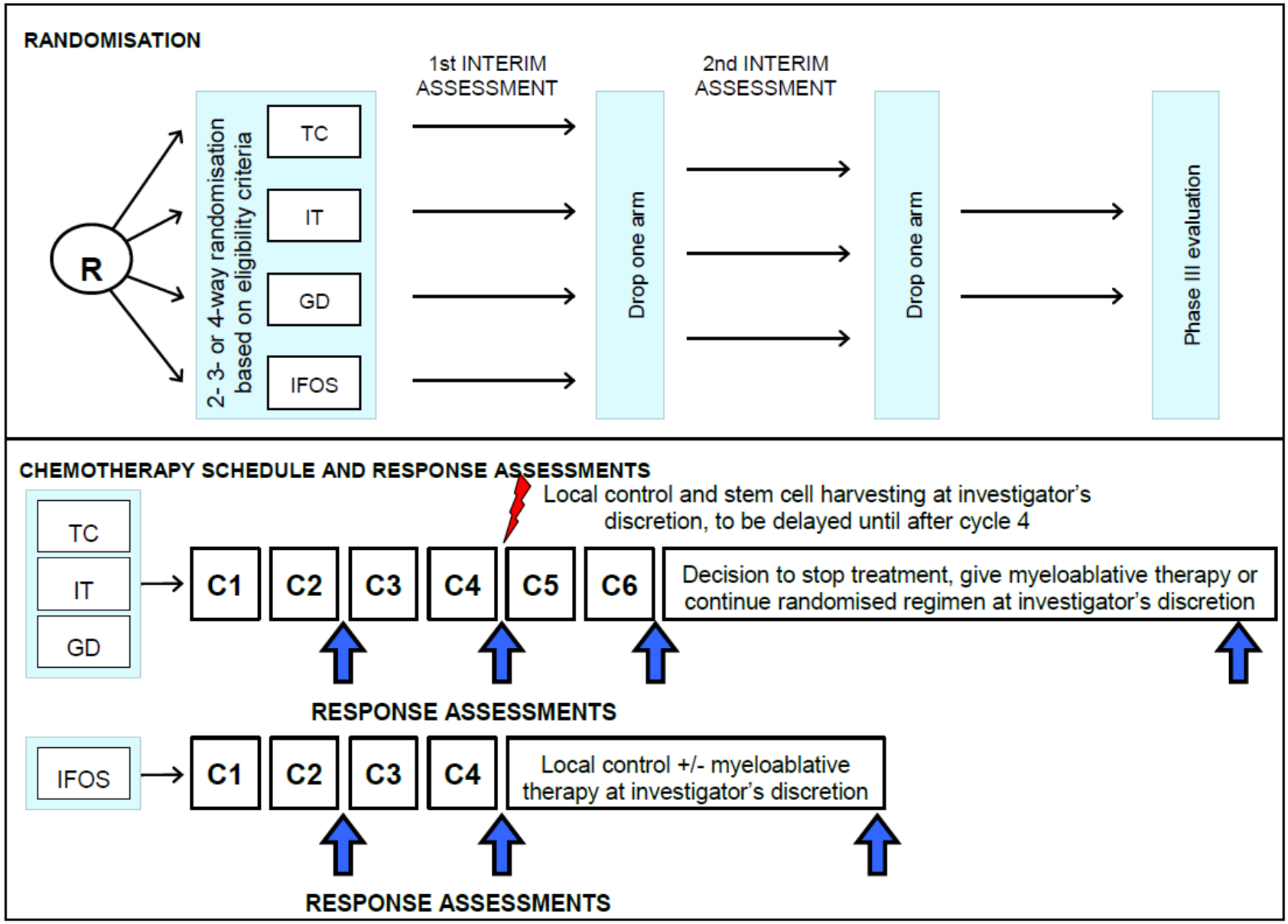
Advantage of MAMS

- Fewer patients
- Less overall time
  - Randomised from the start
  - Concurrent (not sequentially)
  - No delay between Phase II & Phase III assessment
  - Fewer applications for finance and approvals
- Increased flexibility
  - Focus trial resources on more active arms
- Reduced costs
  - Limited resources trial
  - Responsibility to use fairly and efficiently
  - Value

rEECur

International Randomised Controlled Trial of Chemotherapy for the Treatment of Recurrent and Primary Refractory Ewing Sarcoma

Trial Schema



Phase II

- First assessment 50pt in 4 arms
- Second assessment further 25 in each of the 3 remaining arms
- Analysis based on objective response of 40%
  - With a true difference of 15% in OR between best and worse arms there is a 2-15% chance of dropping the best arm
  - If 10% OR difference 8-18% chance
  - If 5% OR difference up to 25% chance

Phase III

- Target 400pts 200 in each arm
- Likelihood Bayesian approach

Phase III

1-yr EFS Arm A	Improvement	1-yr EFS Arm B	Total Events	Hazard Ratio (HR)	lnHR	P(HR < 1.00) (%)	P(HR < 0.87) (%)	P(HR < 0.76) (%)
0.30	0.00	0.30	280	1.00	0.00	50	13	1
0.30	0.05	0.35	270	0.87	-0.14	87	50	13
0.30	0.10	0.40	260	0.76	-0.27	99	86	50

Table 5. Observed HRs scenarios and associated probabilities

- If arm A is 10% better than arm B you have a 99% chance that it is at least better, an 86% chance it is at least 5% better and a 50% chance it is at least 10% better

Conclusions

- It’s all about response
- Accurately define response
  - What, how, when, how big
- Conventional phase 2 studies rely on an accurate estimation of the size of response you expect to see
- Randomised phase 2 studies give some reassurances around control group responses
- New trial designs (MAMS) allow several agents to be tested together and give phase 1, 2 and phase 3 endpoints

Questions?