Sub-group C luster-Based Adaptive Designs for Precision Medicine (S C UBA)

Yuan Ji, PhD

Program for Computational Genomics & Medicine
NorthShore University HealthSystem
Department of Public Health Sciences
The University of Chicago

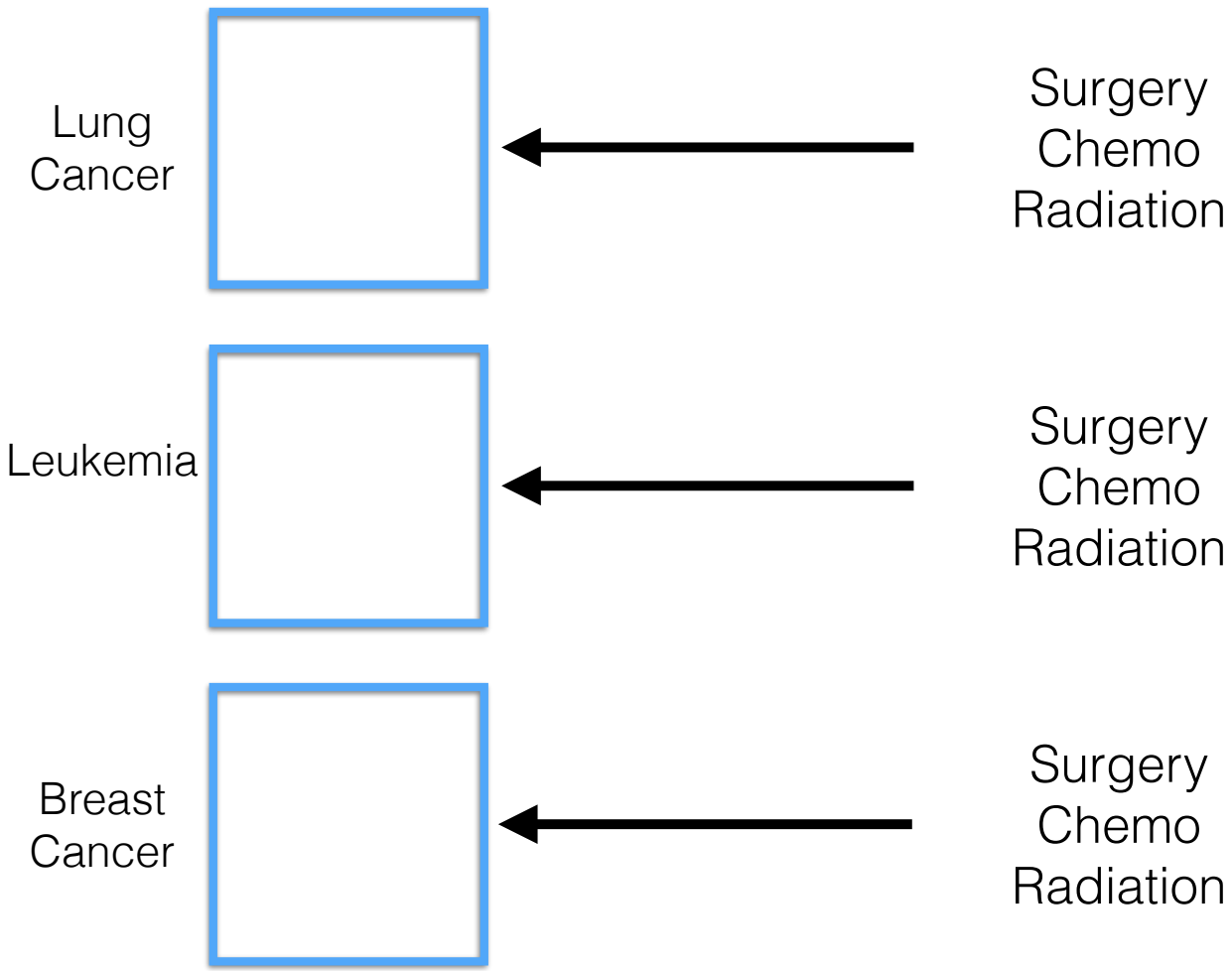
5.8.2015

SUBA and SCUBA

SUBA design (Xu et al. 2014, *Statistics in Biosciences*)

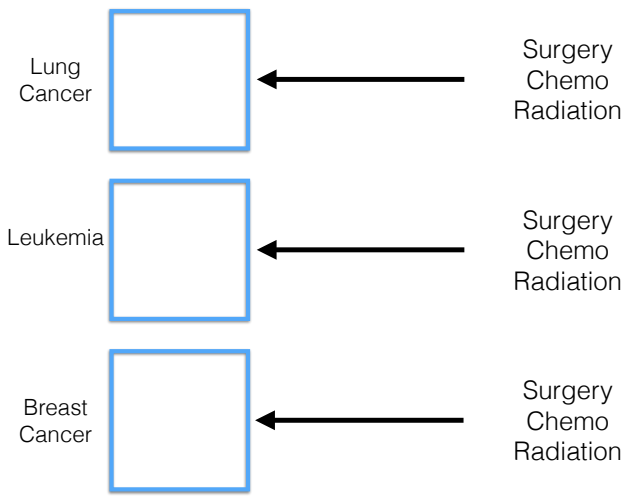
SCUBA design (Guo, Catenacci, and Ji, 2015. To be submitted)

One-size Fit All Cancer Treatment



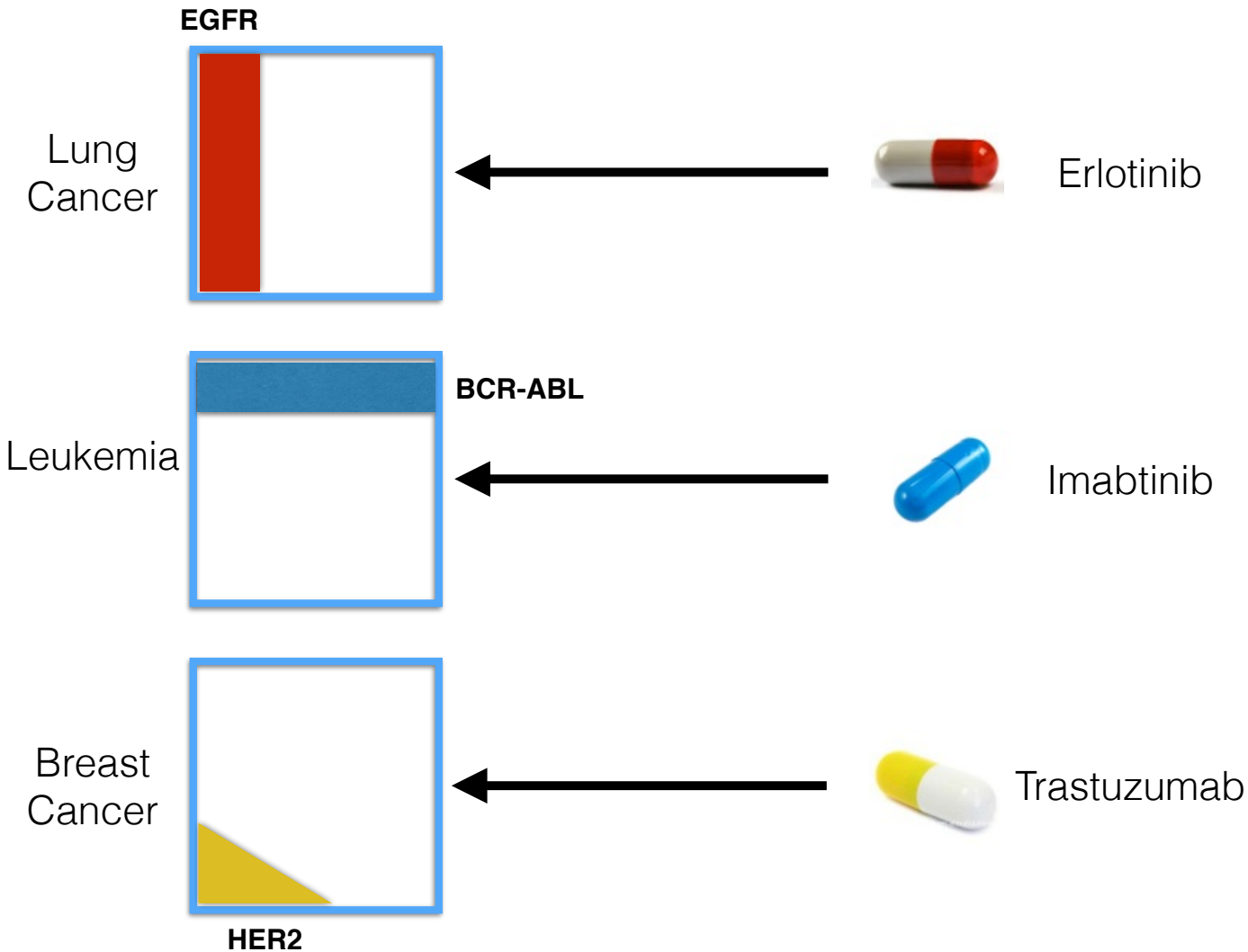
One-size Fit All Cancer Treatment – 2

The problem of one-size fit all:



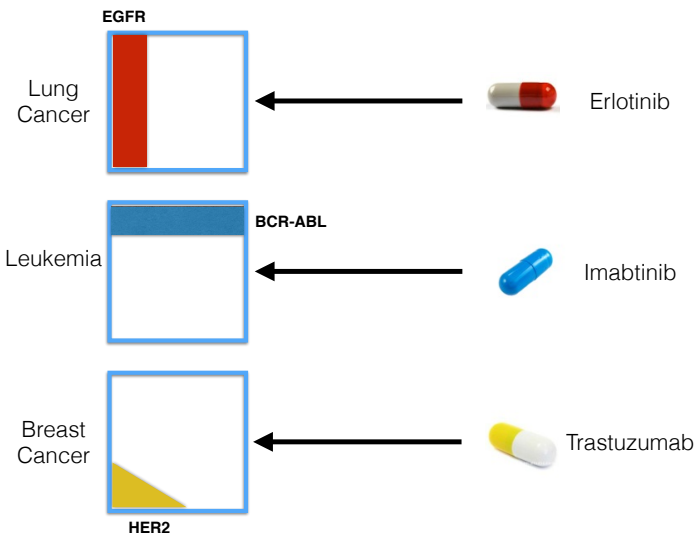
- Treat the "phenotypes" with brute force
- Severe side effects and poor quality of life
- Unpredictable prognosis
- Risk of over-treatment

Targeted Cancer Treatment



Targeted Cancer Treatment – 2

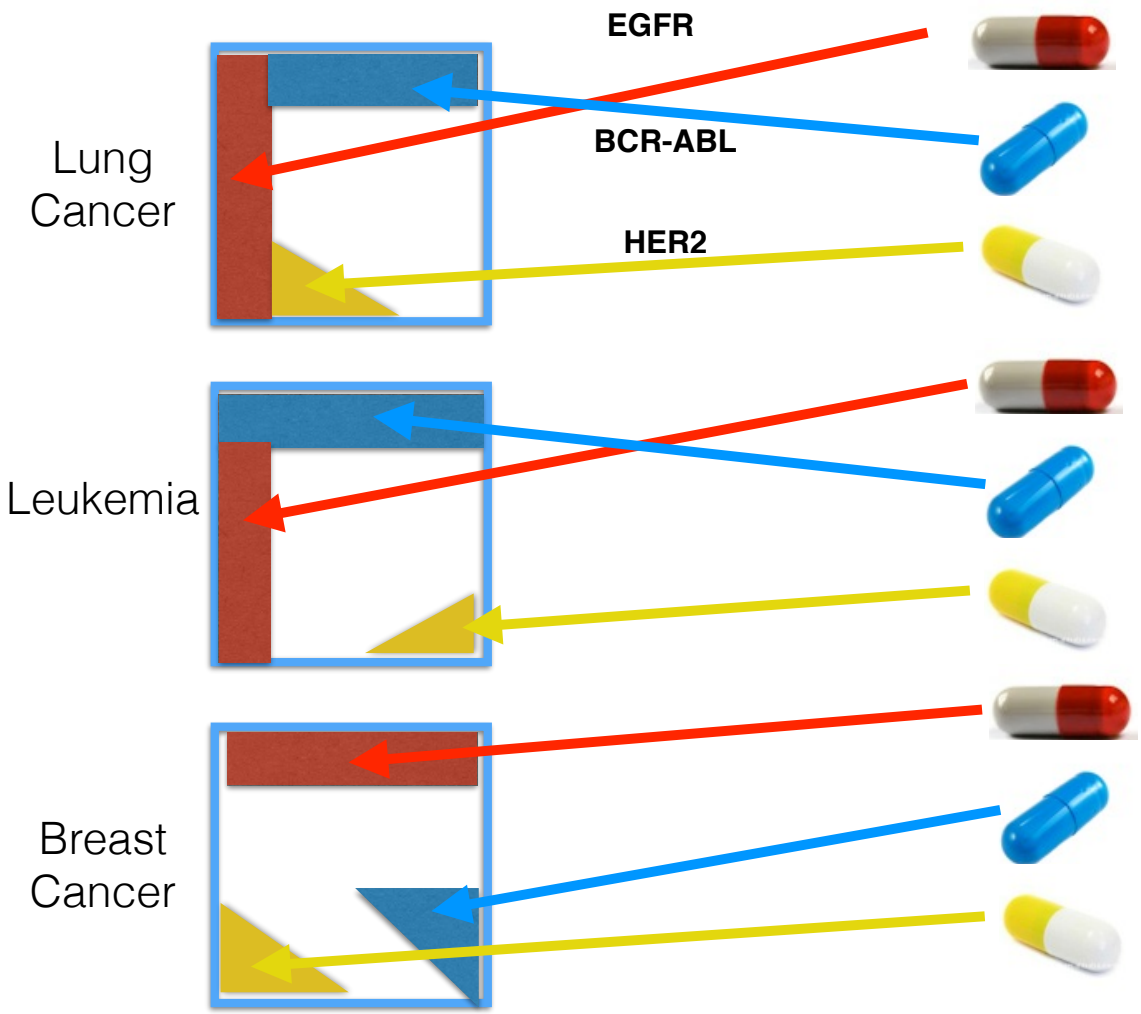
The benefits of being on target:



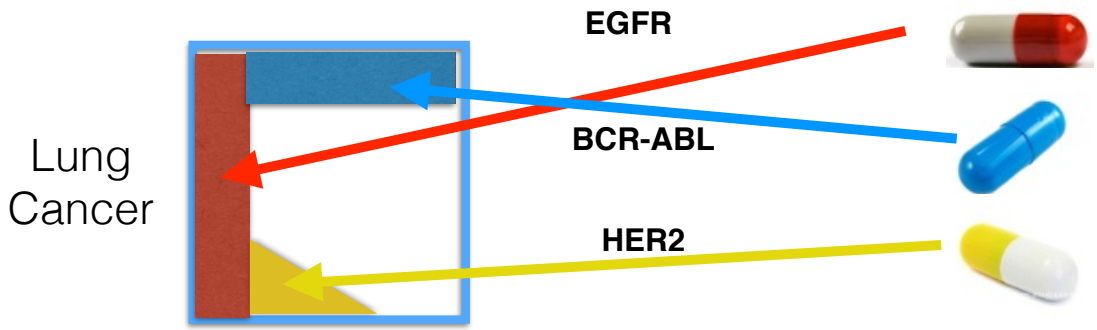
- Treat the "genotypes" that are causal of phenotypes
- Typically mild side effects and high quality of life
- Predictable prognosis
- Less chance of over-treatment

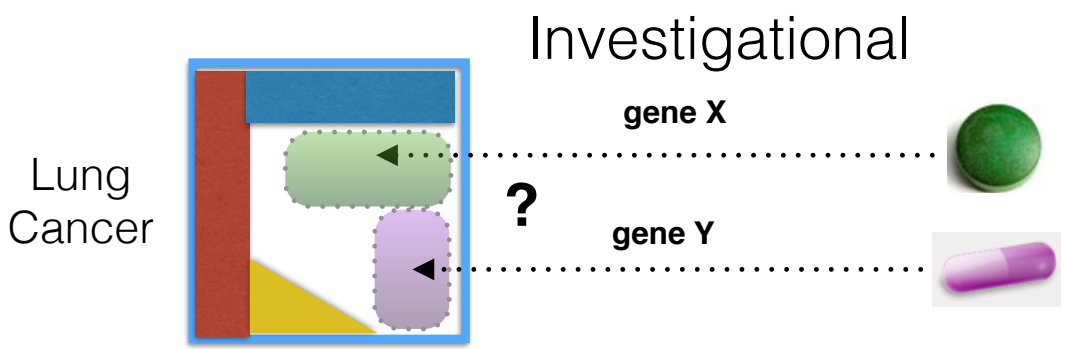
Precision Cancer Care

Disease agnostic; Genotype specific (e.g., NCI MATCH trial)



Patients with NO actionable genotypes





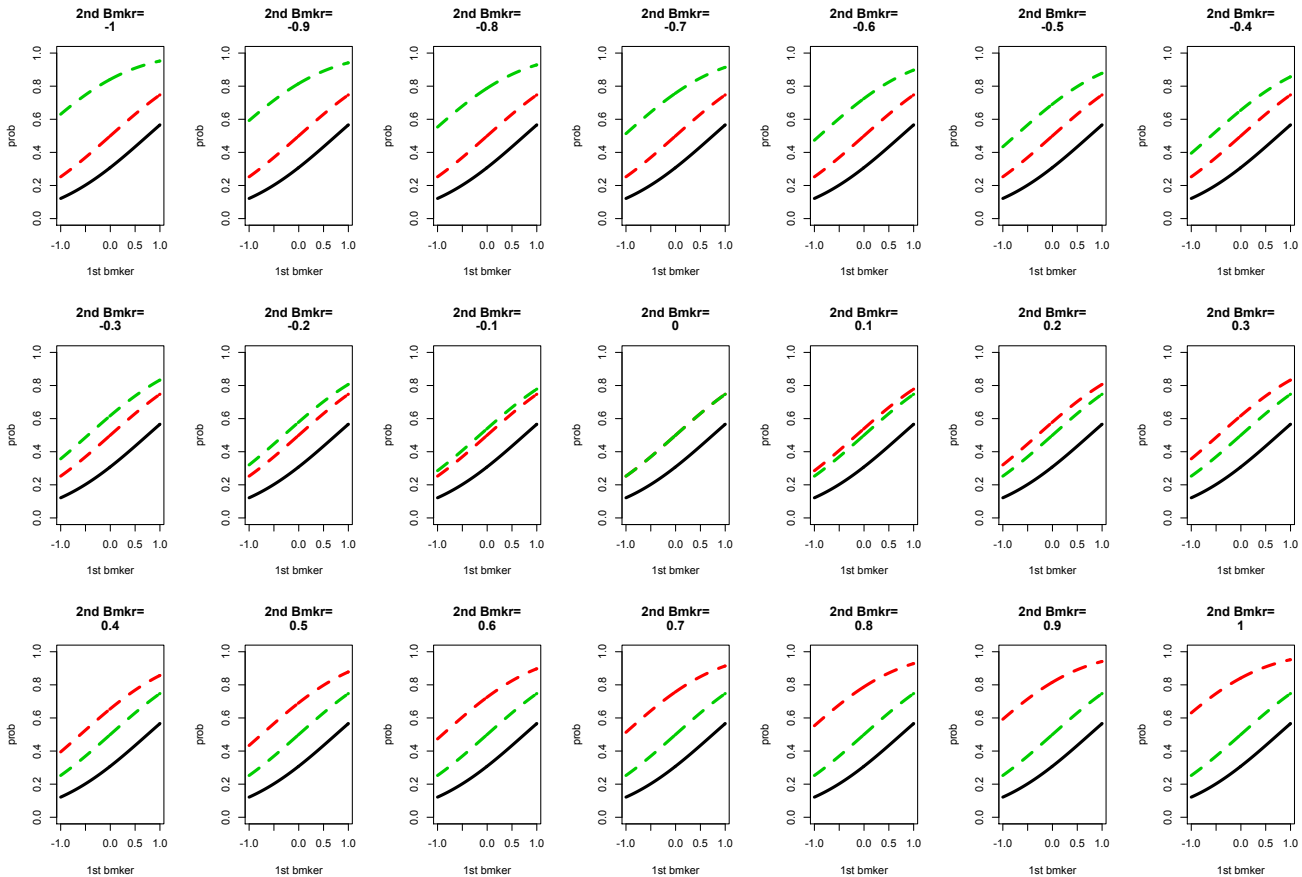
When to use SCUBA

Which treatment is the best depends on status of biomarkers X

When to use SCUBA

Which treatment is the best depends on status of biomarkers X

A hypothetical example

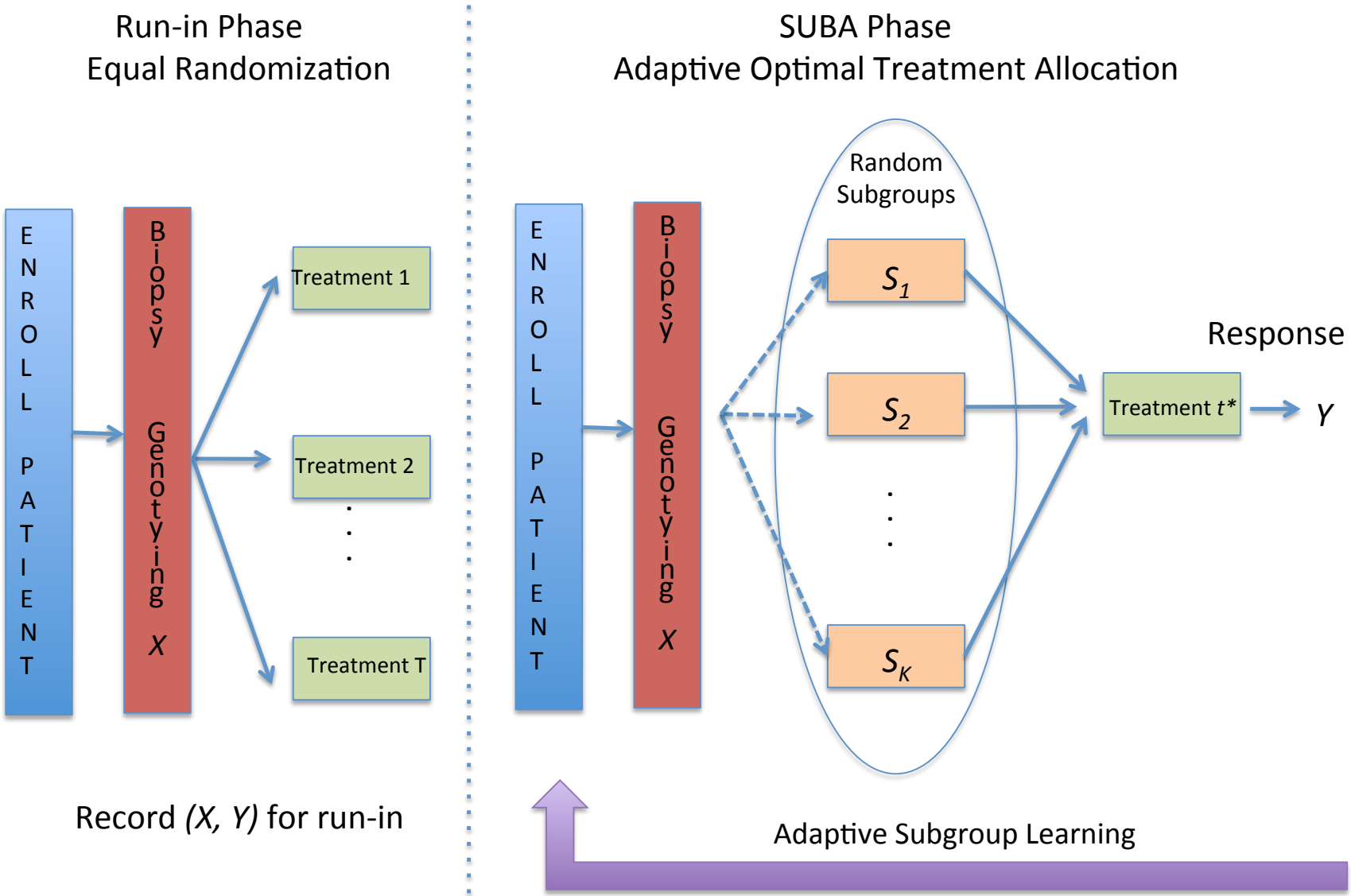


Trial Setup for SCUBA

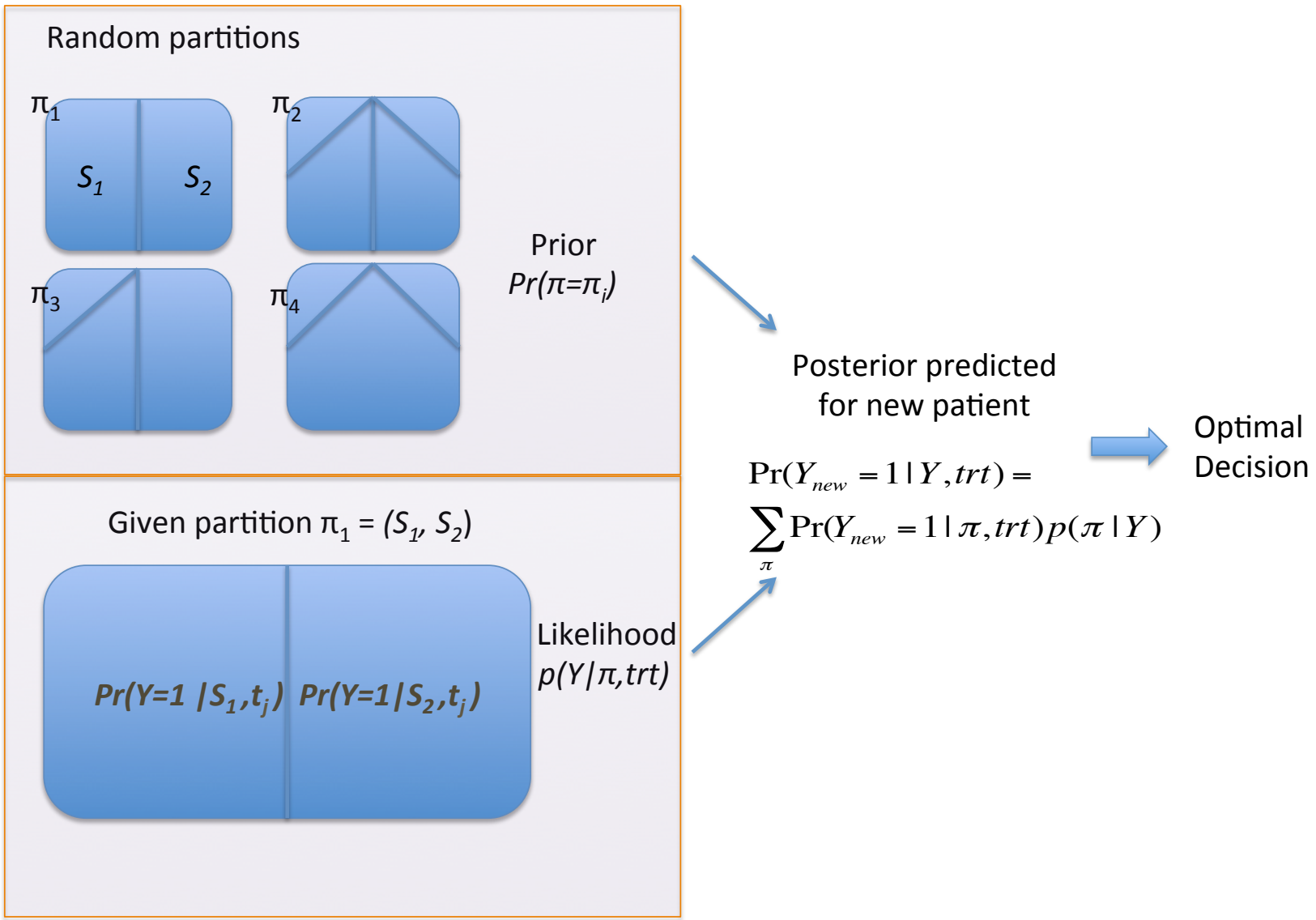
- Patients without known targeted drugs (e.g., relapsed patients out of options)
- A set of relevant biomarkers (or PCs) $\boldsymbol{X} = (X_1, X_2, ...X_p)$, p small
- A set of candidate drugs $(t_1, t_2, ..., t_T)$, $T \geq 1$.

Goal: find a rule that allocates patient subgroup $S_k(\boldsymbol{X})$ to drug t_k, such that the response rate under the rule is better than standard strategy , such as treating ALL patients with drug t_k or randomization between different drugs (in a trial) .

Overview of SCUBA



Overview of Probability Model and Inference



CART-type model for binary outcomes

Consider binary outcome $y_i \in \{0, 1\}$ where 0 and 1 denotes no response and response.

$[\Pi]$

Let $\Pi = (B_1, B_1, \ldots, B_M)$ be a random partition of $\mathbf{X} = R^k$;

$[\theta \mid \Pi]$:

$$\theta_{j,m} \mid \Pi \overset{iid}{\sim} \text{Beta}(a, b) \qquad j = 1, 2, 3, \;\; m = 1, \ldots, M$$

$[Y|X, t, \Pi, \theta]$:

$$Y_i \mid X_i, t_i = j, \Pi, \theta \sim Bernoulli(\theta_{j,m_{X_i}}), \quad m_{X_i} = (m : X_i \in B_m)$$

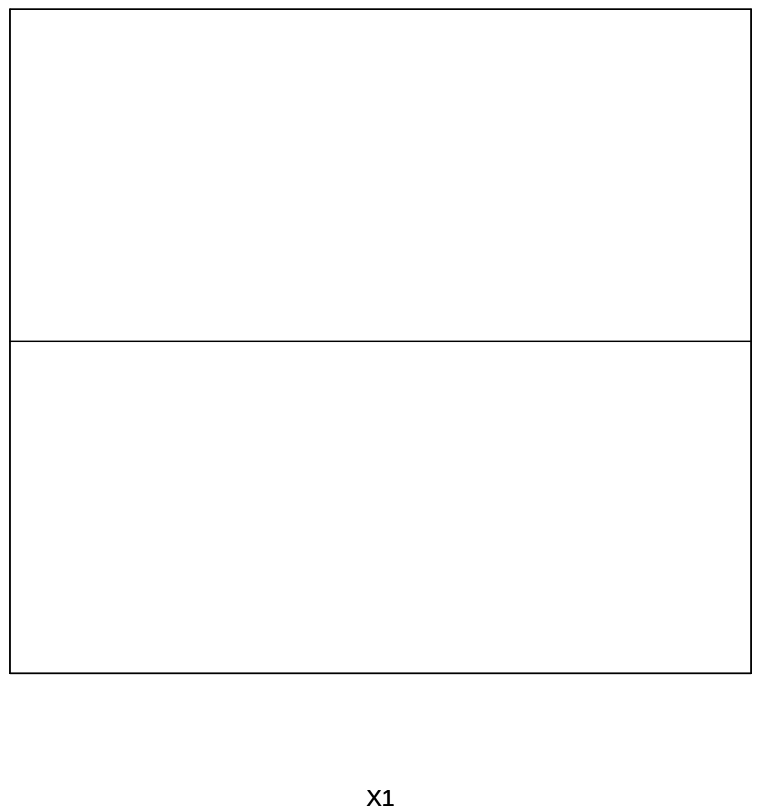
A simple random partition $P(\Pi)$
is constructed by randomly selecting one biomarker and partition the patient groups into half according to the median.

Prior on Π

Choose one X_i with probability p_i ($\sum_i p_i = 1$), and with probability q to split the space by $\mathbb{I}(X_i > median(X_i))$. Do the same for the new subset if the split does occur. Repeat 3 times.

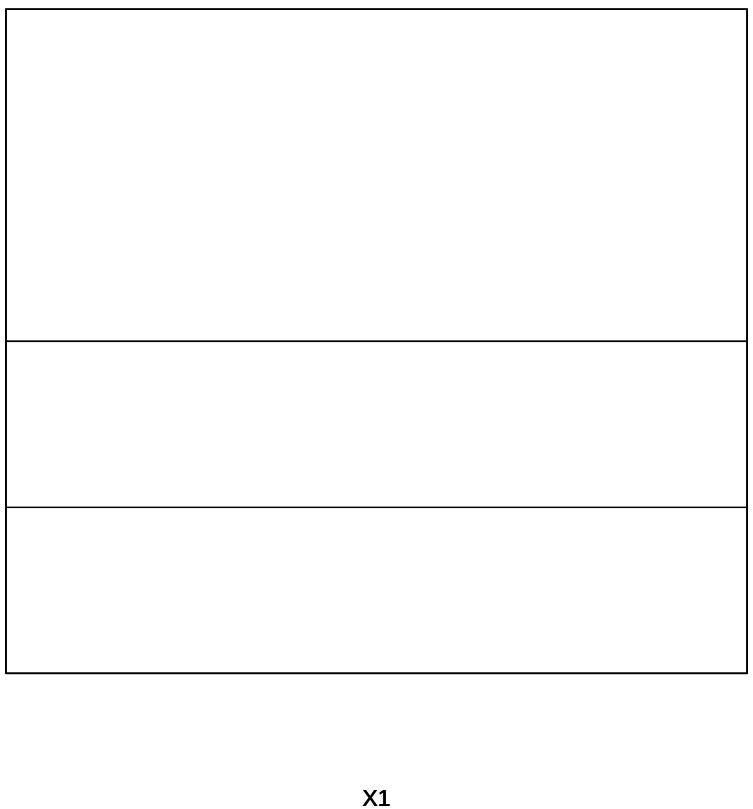
Prior on Π

Choose one X_i with probability p_i ($\sum_i p_i = 1$), and with probability q to split the space by $\mathbb{I}(X_i > median(X_i))$. Do the same for the new subset if the split does occur. Repeat 3 times.



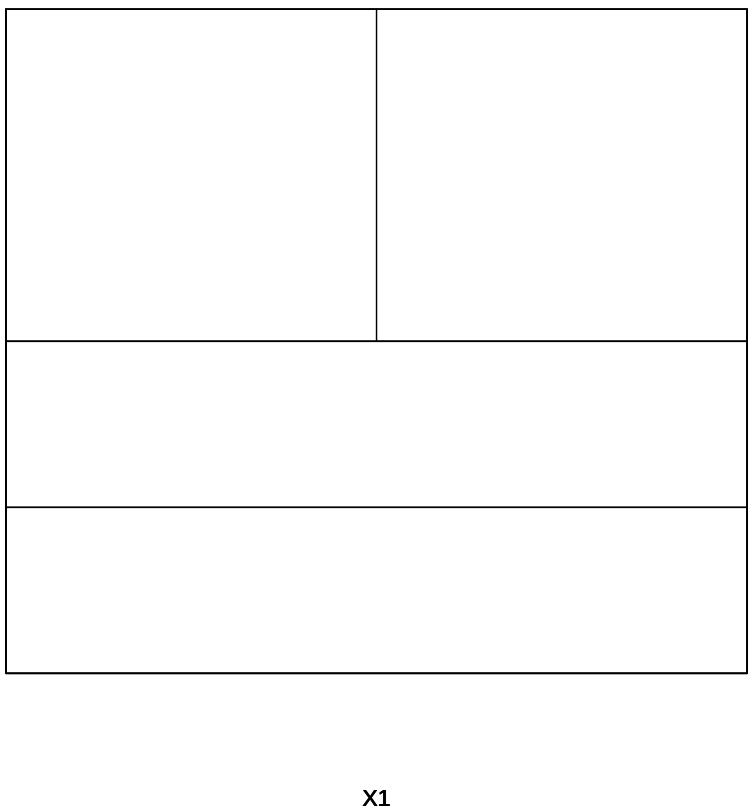
Prior on Π

Choose one X_i with probability p_i ($\sum_i p_i = 1$), and with probability q to split the space by $\mathbb{I}(X_i > median(X_i))$. Do the same for the new subset if the split does occur. Repeat 3 times.



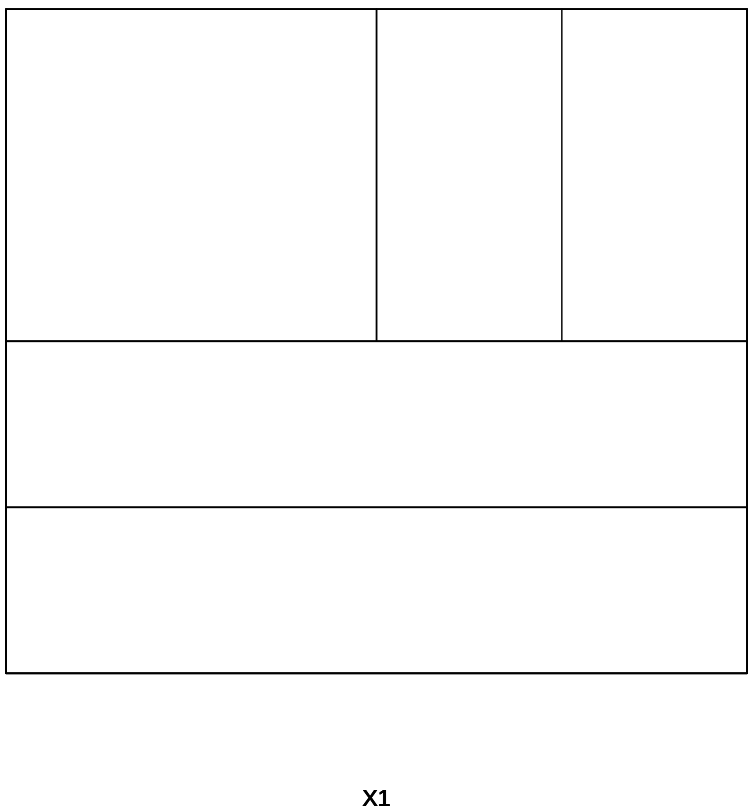
Prior on Π

Choose one X_i with probability p_i ($\sum_i p_i = 1$), and with probability q to split the space by $\mathbb{I}(X_i > median(X_i))$. Do the same for the new subset if the split does occur. Repeat 3 times.



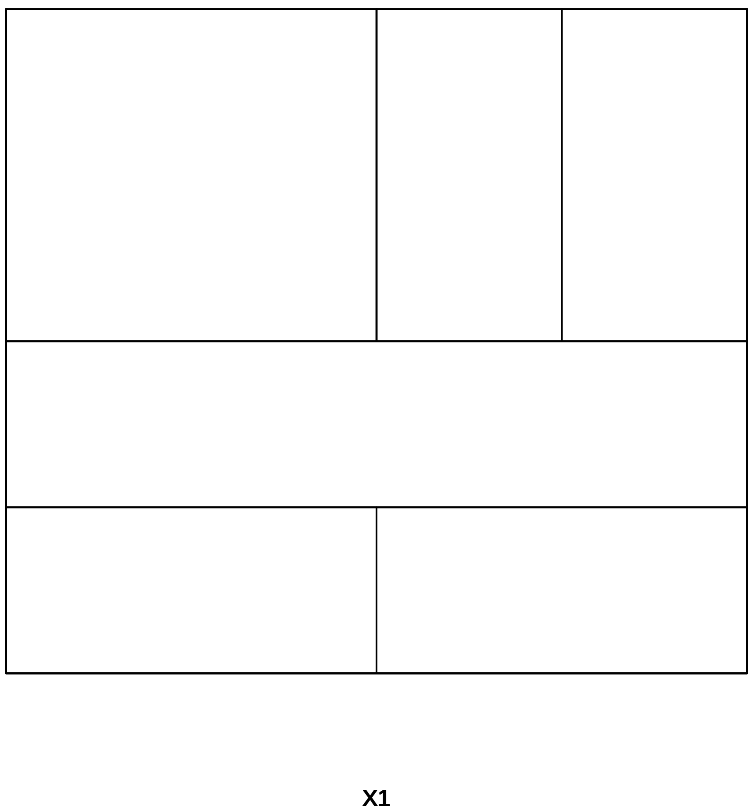
Prior on Π

Choose one X_i with probability p_i ($\sum_i p_i = 1$), and with probability q to split the space by $\mathbb{I}(X_i > median(X_i))$. Do the same for the new subset if the split does occur. Repeat 3 times.



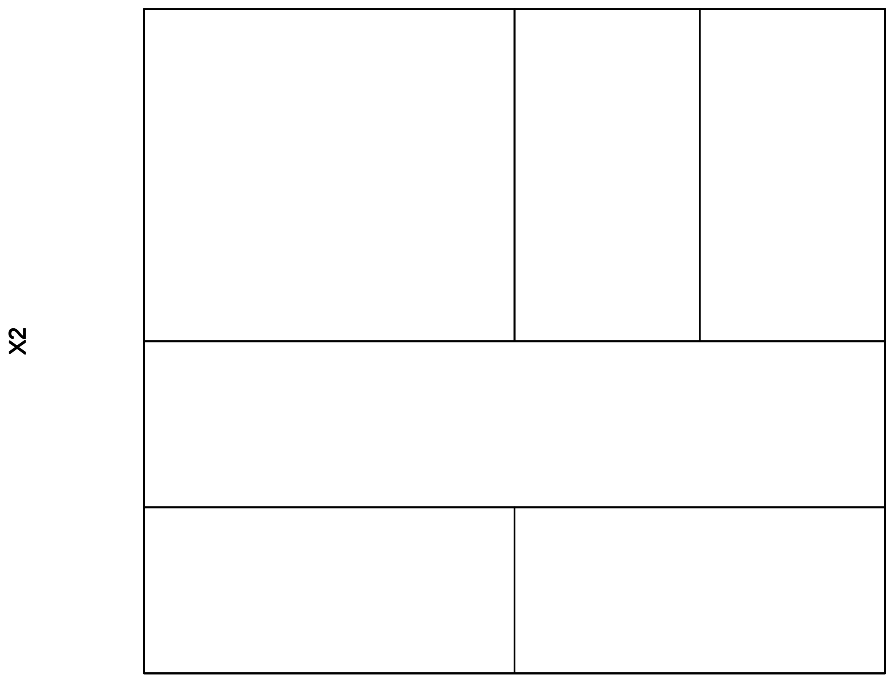
Prior on Π

Choose one X_i with probability p_i ($\sum_i p_i = 1$), and with probability q to split the space by $\mathbb{I}(X_i > median(X_i))$. Do the same for the new subset if the split does occur. Repeat 3 times.



Prior on Π

Choose one X_i with probability p_i ($\sum_i p_i = 1$), and with probability q to split the space by $\mathbb{I}(X_i > median(X_i))$. Do the same for the new subset if the split does occur. Repeat 3 times.



Prior probability $p_2 q \times p_2 q p_1 q \times p_1^2 q^2 (1 - p_1 - p_2)^2$

Subgroup-Based Trial Design

Let N be the total sample size. For patient i, let x_i be the biomarker profile, t_i be the treatment allocation, and y_i be the response outcome.

1 An initial run-in with an equal randomization of $n \leq N$ patients.

3 Compute for patient $n+1$,

$$q_{n+1}(t) = Pr(y_{n+1} = 1 \mid \mathbf{y_n}, \mathbf{x_n}, \mathbf{t_n}, x_{n+1}, t_{n+1} = t) =$$

$$\int Pr(y_{n+1} = 1 \mid x_{n+1}, t_{n+1} = t, \pi, \theta) p(\pi, \theta \mid \mathbf{y_n}, \mathbf{x_n}, \mathbf{t_n}) d\theta$$

.

4 Allocate patient $n+1$ to treatment $t* = \arg\max_t q_{n+1}(t)$.

5 Update the observed data as $(\mathbf{y_{n+1}}, \mathbf{x_{n+1}}, \mathbf{t_{n+1}})$, and repeat steps 2-4 for patient $n+2$, $n+3$, ...N.

Subgroup-Based Trial Design

Let N be the total sample size. For patient i, let x_i be the biomarker profile, t_i be the treatment allocation, and y_i be the response outcome.

1. An initial run-in with an equal randomization of $n \leq N$ patients.
2. Fit a Bayesian model $\prod_i p(y_i \mid x_i = x, t_i = t, \pi, \theta) \cdot p(\pi)p(\theta)$ to the data of n patients from step 1, denoted as $(\mathbf{y_n}, \mathbf{x_n}, \mathbf{t_n})$.
3. Compute for patient $n+1$,

$$q_{n+1}(t) = Pr(y_{n+1} = 1 \mid \mathbf{y_n}, \mathbf{x_n}, \mathbf{t_n}, x_{n+1}, t_{n+1} = t) =$$

$$\int Pr(y_{n+1} = 1 \mid x_{n+1}, t_{n+1} = t, \pi, \theta) p(\pi, \theta \mid \mathbf{y_n}, \mathbf{x_n}, \mathbf{t_n}) d\theta$$

.

4. Allocate patient $n+1$ to treatment $t* = \arg\max_t q_{n+1}(t)$.
5. Update the observed data as $(\mathbf{y_{n+1}}, \mathbf{x_{n+1}}, \mathbf{t_{n+1}})$, and repeat steps 2-4 for patient $n+2$, $n+3$, ...N.

Subgroup-Based Trial Design

Let N be the total sample size. For patient i, let x_i be the biomarker profile, t_i be the treatment allocation, and y_i be the response outcome.

1. An initial run-in with an equal randomization of $n \leq N$ patients.

2. Fit a Bayesian model $\prod_i p(y_i \mid x_i = x, t_i = t, \pi, \theta) \cdot p(\pi)p(\theta)$ to the data of n patients from step 1, denoted as $(\mathbf{y_n}, \mathbf{x_n}, \mathbf{t_n})$.
3. Compute for patient $n+1$,

$$q_{n+1}(t) = Pr(y_{n+1} = 1 \mid \mathbf{y_n}, \mathbf{x_n}, \mathbf{t_n}, x_{n+1}, t_{n+1} = t) =$$

$$\int Pr(y_{n+1} = 1 \mid x_{n+1}, t_{n+1} = t, \pi, \theta) p(\pi, \theta \mid \mathbf{y_n}, \mathbf{x_n}, \mathbf{t_n}) d\theta$$

.

4. Allocate patient $n+1$ to treatment $t* = \arg\max_t q_{n+1}(t)$.

5. Update the observed data as $(\mathbf{y_{n+1}}, \mathbf{x_{n+1}}, \mathbf{t_{n+1}})$, and repeat steps 2-4 for patient $n+2$, $n+3$, ...N.

Subgroup-Based Trial Design

Let N be the total sample size. For patient i, let x_i be the biomarker profile, t_i be the treatment allocation, and y_i be the response outcome.

1. An initial run-in with an equal randomization of $n \leq N$ patients.

2. Fit a Bayesian model $\prod_i p(y_i \mid x_i = x, t_i = t, \pi, \theta) \cdot p(\pi)p(\theta)$ to the data of n patients from step 1, denoted as $(\mathbf{y_n}, \mathbf{x_n}, \mathbf{t_n})$.
3. Compute for patient $n+1$,

$$q_{n+1}(t) = Pr(y_{n+1} = 1 \mid \mathbf{y_n}, \mathbf{x_n}, \mathbf{t_n}, x_{n+1}, t_{n+1} = t) =$$

$$\int Pr(y_{n+1} = 1 \mid x_{n+1}, t_{n+1} = t, \pi, \theta) p(\pi, \theta \mid \mathbf{y_n}, \mathbf{x_n}, \mathbf{t_n}) d\theta$$

.

4. Allocate patient $n+1$ to treatment $t* = \arg\max_t q_{n+1}(t)$.

5. Update the observed data as $(\mathbf{y_{n+1}}, \mathbf{x_{n+1}}, \mathbf{t_{n+1}})$, and repeat steps 2-4 for patient $n+2$, $n+3$, ...N.

Subgroup-Based Trial Design

Let N be the total sample size. For patient i, let x_i be the biomarker profile, t_i be the treatment allocation, and y_i be the response outcome.

1. An initial run-in with an equal randomization of $n \leq N$ patients.

2. Fit a Bayesian model $\prod_i p(y_i \mid x_i = x, t_i = t, \pi, \theta) \cdot p(\pi)p(\theta)$ to the data of n patients from step 1, denoted as $(\mathbf{y_n}, \mathbf{x_n}, \mathbf{t_n})$.
3. Compute for patient $n+1$,

$$q_{n+1}(t) = Pr(y_{n+1} = 1 \mid \mathbf{y_n}, \mathbf{x_n}, \mathbf{t_n}, x_{n+1}, t_{n+1} = t) =$$

$$\int Pr(y_{n+1} = 1 \mid x_{n+1}, t_{n+1} = t, \pi, \theta) p(\pi, \theta \mid \mathbf{y_n}, \mathbf{x_n}, \mathbf{t_n}) d\theta$$

.

4. Allocate patient $n+1$ to treatment $t* = \arg\max_t q_{n+1}(t)$.

5. Update the observed data as $(\mathbf{y_{n+1}}, \mathbf{x_{n+1}}, \mathbf{t_{n+1}})$, and repeat steps 2-4 for patient $n+2$, $n+3$, ...N.

A Breast Cancer Trial

- Patients eligible to the trial are
  - have undergone neoadjuvant systemic therapy (NST) and surgery
  - have their protein biomarkers measured (through biopsy samples) at the end of NST but before surgery

A Breast Cancer Trial

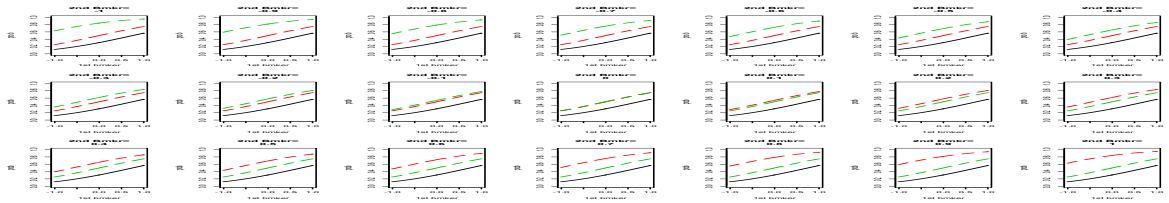
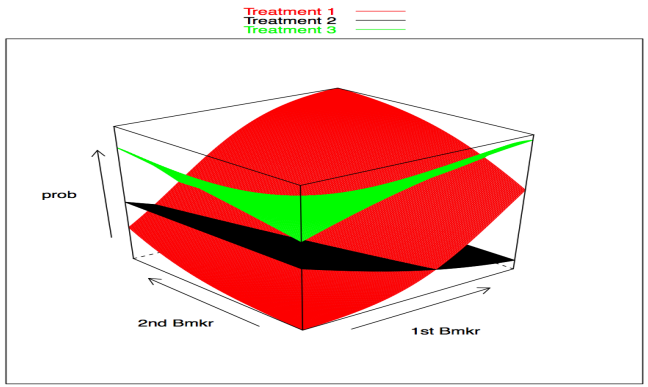
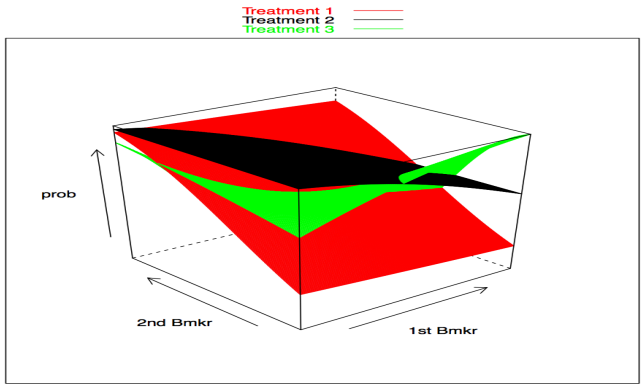
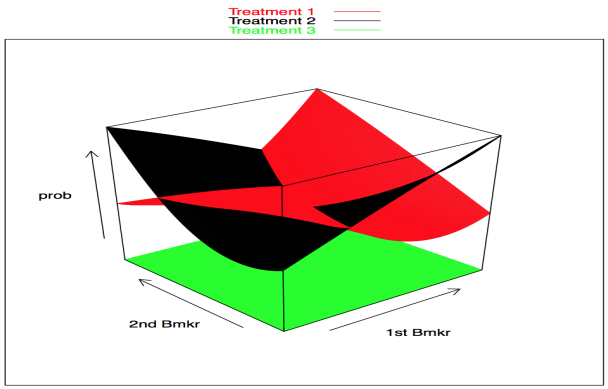
- Patients eligible to the trial are
  - have undergone neoadjuvant systemic therapy (NST) and surgery
  - have their protein biomarkers measured (through biopsy samples) at the end of NST but before surgery
- Three candidate treatments
  - Poly (ADP-ribose) polymerase (PARP) inhibitor – DNA repair and programmed cell death
  - PI3K pathway inhibitor – cell growth, proliferation, differentiation, motility, survival and intracellular trafficking
  - Cell cycle inhibitor
- About 300 patients had expression measurements for a number of proteins from MAPK and PI3K pathways.

Simulation Setup

Basic setup

- Samp size $N = 300$, run-in phase $n = 100$ (equal randomization), $T = 3$ treatment arms
- Six scenarios, 1,000 simulated trial per scenario.
- Compare to ER, AR (outcome adaptive), and probit-reg designs.

Simulation Scenarios

Sc 1-2	
Sc 3	 
Sc 4-5	
Sc 6	all treatments are the same regardless of X.

Operating Characteristics – All scenarios

Sc		ER			AR			Reg			SUBA		
	S*	1	2	3	1	2	3	1	2	3	1	2	3
1	/	66.76	66.60	66.64	83.02	65.35	51.63	119.46	70.13	10.41	**177.11**	18.67	4.22
2	S_1	33.49	33.09	33.24	33.37	33.19	33.25	35.24	32.88	31.69	**72.57**	18.37	8.88
	S_3	33.27	33.51	33.40	33.41	33.25	33.53	35.42	33.01	31.76	8.63	17.79	**73.77**
3	S_1	19.49	19.09	19.29	22.21	17.63	18.03	18.65	16.40	22.81	**41.11**	8.94	7.82
	S_2	25.23	25.17	25.35	21.13	26.81	27.80	24.10	21.86	29.79	13.67	**35.91**	26.17
	S_3	22.05	22.34	22.00	24.61	20.52	21.26	21.27	18.99	26.12	11.33	11.54	**43.52**
4	S_1	33.26	33.11	33.44	43.01	42.32	14.49	51.81	48.00	0	**52.76**	46.96	0.10
	S_2	33.50	33.49	33.20	42.32	43.46	14.41	51.75	48.44	0	50.78	**49.29**	0.11
5	S_1	33.26	33.11	33.44	39.14	38.49	22.19	51.51	48.25	0.05	**51.13**	47.05	1.63
	S_2	33.50	33.49	33.20	38.29	39.32	22.58	51.22	48.92	0.05	47.07	**51.53**	1.59
6	/	66.76	66.60	66.64	66.66	66.89	66.46	65.04	67.84	67.12	66.90	64.20	68.90

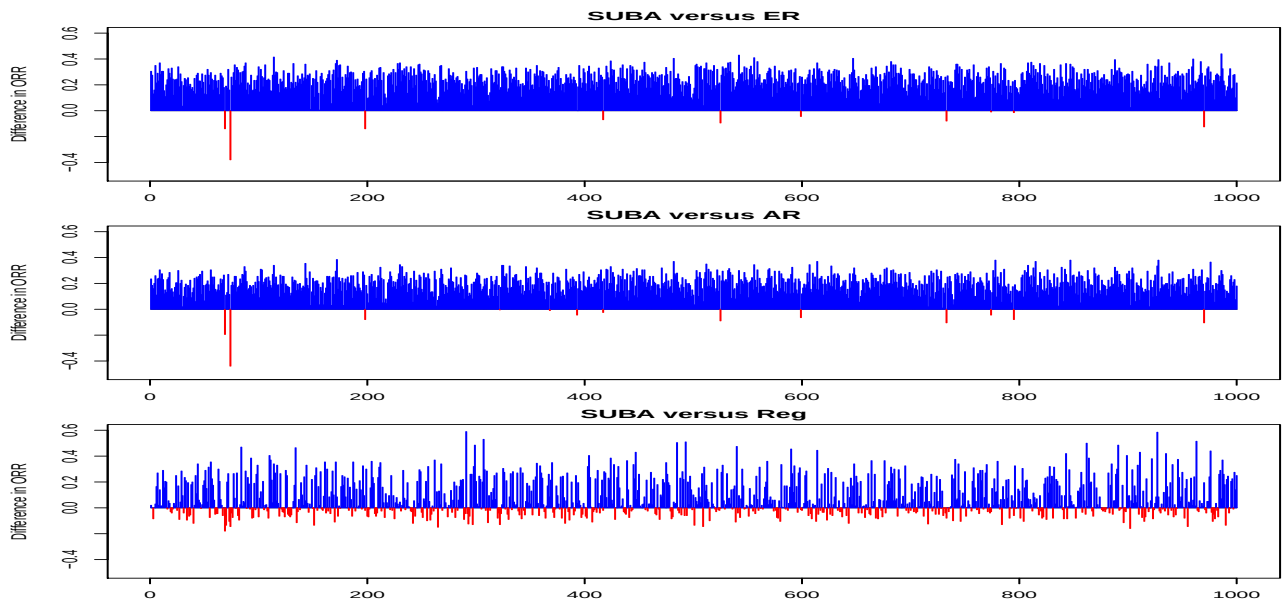
*: S_t is the subset of of biomarker space X in which the t-th treatment has the highest response rate.

Operating Characteristics – All scenarios

Define the overall response rate (ORR) as the proportion of responders among those patients who are treated after the run-in phase

$$\text{ORR} = \frac{1}{N-n} \sum_{i=n+1}^{N} I(y_i = 1),$$

- Plot ORR(SUBA) - ORR(design): Sc 1 Sc 2 Sc 3 Sc 4 Sc 5 Sc 6

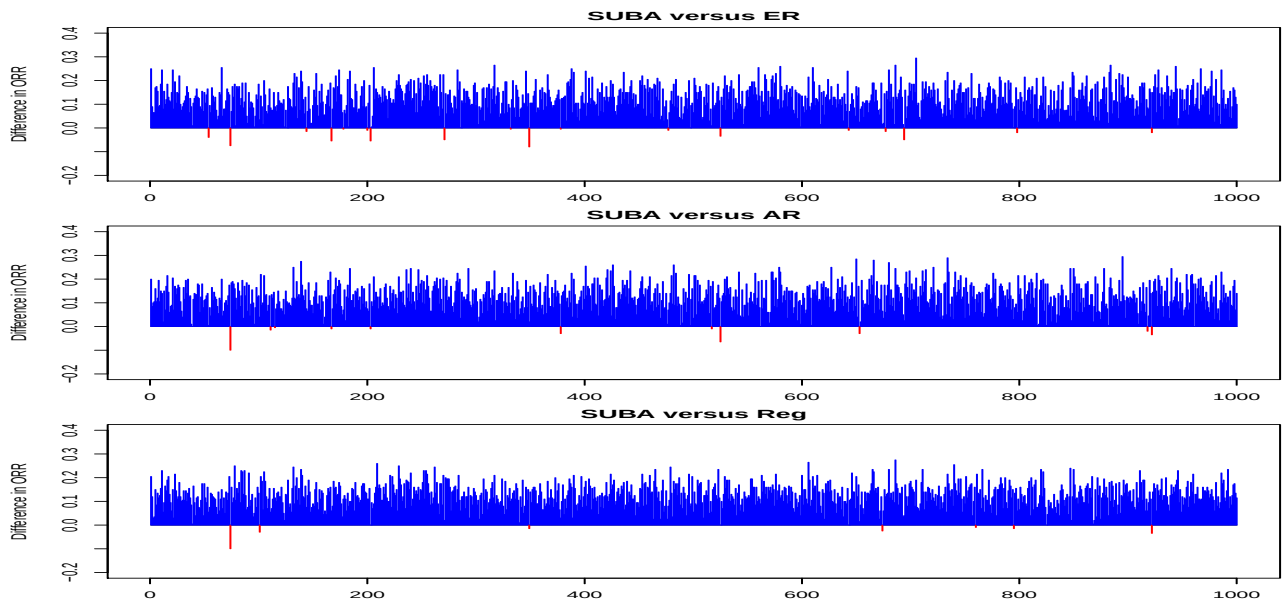


Operating Characteristics – All scenarios

Define the overall response rate (ORR) as the proportion of responders among those patients who are treated after the run-in phase

$$\text{ORR} = \frac{1}{N-n} \sum_{i=n+1}^{N} I(y_i = 1),$$

- Plot ORR(SUBA) - ORR(design): Sc 1 Sc 2 Sc 3 Sc 4 Sc 5 Sc 6

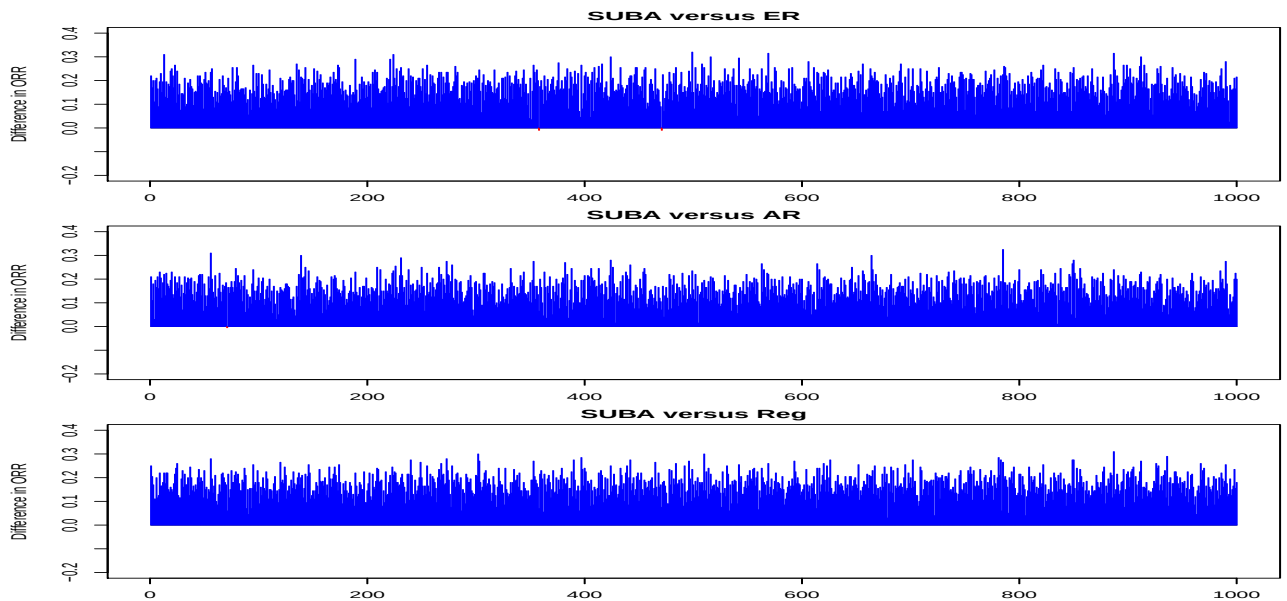


Operating Characteristics – All scenarios

Define the overall response rate (ORR) as the proportion of responders among those patients who are treated after the run-in phase

$$\text{ORR} = \frac{1}{N-n} \sum_{i=n+1}^{N} I(y_i = 1),$$

- Plot ORR(SUBA) - ORR(design): Sc 1 Sc 2 Sc 3 Sc 4 Sc 5 Sc 6

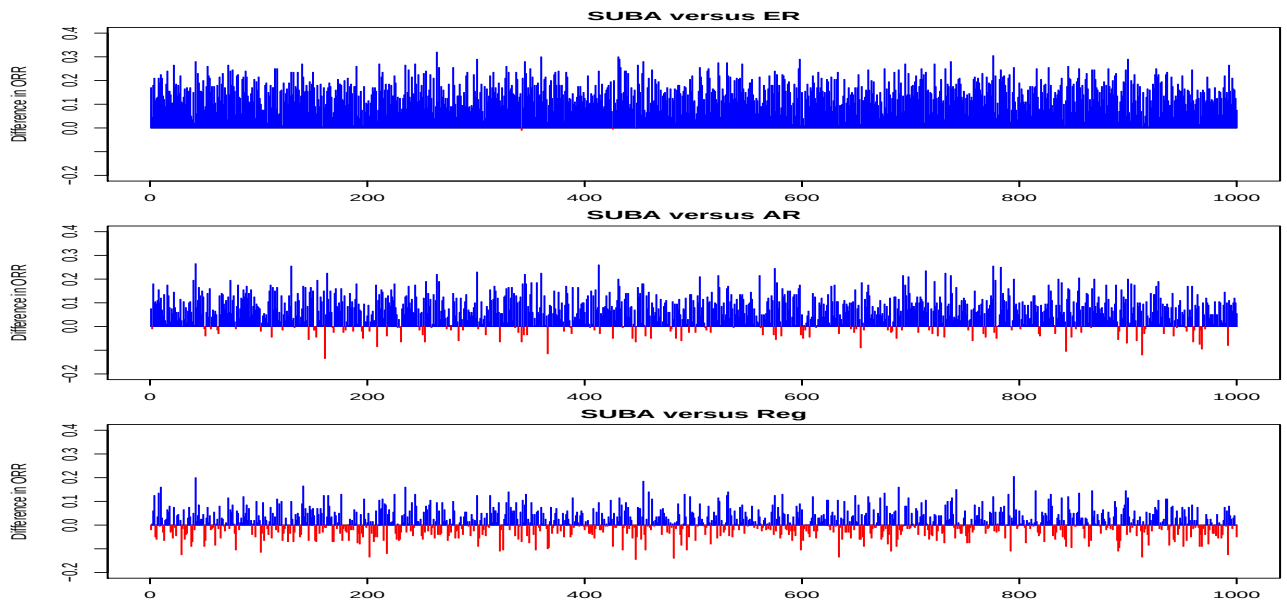


Operating Characteristics – All scenarios

Define the overall response rate (ORR) as the proportion of responders among those patients who are treated after the run-in phase

$$\text{ORR} = \frac{1}{N-n} \sum_{i=n+1}^{N} I(y_i = 1),$$

- Plot ORR(SUBA) - ORR(design): Sc 1 Sc 2 Sc 3 Sc 4 Sc 5 Sc 6

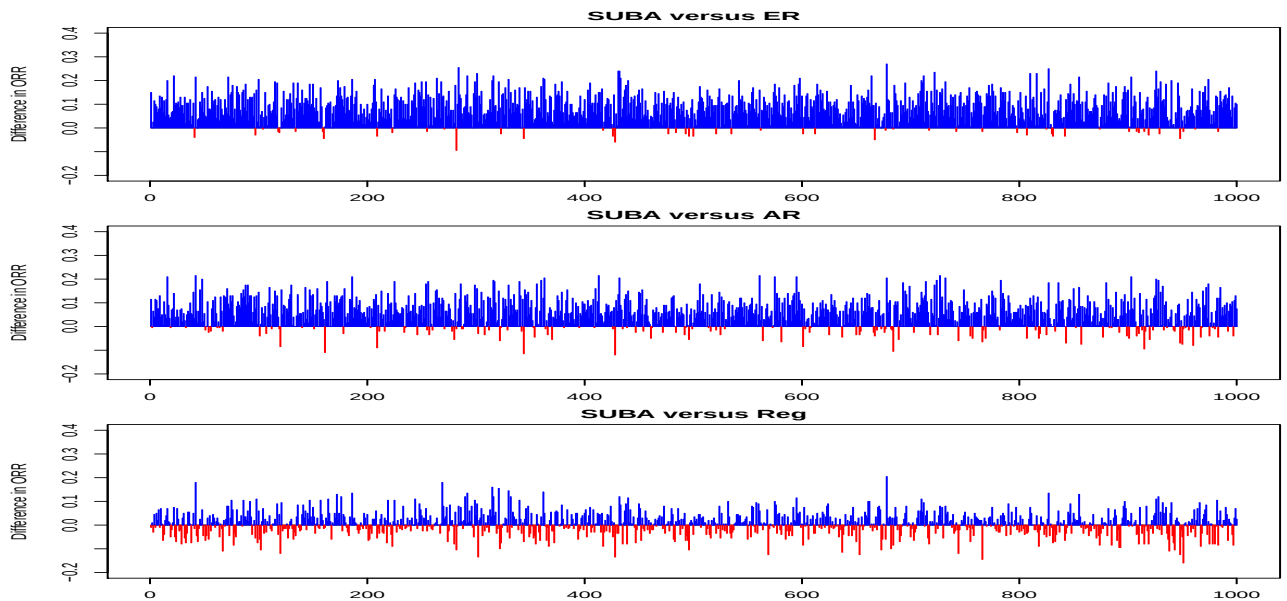


Operating Characteristics – All scenarios

Define the overall response rate (ORR) as the proportion of responders among those patients who are treated after the run-in phase

$$\text{ORR} = \frac{1}{N-n} \sum_{i=n+1}^{N} I(y_i = 1),$$

- Plot ORR(SUBA) - ORR(design): Sc 1 Sc 2 Sc 3 Sc 4 Sc 5 Sc 6

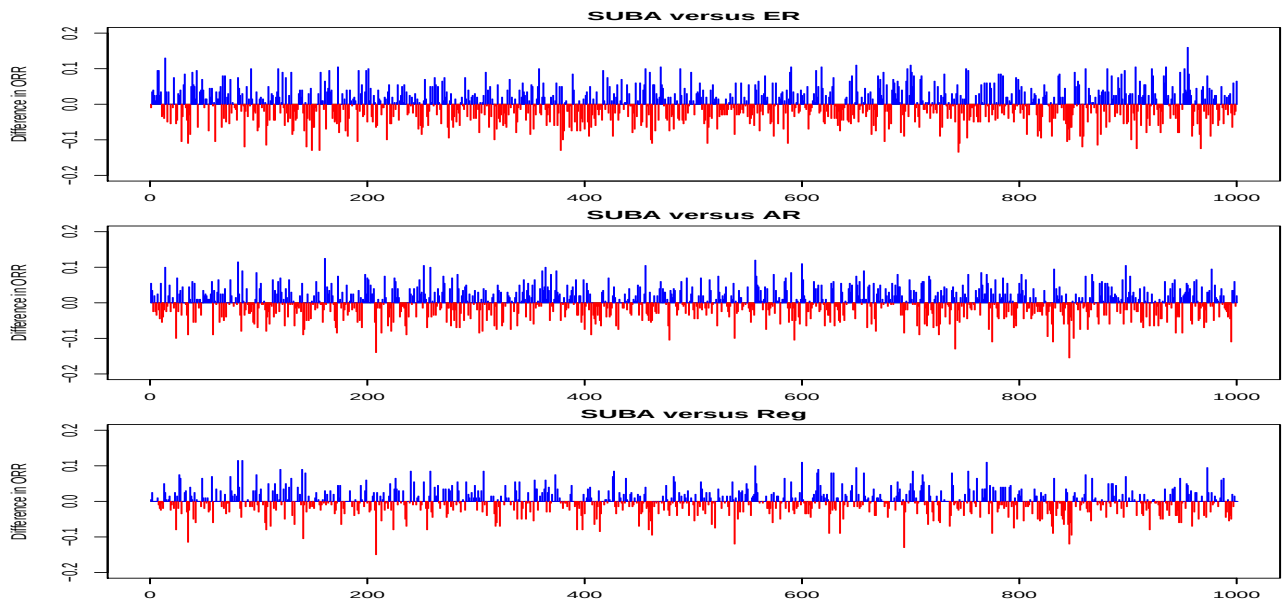


Operating Characteristics – All scenarios

Define the overall response rate (ORR) as the proportion of responders among those patients who are treated after the run-in phase

$$\text{ORR} = \frac{1}{N-n} \sum_{i=n+1}^{N} I(y_i = 1),$$

- Plot ORR(SUBA) - ORR(design): Sc 1 Sc 2 Sc 3 Sc 4 Sc 5 Sc 6



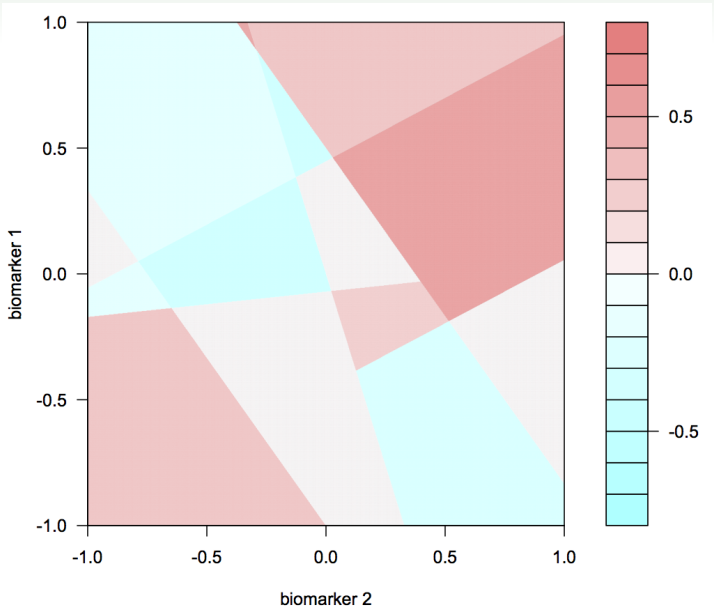
Bayesian nonparametric modeling for Clustering (SCUBA)

Model extension A nonparatric Bayesian model using Dirichlet process priors

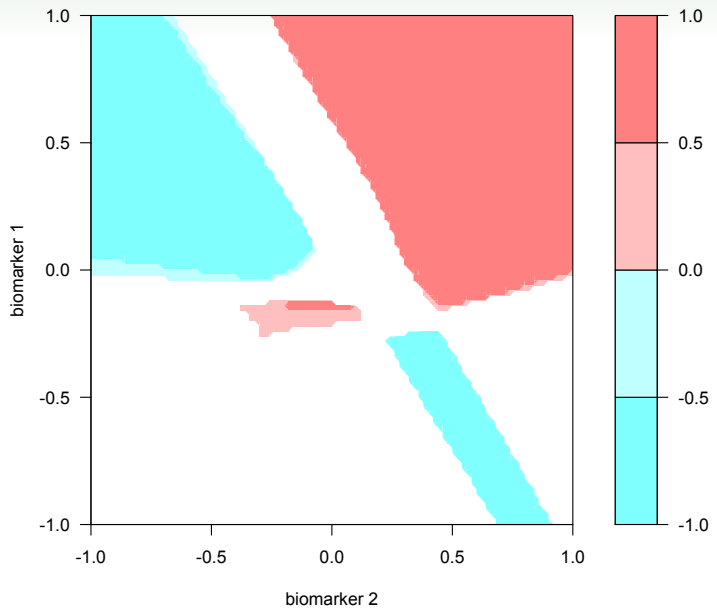
Flexible boundaries Allowing a varying number of boundaries

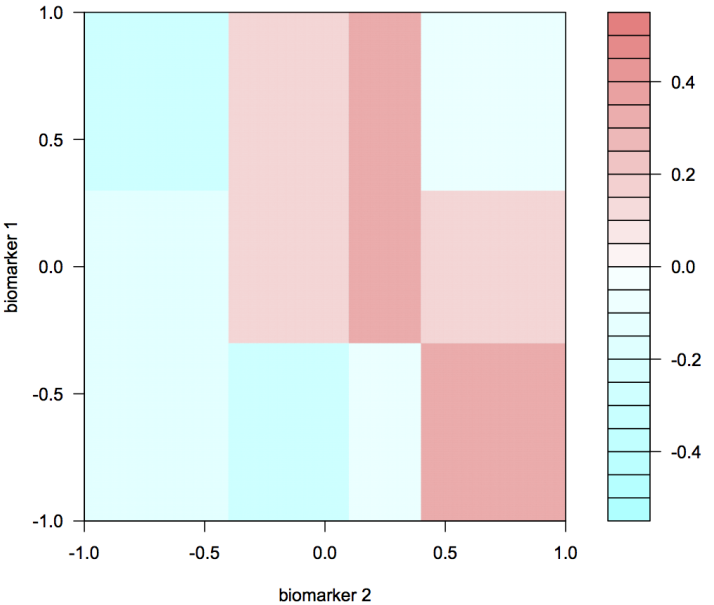
Precision medicine Report subgroup-treatment pairs for confirmatory studies

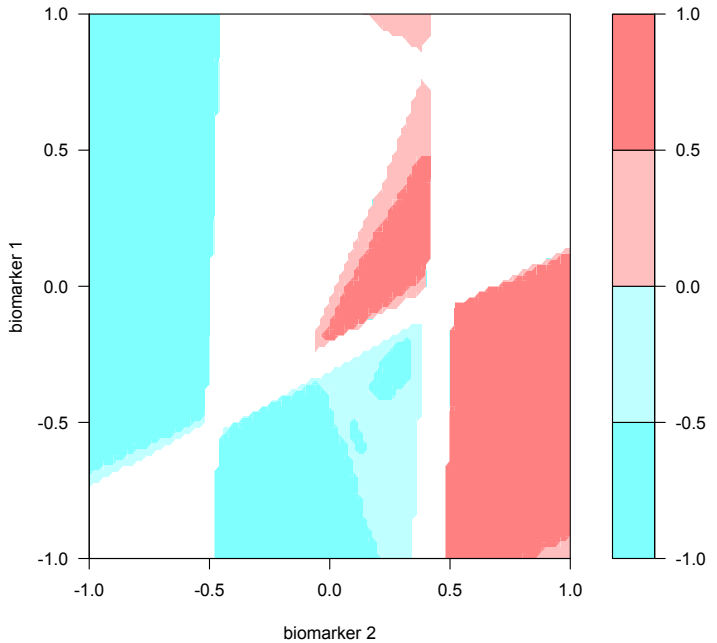
Truth



SCUBA estimate







Conclusions

SCUBA is about precision medicine and targeted therapy.

- Precision medicine: Response to treatment (its order) is assumed to depend on X – biomarkers.
- Adaptive learning based on Bayesian hierarchical models
- Subgroup-treatment pair report with confidence – multiple confirmatory trials for targeted drugs/companion diagnositics