Bayesian Statistics with Applications in Clinical Trials in the 21st Century

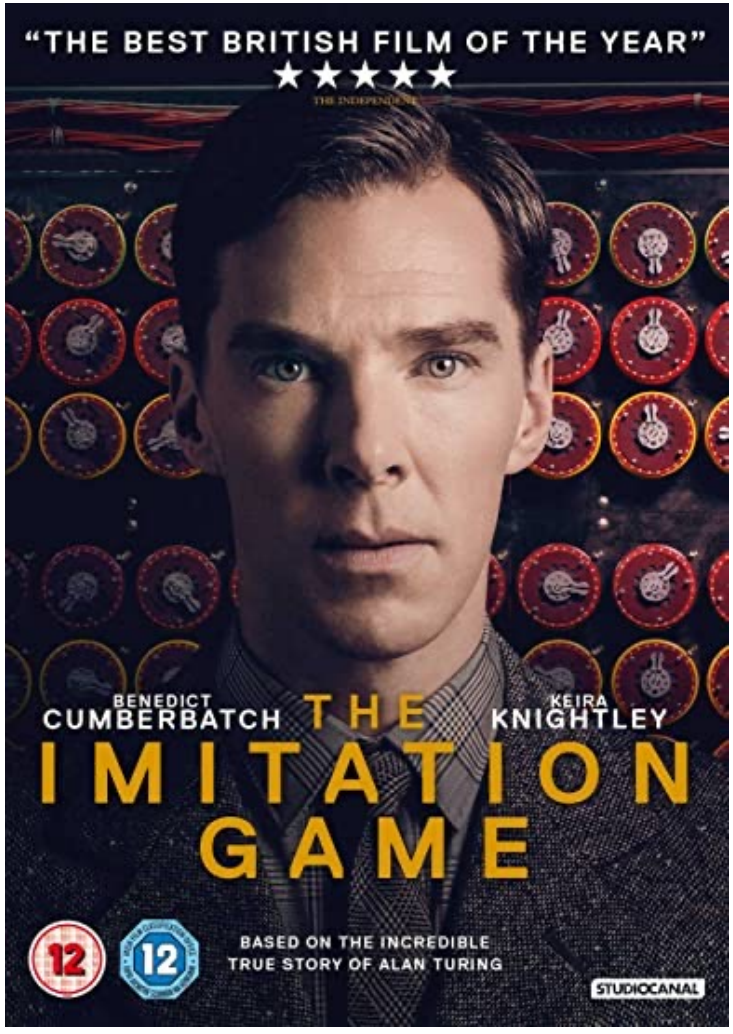
Yuan Ji, PhD
Professor of Biostatistics
The University of Chicago

Conflict of Interests

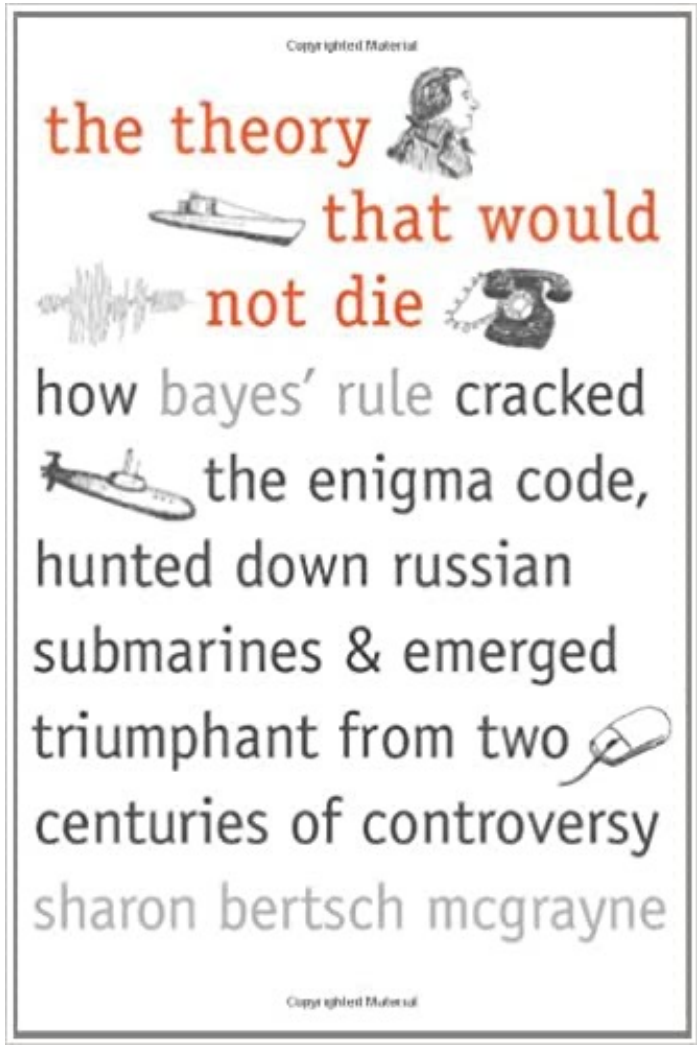
- Yuan Ji is a co-founder of Bayesoft
  - Also, an IDMC member for Astellas and BI
  - Executive Advisor for Cytel

Outline

1. We are all Bayesians
2. Bayesian Applications are Everywhere
3. Bayesian Decision Making – Late Phase Clinical Trials
4. Modern Bayesian Trials
5. Bayesian designs for early-phase trials
   - Dose-Finding in the Era of Dose Optimization
   - MUCE: A 2-dimensional Bayesian Basket Trial Design
6. Cytel's Software: East Bayes



https://www.google.com/url?sa=i&url=https%3A%2F%2Fwww.amazon.com%2FThe-Imitation-Game-DVD%2Fdp%2FB00PC1FD9U&psig=AOvVaw2JFd0ehnaZBxQ0Il3jhbMt&ust=1645290463649000&source=images&cd=vfe&ved=0CAsQjRxqFwoTCLi7wOHeifYCFQAAAAAdAAAAABAD



Bayesian Essentials: Borrow & Learn

✓ What should be the number in the middle?

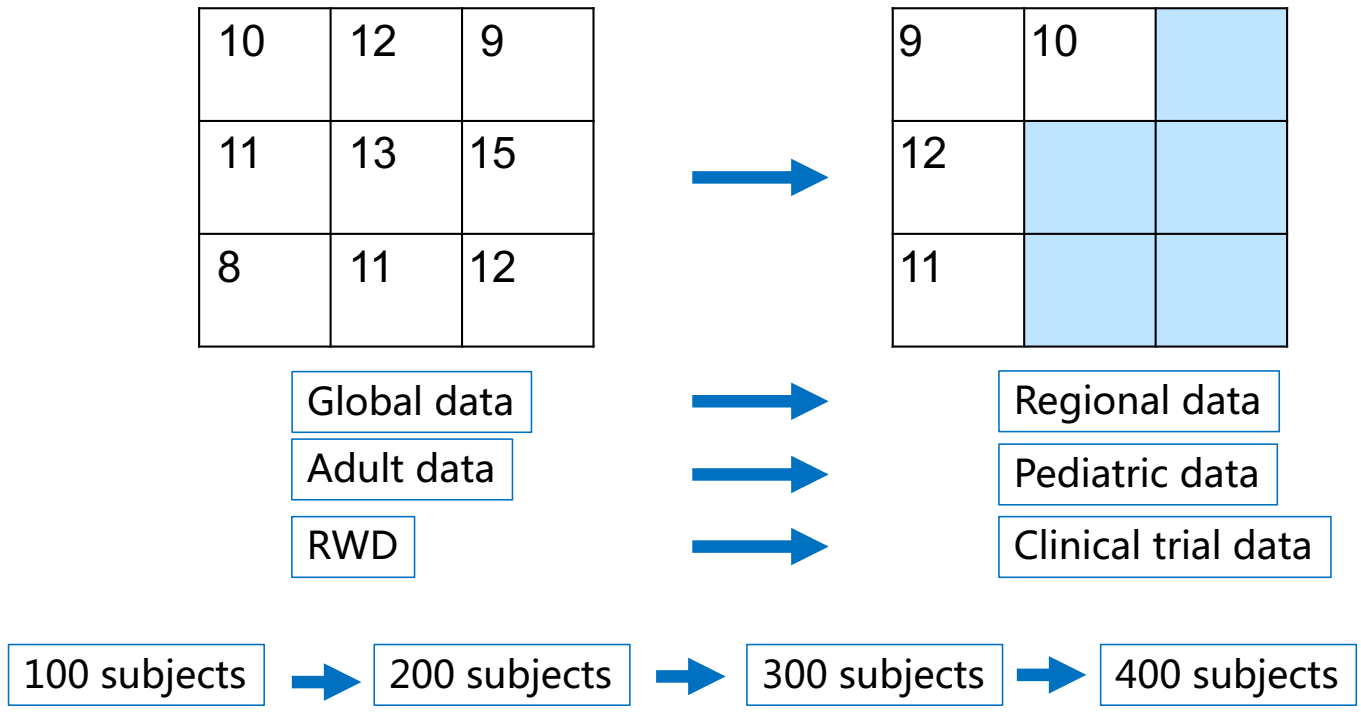
A. 10
B. 20
C. -100
D. 0.1

10	12	9
11	?	10
10	11	12

-200	0.2	12
-150	?	10
-120	0.05	9

Note: no prior information or context here

Bayesian Essentials: Borrow & Learn



Bayesian inference and applications

Inference: refers to the statistical procedure in which estimation of unknown parameters is conducted using information from the observed data.

Bayesian inference is applied in a wide variety of fields

- Medicine: psycology; disease modeling; clinical trials
- Biology: genomics; brain function; imaging analyses
- Law: Courtroom analysis of evidence
- Finance: prediction of financial market
- Politics: presidential election prediction (Nate Silver and now `https://fivethirtyeight.com`)

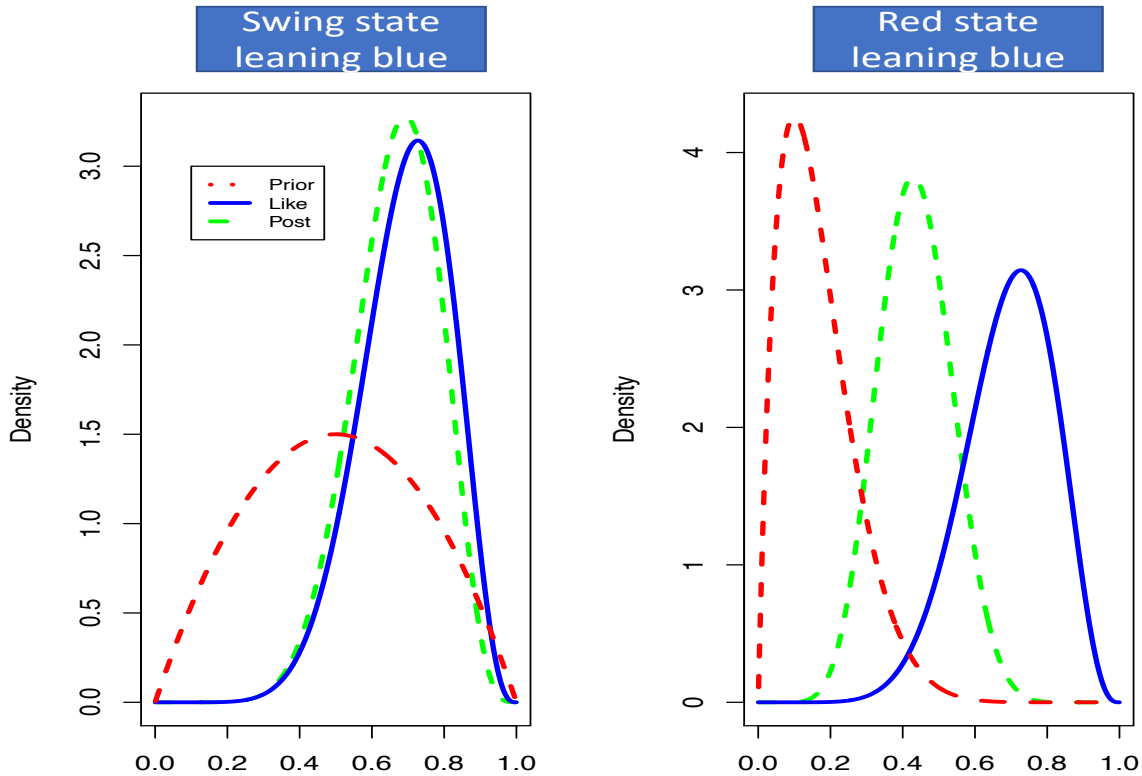
Nate Silver's 2012 Presidential Election Prediction

On Nov. 7, 2012, Nate Silver predicted the presidential election results on all US 50 states with 100% accuracy!

For the Nate-haters, here's the 538 prediction and actual results side by side pic.twitter.com/jbny4pRX

Michael Cosentino 15 hours ago

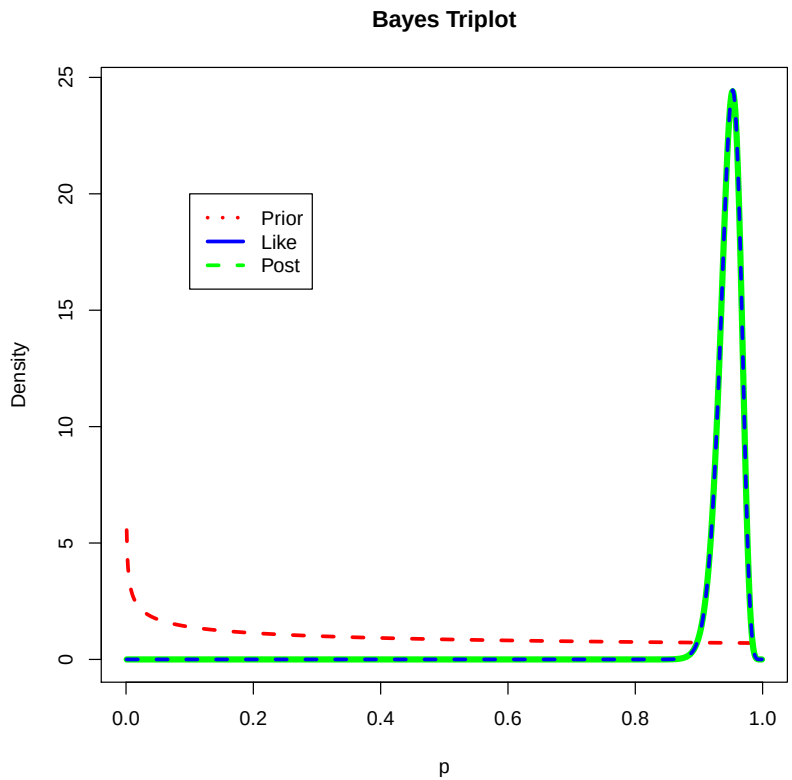
Example 1: Posterior probability of Election



Bayesian posterior probabilities (green) is a compromise of prior (red) and observed data (likelihood)

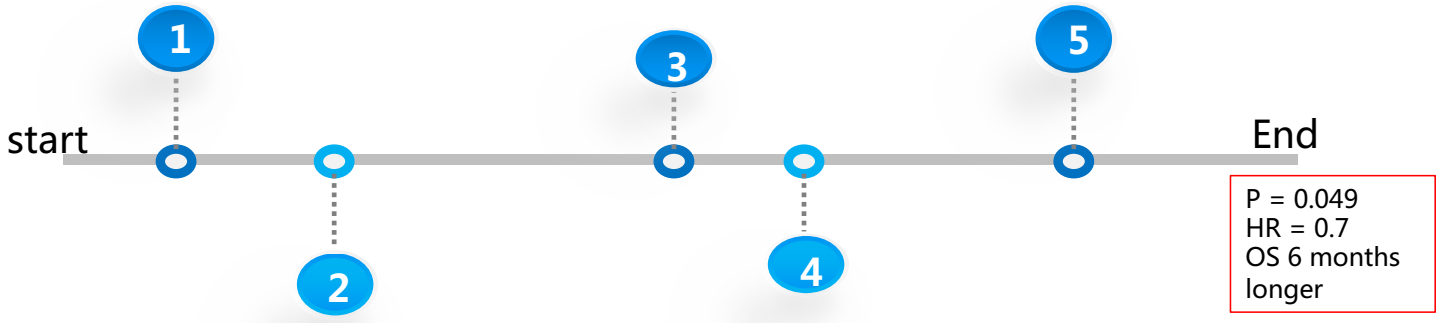
Example 2: Pfizer COVID-19 Vaccine

The Pfizer/BioNTech vaccine is based on a Bayesian trial. The posterior shows a highly efficacious vaccine based on the proposed Bayesian model and the observed data.

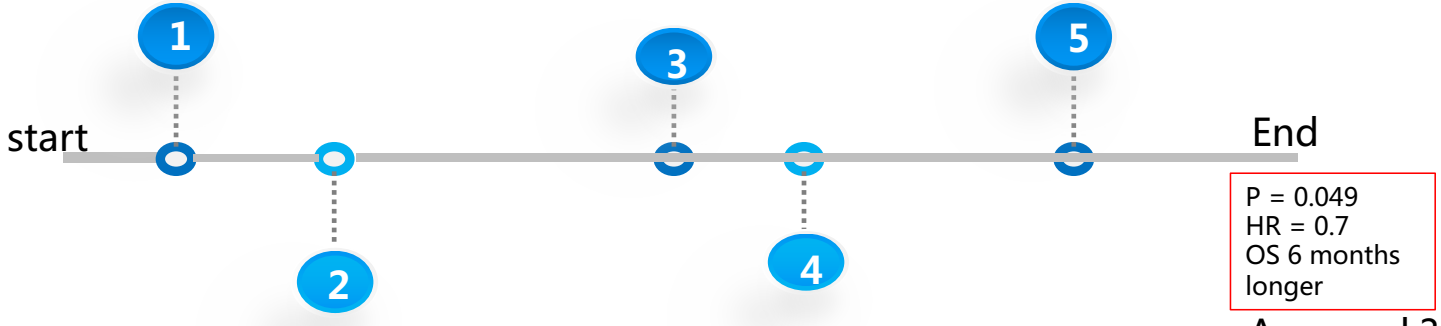


Placebo: 162 COVID-19 cases; Vaccine: 8 COVID-19 cases

Questions for Sequential Designs and Multiple Comparisons



Retrospective sequential analysis
No alpha spending



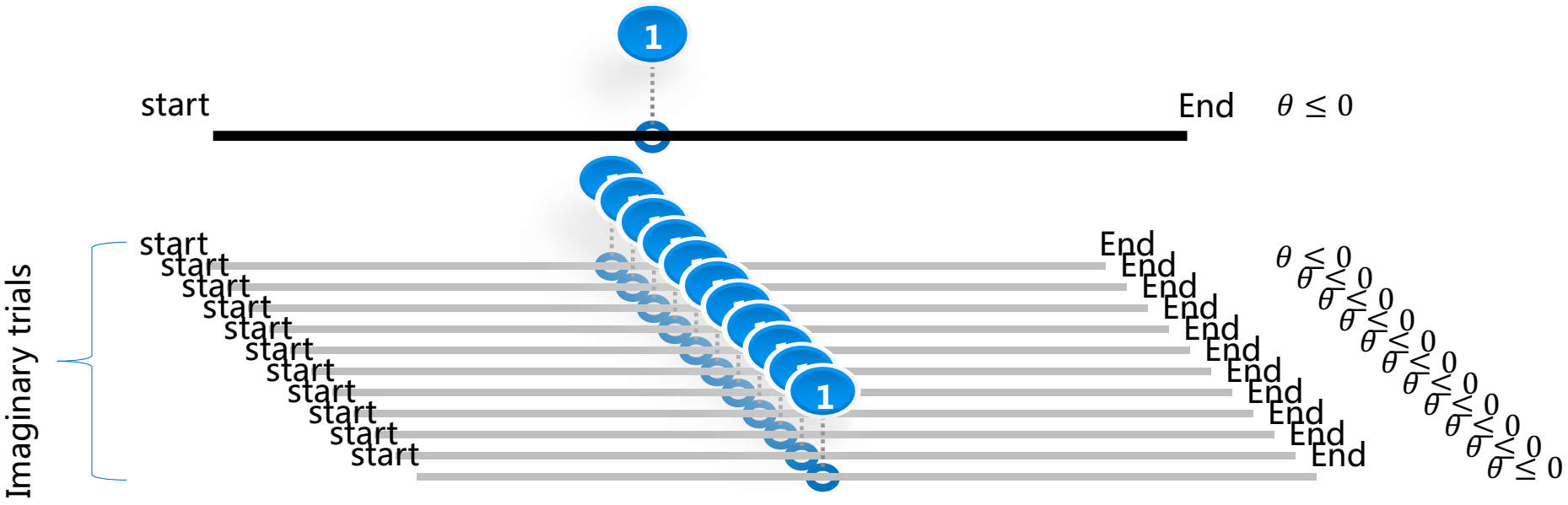
Perspective sequential analysis
Alpha spending

Sequential decision making during interim analysis for clinical trials

- Interim analyses (IA) allow modification of trial conduct (including terminating trials) based on accumulating data from a clinical trial
- Example: Pfizer COVID-19 vaccine trial (Polack et al., 2020)
  - It's a Bayesian trial
- Frequentist sequential designs are concerned about the overall Type I error rate
- Bayesians, by definition, use posterior inference; and Type I error rate is not a statistics of concern

Type I error rate – what is it?

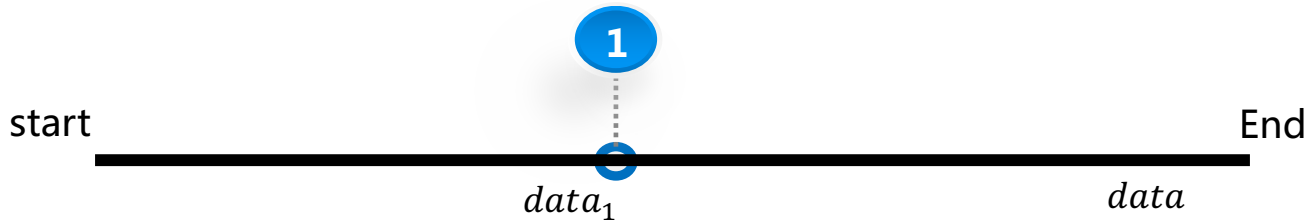
- Hypothesis testing of treatment effect θ $H_0: \theta \leq 0 \; v.s. \; H_1: \theta > 0$



The frequency that H_0 is rejected across all the "imaginary" trials = Type I error rate

Bayesians' approach for assessing error rate

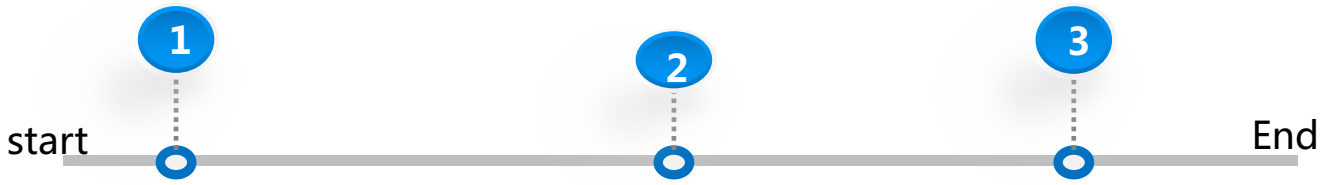
- Hypothesis testing of treatment effect θ $H_0: \theta \leq 0 \; v.s. \; H_1: \theta > 0$



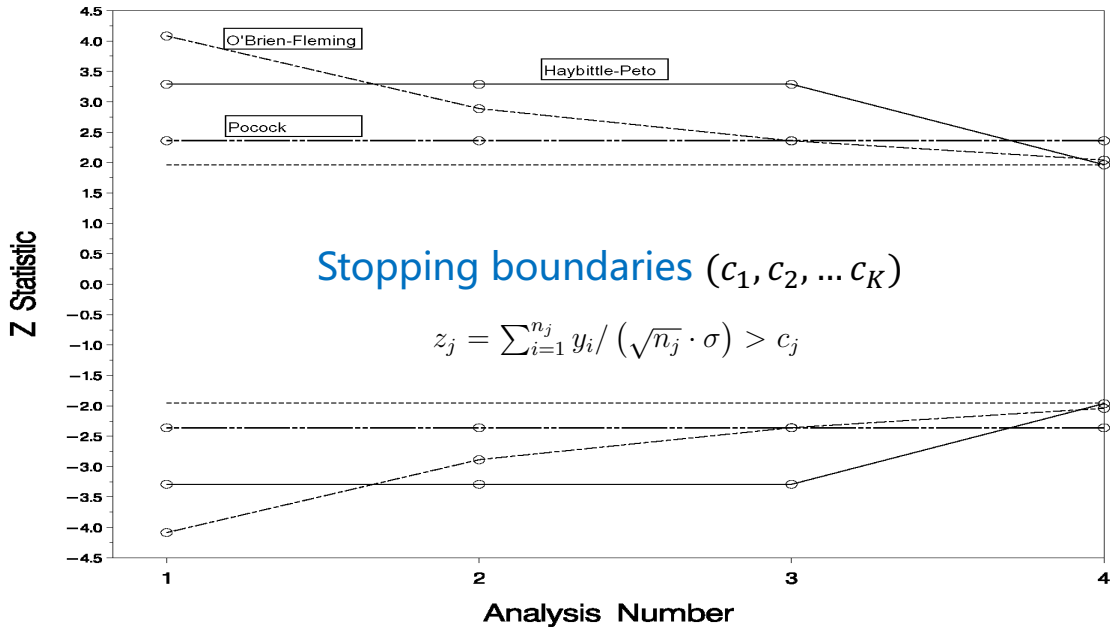
1. Compute $\xi_1 = \Pr(H_1|data_1)$ at IA;
2. If $\xi_1 > 1 - \alpha$, stop the trial due to efficacy and reject H_0;
3. The probability of making the wrong rejection is $(1 - \xi_1) < \alpha$

A smaller α makes it less likely to make a wrong rejection for the trial.

Frequentist Sequential Trial Designs: α-Spending Boundaries



IA and final: $z_1 > c_1$ $z_2 > c_2$ $z_3 > c_3$ $z_4 > c_4$

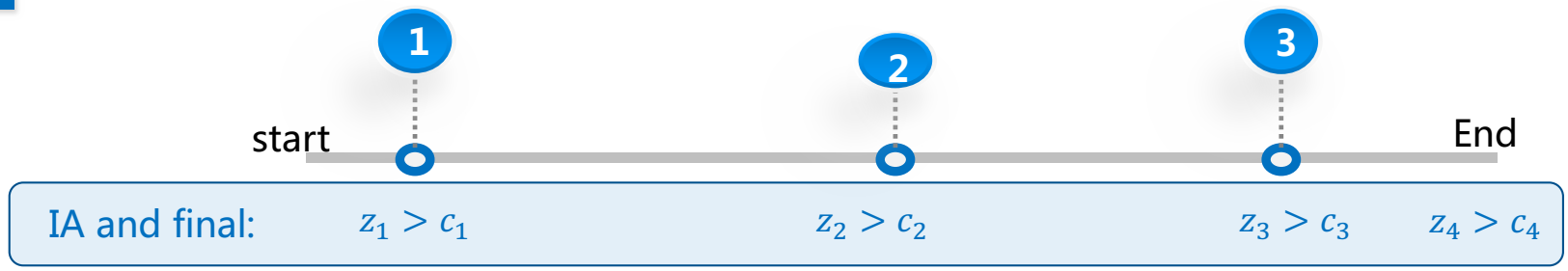


Stopping boundaries $(c_1, c_2, \ldots c_K)$

$$z_j = \sum_{i=1}^{n_j} y_i / \left(\sqrt{n_j} \cdot \sigma\right) > c_j$$

$$H_0: \theta \leq 0 \; v.s. \; H_1: \theta > 0$$

Frequentist Sequential Trial Designs: α -Spending Functions



α-spending functions so that $h(0) = 0$ and $h(1) = \alpha$ (Lan and DeMets, 1983)

$$h_1(u) = \alpha \log\left(1 + (e-1)u\right),$$

$$h_2(u) = 2 - 2\Phi\left(q_{\alpha/2}/\sqrt{u}\right),$$

$$h_3(u) = \alpha u^b \quad \text{for} \quad b > 0.$$

Conditional power (Lan et al., 1982):

$$\mathrm{CP}_j(\theta) = \Pr(z_K > q_\eta \mid \theta, \boldsymbol{y}_j)$$

Bayesian Sequential Designs -- Overview

- A full Bayesian is not concerned about type I error rates;

- Therefore, a "pure" Bayesian would simply stop the trial if
  - $\xi_1 = \Pr(H_1|data_1)$ is too high (stopping for efficacy), or
  - $\xi_0 = \Pr(H_0|data_1)$ is too high (stopping for futility)
  - at any IA or final analysis

- Question: Does the Bayesian procedure inflate the type I error rate?
  - Answer: It does;
  - Because the Bayesian procedure would have a higher probability of making the wrong rejection in the same "imaginary" trials if multiple IAs are performed

- The real question is: what does controlling type I error rate do to the drug approval system and human health care? – Research needed

Bayesian Sequential Designs: A Review

- Bayesian designs assume a probability model and use model-based inference for parameter estimation.

- Some Bayesian designs use posterior probability (pp) for decision making
  - Mixed Frequentist-Bayesian approaches
  - Subjective Bayesian approaches
  - Calibrated Bayesian approaches

- Some Bayesian designs use posterior predictive probability (ppp) for decision making

- Other Bayesian designs use decision theoretical (DT) approaches

A simple model to illustrate the Bayesian designs

- Suppose that a total of K analyses, including $(K-1)$ interim analyses and a final analysis, are planned.
- At the jth analysis, data $y_i \sim N(\theta, \sigma^2), i = 1, \ldots, n_j, iid$
- θ – treatment effect; σ^2 – known (for simplicity)
- Planned max sample size n_K; sample size at jth analysis, n_j

Likelihood: $y_i \sim N(\theta, \sigma^2), i = 1, \ldots, n_j$

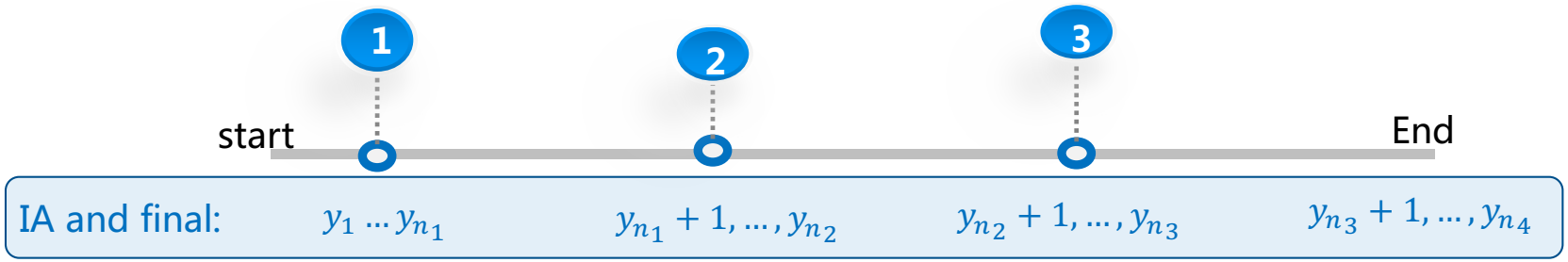
Prior: $\theta \sim N(\mu, \nu^2)$

Posterior:

$$\theta \mid \boldsymbol{y}_j \sim \mathrm{N}\left(\frac{\mu\nu^{-2} + \bar{y}_j n_j \sigma^{-2}}{\nu^{-2} + n_j\sigma^{-2}}, \frac{1}{\nu^{-2} + n_j\sigma^{-2}}\right)$$

Hypothesis:

$$H_0: \theta \leq 0 \; vs \; H_1: \theta > 0$$



K=4

Bayesian designs based on pp: **Mixed Frequentist-Bayesian Designs**

If $PP_j = \Pr(\theta > 0 \mid y) > \gamma_j$, for some threshold γ_j, H_0 is rejected, the trial stopped, and the efficacy of the drug is declared. This is equivalent to

$$Z_j > q_{1-\gamma_j}\sqrt{1+\frac{\nu^{-2}}{n_j\sigma^{-2}}} - \frac{\mu\nu^{-2}}{\sqrt{n_j\sigma^{-2}}}$$

where $q_{1-\gamma_j}$ is the $(1-\gamma_j)$th – upper quantile of $N(0,1)$.

Likelihood $f(y|\theta)$: $y_i \sim N(\theta, \sigma^2)$

Prior $\pi(\theta)$: $\theta \sim N(\mu, \nu^2)$

Posterior $p(\theta|y)$:

$$\theta \mid \boldsymbol{y}_j \sim \mathrm{N}\left(\frac{\mu\nu^{-2}+\bar{y}_j n_j\sigma^{-2}}{\nu^{-2}+n_j\sigma^{-2}}, \frac{1}{\nu^{-2}+n_j\sigma^{-2}}\right)$$

Hypothesis:

$$H_0: \theta \le 0 \; vs \; H_1: \theta > 0$$

Mixed Frequentist Bayesian Designs : Adjust $\pi(\theta)$ and γ_j to achieve overall type I error rate

Examples 1: if $\boldsymbol{\pi(\theta) \propto 1}$, then Bayesian = Frequentist, and $\boldsymbol{\gamma_j}$ may be set to any frequentist boundaries, e.g., O'Brian & Fleming

Example 2: if $\boldsymbol{\pi(\theta) = N(0, 0.054^2)}$and $\boldsymbol{\gamma_j \equiv 0.95}$, then (6) controls type I error rate at 0.05.

Bayesian designs based on pp: **Subjective (Full) Bayesian Designs**

If $PP_j = \Pr(\theta > 0 \mid y) > \gamma_j$, for some threshold γ_j, H_0 is rejected, the trial stopped, and the efficacy of the drug is declared. This is equivalent to

$$Z_j > q_{1-\gamma_j}\sqrt{1+\frac{\nu^{-2}}{n_j\sigma^{-2}}} - \frac{\mu\nu^{-2}}{\sqrt{n_j\sigma^{-2}}}$$

where $q_{1-\gamma_j}$ is the $(1-\gamma_j)$th – upper quantile of $N(0,1)$.

Likelihood $f(y|\theta)$: $y_i \sim N(\theta, \sigma^2)$

Prior $\pi(\theta)$: $\theta \sim N(\mu, \nu^2)$

Posterior $p(\theta|y)$:

$$\theta \mid \boldsymbol{y}_j \sim \mathrm{N}\left(\frac{\mu\nu^{-2}+\bar{y}_j n_j\sigma^{-2}}{\nu^{-2}+n_j\sigma^{-2}}, \frac{1}{\nu^{-2}+n_j\sigma^{-2}}\right)$$

Hypothesis:

$$H_0: \theta \le 0 \; vs \; H_1: \theta > 0$$

Subjective (Full) Bayesian Designs : Subjectively decide $\pi(\theta)$ and γ_j based on prior knowledge and risk-benefit assessment

Examples: if the loss of wrongly rejecting the H_0 is 19 times the loss of wrongly rejecting H_1, then stop the trial at any interim as long as $\Pr(\theta > 0 \mid y_j) > 0.95$, i.e., $\boldsymbol{\gamma_j \equiv 0.95}$.

The type I error rate is inflated: the overall type I error rates are
$\alpha = 0.05, 0.08, 0.13, 0.17, 0.31$, and 1, for
$K = 1, 2, 5, 10, 100$, and ∞.

Bayesian designs based on pp:
Subjective Bayesian Designs (con't)

Subjective Bayesian Designs : Subjectively decide $\pi(\theta)$ and γ_j based on prior knowledge and risk-benefit assessment

Examples:

if the loss of wrongly rejecting the H_0 is 19 times the loss of wrongly rejecting H_1, then stop the trial at any interim as long as $\Pr(\theta > 0 \mid y_j) > 0.95$, i.e., $\boldsymbol{\gamma_j \equiv 0.95}$.

The type I error rate is **inflated**: the type I error rate $\alpha = 0.05, 0.08, 0.13, 0.17, 0.31$, and 1, for $K = 1, 2, 5, 10, 100$, and ∞.

A few notes:

1. "It is entirely appropriate to collect data until a point has been proven or disproven, or until the data collector runs out of time, money, or patience."(Edwards et al., 1963)
2. Harrell (2020a) gave an intriguing example.
3. The simple procedure of rejecting H_0 when $PP_j = \Pr(\theta > 0 \mid y) > 0.95$ at any interim analysis does not have the problem of the p −value: when sample size is large enough, $p < 0.05$ with probability 1. Not for PP_j.

Bayesian designs based on pp: **Calibrated Bayesian Designs**

Likelihood $f(y|\theta)$: $y_i \sim N(\theta, \sigma^2)$

Prior $\pi(\theta)$: $\theta \sim N(\mu, \nu^2)$

Posterior $p(\theta|y)$:

$$\theta \mid \boldsymbol{y}_j \sim \mathrm{N}\left(\frac{\mu\nu^{-2} + \bar{y}_j n_j \sigma^{-2}}{\nu^{-2} + n_j\sigma^{-2}}, \frac{1}{\nu^{-2} + n_j\sigma^{-2}}\right)$$

Hypothesis:

$$H_0: \theta \le 0 \; vs \; H_1: \theta > 0$$

False discovery rate (FDR) and false positive rate (FPR):

$$\mathrm{FDR} = \frac{\int_{\boldsymbol{y}_K \in \Gamma} \int_{\theta \le 0} f_0(\boldsymbol{y}_K \mid \theta)\pi_0(\theta)\mathrm{d}\theta\mathrm{d}\boldsymbol{y}_K}{\int_{\boldsymbol{y}_K \in \Gamma} f_0(\boldsymbol{y}_K)\mathrm{d}\boldsymbol{y}_K},$$

Wrong approvals among -- all approvals

$$\mathrm{FPR} = \frac{\int_{\theta \le 0} \int_{\boldsymbol{y}_K \in \Gamma} f_0(\boldsymbol{y}_K \mid \theta)\pi_0(\theta)\mathrm{d}\boldsymbol{y}_K\mathrm{d}\theta}{\int_{\theta \le 0} \pi_0(\theta)\mathrm{d}\theta}.$$

-- all ineffective drugs

Here, $\Gamma = \{y_K: \exists j \in \{1, \dots, K\} \; such \; that \; \Pr(\theta > 0 | y_j) > \gamma_j\}$

Calibrated Bayesian Designs : Calibrate $\pi(\theta)$ and γ_j according to the operating characteristics of the Bayesian designs (not just type I error rate)

Examples: if the true model and the assumed model are the same, i.e., $f_0(y|\theta)\pi_0(\theta) = f(y|\theta)\pi(\theta)$, then (6) guarantees that

$$(8) \qquad \mathrm{FDR} \le 1 - \gamma_{\min}, \quad \text{and} \quad \mathrm{FPR} \le \frac{(1-\gamma_{\min}) \cdot \int_{\theta>0} \pi(\theta)\mathrm{d}\theta}{\gamma_{\min} \cdot \int_{\theta \le 0} \pi(\theta)\mathrm{d}\theta},$$

where $\gamma_{min} = \min\{\gamma_1, \dots, \gamma_K\}$.

How to calibrate Bayesian designs?

- Through computer simulations:
  1. For each plausible true model $f_0(y|\theta)\pi_0(\theta)$, generate S hypothetical trials with treatment effects $\{\theta^{(1)}, \theta^{(2)}, \cdots, \theta^{(S)}\}$ and outcomes $\{y^{(1)}, y^{(2)}, \cdots, y^{(S)}\}$, for large S.
  2. Compute

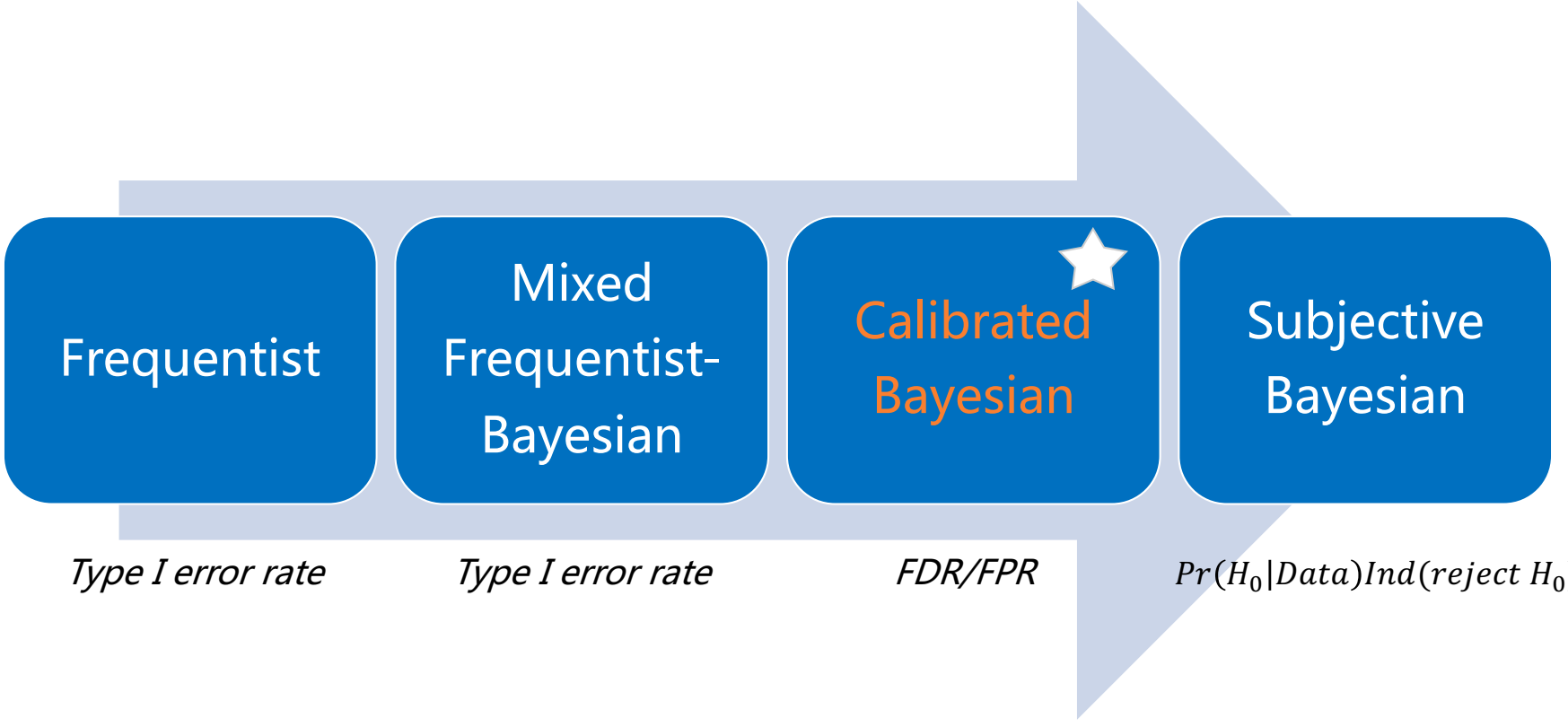
$$\widehat{\text{FDR}} = \frac{\sum_{s=1}^{S} \mathbf{1}\left(\boldsymbol{y}_K^{(s)} \in \Gamma, \theta^{(s)} \leq 0\right)}{\sum_{s=1}^{S} \mathbf{1}\left(\boldsymbol{y}_K^{(s)} \in \Gamma\right)}, \quad \text{and}$$

$$\widehat{\text{FPR}} = \frac{\sum_{s=1}^{S} \mathbf{1}\left(\boldsymbol{y}_K^{(s)} \in \Gamma, \theta^{(s)} \leq 0\right)}{\sum_{s=1}^{S} \mathbf{1}\left(\theta^{(s)} \leq 0\right)}.$$

  3. Calibrate prior and threshold γ_j's so that the $\widehat{\text{FDR}}$ and $\widehat{\text{FPR}}$ do not exceed prespecified levels for every plausible $f_0(y|\theta)\pi_0(\theta)$.

Note: Frequentist type I error rate is equivalent to FPR when $\pi_0(\theta)$ is a point mass at the null value.

The Bayes-Frequentist Spectrum



Bayesian Designs Possess Some Advantages

- If needed, frequentist properties (such as type I error rate) can be achieved

- Bayesian inference (e.g., estimation of treatment effects) obeys the likelihood principle; Frequentist inference, e.g., confidence interval and p-value, are affected by unrealized events

  - The confidence interval in a sequential trial may be counter-intuitive (Freedman et al., 1994; Rosner and Tsiatis, 1988)

  - Frequentist inference can be seriously affected by unexpected events that resulting in deviation of the sequential design, such as outbreak of COVID-19

  - Bayesian credible intervals are indifferent to stopping rules, no matter what they are, since stopping time and treatment effects are independent given the observed data (Hendriksen et al., 2021).

Likelihood Principle and Frequentist example

The Likelihood Principle. *All the statistical evidence about θ arising from an experiment is contained in the likelihood function for θ given y. Two likelihood functions for θ (from the same or different experiments) contain the same statistical evidence about θ if they are proportional to one another.*

- Imagine that a single-arm trial has been conducted, and 200 outcomes have been recorded that result in a z-statistic of $z_1 = 1.75$.
- Two investigators A and B, who used the same probability model but had different plans about the next step.
  - Investigator A planned a second stage for the trial to enroll 200 more patients if $z_1 \leq 1.88$ (the Pocock stopping boundary),
  - Investigator B did not plan to enroll any more patients.
  - According to the LP, the evidence about θ provided by the 200 observations is the same.
  - However, investigator A cannot claim statistical significance using the Pocock design after 200 observations (and may fail again after all 400 observations),
  - while investigator B can using a fixed design with 200 patients ($z_1 > q_{0.05} = 1.645$).
  - In other words, these investigators can reach completely different conclusions about the effectiveness of the drug with the exact same data.

Likelihood Principle and Bayesian example

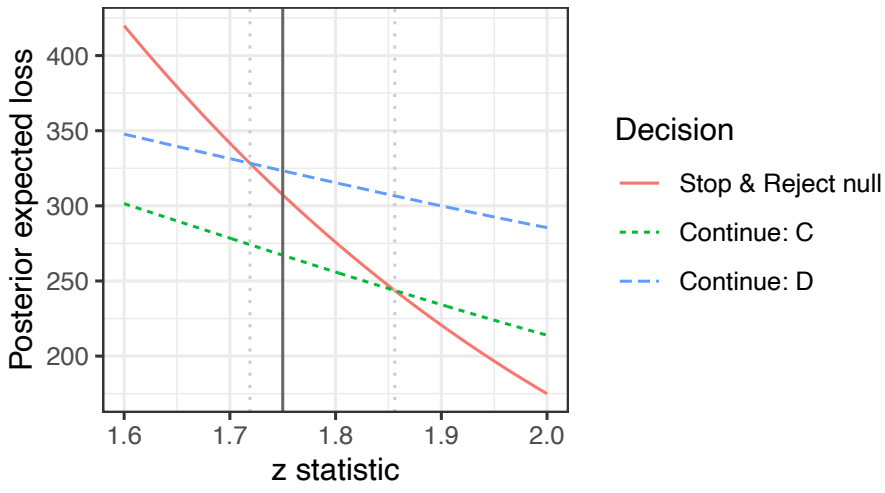


FIGURE 1. Posterior expected losses, as functions of the z-statistic, for possible decisions that can be made by investigators C and D at an interim analysis after 200 observations. The trial has a maximum sample size of 400 patients. Investigator C planned another interim analysis after 300 observations, while investigator D did not plan to conduct any additional interim analysis.

A Bayesian decision-theoretic design may lead to two different decisions for investigators C and D, who did and did not plan an interim analysis, respectively, based on the same 200 observations.

Zhou and Ji (2022)

LP only provides guidance on the estimation, not decision making.

Numerical Results: Stopping boundaries

TABLE 1. Stopping boundaries for the z-statistics given by several frequentist and Bayesian sequential designs. The single-arm trial in Section 1 is considered with $K = 5$ analyses, a maximum sample size of $n_K = 1000$, and equal group sizes ($n_j = 200j$). The design parameters are calibrated such that the type I error rate at $\theta = 0$ is $\alpha = 0.05$ for every design.

Analysis	1	2	3	4	5
No. of patients	200	400	600	800	1000
Frequentist designs:					
Pocock	2.12	2.12	2.12	2.12	2.12
O'Brien-Fleming	3.92	2.77	2.26	1.96	1.75
Error spending h_1	2.18	2.14	2.11	2.09	2.07
Error spending h_2	4.23	2.89	2.30	1.96	1.74
Error spending h_3	2.33	2.22	2.12	2.03	1.96
Stochastic curtailment	5.38	3.65	2.82	2.27	1.65
Bayesian designs:					
Posterior probability (ver. 1)	2.71	2.24	2.06	1.97	1.91
Posterior probability (ver. 2)	2.13	2.12	2.12	2.12	2.12
Posterior predictive probability	2.50	2.26	2.18	2.11	1.84
Decision-theoretic	2.33	2.22	2.15	2.09	1.91

For stopping boundaries based on posterior probabilities,

Version 1 refers to setting $\gamma_j \equiv 0.95$ and $\pi(\theta) = N(0, 0.054^2)$,which leads to a type I error rate of 0.05.

Version 2 sets $\pi(\theta) = N(0,1)$ and $\gamma_j \equiv 0.983$, which also leads to a type I error rate of 0.05.

Simulation Results

Likelihood $f(y|\theta)$: $y_i \sim N(\theta, 1)$ Prior $\pi(\theta)$: $\theta \sim N(0, \nu^2)$

True model: $f(y|\theta)$: $y_i \sim N(\theta, 1)$ Prior $\pi(\theta)$: $\theta \sim N(0, \nu_0^2)$

If $PP_j = \Pr(\theta > 0 | y_j) > 0.95$, then efficacy is declared and trial stopped.

Operating Characteristics

- When the Bayesian models are correctly specified, frequentist properties are preserved without the need to adjusting for multiple comparison
- When the Bayesian models are missepecified, type I error rates can be inflated when K is large
- If no more than K=10 IAs are planned, recommend setting $\nu < 1$ to achieve desirable frequentist properties.

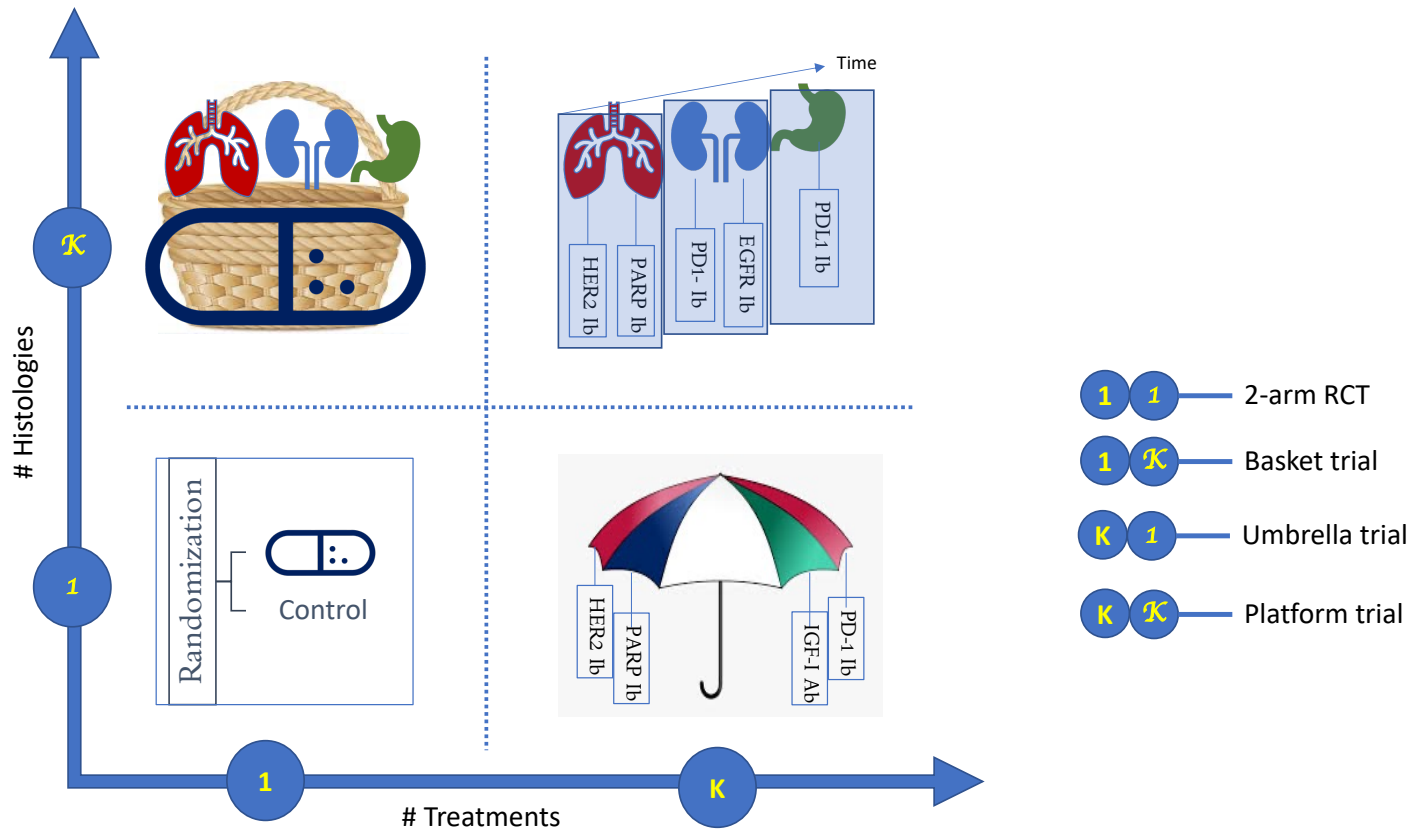
Discussion

- Multiplicity: Bayes vs Frequentist
- Two-arm randomized trials
- Estimation of false approval rate from all approved drugs
- Future and ideal approval process
- Practical Impediment
  - Operation and human bias
  - Unblinding

Master Protocols

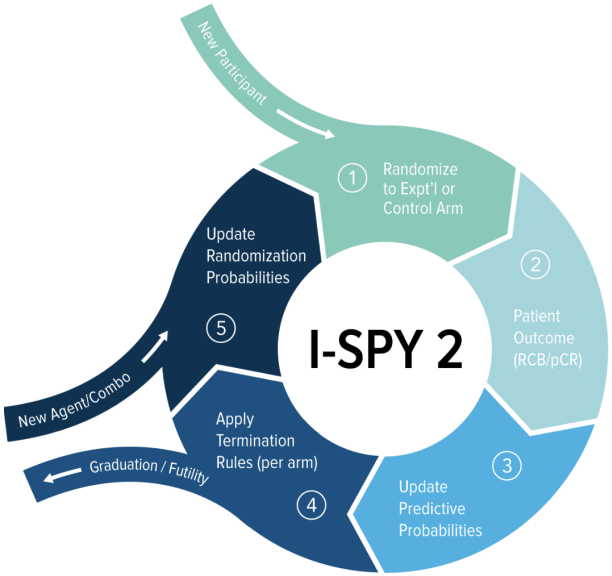
Modern Clinical Trials with Master Protocols

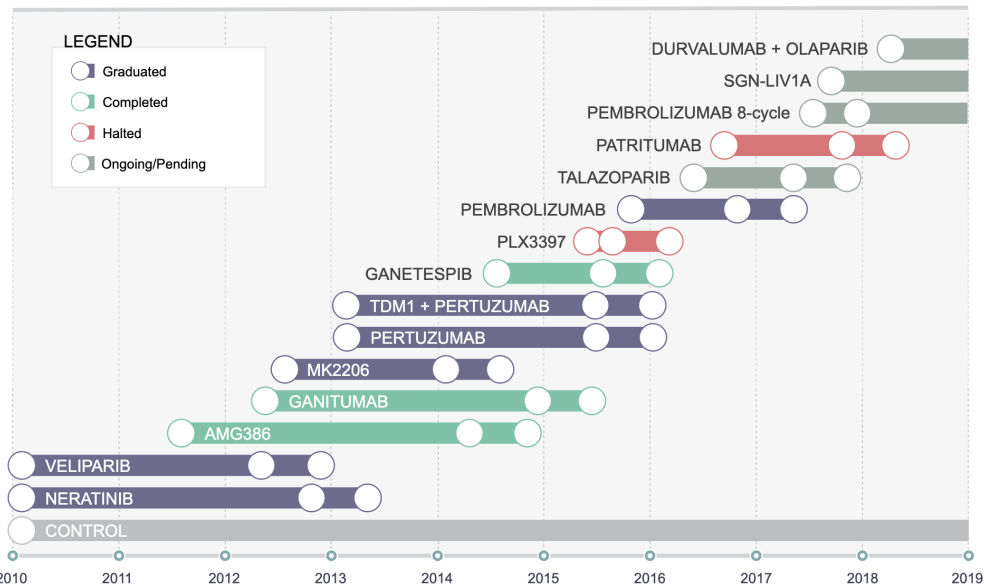
- Basket trials: 1 treatment and K histologies
- Umbrella trials: K treatments and 1 histology
- Platform trials: ≥ 1 treatments, ≥ 1 histologies, over time



Tsimberidou, Müller, Ji (2020)

I-SPY2: A Successful Platform Trials in 21st Century



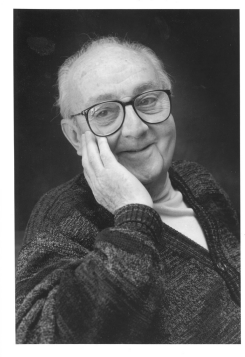


- Adaptive platform trial;
- RCB 0 or pCR endpoint; Neoadjuvent
- A common control
- Adaptive randomization
- No sample size – add new arms, graduate existing arms adaptively
- Bayesian predictive probability
- A few promising drugs graduated; 3 received accelerated approval
- Reference: `ispytrials.org`

Bayesian designs for early-phase trials

All models are wrong, but some are useful.

-- George Box



(1919-2013)

Oncology drugs with post-market dose modification

Examples of Drugs Whose Doses or Schedules Were Modified for Safety or Tolerability after Approval.*

Drug	Initial Dose and Trials	Modified or Added Dose and Trials	Reason for Modified or Added Dose
Small-molecule drugs			
Ceritinib (Zykadia)	750 mg PO daily fasted (ASCEND-1)	450 mg PO daily with food (ASCEND-8)	Reduce gastrointestinal toxic effects
Dasatinib (Sprycel)	70 mg PO twice daily (CA180013, CA180005, CA180006, and CA180015)	100 mg PO daily (CA180034)	Reduce hematologic toxic effects and fluid retention
Niraparib (Zejula)	300 mg PO daily (NOVA)	200 mg PO daily (PRIMA)	Reduce thrombocytopenia in patients with a lower platelet count or lower body weight
Ponatinib (Iclusig)	45 mg PO daily (PACE)	45 mg PO daily, then 15 mg PO daily once ≤1% *BCR-ABL* is achieved (OPTIC)	Reduce vascular occlusive events
Chemotherapy			
Cabazitaxel (Jevtana)	25 mg/m² IV every 3 wk (TROPIC)	20 mg/m² IV every 3 wk (PROSELICA)	Reduce hematologic toxic effects and infections
Antibody–drug conjugates			
Gemtuzumab ozogamicin (Mylotarg)	9 mg/m² IV on days 1 and 15 (Study 201, Study 202, and Study 203)	3 mg/m² IV on days 1, 4, and 7 (Mylofrance-1)	Reduce veno-occlusive disease and treatment-related mortality

* Adapted from the Food and Drug Administration.[2] IV denotes intravenous, and PO by mouth.

Shah et al., 2021, NEJM

- All the listed drugs had to **reduce** their dose or schedule **due to toxicity**

Sotorasib (Lumakras) for NSCLC

Approved in May 2021 for patients with NSCLCs harboring KRAS p.G12C mutation (based on a phase 2 trial)

The first drug successfully targets KRAS, a historically "undruggable" and yet important cancer biomarker

However, a postmarketing trial is required by FDA to further explore lower doses than the approved one

This is due to lack of sufficient dose exploration in early-phase development (e.g., phase 1 with small sample size; dose selection under MTD-regime)

New Era of Dose Optimization: Challenges

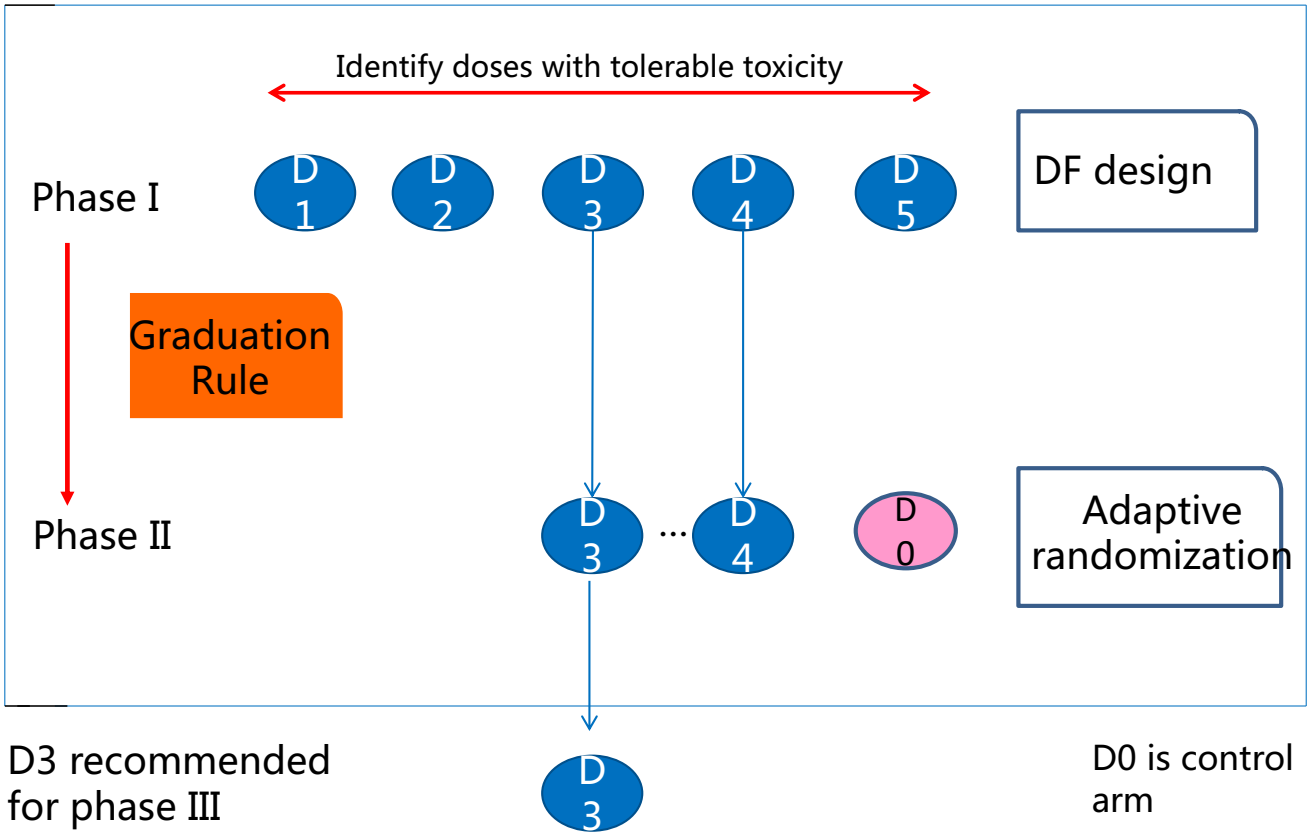
Higher doses might not have better therapeutic activity	DLT may not be observed at clinically active doses	Serious toxic effects may occur after several cycles of drug usage
• MTD is no longer the optimal dose	• Dose escalation and dose selection challenges	• Delayed toxicity/efficacy

New Era of Dose Optimization: Solutions

MTD not optimal	DLT may not be observed at clinically active doses	Delayed toxicity/efficacy
• Eff/Tox dose finding (or Biomarker/Tox)	• Eff/Tox/Exposure dose finding	• Time-to-event (TITE) or Probability of decision (PoD) modeling

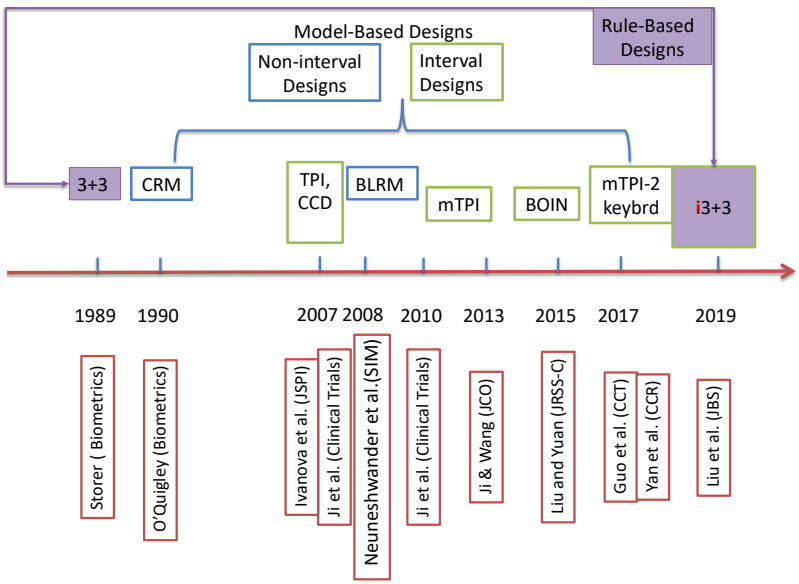
The SEARS Design (Pan et al., 2014, Clinical Trials)

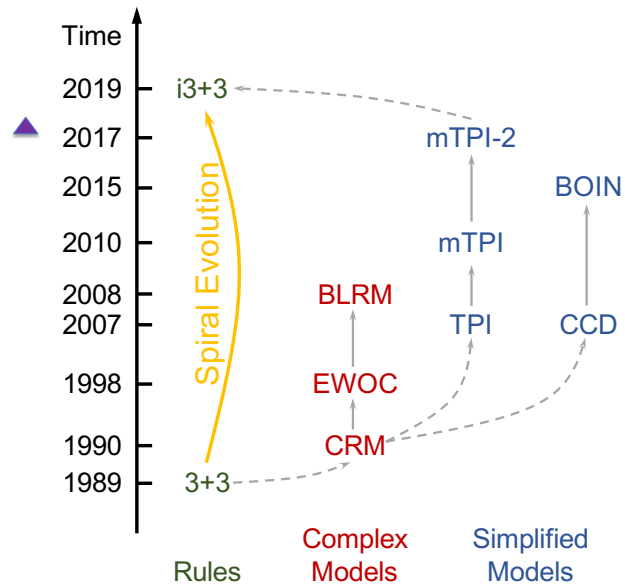
Seamless phase 1/2 oncology dose optimization trials



Dose-finding designs over last 30 years

- So many designs are available now. Which one to use?





MUCE: A New Bayesian Basket Trial Designs

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

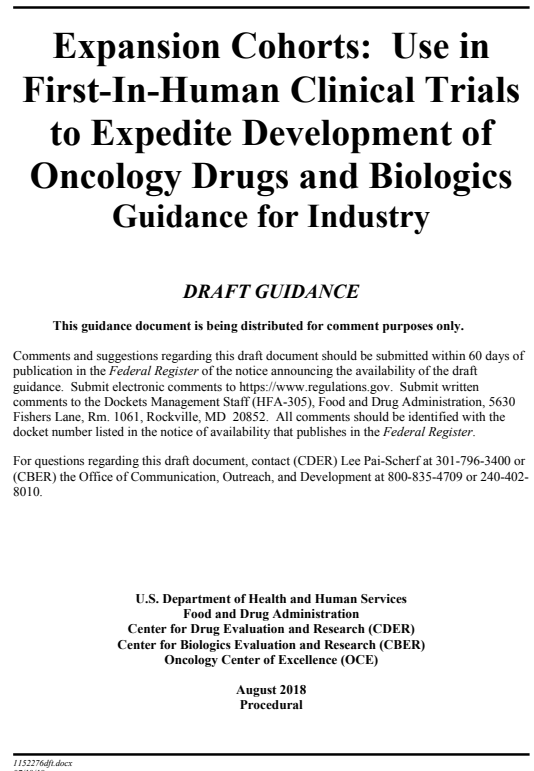
Master protocol phase II/III trials Each arm is a subgroup of patients defined by biomarker status, a different drug, or a mini two-arm subtrial

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.

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)

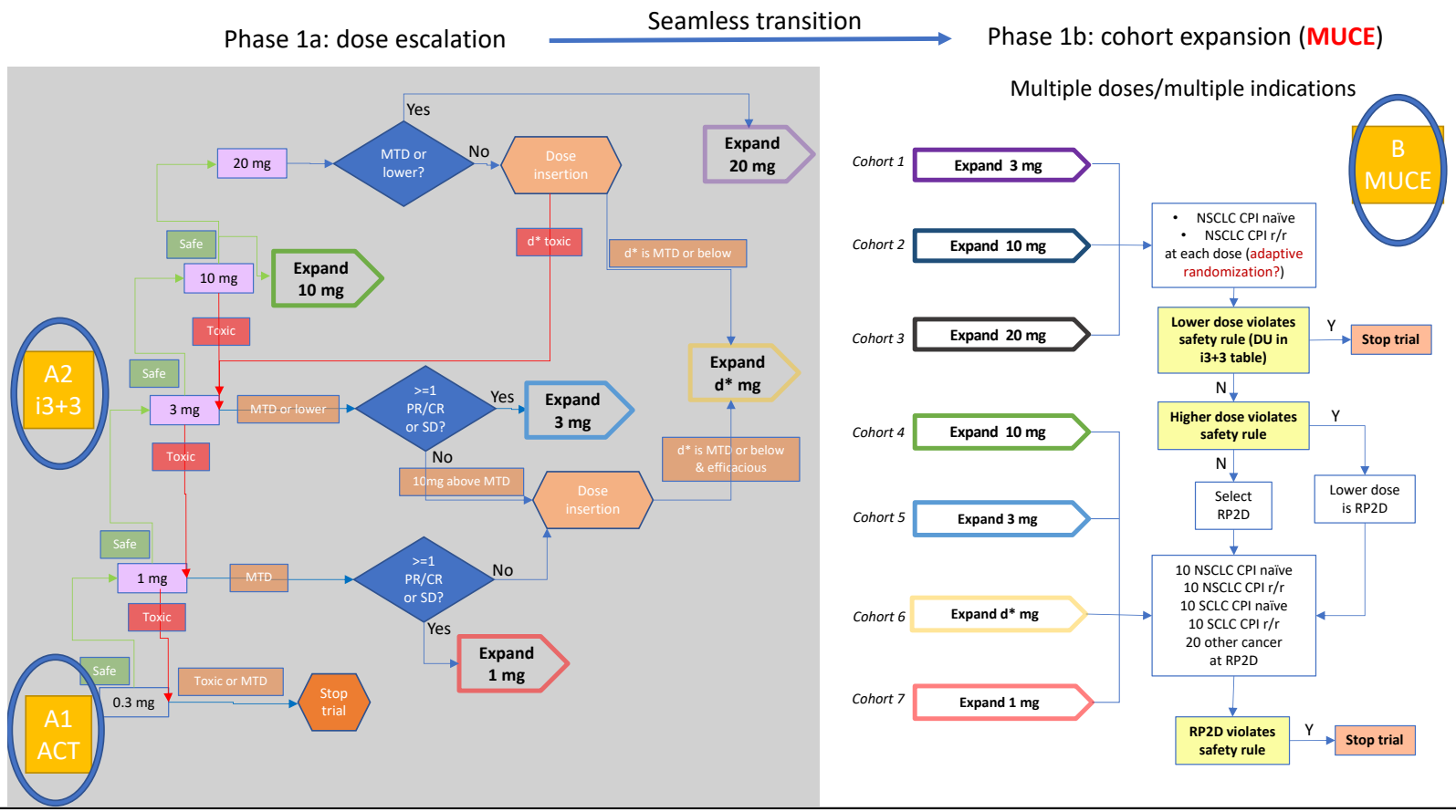
August 2018
Procedural

115227dft.docx
07/18/18

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



The MUCE design/method (for data analysis)

Lyu et al. (2021): `https://arxiv.org/abs/2006.07785`

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 hyperpriorallow p to be random and realizes "multiplicity control" – a smaller valuemore stringentcontrol.

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.
- Futility stopping: Stop for futility at interim analysis if $Pr(\lambda_{ij} = 1 \mid data) < v_1$
- Efficacy stopping: 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

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

Case Study: 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.

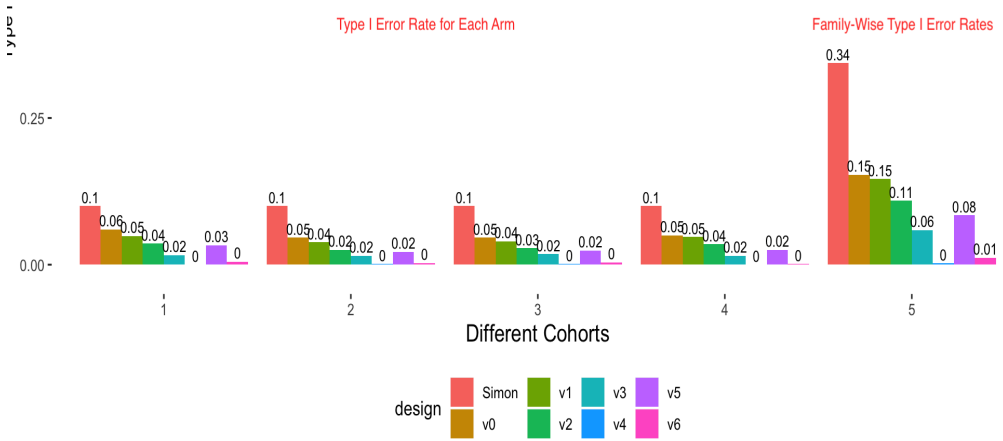
Subgroups		Arm 1	Arm 2	Arm 3	Total sample size
Assumptions	Endpoint	pCR	ORR	ORR	
	Historical vs Expected	0.05 vs 0.2	0.4 vs 0.5	0.15 vs 0.3	
	Alpha	0.05	0.20	0.05	
	Power	0.80	0.80	0.80	
Simon's 2-stage design		N=29 N1*=10	N=81 N1=40	N=55 N1=19	**165**
MUCE design		N=20 N1=10	N=30 N1=15	N=30 N1=15	**80**

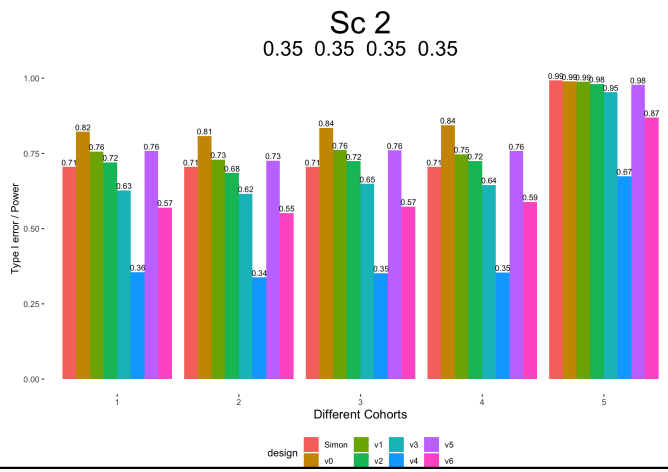
MUCE Multiplicity Control – How it is done?

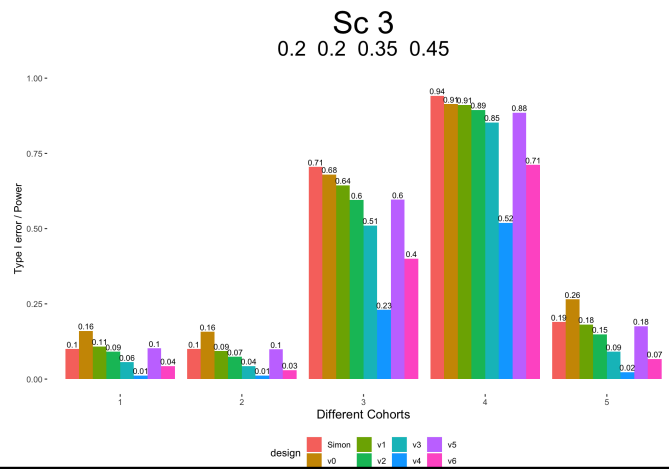
Consider the following hyper-parameters in $\xi_0 \sim N(\mu_\xi, \sigma_\xi = 1), \eta_0 \sim N(\mu_\eta \equiv 0, \sigma_\eta = 1)$. Consider 7 versions of the hyper-parameters. Conclude treatment efficacious if $Pr(\lambda_{ij} = 1 \mid data) > v_2 = 0.95$. Note no calibration of v_2 here.

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

$$\xi_i \mid \xi_0 \sim N(\xi_0, 1), \quad \text{and} \quad \eta_j \mid \eta_0 \sim N(\eta_0, 1). \ \Big\} \text{Borrow \& Shrinkage}$$

Hyperprior ξ_0 and η_0

$$\xi_0 \sim N(\mu_\xi, 1), \quad \text{and} \quad \eta_0 \sim N(\mu_\eta, 1) \ \Big\} \text{Multiplicity control}$$

Making μ_ξ and μ_η negative induces multiplicity control!

Summary and Remarks

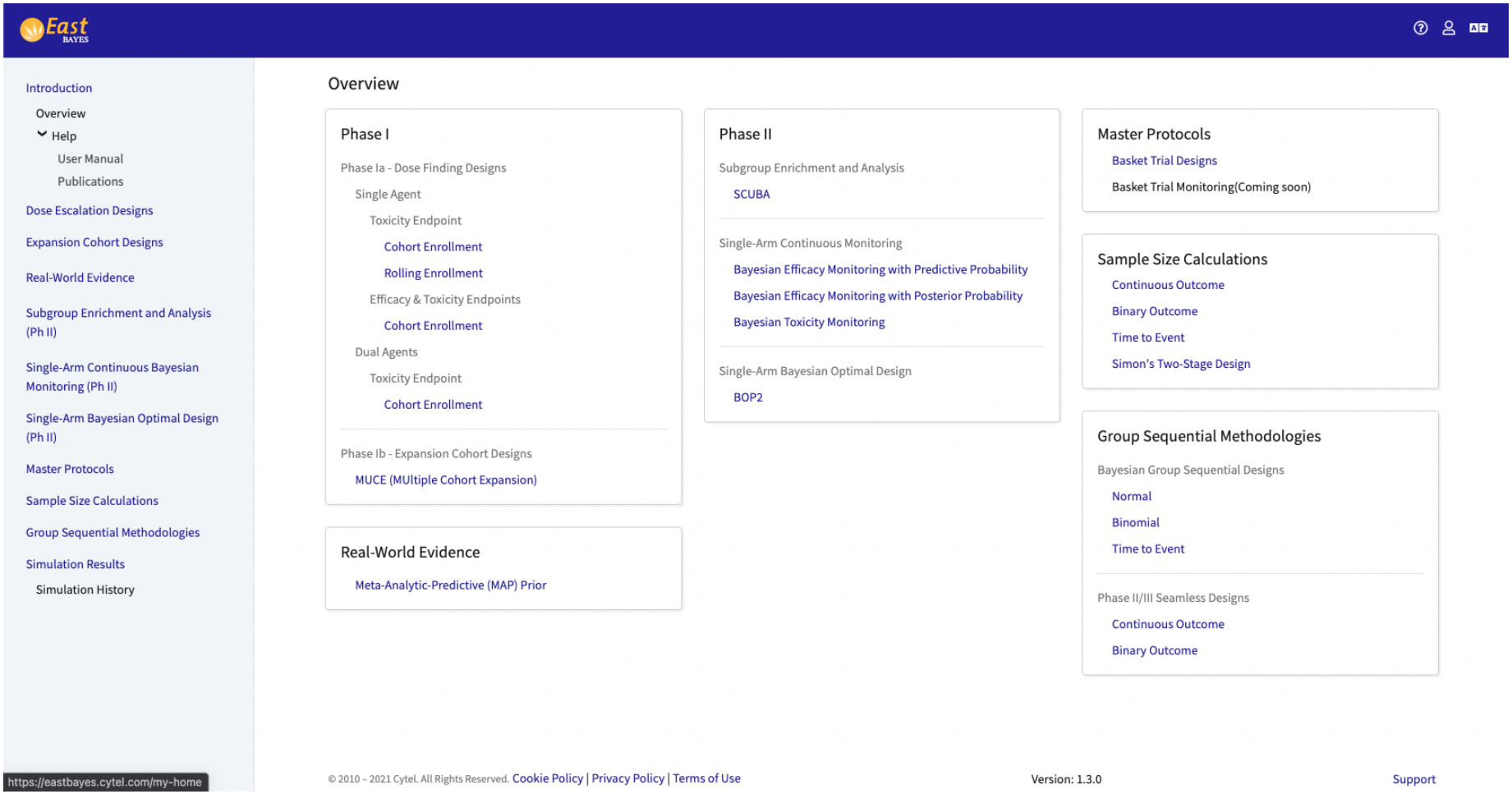
Superior performance MUCE is an advanced Bayesian approach superior to the Simon's 2-stage design for expansion cohorts trials and master protocols: smaller sample size or higher power in frequentist OCs; better control of Type I error rates in global null

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.

East Bayes

Formerly known as "U-Design".



THANK YOU!