WARWICK
THE UNIVERSITY OF WARWICK

Optimal designs for group-sequential biomarker enrichment oncology trials

Deepak Parashar

Statistics and Epidemiology Unit
Warwick Cancer Research Centre

28 April 2016

Collaborators: J. Bowden, C. Starr, L. Wernisch, A. Mander

Oncology Trials (*State-of-the-art*)

- High attrition rate (90%)
- Paradigm shift in cancer research: classify disease on underlying molecular biology
- Measures
  - Integrated translational research: *biological question integrated into study design.*
  - Biomarkers: *pair biological measurements with clinical outcomes.*
  - Stratified / Personalised Medicine: *enhance efficacy, reduce futility, cut costs.*

Biomarkers

- prognostic: predicts how a disease may develop in an individual regardless of the type of treatment.
- predictive: provides an indication of the probable effect of treatment on patient.

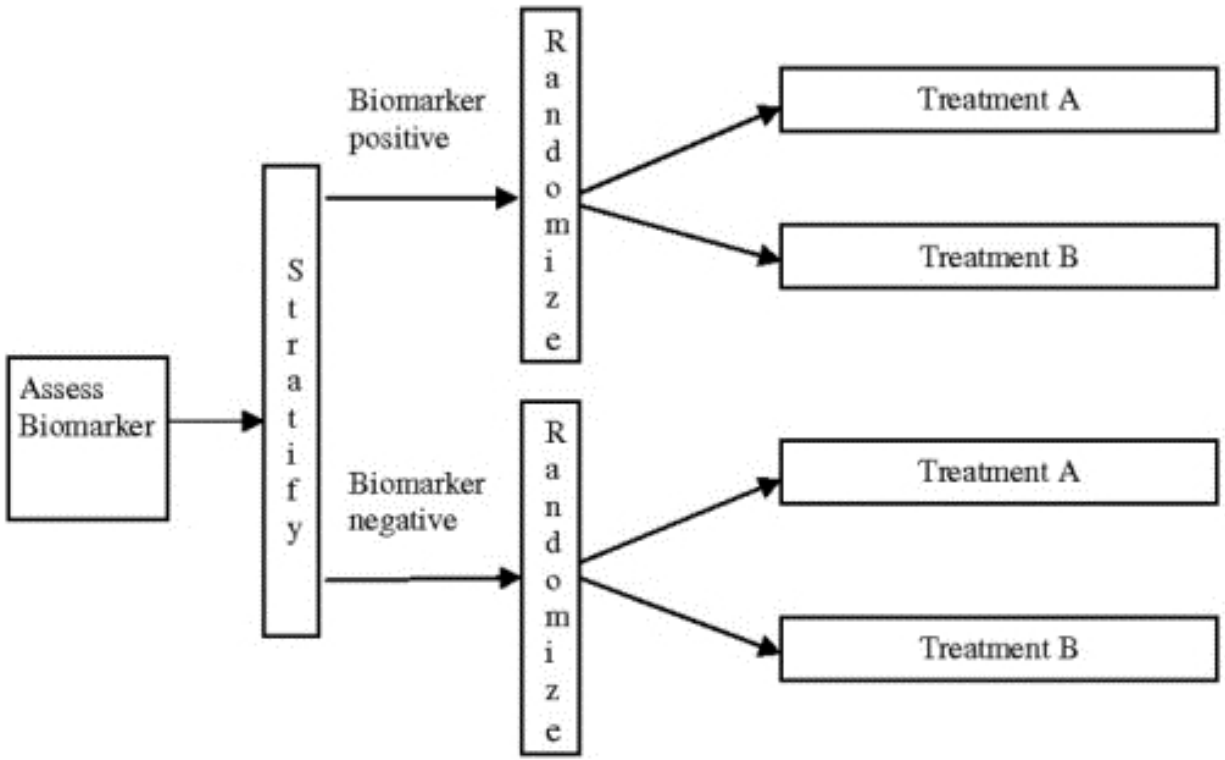
Focus on predictive biomarkers since the goal of the biomarker-based designs is to establish whether the biomarker predicts treatment response.

Biomarkers

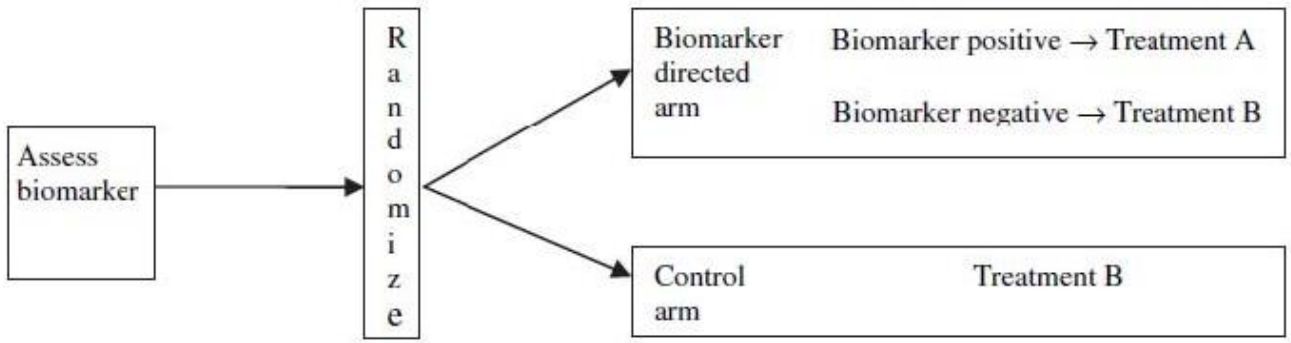
- prognostic: predicts how a disease may develop in an individual regardless of the type of treatment.
- predictive: provides an indication of the probable effect of treatment on patient.

Focus on predictive biomarkers since the goal of the biomarker-based designs is to establish whether the biomarker predicts treatment response.

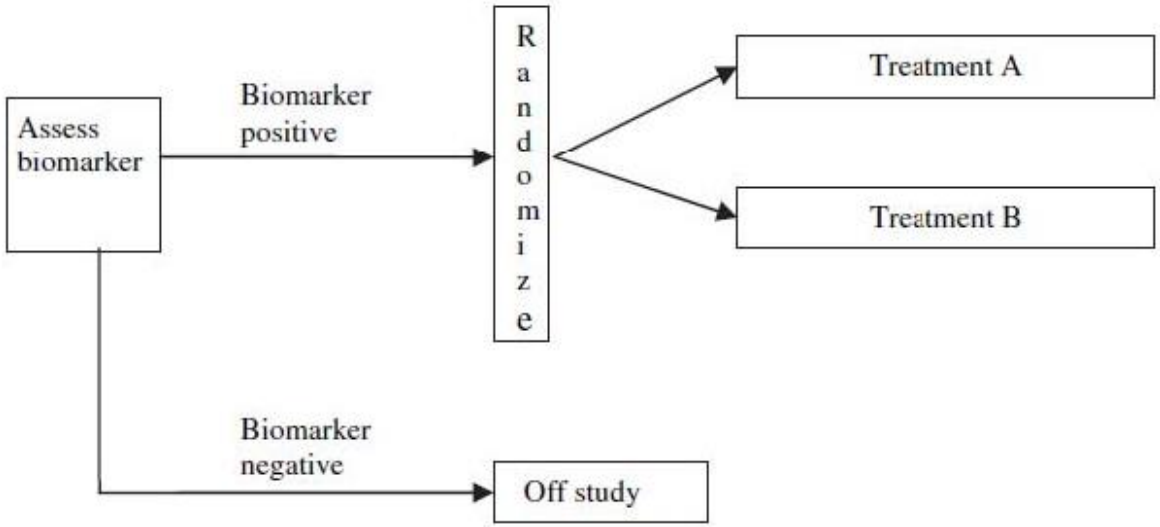
INTERACTION (BIOMARKER-STRATIFIED) DESIGN



BIOMARKER-STRATEGY DESIGN WITH STANDARD CONTROL



TARGETED (SELECTION) / ENRICHMENT DESIGN



Issues with enrichment design

1. No info about biomarker −ve patients.
2. Design may not be efficient if drug has at least some activity in biomarker −ve patients.
3. Effect in biomarker −ve patients may never be known.
4. Study would provide no new clinical evidence w.r.t. biomarker −ve patients.

Simon two-stage design

Hypotheses

$$H_0 : p \leq p_0 \qquad H_1 : p \geq p_1 (> p_0)$$

$$p \sim \text{true response rate}$$

Fix

- $p_0 =$ maximum unacceptable response rate
- $p_1 =$ minimum acceptable response rate
- $\alpha =$ desired type I error
- $\beta =$ desired type II error

X_1, X_2 are the number of responders in stage 1, 2 resp., binomially distributed.

R. Simon, Cont. Clin. Trials, 1989

Simon two-stage design

Strategy

- recruit n_1 patients at start
- if $\leq r_1$ responders at stage 1, stop for futility
- if $> r_1$ responders at stage 1, continue to recruit up to n patients
- reject H_0 if $\geq r$ responders

Probability of rejecting H_0

= 1 - Probability of NOT rejecting H_0

$$= 1 - \left[P(X_1 \leq r_1) + \sum_{i=r_1+1}^{min(r,n_1)} P(X_1 = i) P(X_2 \leq r - i) \right]$$

where $P(X_1 \leq r_1)$ is the probability of early termination, $PET(p)$

Note: stop for efficacy if $\geq r$ responders at first stage

Simon two-stage design

Strategy

- recruit n_1 patients at start
- if $\leq r_1$ responders at stage 1, stop for futility
- if $> r_1$ responders at stage 1, continue to recruit up to n patients
- reject H_0 if $\geq r$ responders

Probability of rejecting H_0

$= 1$ - Probability of NOT rejecting H_0

$$= 1 - \left[P(X_1 \leq r_1) + \sum_{i=r_1+1}^{min(r,n_1)} P(X_1 = i)P(X_2 \leq r - i)\right]$$

where $P(X_1 \leq r_1)$ is the probability of early termination, $PET(p)$

Note: stop for efficacy if $\geq r$ responders at first stage

Simon two-stage design

Expected sample size

$$E(n|p) = n_1 + [1 - PET(p)](n - n_1)$$

OPTIMALITY $\sim$ acceptable design (satisfying the desired α and β error rates) with smallest expected sample size under H_0

Example: Adaptive Biomarker Enrichment

Phase II targeted cancer therapy

- Determine whether drug has activity only in target population or the general population
- Outcome is (RECIST) tumour response
- Single-arm trial
- Enrichment adaptation based on Simon two-stage design

Jones & Holmgren, Cont. Clin. Trials 2007

Hypotheses (Group-sequential)

$$H_0^- : p^- = p_0^-, \qquad H_0^+ : p^+ = p_0^+$$
$$H_1^- : p^- = p_1^-, \qquad H_1^+ : p^+ = p_1^+$$

Assume $p^- < p^+$

- Conclude efficacy in unselected if we reject H_0^-
- Conclude efficacy in biomarker positive if we reject H_0^+

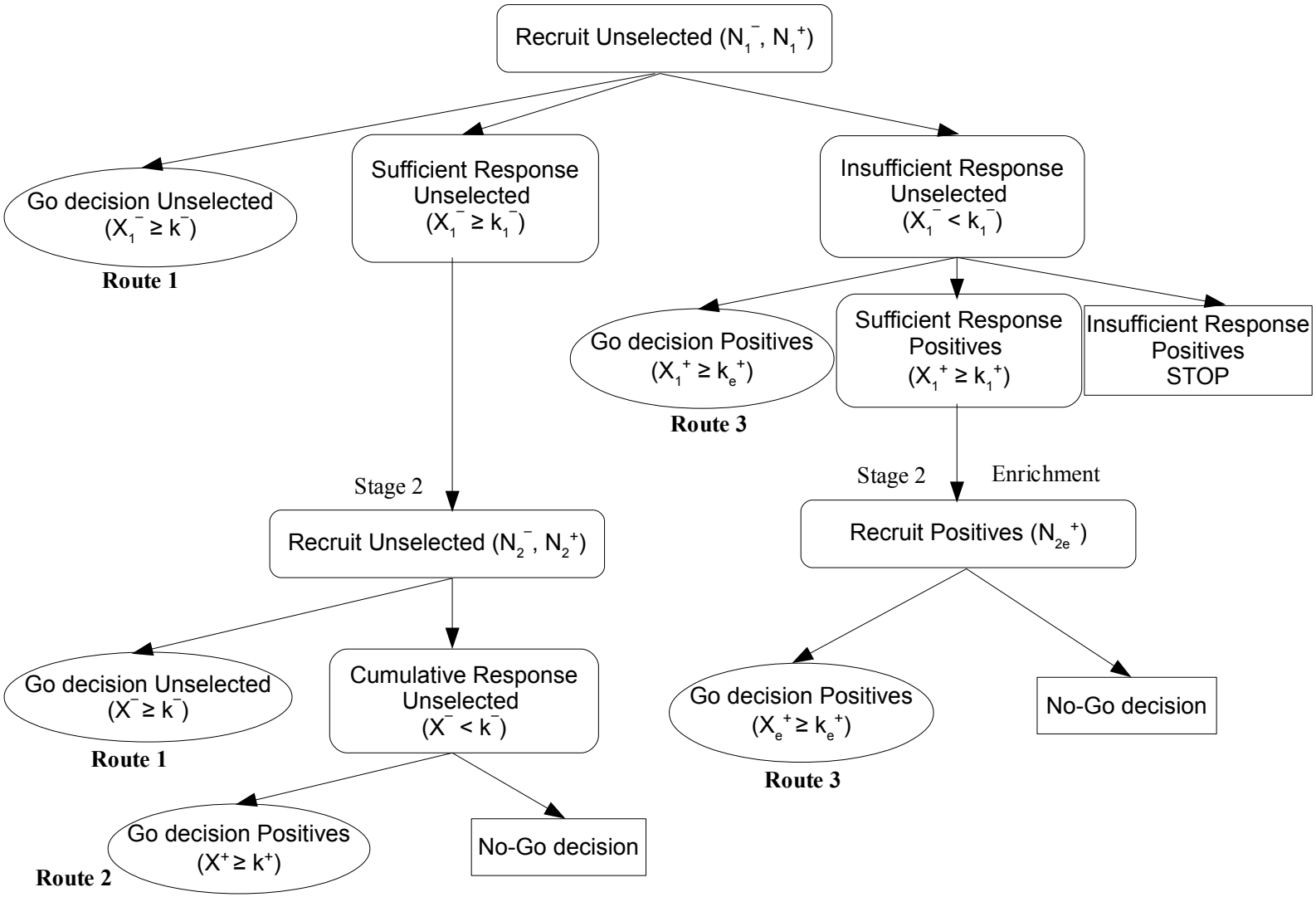
Hypotheses (Group-sequential)

$$H_0^- : p^- = p_0^-, \qquad H_0^+ : p^+ = p_0^+$$
$$H_1^- : p^- = p_1^-, \qquad H_1^+ : p^+ = p_1^+$$

Assume $p^- < p^+$

- Conclude efficacy in unselected if we reject H_0^-
- Conclude efficacy in biomarker positive if we reject H_0^+

Design Schematic



Issues with J & H Design

1. Hypotheses testing not properly addressed
2. Type I error and power calculations wrong
3. Expected sample size formulae do not take into account early stopping for efficacy
4. Designs not optimal

Our work rectifies these issues and provides a robust framework for adaptive enrichment

Error Probabilities

Type I error

- $R_1(p^-) \sim$ probability of rejecting H_0^- via Route (1)
- $R_2(p^-, p^+) \sim$ probability of rejecting H_0^+ via Route (2) (non-monotonic)
- $R_3(p^-, p^+) \sim$ probability of rejecting H_0^+ via Route (3) (Enrichment)

$$R_{23}() = R_2() + R_3()$$

$$R_{123}() = R_1() + R_2() + R_3() \leq \alpha$$

Error Probabilities

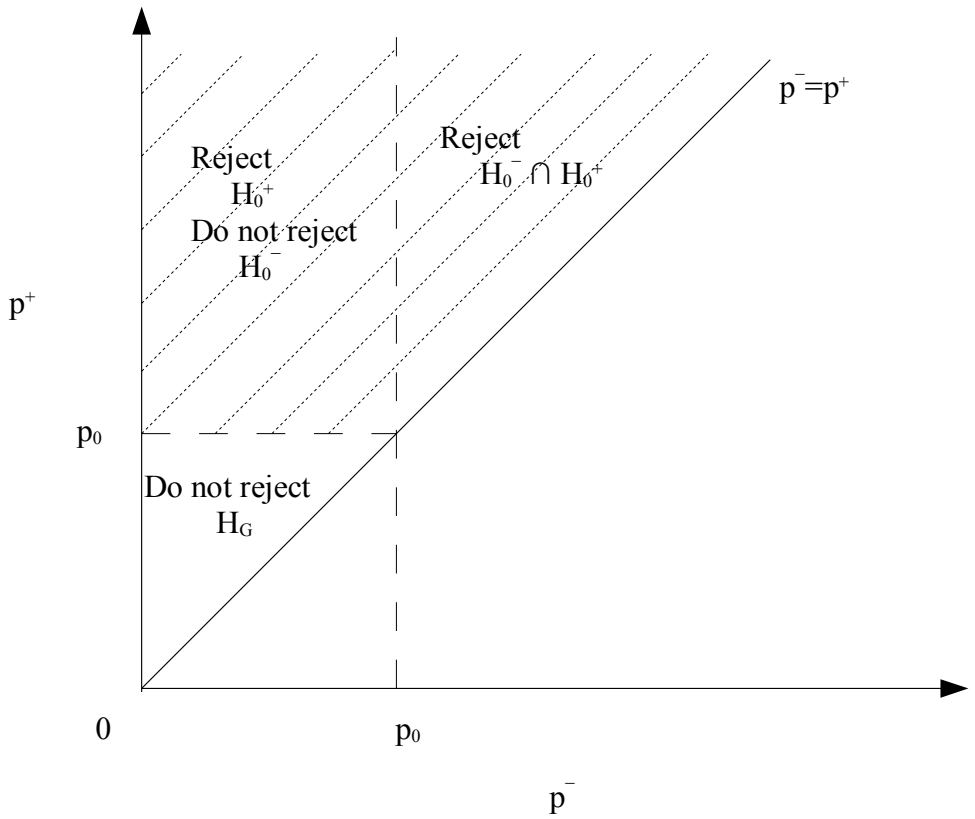
Type I error

- $R_1(p^-) \sim$ probability of rejecting H_0^- via Route (1)
- $R_2(p^-, p^+) \sim$ probability of rejecting H_0^+ via Route (2) (non-monotonic)
- $R_3(p^-, p^+) \sim$ probability of rejecting H_0^+ via Route (3) (Enrichment)

$$R_{23}() = R_2() + R_3()$$

$$R_{123}() = R_1() + R_2() + R_3() \leq \alpha$$

Parameter space of (p^-, p^+)



Generic null hypotheses: $H_0^- : p^- \leq p_0$, $H_0^+ : p^+ \leq p_0$
Global null $H_G : H_0^- \bigcap H_0^+$; $p^- \leq p^+$

The formulae

$$R_1(p^-) = \left(\sum_{i=k_1^-}^{min(N_1^-, k^- - 1)} P(X_2^- \geq k^- - i) P(X_1^- = i) \right) + P(X_1^- \geq k^-)$$

$$R_2(p^-, p^+) = P(X^+ \geq k^+) \left(\sum_{i=k_1^-}^{min(N_1^-, k^- - 1)} P(X_2^- < k^- - i) P(X_1^- = i) \right)$$

$$R_3(p^-, p^+) = P(X_1^- < k_1^-) \times \left(\left\{ \sum_{i=k_1^+}^{min(N_1^+, k_e^+ - 1)} P(X_{2e}^+ \geq k_{2e}^+ - i) P(X_1^+ = i) \right\} + P(X_1^+ \geq k_e^+) \right)$$

Errors all the way !

Real World	Outcomes		
	R_0 No Efficacy	R_1 Unselected	R_{23} Positive only
W_0: No Efficacy (p_0^-, p_0^+)	$R_0(p_0^-, p_0^+)$ True negative	$R_1(p_0^-)$ False positive	$R_{23}(p_0^-, p_0^+)$ False positive
W_1: Unselected (p_1^-, p_1^-)	$R_0(p_1^-, p_1^-)$ False negative	$R_1(p_1^-)$ True positive	$R_{23}(p_1^-, p_1^-)$ Wrong positive
W_{23}: Positive only (p_0^-, p_1^+)	$R_0(p_0^-, p_1^+)$ False negative	$R_1(p_0^-)$ Wrong positive	$R_{23}(p_0^-, p_1^+)$ True positive

Power

$$min(R_1(p_1^-), R_{23}(p_0^-, p_1^+)) \geq 1 - \beta$$

Family Wise Error Rates (Type I)

Let

- Family:= set of null hypotheses $\{H_{01}, H_{02}, \ldots\}$
- $V \sim$ number of incorrectly rejected H_0's

Then, FWER

- $= P(V \geq 1) = 1 - P(V = 0)$
- $\sim$ probability of making at least one type I error in the family *(i.e. rejecting any of the null hypotheses)*
- can be weak, strong, with / without Individual Outcome (IO) control *(i.e. 4 options)*

Power

$$min(R_1(p_1^-), R_{23}(p_0^-, p_1^+)) \geq 1 - \beta$$

Family Wise Error Rates (Type I)

Let

- Family:= set of null hypotheses $\{H_{01}, H_{02}, \ldots\}$
- $V \sim$ number of incorrectly rejected H_0's

Then, FWER

- $= P(V \geq 1) = 1 - P(V = 0)$
- $\sim$ probability of making at least one type I error in the family *(i.e. rejecting any of the null hypotheses)*
- can be weak, strong, with / without Individual Outcome (IO) control *(i.e. 4 options)*

Power

$$min(R_1(p_1^-), R_{23}(p_0^-, p_1^+)) \geq 1 - \beta$$

Family Wise Error Rates (Type I)

Let

- Family:= set of null hypotheses $\{H_{01}, H_{02}, \ldots\}$
- $V \sim$ number of incorrectly rejected H_0's

Then, FWER

- $= P(V \geq 1) = 1 - P(V = 0)$
- $\sim$ probability of making at least one type I error in the family *(i.e. rejecting any of the null hypotheses)*
- can be weak, strong, with / without Individual Outcome (IO) control *(i.e. 4 options)*

Proposed Type I error control

Option 1 (Weak FWER): $\sum_i FP_i \leq \alpha$

Option 2 (Strong FWER): $max(\sum_i FP_i, WP_1, WP_2) \leq \alpha$

Option 3 (Weak IH): $max(FP_1, FP_2) \leq \alpha/2$

Option 4 (Strong IH):

- $max(FP_1, FP_2) \leq \alpha/2$
- $max(WP_1, WP_2) \leq \alpha$

Weak FWER

- probability of type I error in R_1 or R_{23}
- reject H_0^- (via R_1) or reject H_0^+ (via R_{23})
- $\sum$ (False positives) $\leq \alpha$

	R_0	R_1	R_{23}
W_0	[ ]	$\{ \sum \leq \alpha \}$	
W_1	[ ]	$\geq 1 - \beta$	[ ]
W_{23}	[ ]	[ ]	$\geq 1 - \beta$

$$R_1(p_0^-) + R_2(p_0^-, p_0^+) + R_3(p_0^-, p_0^+) \leq \alpha$$

Expected Sample Size

$$E(N) = N_1^- + N_1^+ + N_2[1 - PET(p^-)] + N_{2e}^+ P(X_1^- < k_1^-)[(1 - PET(p^+))]$$

where

$$PET(p^-) = P(X_1^- < k_1^-) + P(X_1^- \geq k^-)$$
$$PET(p^+) = P(X_1^+ < k_1^+) + P(X_1^+ \geq k_e^+)$$

Overall probability of early termination

$$PET = P(X_1^- \geq k^-) + P(X_1^- < k_1^-)[P(X_1^+ \geq k_e^+) + P(X_1^+ < k_1^+)]$$

Note: $PET \neq PET(p^-) + PET(p^+)$

Results (J & H)

Operating characteristics given $k_1^- = 2, k_1^+ = 1, k^- = 4, k^+ = 4, k_e^+ = 5$

p^-	p^+	Power: Route 1 *(unselected)*	Power: Routes $2+3$ *(positives)*	$E(N)_{Simon}$	$E(N)_{Adaptive}$	$E(N)_{Adaptive}$ / $E(N)_{Simon}$
0.03	0.03	0.067	0.012	74.61	65.79	0.881
0.03	0.10	0.067	0.373	85.21	76.91	0.902
0.03	0.15	0.067	0.624	88.36	80.21	0.907
0.10	0.15	0.755	0.624	127.66	80.03	0.626
0.10	0.25	0.755	0.807	129.78	80.44	0.619
0.15	0.30	0.952	0.852	136.99	80.10	0.584

Significance, $\alpha = 0.079$

Design

$$(k_1^- k_1^+)/(N_1^- N_1^+) \rightarrow (k_e^+/N_{2e}^+)|(k^- k^+)/(N^- N^+)$$
$$(2\ 1)/(34\ 14) \rightarrow (5/50)\ |\ (4\ 4)/(53\ 27)$$

Optimal design results

Weak FWER ($p_0 = 0.03, \alpha = 0.05, \beta = 0.2$)

p_1^-	p_1^+	ESS	Design $(k_1^- k_1^+)/(N_1^- N_1^+) \rightarrow (k_e^+/N_{2e}^+) \mid (k^- k^+)/(N^- N^+)$
0.10	0.10	110.2	(3 2)/(44 34) → (7/104) | (9 4)/(135 53)
0.10	0.15	77.9	(2 2)/(32 21) → (6/67) | (7 3)/(106 29)
0.10	0.25	59.9	(2 1)/(34 8) → (4/29) | (6 2)/(87 9)
0.15	0.15	46.9	(2 1)/(20 12) → (4/43) | (6 2)/(66 21)
0.15	0.25	32.5	(1 1)/(12 7) → (4/28) | (4 2)/(43 11)
0.15	0.35	27.8	(1 1)/(11 5) → (3/15) | (4 2)/(47 7)
0.25	0.25	18.4	(1 1)/(6 6) → (3/24) | (3 2)/(23 13)
0.25	0.40	13.4	(1 1)/(6 4) → (2/9) | (3 2)/(23 5)

Conclusions

- Extension to randomised Phase II / III trials
- Study different outcomes: PFS, OS
- Weak FWER: smaller study with sufficient control for early phase trial.
- Strong FWER: late stage definitive study.
- Choice of error control: what clinicians / trialists prefer

An optimal stratified Simon two–stage design
D. Parashar, J. Bowden, C. Starr, L. Wernisch, A. Mander
Pharm. Statistics (2016), doi: 10.1002/pst.1742.

WARWICK
THE UNIVERSITY OF WARWICK