MUCE: A Bayesian Design for Clinical Trials of Multiple Arms/Cohorts with Multiplicity Control

Yuan Ji, PhD

Professor of Biostatistics
Dept. of Public Health Sciences
The University of Chicago

Joint work with Laiya Consulting and Tianjian Zhou at CSU

Sept. 24, 2020

Manuscript: https://arxiv.org/abs/2006.07785

COI

- IDMC for Astellas
- Co-founder of Laiya Consulting; now part of Cytel
- Co-founder of Bayesoft

♠ Introduction

Modern Clinical Trials with Multiple Arms

We consider Bayesian designs and analyses for clinical trials with > 2 arms

Randomized phase II/III trials For example, a three-arm trial with two doses of a new drug and a placebo/control arm

Master protocol phase II/III trials Each arm is a subgroup of patients defined by biomarker status, a different drug, or a mini two-arm subtrial

Multiple expansion cohorts phase Ib trials Each arm is a dose/indication combination

The endpoints can be survival, response rate, or even continuous measurement of monitoring biomarkers.

Master Protocols

THE UNIVERSITY OF CHICAGO

Two popular master protocols: Basket & Umbrella Trials

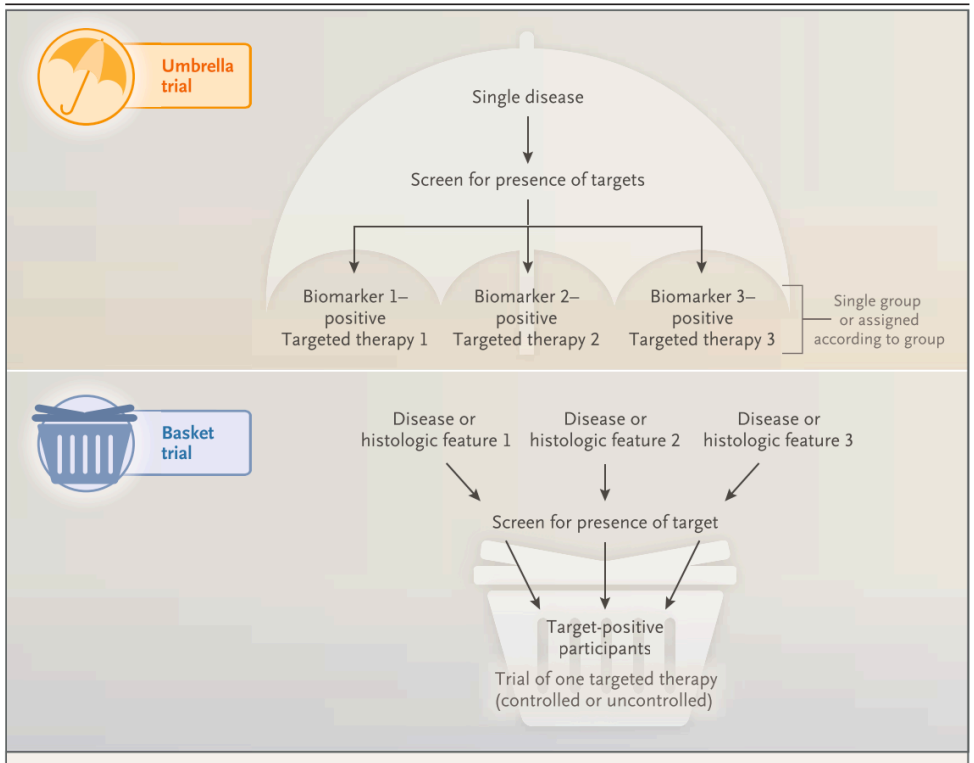


Figure 1. Umbrella Trial and Basket Trial.

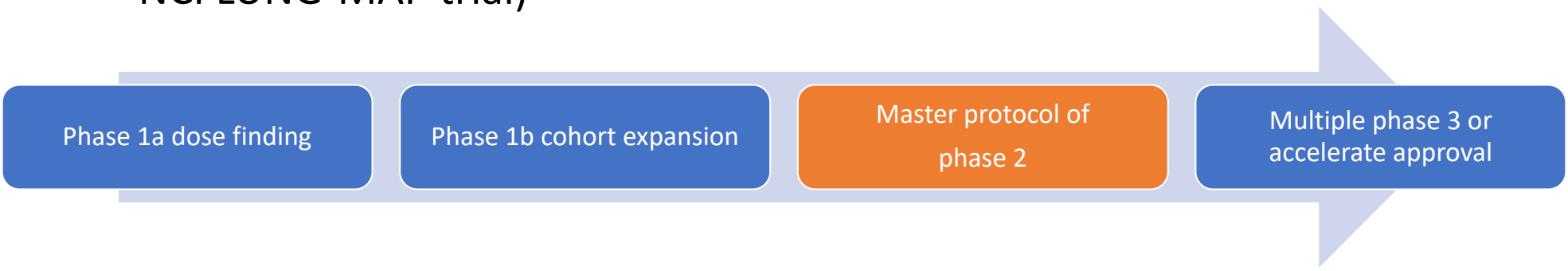
An umbrella trial (top) evaluates various (often biomarker-defined) subgroups within a conventionally defined disease. Patients with the disease are screened for the presence of a biomarker or other characteristic and then assigned to a stratum on the basis of the results. Multiple drugs are studied in the various strata, and the design may be randomized or use external controls depending on the disease. A basket trial (bottom) involves multiple diseases or histologic features (i.e., in cancer). After participants are screened for the presence of a target, target-positive participants are entered into the trial; as a result, the trial may involve many different diseases or histologic features. A master protocol for a basket trial could contain multiple strata that test various biomarker–drug pairs. Statistical approaches are not depicted in this figure.

Woodcock and LaVange, NEJM, 2017

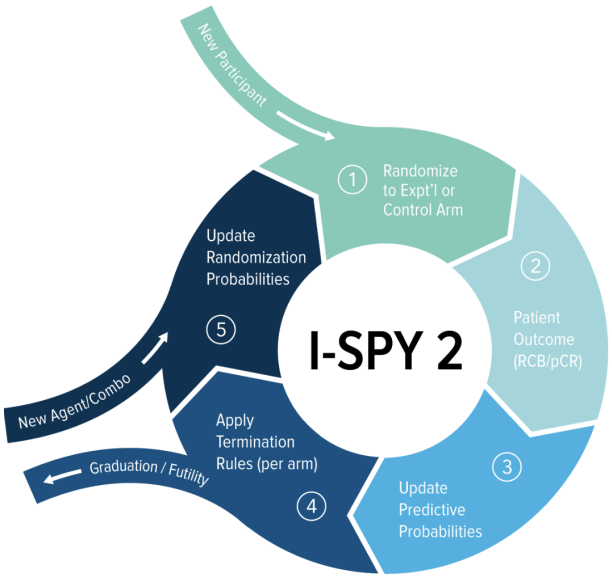
THE UNIVERSITY OF CHICAGO

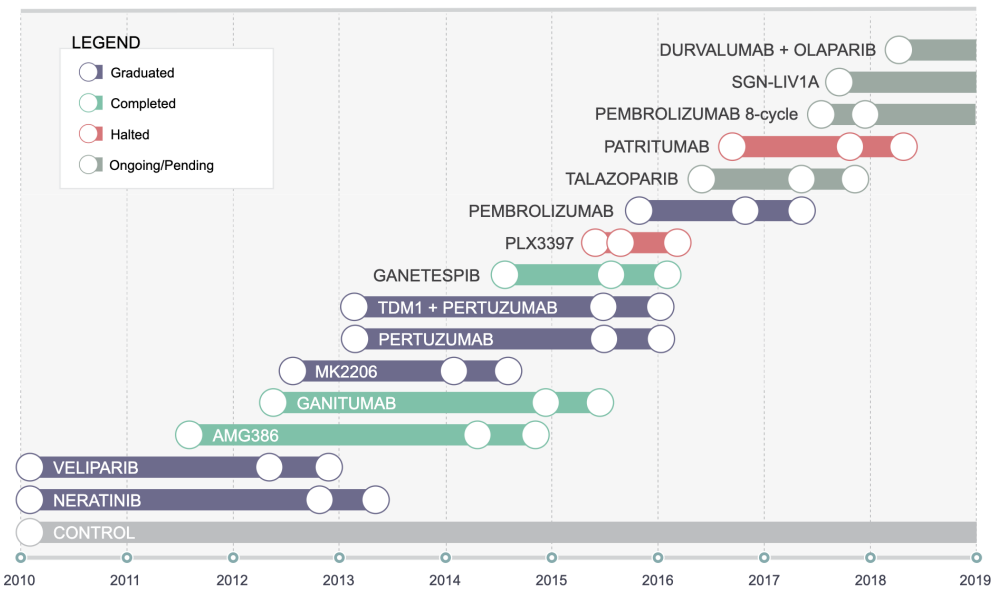
When to use master protocol?

- Each drug has established the RP2D
- Phase 2 master protocol – exploratory or accelerated approval
- One drug (usually targeted or immune) that may work on multiple cancer types with the biomarker (e.g., vemurafenib and BRAF V600 mutation)
- Multiple drugs are available for treating a single cancer type (e.g., NCI LUNG-MAP trial)



I-SPY2



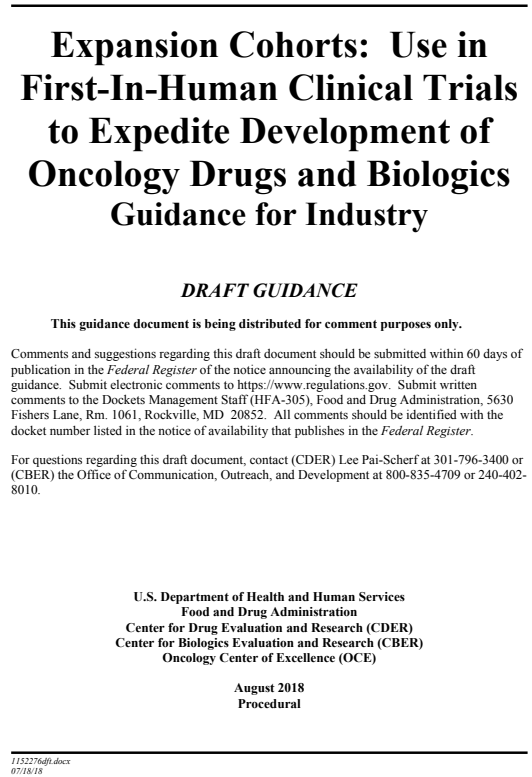


- Adaptive platform trial;
- RCB 0 or pCR endpoint; Neoadjuvent
- A common control
- Adaptive randomization
- No sample size – add new arms, graduate existing arms adaptively
- Bayesian predictive probability
- A few promising drugs graduated; 3 received accelerated approval
- Reference: `ispytrials.org`

Multiple Expansion Cohorts as 2d-basket trials

Multiple expansion cohorts

- A first-in-human (FIH) multiple expansion cohort trial is a FIH trial with an initial dose-escalation phase followed by expansion cohorts on specific doses, indications, schedules, or even drug combinations.
- FDA released a draft guidance on multiple expansion cohorts in FIH trials on August 2018 recommending incorporating multiple expansion cohorts in FIH trials that can "expedite development by seamlessly proceeding from initial determination of a potentially effective dose to individual cohorts that have trial objectives typical of Phase 2 trials.".
- Multiple cohorts expansion might include multiple doses and multiple disease indications, which results in multiple "baskets" ;
- Doses and indications are two factors; Basket trials usually only have one factor – indications



Expansion Cohorts: Use in First-In-Human Clinical Trials to Expedite Development of Oncology Drugs and Biologics
Guidance for Industry

DRAFT GUIDANCE

This guidance document is being distributed for comment purposes only.

Comments and suggestions regarding this draft document should be submitted within 60 days of publication in the *Federal Register* of the notice announcing the availability of the draft guidance. Submit electronic comments to https://www.regulations.gov. Submit written comments to the Dockets Management Staff (HFA-305), Food and Drug Administration, 5630 Fishers Lane, Rm. 1061, Rockville, MD 20852. All comments should be identified with the docket number listed in the notice of availability that publishes in the *Federal Register*.

For questions regarding this draft document, contact (CDER) Lee Pai-Scherf at 301-796-3400 or (CBER) the Office of Communication, Outreach, and Development at 800-835-4709 or 240-402-8010.

U.S. Department of Health and Human Services
Food and Drug Administration
Center for Drug Evaluation and Research (CDER)
Center for Biologics Evaluation and Research (CBER)
Oncology Center of Excellence (OCE)

August 2018
Procedural

115227dft.docx
07/18/18

Some initial feedbacks to the draft guidance

- Biotechnology Innovation Organization (BIO) says FDA "should avoid mentioning specific statistical approaches," such as Simon's two-stage design , as the group believes there are other approaches "that potentially offer greater flexibility while still maintaining rigor."
- ASCO also calls for changes to the draft guidance, including expanding the guidance to cover trials with a single expansion cohort in addition to trials with multiple expansion cohorts and developing more specific requirements for transitioning from the dose escalation phase to the dose expansion phase.
- Also, calls to expand to non-oncology by others.

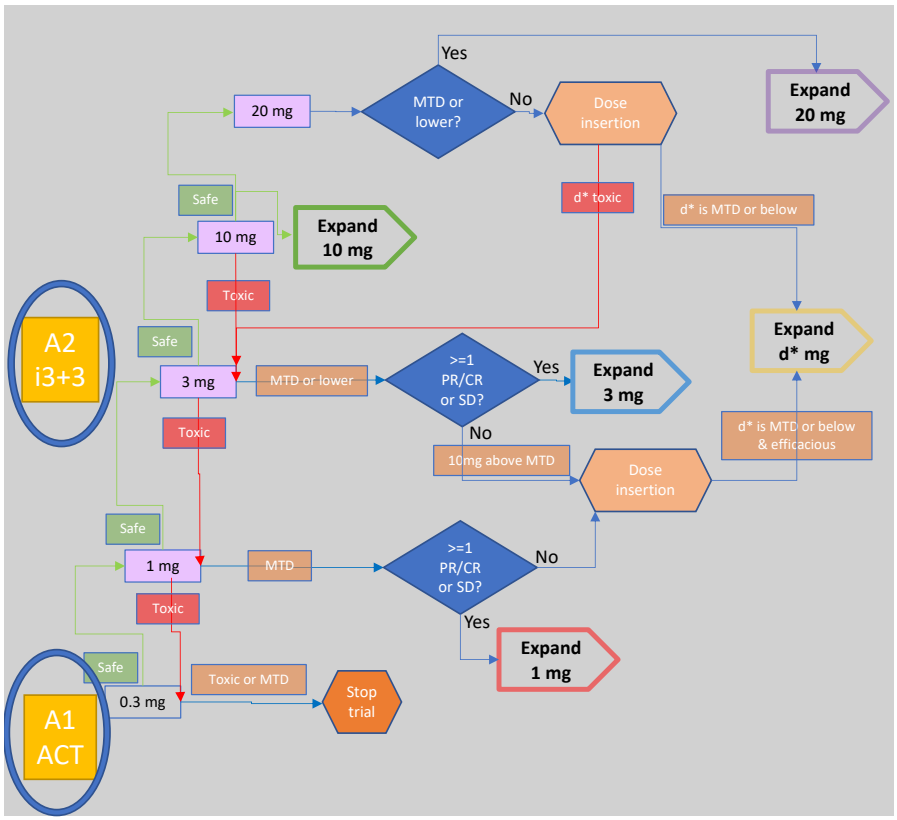
`https://www.raps.org/news-and-articles/news-articles/2018/10/industry-proposes-changes-to-fdas-fih-expansion-c`

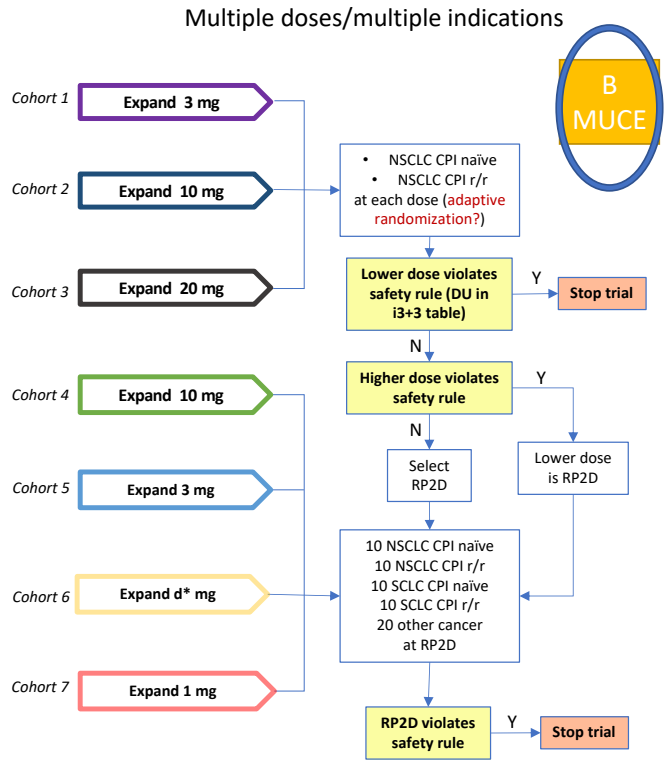
A real use-case for a seamless phase 1a and 1b dose escalation/expansion cohort trial

The i3+3 design (Liu et al., 2020) for dose escalation and MUCE for expansion cohorts

Phase 1a/1b seamless design

Phase 1a: dose escalation → Seamless transition → Phase 1b: cohort expansion (**MUCE**)





Goal

We propose a

Bayesian Hierarchical Model Framework

for Multi-arm trials, e.g., multiple expansion cohorts or master protocols

using hypothesis testing to quantify error rates and report multiplicity

We first review a popular Frequentist Design: Simon's 2-stage design

An example of the Simon's 2-stage design

Suppose we test a new drug at a selected dose on four indications, T_i, with reference object response rate (ORR) for all indications is $p_0 = 0.2$, and the target ORR is $p_i = 0.35$, $i = 1, 2, 3, 4$.

Simon's 2-stage design (r_1, n_1, r, n)

treat n_1 patients first, stop if $\leq r_1$ patients respond; otherwise, treat $(n - n_1)$ patients in the second stage, and reject H_0 if $> r$ patients (across both stages) respond.

$\alpha = 0.1$, $\beta = 0.2$ for each cohort:

$$r_1 = 2, \quad n_1 = 13, \quad r = 12, \quad n = 46$$

An example of the Simon's 2-stage design

Suppose we test a new drug at a selected dose on four indications, T_i, with reference object response rate (ORR) for all indications is $p_0 = 0.2$, and the target ORR is $p_i = 0.35$, $i = 1, 2, 3, 4$.

Simon's 2-stage design (r_1, n_1, r, n)

treat n_1 patients first, stop if $\leq r_1$ patients respond; otherwise, treat $(n - n_1)$ patients in the second stage, and reject H_0 if $> r$ patients (across both stages) respond.

$\alpha = 0.1$, $\beta = 0.2$ for each cohort:

$$r_1 = 2, \quad n_1 = 13, \quad r = 12, \quad n = 46$$

Weak Rejection efficacy boundary $12/46 \approx 0.26$ much smaller than 0.35  ???

No Learning Each arm does not share information with others

No Multiplicity Control! No multiplicity control – family-wise type I error rate can be much higher (see later)

Recent work: Jin and Ying (2020; Statistics in Medicine; Bayesian enhancement two-stage design with error controlfor phase II clinical trials.

Bayesian hierarchical models may help – if done right

- Why multiple expansion cohorts: the drug might be efficacious with different doses and on multiple cancer (sub)types (e.g., Check-point inhibitors, NTRK-inhibitors, combo therapies).
- Information between different indications with similar mechanism can be borrowed to increase statistical efficiency.
- How to quantify how much information should be borrowed?

Let's look at two existing Bayesian methods!

♠ Existing Bayesian Methods

The Bayesian hierarchical model (BHM) approach (Berry et al., 2013)

No existing Bayesian methods for cohort expansion; but some Bayesian approaches for basket trials

- p_i: the true and unknown response probabilities for arm i;
  n_i and y_i: the number of patients and responders at arm i
  Test each arm i by two hypotheses $H_{0i} : p_i \leq p_0$ vs $H_{1i} : p_i \geq p_1$

Berry's BHM

Likelihood $y_i \mid n_i, p_i \sim Bin(n_i; p_i)$
Parameter transformation $\theta_i = log(\frac{p_i}{1-p_i})$
Prior $\theta_i \mid \theta \sim N(\theta, \sigma^2)$
Hyperprior $\theta \sim N(\theta_0, \sigma_0^2), \quad \sigma^2 \sim Inv - Gamma(\alpha_0, \beta_0)$

- The prior construction assumes the response rates p_i's or θ_i's across arms are exchangeable. This allows borrowing information across arms.
- The response rates shrinks to a common value, θ_0, due to borrowing.
- The degree of shrinkage or information borrowing is controlled by the variance parameter σ^2.

Decision Making under BHM (Berry et al., 2013)

- Interim futility stopping based on

$$Pr(p_i > \frac{p_0 + p_1}{2} \mid data) < \phi_1$$

  – note the interesting choice of $\frac{p_0+p_1}{2}$.

- Reject H_{0i} at the end of the trial if

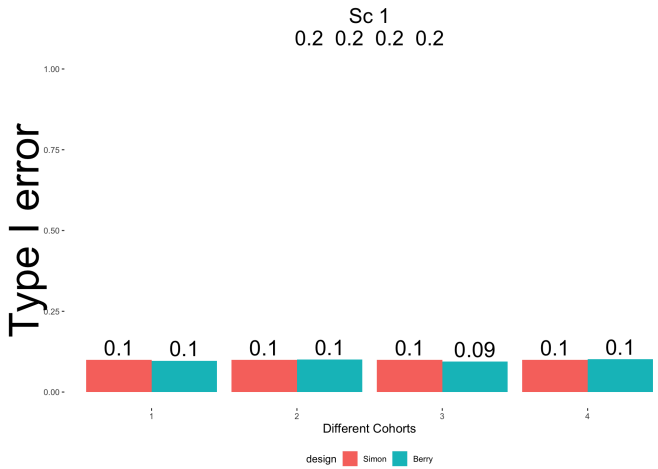
$$Pr(p_i > p_0 \mid data) > \phi_2$$

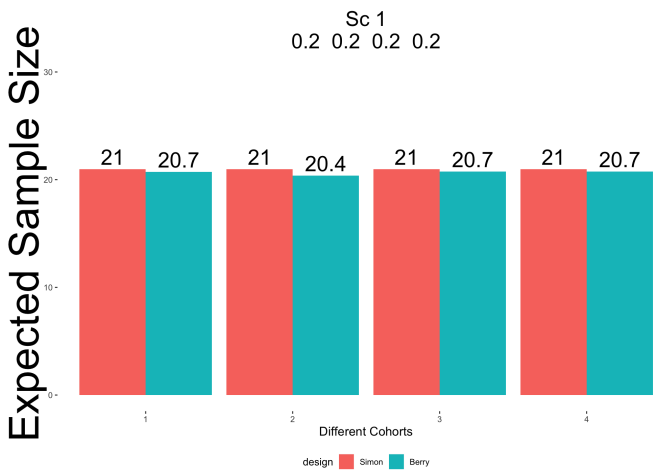
- ϕ_1 and ϕ_2 : tuning parameters determined through simulation studies to generate desirable frequentist operating characteristics (OCs)
  - One issue is "how" and "what OCs are desirable" – a common issue for all "calibrated" Bayesian designs

Berry's BHM may increase power

Scenario 1 (Global Null)

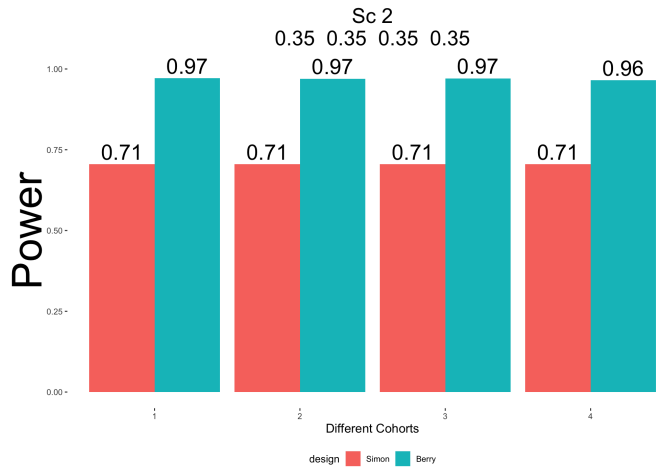
$$p_1 = p_2 = p_3 = p_4 = 0.2$$

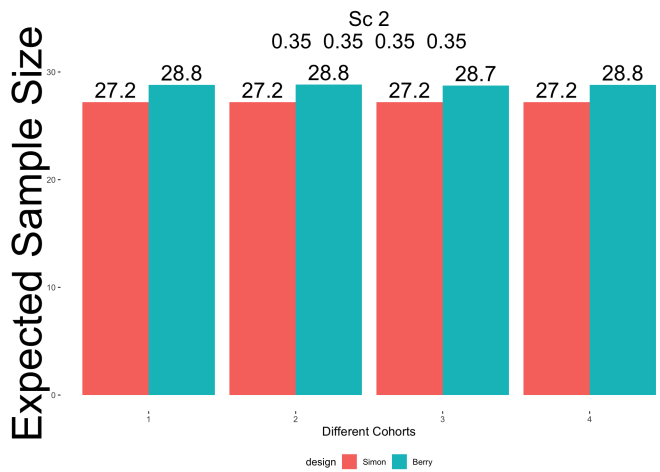




Scenario 2 (Global Alternative)

$$p_1 = p_2 = p_3 = p_4 = 0.35$$



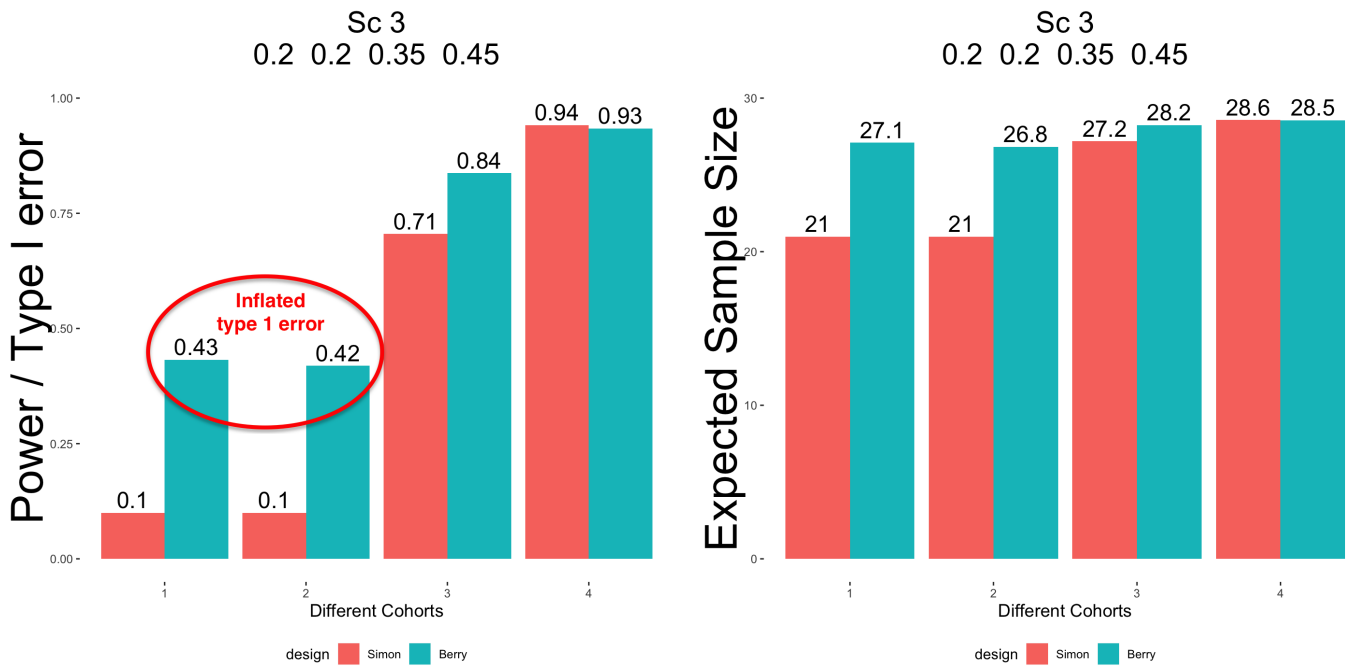


Berry's BHM may also increase Type I error rate

- The model could inflate the type I error rate by wrongly borrowing:
  - Essentially, the model shrinks the response rates of all the arms to a single value , the "elasticity" of which determined by σ^2

Scenario 3

$$p_1 = p_2 = 0.2, p_3 = 0.35, p_4 = 0.45$$



Challenges with BHMs

Existing BHM Approaches

Likelihood $y_i \mid n_i, p_i \sim Bin(n_i; p_i)$
Parameter transformation $\theta_i = log(\frac{p_i}{1-p_i})$

Prior $\theta_i \mid \theta \sim N(\theta, \sigma^2)$
Hyperprior $\theta \sim N(\theta_0, \sigma_0^2)$, $\sigma^2 \sim Inv - Gamma(\alpha_0, \beta_0)$ (or empirical Bayes)

Reject H_{0i} at the end of the trial if

$$Pr(p_i > p_0 \mid data) > \phi_2$$

- The BHM framework is for one "factor": treatment arms;
- In multiple expansion cohort studies, there could be two "factors": doses and indications
- It is desirable to quantify the estimate error rates of the Bayesian decisions – what is the probability of wrong rejections?
- Multiplicity control in the BHM is lacking.

Multiplicity control is CRITICAL!

In early-phase multi-arm trials, sponsors should be more conservative (than FDA) to control multiplicity – thereby control risks and cost! (Drug development has a miserable success rate!)

- For non-registration trials, sponsors MUST control multiplicity – to avoid conducting failed big late-phase trials!

- Most existing Bayesian designs for basket trials do not explicitely model the hypotheses
- Quantify the error rates with Bayesian hypothesis testing framework

The MUCE design/method (for data analysis)

MUCE – Basket trial designs with multiplicity control

$H_{1i}: \theta_{N,i} > \theta_{C,i}$ and $H_{0i}: \theta_{N,i} \leq \theta_{C,i}$.

Bayesian hierarchical model for multiplicity control

Likelihood $Y \mid \theta_{N,i}, \theta_{C,i} \sim f(\cdot; \theta_{N,i}, \theta_{C,i})$,

Prior for θ

$$(\theta_{N,i}, \theta_{C,i}) \mid H_{1i} \sim f_1(\cdot) I(\theta_{N,i} > \theta_{C,i})$$

$$(\theta_{N,i}, \theta_{C,i}) \mid H_{0i} \sim f_0(\cdot) I(\theta_{N,i} \leq \theta_{C,i})$$

Prior for H_{1i} $H_{1i} \mid p \sim Bern(p)$ – the prior probability that H_{1i} is true is p.

Hyperprior for p $p \sim Beta(a, b)$

Decision Rule

Reject H_{0i} if $Pr(H_{1i} \mid data) > v$. Here $(1 - v)$ is the conditional (posterior) probability of H_{0i}. It is the "Bayesian type I error rate" for arm i if the decision is to reject H_{0i}.

The priors for H_{1i} and hyperprior allow p to be random and realizes "multiplicity control" – a smaller value more stringent control.

Application to multiple expansion cohort studies (as a two-dimensional basket trial)

Expansion cohorts: each cohort consists of a dose level and an indication (biomarker subgroups; different cancer types)
Let (i, j) denote the cohort for dose level i, $i = 1, \ldots, I$, and indication j, $j = 1, \ldots, J$,

- p_{ij} : the true and unknown probability of efficacy at cohort (i, j)
- n_{ij} : number of patients treated at cohort (i, j)
- y_{ij} : number of responders at cohort (i, j)

Whether a cohort (i, j) is promising or not can be tested by two hypotheses,

$$H_{0,ij} : p_{ij} \leq p_{0j} \; vs \; H_{1,ij} : p_{ij} > p_{0j}$$

where p_{0j} is the reference response rate for indication j.

MUCE BHM models

Let λ_{ij} be the indicator of the two hypotheses:
$\{\lambda_{ij} = 1\}$: $H_{1,ij}$ is true , or $\{\lambda_{ij} = 0\}$: $H_{0,ij}$ is true

BHM with multiplicity control

likelihood $f(y \mid \theta)$ $y_{ij} \mid n_{ij} \sim Bin(n_{ij}, p_{ij} = logit^{-1}(\theta_{ij}))$

Prior for θ $\theta_{ij} \mid \lambda_{ij} = 1 \sim f_1(\theta_{ij}) I(p_{ij} > p_{0j})$
$\theta_{ij} \mid \lambda_{ij} = 0 \sim f_0(\theta_{ij}) I(p_{ij} \leq p_{0j})$

Latent Probit Score $\lambda_{ij} = I(Z_{ij} > 0)$

Prior $Z_{ij} \mid (\xi_i, \eta_j)$ $Z_{ij} \sim N(\xi_i + \eta_j, 1)$

Priors ξ_i and η_j

$$\left.\begin{array}{l} \xi_i \mid \xi_0 \sim N(\xi_0, 1), \\ \eta_j \mid \eta_0 \sim N(\eta_0, 1). \end{array}\right\} \text{Borrow \& Shrinkage}$$

Hyperprior ξ_0 and η_0

$$\left.\begin{array}{l} \xi_0 \sim N(\mu_\xi, 1), \\ \eta_0 \sim N(\mu_\eta, 1) \end{array}\right\} \text{Multiplicity control}$$

Intuitive Decision Rules

- Use $Pr(\lambda_{ij} = 1 \mid data)$ to make inference, which directly quantifies the posterior probability of each hypothesis.

Optional Stop for futility at interim analysis if $Pr(\lambda_{ij} = 1 \mid data) < v_1$

- Declare arm (i, j) efficacious (i.e, reject $H_{0,ij}$) at the end of the trial if

$$Pr(\lambda_{ij} = 1 \mid data) > v_2$$

- v_2: directly controls the "Bayesian type I error probability, which is $< (1 - v_2)$.
- Denote $\xi_{ij} = Pr(\lambda_{ij} = 1 \mid data)$. Bayesian family-wise error rate is

$$1 - Pr(\cap_{\{(i,j):\xi_{ij} > v_2\}} \{\lambda_{ij} = 1\} \mid data)$$

and Bayesian false discovery rate is

$$\frac{\sum_{(i,j):\xi_{ij} > v_2} (1 - \xi_{ij})}{(\# : \xi_{ij} > v_2)}.$$

♠ Results

Case 1: MUCE and Simon's 2-stage

- Consider a phase 1b trial for an IO agent with four indications of interest
- reference ORR: 0.2, target ORR: 0.35
- Aiming for a phase 1b trial with 100 - 120 patients
- FDA draft guidance: Simon's 2-stage design
- Simon's 2-stage design for a single arm under $\alpha = 0.1$ and $\beta = 0.3$:

$$r_1 = 2, \quad n_1 = 13, \quad r = 8, \quad n = 29$$

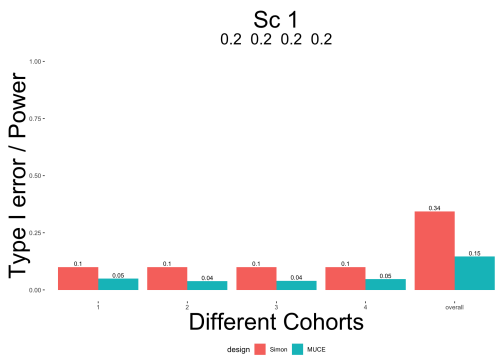
- Simon's 2-stage design will use $29 \times 4 = 116$ patients – might look OK.. But
- The family-wise type 1 error rate for the Simon's 2-stage design in the global null scenario is $1 - (1 - 0.1)^4 = 0.34$

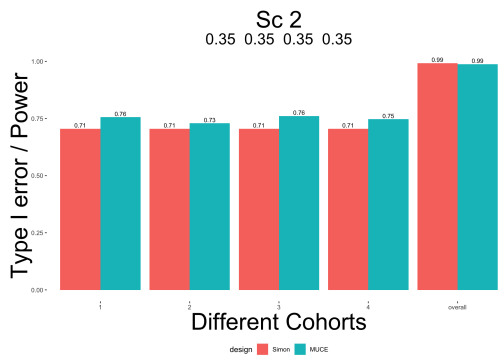
Case 1: MUCE and Simon's 2-stage: Smaller (frequentist) Type I

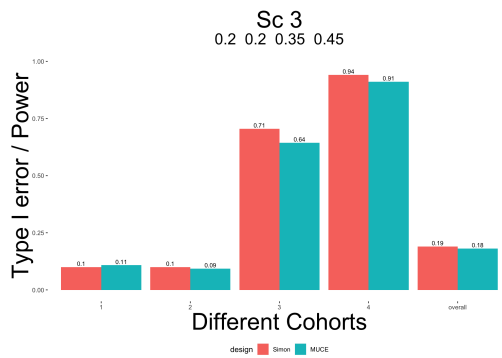
Consider three scenarios that might arise:

- Sc 1 (global null): 0.2, 0.2, 0.2, 0.2
- Sc 2 (global alternative): 0.35, 0.35, 0.35, 0.35
- Sc 3 (mixed 2 null, 2 alternative): 0.2, 0.2, 0.35, 0.35
- MUCE : with maxim sample size =29 and the efficacy threshold at the end of the trial for MUCE is $v_2 = 0.95$. (Bayesian type I for each arm < 0.05)
- MUCE performs better than the Simon's 2-stage design in both the global null (family-wise type I = 0.15) and global alternative; comparable at the mixed scenario (all based on frequentist OCs)

red – Simon; blue – MUCE







Case 1: MUCE and Simon's 2-stage: Smaller sample size if matching types I/II error rates

- If we are aiming for a design with comparable frequentist type 1 error rate and power to the Simon's 2-stage design, MUCE can save the sample size!
- Simon's 2-stage design for a single arm under $\alpha = 0.1$ and $\beta = 0.3$:

$$r_1 = 2, \quad n_1 = 13, \quad r = 8, \quad n_{simon} = 29$$

This gives family-wise type I error rate = 0.34 for the global null.

- MUCE sample size: $n_{MUCE} = 16$ patients per arm
- Both designs have 70% power fro each arm in the global alternative
- MUCE sample size goes up and near Simon's when half of the arms are truly efficacious and half are not (work in progress)

Case 2: A Challenging Seamless Phase 1a/1b Trial

- In 2018 we designed a phase 1a/1b seamless trial with an IO agent, where up to three doses may be graduated from the phase 1a dose-escalation trial to the phase 1b expansion cohorts trial based on safety and activity outcome
- Four indications (disease types) are of interest — resulting in up to 12 cohorts
- Phase 1b sample size:100 - 120 patients
- Reference ORR: 0.2 vs. Target ORR: 0.5
- Simon's 2-stage design for a single arm under $\alpha = 0.05$ and $\beta = 0.3$:

$$r_1 = 1, \quad n_1 = 5, \quad r = 5, \quad n = 14$$

- This results in a sample size $14 \times 12 = 168$, and a family-wise type 1 error rate

$$1 - (1 - 0.05)^{12} = 0.46!!!$$

– more exclamation marks.  Simon's design won't work here with the limited resources.

Case 2: A Challenging Seamless Phase 1a/1b Trial

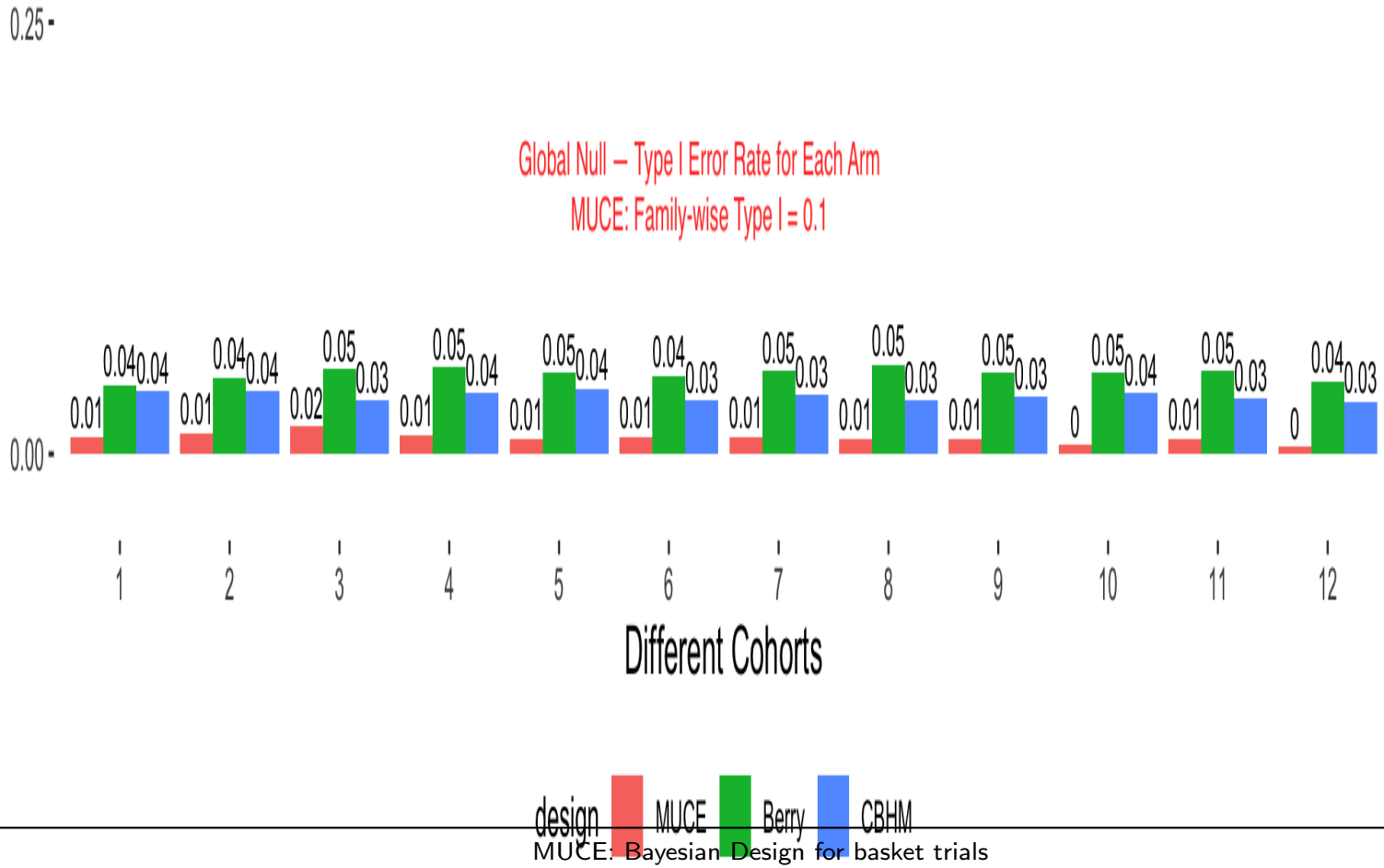
- The MUCE design for phase 1b with 10 patients per cohort. Conduct simulation to examine the performance of 5 scenarios
- Consider 5 scenarios:

Table 1: Scenarios 3.1 to 3.5. Entries are efficacy rates for three arms in four indications. The reference rate is 0.2 for all arms.

Sc	Dose level	Dosage	Ind1	Ind2	Ind3	Ind4
3.1 (NULL)	1	3mg	0.2	0.2	0.2	0.2
	2	10mg	0.2	0.2	0.2	0.2
	3	20mg	0.2	0.2	0.2	0.2
3.2	1	3mg	0.5	0.5	0.5	0.5
	2	10mg	0.5	0.5	0.5	0.5
	3	20mg	0.5	0.5	0.5	0.5
3.3	1	3mg	0.3	0.3	0.3	0.3
	2	10mg	0.4	0.4	0.4	0.4
	3	20mg	0.5	0.5	0.5	0.5
3.4	1	3mg	0.5	0.5	0.2	0.2
	2	10mg	0.5	0.5	0.2	0.2
	3	20mg	0.5	0.5	0.2	0.2
3.5	1	3mg	0.3	0.3	0.2	0.2
	2	10mg	0.4	0.4	0.2	0.2
	3	20mg	0.5	0.5	0.2	0.2

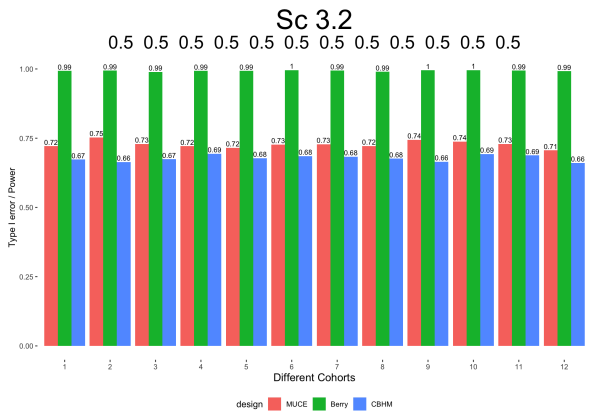
Case 2: Bayesian Designs – Frequentist Type I Error

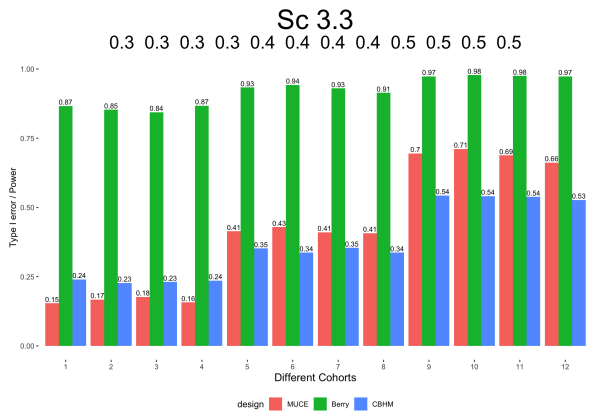
- Compare the performance of MUCE and the other two Bayesian basket trial design: Berry's method and CBHM (Chu and Yuan, 2018).
- The family-wise type 1 error rate under the global null scenario 3.1 is controlled at 0.1 for all three methods.

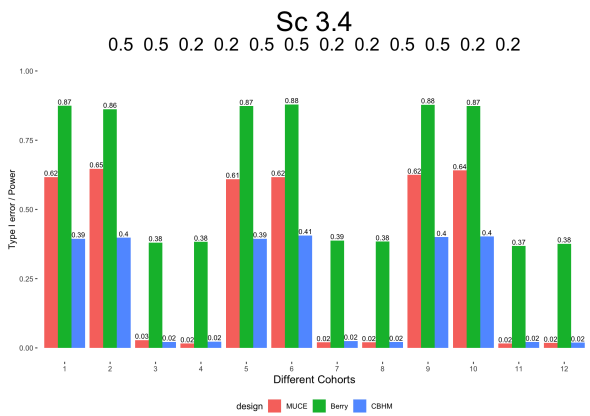


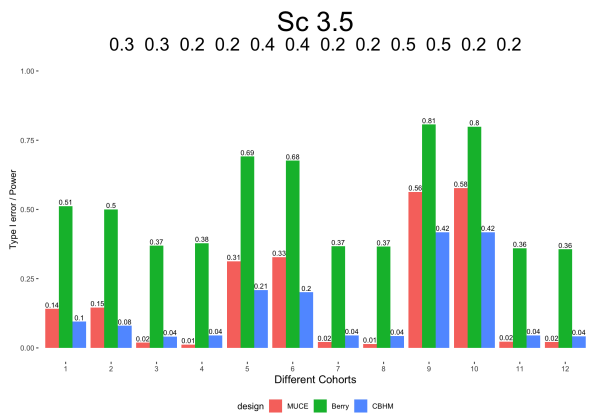
Case 2: Bayesian Designs – Power

- Berry's method has the highest power and the highest type I error rates for the null arms. The type I error rates might be too high in some cases
- CBHM has lower power than MUCE;
- MUCE appears to be the best method with smaller Type I and higher power – thanks to the two-way modeling









Case 3: Sample size reduction

An ongoing oncology trial in Gastric cancer of three expansion cohorts, single dose, three different H_0 and H_1's with different desired α and power.

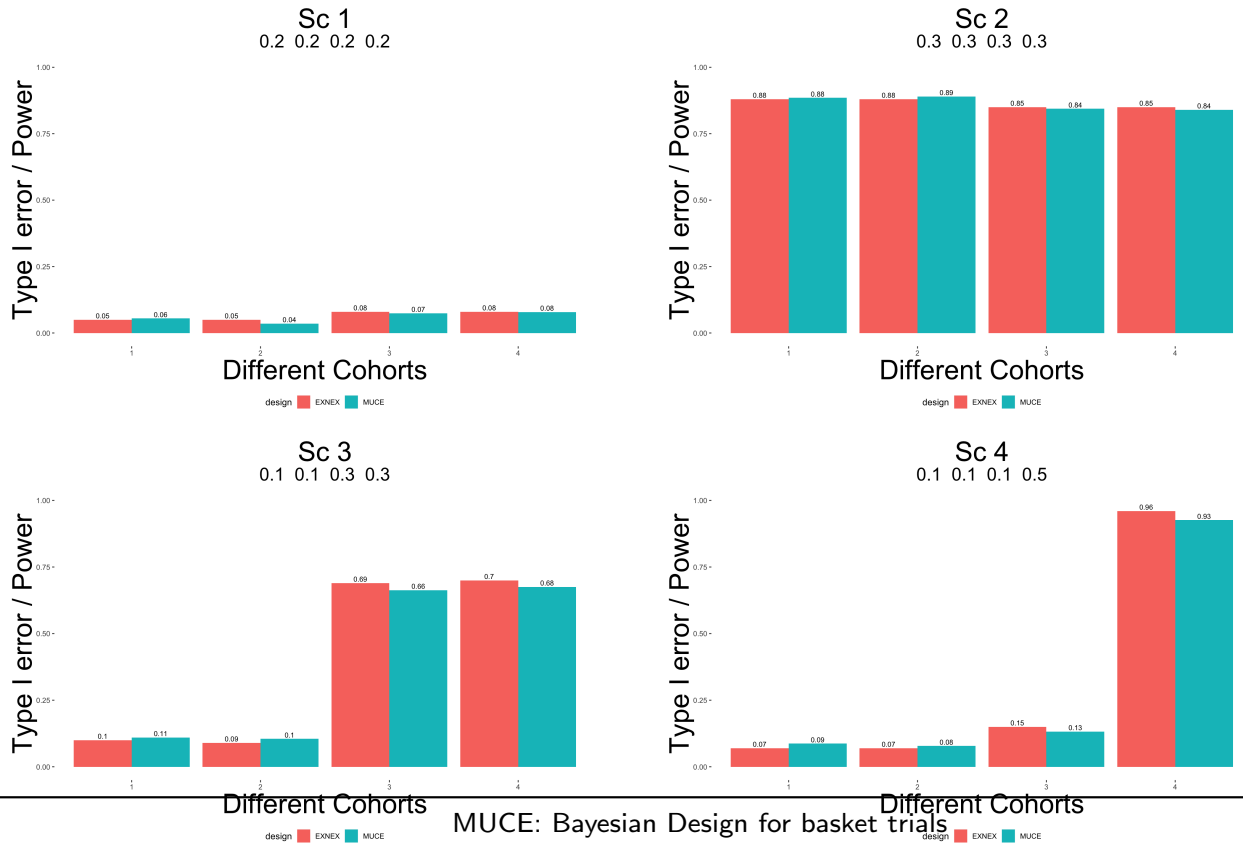
Compared to Simon's 2-stage design, MUCE cuts the sample size by half with similar type I error rate and power requirement.

Subgroups		Arm 1	Arm 2	Arm 3	Total sample size
Assumptions	Endpoint	pCR	ORR	ORR	
	Historical vs Expected	0.05 vs 0.2	0.4 vs 0.5	0.15 vs 0.3	
	Alpha	0.05	0.20	0.05	
	Power	0.80	0.80	0.80	
Simon's 2-stage design		N=29 N1*=10	N=81 N1=40	N=55 N1=19	**165**
MUCE design		N=20 N1=10	N=30 N1=15	N=30 N1=15	**80**

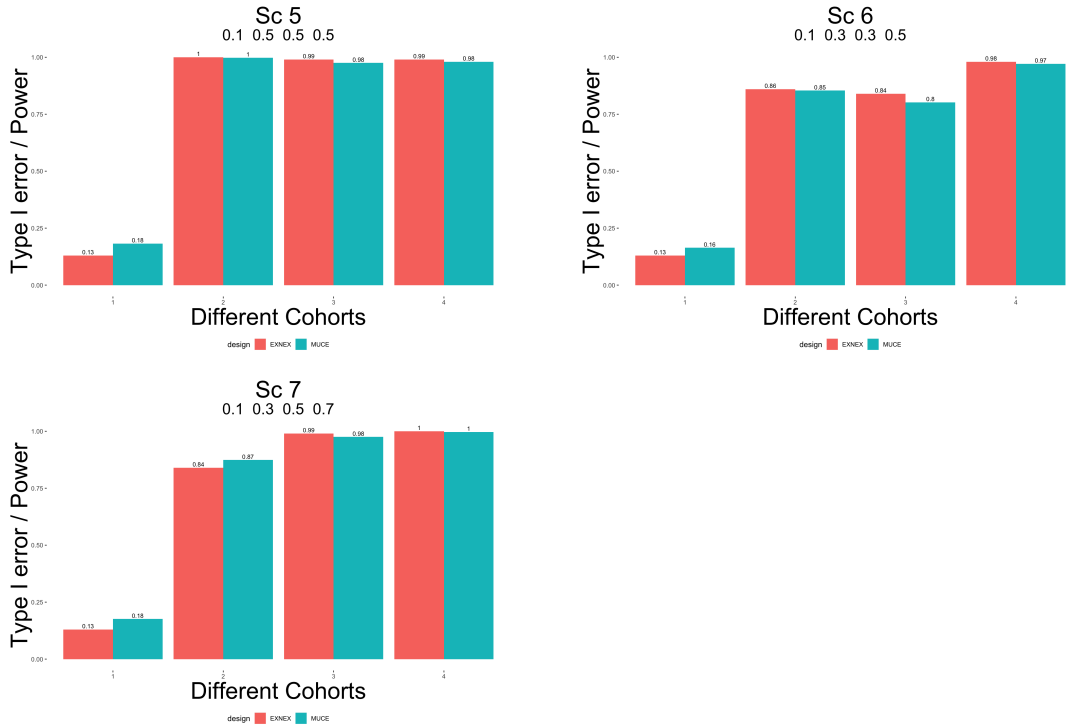
Comparison with EXNEX (1)

Scenarios from EXNEX in Neuenschwander et al. (2016): *Robust exchangeability designs for early phase clinical trials with multiple strata*

- Four indication arms: sample sizes = 20 for indications 1 and 2, and 10 for indications 3 and 4
- reference rate = 0.2.
- the type I error rate in each arm for the global null scenario (Sc 1) is matched



Part 3: Comparison with EXNEX (2)



MUCE is comparable to EXNEX based on scenarios adapted from EXNEX paper

More simulations have been conducted

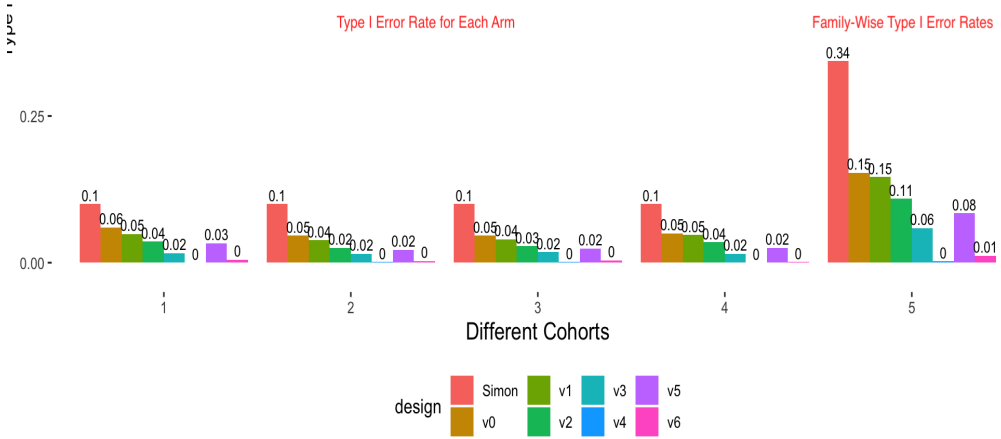
EXNEX requires prespecification of the number of subpopulations and the "proportions" of all them – ideally, should be estimated.

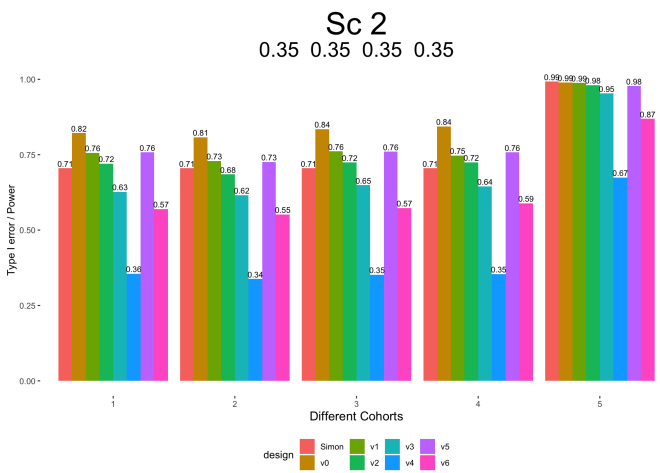
MUCE Multiplicity Control – How it is done?

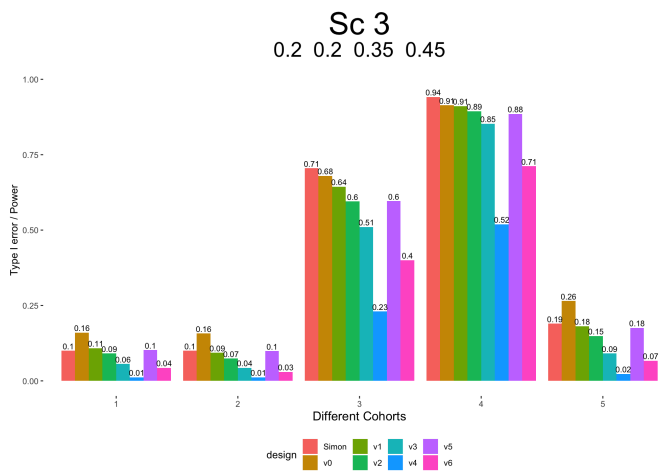
Consider the following hyper-parameters in $\xi_0 \sim N(\mu_\xi, \sigma_\xi = 1), \eta_0 \sim N(\mu_\eta \equiv 0, \sigma_\eta = 1)$. Consider 7 versions of the hyper-parameters. Conclude treatment efficacious if $Pr(\lambda_{ij} = 1 \mid data) > v_2 = 0.95$. Note no calibration of v_2 here.

v0: $\mu_\xi = 0; \sigma_\xi = \sigma_\eta = 2.5$
v1: $\mu_\xi = 0; \sigma_\xi = \sigma_\eta = 1$
v2: $\mu_\xi = -3; \sigma_\xi = \sigma_\eta = 1$
v3: $\mu_\xi = -6; \sigma_\xi = \sigma_\eta = 1$

prior 4: $\mu_\xi = -10; \sigma_\xi = \sigma_\eta = 1$
prior 5: $\mu_\xi = -3; \sigma_\xi = \sigma_\eta = 2.5$
prior 6: $\mu_\xi = -10; \sigma_\xi = \sigma_\eta = 2.5$







MUCE: Changing the mean (μ_ξ, μ_η) gives different level of multiplicity control

Recall the full model of MUCE. Different arms can have different endpoints!

BHM with multiplicity control

likelihood $f(y \mid \theta)$ $y_{ij} \mid n_{ij} \sim Bin(n_{ij}, p_{ij} = logit^{-1}(\theta_{ij}))$

Prior for θ $\theta_{ij} \mid \lambda_{ij} = 1 \sim f_1(\theta_{ij}) I(p_{ij} > p_{j0})$

$\theta_{ij} \mid \lambda_{ij} = 0 \sim f_0(\theta_{ij}) I(p_{ij} \leq p_{j0})$

Latent Probit Score $\lambda_{ij} = I(Z_{ij} > 0)$

Prior $Z_{ij} \mid (\xi_i, \eta_j)$ $Z_{ij} \sim N(\xi_i + \eta_j, 1)$

Priors ξ_i and η_j

$$\left.\begin{array}{l} \xi_i \mid \xi_0 \sim N(\xi_0, 1), \\ \eta_j \mid \eta_0 \sim N(\eta_0, 1). \end{array}\right\} \text{Borrow \& Shrinkage}$$

Hyperprior ξ_0 and η_0

$$\left.\begin{array}{l} \xi_0 \sim N(\mu_\xi, 1), \\ \eta_0 \sim N(\mu_\eta, 1) \end{array}\right\} \text{Multiplicity control}$$

Making μ_ξ and μ_η negative induces multiplicity control!

Summary and Remarks

Superior performance MUCE is an advanced Bayesian approach superior to the Simon's 2-stage design for expansion cohorts trials and master protocols: smaller sample size or higher power in frequentist OCs; better control of Type I error rates in global null

Multiplicity control Compared to existing Bayesian methods, MUCE can formally adjust the estimated error rates for the decisions based on posterior inference.

2d-basket MUCE is capable of dealing with flexible borrowing from multiple doses and multiple indications.

References

- Berry, S. M., Broglio, K. R., Groshen, S., & Berry, D. A. (2013). Bayesian hierarchical modeling of patient subpopulations: efficient designs of phase II oncology clinical trials. Clinical Trials, 10(5), 720-734.
- Chu, Y., & Yuan, Y. (2018). A Bayesian basket trial design using a calibrated Bayesian hierarchical model. Clinical Trials, 15(2), 149-158.
- FDA draft guidance (2018). Expansion Cohorts: Use in First-In-Human Clinical Trials to Expedite Development of Oncology Drugs and Biologics Guidance for Industry.
- Lyu, J., Zhou, T., Yuan, S., Guo, W., Ji, Y. (2020) MUCE: Bayesian Hierarchical Modeling for the Design and Analysis of Phase 1b Multiple Expansion Cohort Trials. `arXiv:2006.07785`