CERSI
Johns Hopkins University

FDA U.S. FOOD & DRUG ADMINISTRATION

FDA Center for Drug Evaluation and Research and Johns Hopkins Center of Excellence in Regulatory Science and Innovation (CERSI) Workshop

Addressing Challenges in the Design and Analysis of Rare Disease Clinical Trials: Considerations and Tools

May 2 – 3, 2023

CERSI
Johns Hopkins University

FDA U.S. FOOD & DRUG ADMINISTRATION

May 3, 2023

Design and Analysis Methods for Clinical Trials for Rare Diseases

CERSI
Johns Hopkins University

FDA U.S. FOOD & DRUG ADMINISTRATION

Welcome

Dionne Price, PhD

Deputy Director

Office of Biostatistics, Office of Translational Sciences

Center for Drug Evaluation and Research, FDA

CERSI
Johns Hopkins University

FDA U.S. FOOD & DRUG ADMINISTRATION

Session 1:
Adaptive Designs in Small Populations

Moderator: Michael Rosenblum, PhD
Professor of Biostatistics
Johns Hopkins Bloomberg School of Public Health

SMART Design and Bayesian Methods for Rare Disease Trials

KELLEY M KIDWELL, PHD

UNIVERSITY OF MICHIGAN, DEPARTMENT OF BIOSTATISTICS

MAY 2023

Challenges in Rare Disease Research

- Small patient numbers
- Even smaller number of endpoint events
- Challenging to run separate **dose finding** trial and **confirmatory** trial
- Difficult to meet “standard” **Frequentist benchmarks** (80% power, 5% type I error)

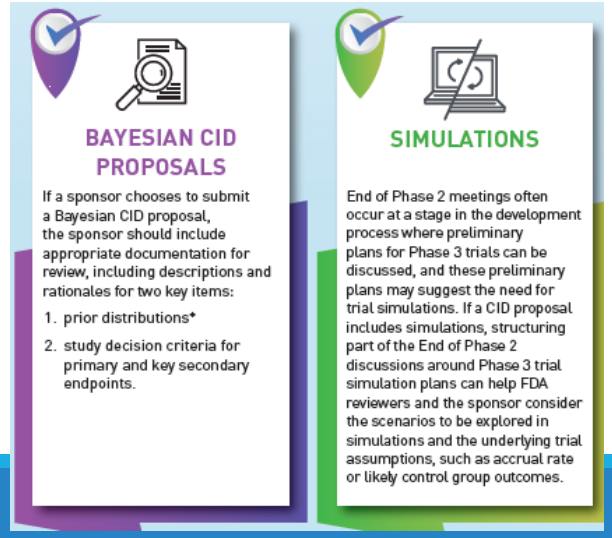
Need for clinical trial innovation

DESIGN

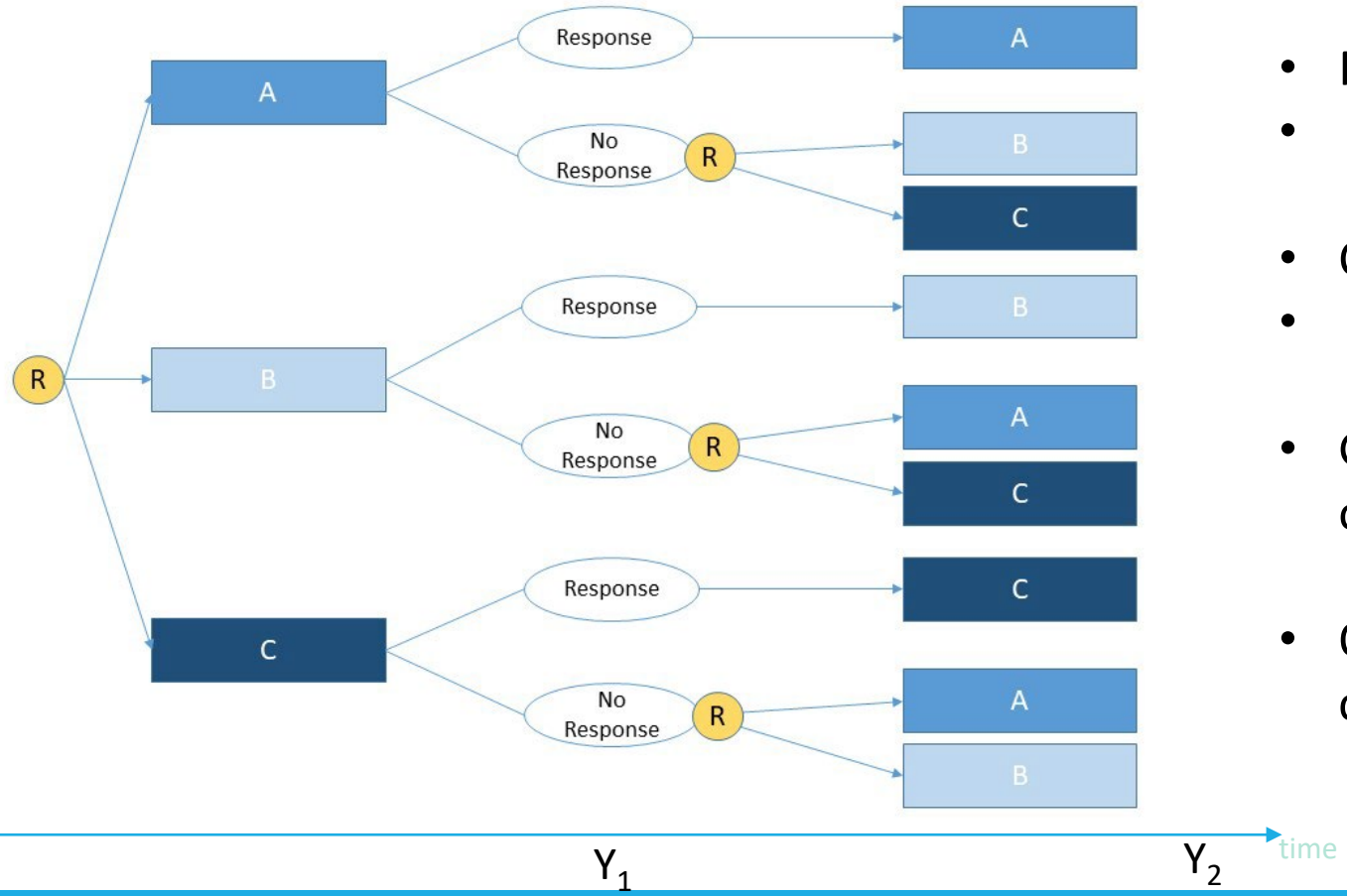
- Minimal **size** while providing robust evidence
- **Benefit participants**
  - maximize chance of receiving therapy
  - minimize number receiving placebo
- Consider **more than 1 dose or treatment** and confirm its efficacy

ANALYSIS

- Provide estimates with **clinical interpretability**: probability of clinical meaningful treatment benefit
- Incorporate **external data** (natural history studies, previous trials)



snSMART Design: 3 active treatments



- Motivated by trial in isolated skin vasculitis
- ARAMIS (NCT02939573)

- Comparative effectiveness study
- no placebo = increased recruitment

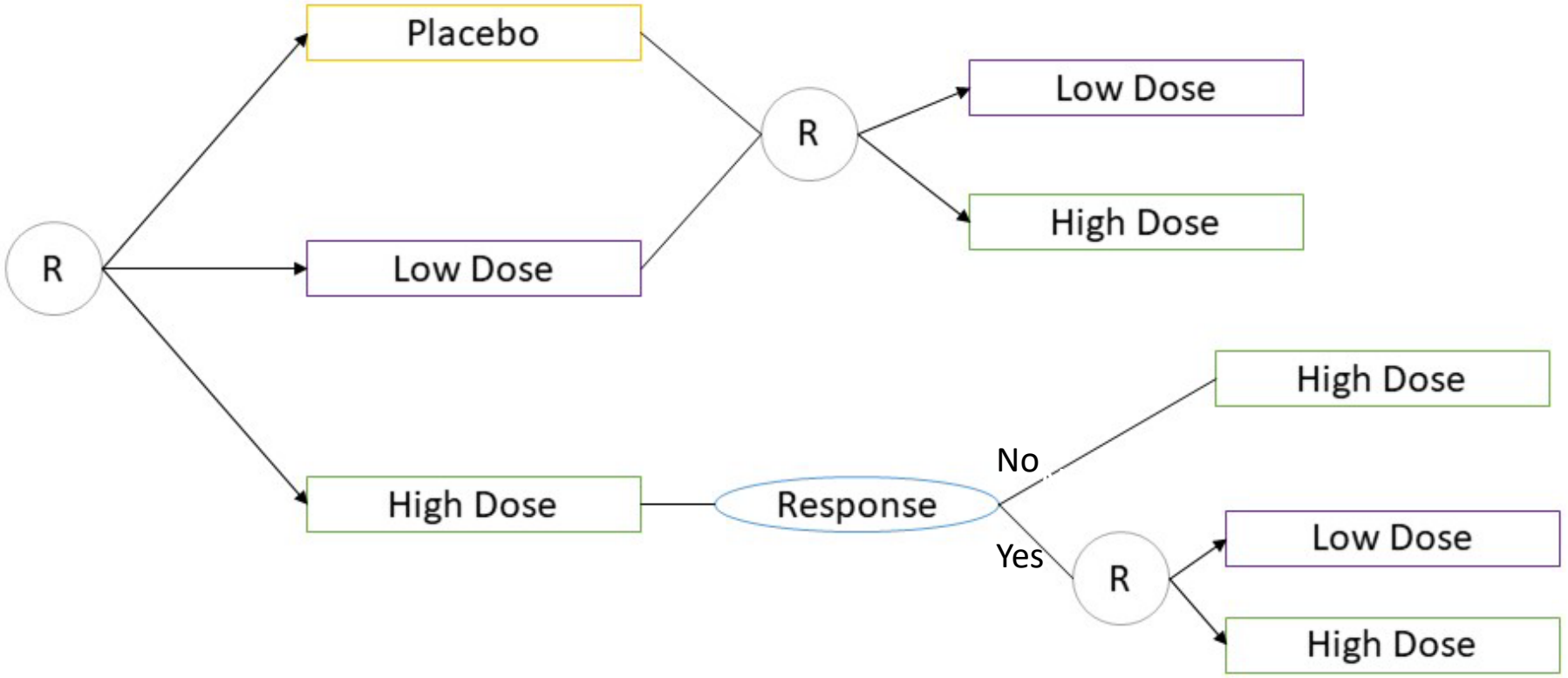
- Goal: Estimate the first stage treatment effect of A, B, C using data from stages 1 and 2

- Outcome: binary (response rate) or continuous (score)

snSMART Design

- small sample (n), Sequential, Multiple Assignment, Randomized Trial
- A type of **multi-stage,** randomized design where individuals are randomized to a set of treatment options and may be **re-randomized based on response** to initial treatment
- **All** participants receive active (or some dose of) treatment
- Obtain **more information from smaller number** of participants
- Ability to **stay on treatment if responding**, **switch** to different treatment if not responding
- Appropriate for rare diseases or disorders that are **chronic, relatively stable** over the 2 stages of the trial
- Restricted crossover design

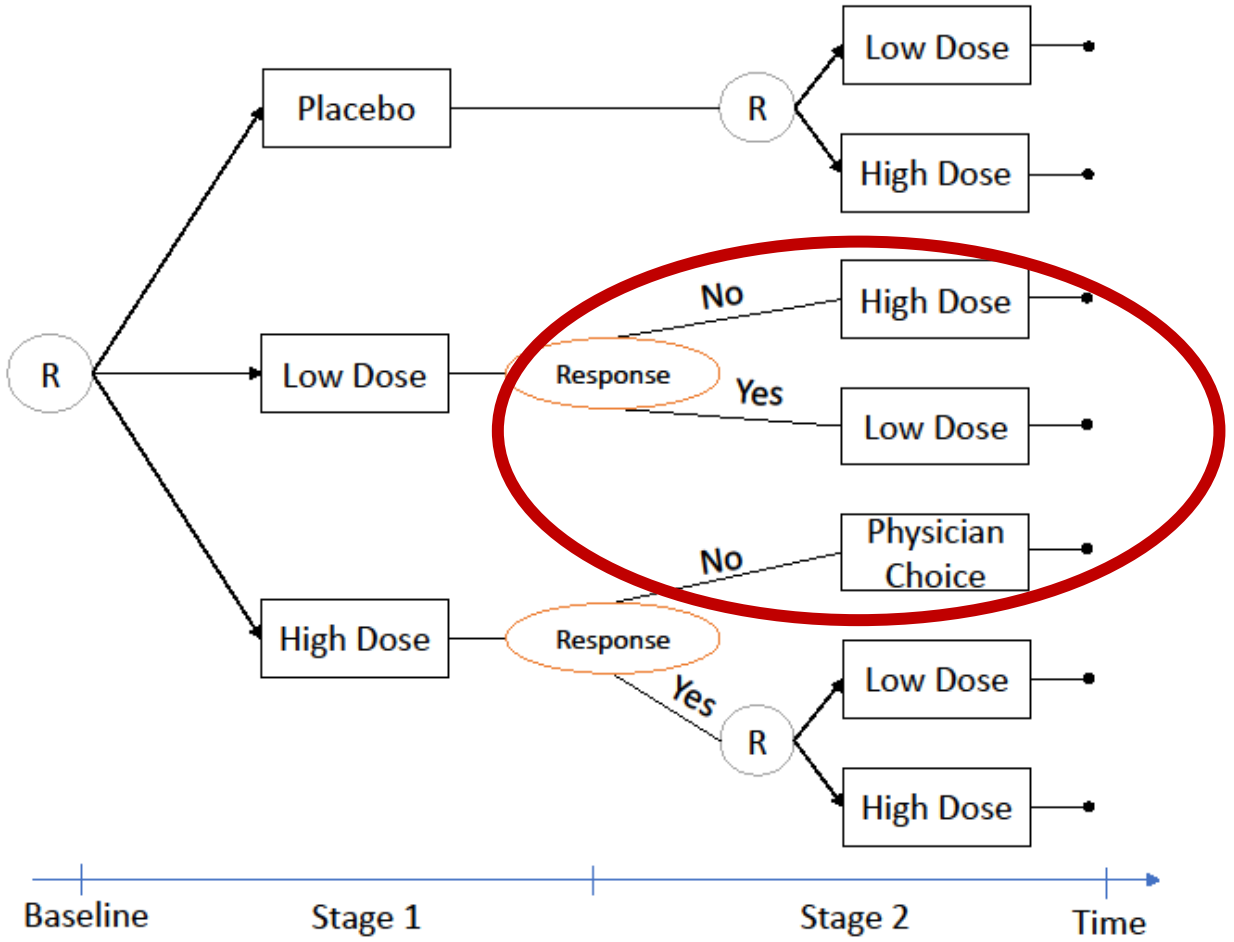
snSMART Dose Design 1



Fang, F, Hochstedler, KA, Tamura, RN, Braun, TM, Kidwell, KM. Bayesian methods to compare dose levels with placebo in a small n, sequential, multiple assignment, randomized trial. *Statistics in Medicine*. 2021; 40: 963– 977

Fang, F, Tamura, RN, Braun, TM, Kidwell, KM. (2022) Comparing Dose Levels to Placebo using a Continuous Outcome in a Small n, Sequential, Multiple Assignment, Randomized trial (snSMART), Statistics in Biopharmaceutical Research.

snSMART Dose Design 2



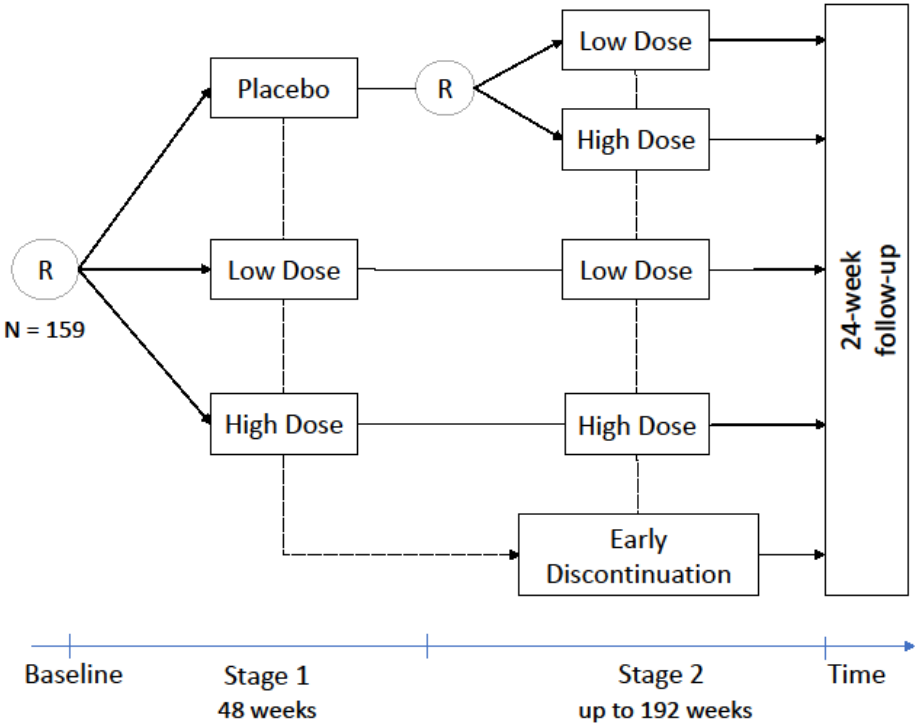
More patient-centered re-randomization

Motivating Setting: DMD

SPITFIRE: 2-phase, placebo-controlled study (NCT03039686) of 2 dose levels treatment in ambulatory boys with DMD

- Only placebo group re-randomized in period 2
- Only stage 1 data used in primary analysis

<u>Outcome</u>: change from baseline to week 48 in 6-minute walk distance or NSAA score



2-stage trial in DMD, similar to an snSMART

Advantages of snSMART with dose levels

1. Many participants will receive a higher dose treatment in stage 2 or switch to other treatments that might be more suitable for them
   - **Engagement** and **retention**
2. Design allows for **both dose-finding and confirmation** of the dose effect to register the drug within one trial
   - often proceed with the highest dose, ignoring that low dose could be just as effective and more tolerable
3. Analysis can incorporate expert opinion or **external co-data**
   - Efficiency of treatment estimates
   - Decrease # on placebo

snSMART Bayesian Analysis

Goal	Provide	Shift	Incorporate
Estimate the first stage response rates (or mean outcomes) of each treatment by pooling data from both stages of the trial	Credible intervals of effect or difference between treatment effects: contain the true effect with some particular probability	Focus away from significance/p-values	Expert opinion, historical data, or co-data to increase precision

Bayesian Framework

We don’t know what the population parameters/true values(e.g. response rates) are

- random (they can change)

We take our best guess at the response rates based on our current knowledge (expert, registry, prior trials)

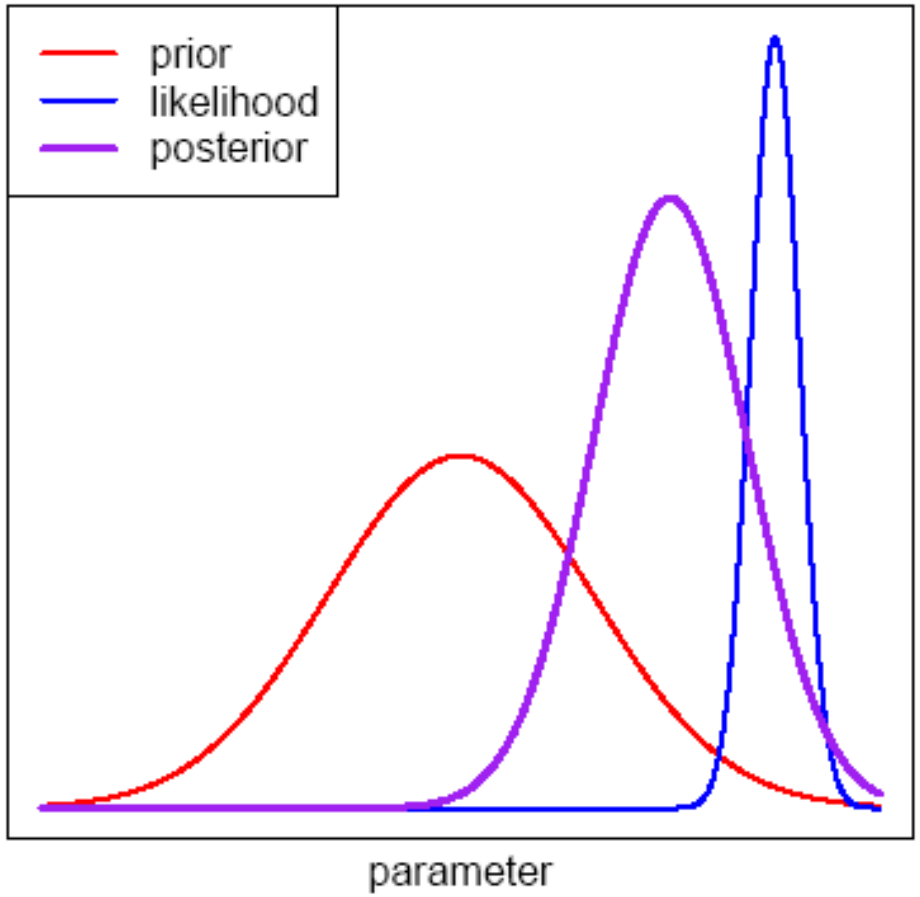
- **PRIOR**

We collect data to observe the response rates (trial)

- **LIKELIHOOD**

We combine our **PRIOR** & **LIKELIHOOD** for updated estimates of the response rates (results)

- **POSTERIOR**



Bayesian Analysis Approach

Prior Distributions: Informed by clinician investigators, historical data

- Informative, usually few people's worth of info
- Mixture approach: informative prior informed by expert & non-informative prior
- Test sensitivity of results given different prior distributions

Likelihood: Joint Model of current snSMART trial data

- Model the first stage outcome simply
- Model the second stage outcome based on the first stage treatment and outcome and second stage treatment
  - Augment expected outcome from stage 1 – can add potential resistance to drug
  - Link the outcome from stage 1 to stage 2 using linkage parameters – induces within patient correlation

Choosing External Data

- Careful choice of control data
- **Pocock criteria** to assess similarity between external control and trial control
  - Inclusion/exclusion criteria
  - Endpoint definition
  - Control treatment
  - Distribution of demographic criteria
- Number of external control patients/**Effective Sample Size** not to exceed the number on control in the trial
- Can allow lower number of participants on placebo in current trial

Model Assumptions

1. Does not control for any patient or disease characteristics (covariates/potential confounders)
2. Often make simplifying assumptions about linkage parameters
   - Second stage outcome is related to first stage outcome similarly across all treatments
3. Washout period between treatments, no carryover effect
4. 1 endpoint of interest
5. No to low missing data

Test our models' sensitivity to these assumptions & developing extensions

Results from Models

- Compared to one stage design analyses or joint stage frequentist analyses, our **Bayesian Joint Stage Models (BJSM)** provide treatment effects that
  - have low to no bias
  - are more efficient (lower variance)

- When assumptions are violated, BJSM are robust and maintain good results

- Can **test sensitivity of BJSM** to assumptions
  - via simulations in design phase
  - via comparing to Frequentist model in the analysis phase

Robust MAC-snSMART model DMD Data

- Re-analysis of study results from SPITFIRE- simulated 1st and 2nd stage data based on summary data
- Incorporated CINRG natural history co-data
- Simulated 30,000 realizations
- Outcome: 6 meter walk distance 95% credible interval: shorter intervals

	Difference Low-Placebo 95% CI	Difference High-Placebo 95% CI
1st stage traditional approach	1.8 (-22.6, 26.0)	11.5 (-12.5, 35.4)
BJSM	1.8 (-16.6, 19.5)	11.4 (-5.8, 28.6)
Robust MAC	1.6 (-15.6, 19.1)	10.9 (-6.4, 28.1)

How do investigators size an snSMART 1

For 3 active comparators and binary outcome

- Rshiny Applet
- 80% probability for the 90% credible interval of the difference between the best and second best treatment to exclude 0

Scenario	Response Rate			Sample Size per arm	True Power
	π_A	π_B	π_C		
1	0.25	0.25	0.50	28	0.78
2	0.20	0.20	0.40	46	0.81
3	0.30	0.30	0.50	48	0.82

How do investigators size an snSMART 2

For placebo, high and low doses and continuous outcome

- Rshiny Applet coming soon
- Find n such that the credible interval of the difference between low dose and placebo rules out 0 with desired probability (power)

Scenario	One stage Design		snSMART Design		N/ N_{1Freq}	N/ N_{1Bayes}
	N_{1Freq}	N_{1bayes}	N	Power		
1	50	46	31	0.81 (0.80-0.82)	0.62	0.67
2 (↑ correlation)	50	46	20	0.80 (0.79-0.81)	0.40	0.43
3 (↑ var on trt est)	50	50	34	0.81 (0.80-0.82)	0.68	0.68

Sample size for an snSMART using the Bayesian Joint Stage Model reduces sample size from a 1 stage design by 15-60%

How do investigators analyze data?

All our current methods, Bayesian and Frequentist, are available in **R package snSMART**

- https://cran.r-project.org/web/packages/snSMART/index.html

8 papers & counting: Statistics in Medicine, Journal of Biopharmaceutical Statistics, Journal of the Royal Statistical Society Series C, Contemporary Clinical Trials, Orphanet Journal of Rare Diseases

ARAMIS: NCT02939573

MISTIC: NCT04898231

Summary

- snSMART design & Bayesian joint stage models fit under **Complex Innovative Design** for comparative effectiveness & confirmatory drug comparison
  - For chronic, stable rare diseases
  - Design has potential to aid in recruitment and retention
- Design and analysis that can both **dose-find and confirm** the best dose level
- Using 2 stage design and Bayesian framework allows for more **efficient**, **unbiased** treatment effect estimates
- We have developed **software to disseminate** these methods in hopes the design will aid in identifying more effective treatments for many rare diseases

Acknowledgements

Work is funded by

- PCORI ME-1507-3118, PI Kidwell
- FDA BAA 75F40120C00195, PI Kidwell

Team

- Thomas Braun, Roy Tamura, Boxian Wei, Yan-Cheng Chao, Fang Fang, Sidi Wang, Satrajit Roychoudhury

References

Wei, Braun, Tamura, Kidwell. A Bayesian analysis of small n sequential multiple assignment randomized trials (snSMARTs). (2018). *Statistics in Medicine*. 37: 3723-32.

Wei, B., Braun, T., Tamura, R., Kidwell, K.M. Sample size considerations to find the optimal treatment in a small n sequential multiple assignment randomized trial in the setting of a rare disease. (2020). *Journal of Biopharmaceutical Statistics*. 30:6. 1109-1120.

Chao, Y.C., Braun, T.M., Tamura, R.N., Kidwell, K.M. A Bayesian two-step dropping rule for small n sequential multiple assignment randomized trials. (2020). *Journal of the Royal Statistical Society Series C*. 69: 663-680.

Chao, Y.C., Trachtman, H., Gipson, D., Spino, C., Braun, T.M., Kidwell, K.M. Dynamic Treatment Regimens in Small n, Sequential, Multiple Assignment, Randomized Trials: An Application in Focal Segmental Glomerulosclerosis. (2020). *Contemporary Clinical Trials*. 92:105989.

Hartman, H., Tamura, R.T., Schipper, M., Kidwell, K.M. Small n sequential multiple assignment trial with continuous intermediate and overall outcomes. (2021). *Statistics in Medicine*. 40: 312-326.

Fang, F., Hochstedler, K.A., Tamura, R.N., Braun, T.M., Kidwell, K.M. A Bayesian Analysis of small n, sequential, multiple assignment, randomized trial designs for the registration of a drug in rare diseases. (2021). *Statistics in Medicine*. 40: 963-977.

Kidwell KM, Roychoudhury S, Wendelberger B, Scott J, Moroz T, Yin S, Majumder M, Zhong J, Huml RA, Miller V. Application of Bayesian methods to accelerate rare disease drug development: scopes and hurdles. Orphanet *J Rare Dis*. (2022) May 7;17(1):186.

Fang, F., Tamura, R.T., Braun, T.M., Kidwell, K.M. Comparing Dose Levels to Placebo using a Continuous Outcome in a small n,Sequential, Multiple Assignment, Randomized Trial (snSMART). (In Press). *Journal of Biopharmaceutical Statistics*.

Thank you!

Adaptive Enrichment Designs in Rare Disease Settings

Noah Simon

May, 2023

The following ideas have evolved from discussion with

A certain Richard Simon

Targeted Treatments

Diseases are often somewhat heterogeneous in mechanistic cause

New treatments commonly target only a subset of people with a disease (from a particular mechanism)

In some cases, characterizing exactly who we think will benefit from treatment before running a pivotal trial is impossible

Targeted Treatments

In such cases, you might

- Enroll broadly (all-comers design)
- Make a best guess and restrict enrollment (enrichment design)

Adaptive Enrichment

Adaptive Enrichment designs provide a happy medium:

The trial begins with broad eligibility...

As evidence accumulates on who benefits from treatment,

enrollment criteria are modified

Modifications will use outcomes and tx assignments from earlier pts

Oncology Examples

Checkpoint inhibitors have been very effective

(in particular targeting PD1/PD-l1)

Treatment effect is often observed to increase with PD1 expression;

However, treatment may be effective even in low-expressors.

Adaptive enrichment can be (/is being) used to help run trials that leverage this partial knowledge

Oncology Examples

Cetuximab is a common cancer treatment that targets EGFR

Pivotal trial in colorectal cancer did not initially find significance

(it was an all-comers design)

Retrospective analysis showed:

KRAS wildtype $\rightarrow$ strong response!

(KRAS mutant tumors are mechanistically different)

Enrichment design was not used as it was unclear whether to restrict eligibility based on EGFR expression, or KRAS mutation status[1]

[1] among other reasons

Oncology Examples

In these cases there is a clear molecular target...

but it is hard apriori to specify the "right" subgroup

These are prime choices for adaptive enrichment

Do not want more than a handful of candidate features...

with strong apriori scientific relevance!

(ideally only 1 feature!)

A Rare Disease Example

Cystic Fibrosis (CF) is a genetic disease that results from dysfunction of the CFTR gene/protein

There are many different mutations in the gene that can cause various types of dysfunction (which are all termed CF)

Recent breakthrough treatments provide small-molecule replacements for certain types of dysfunction

---

These *modulators* have been extremely successful for treating certain well-understood, common mutations

Theratyping

Growing evidence that...

modulators[2] also work for some rare CF mutations

To evaluate suitability for a given mutation...

can run an in-vitro screen, and look at activity

Activity measure is continuous

How much activity is "enough"??

[2] in particular, *Trikafta*

Theratyping

Is it a good fit for adaptive enrichment?

A potentially very effective therapy, but only for a subset of people

Are there good alternatives for those pts to try?

Do we have a larger potential pool of patients than we can likely enroll/treat?

Theratyping Motivated Simulation Study

Supposing this is a good fit for adaptive enrichment...

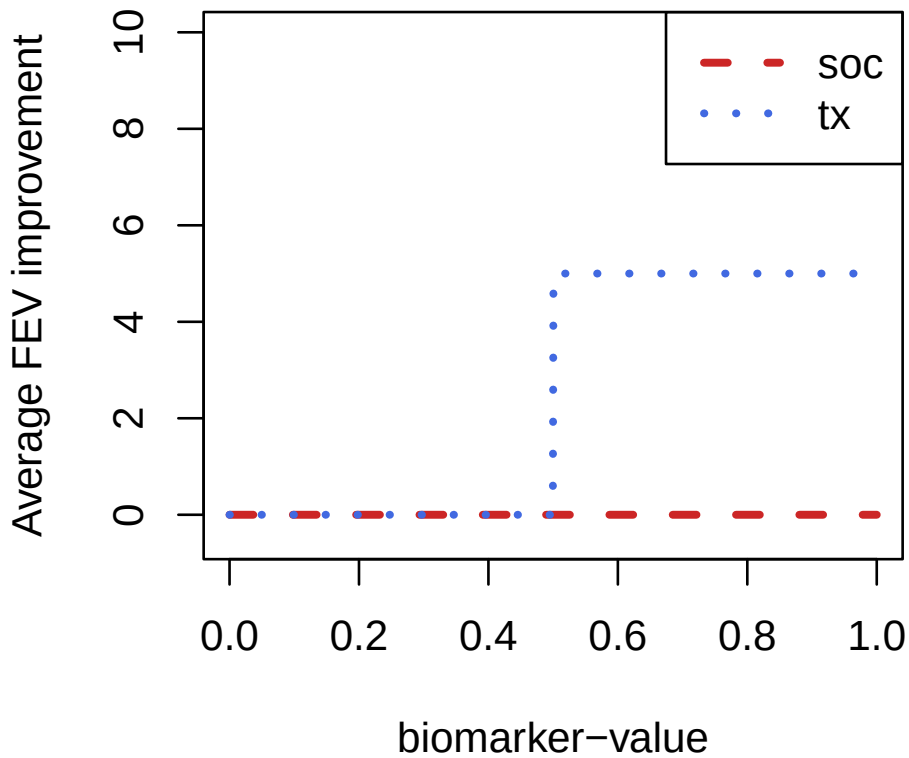
How big is the potential gain in efficiency?

Sim params roughly based on values from pivotal trial of Trikafta[3]

[3] in folks with single F508del allele

Simulation Scenario

Biomarker, x, is $U[0, 1]$, with tx effect generated as



(we vary the height and x-value of the jump)

Simulation Scenario

All trials have 60 patients (randomized 30+30 to new tx/control)

Adaptive trial is run in 2 blocks of 30

First block includes everyone

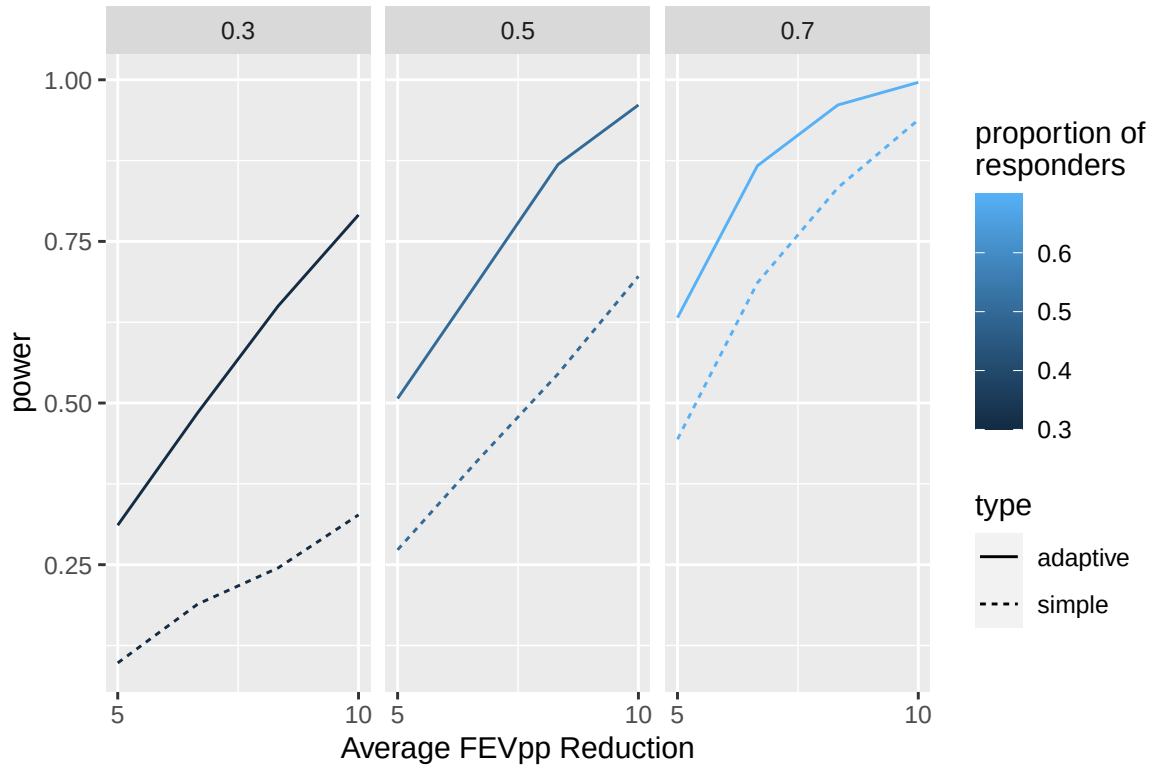
Second block restricts to subgroup $\{x \geq \hat{x}\}$

the group with most statistical evidence of improvement

Hypothesis test combines p-value from block 1 and block 2[4]

[4] There is strong control of the type 1 error, but the formal null hypothesis is a bit subtle here

Simulation Results



Takeaways

In this case, can have large improvement in power

Likely not *perfectly* identifying "optimal" subgroup...

But, have statistical evidence to justify use in ppl w/ large x!

Discussion

Don't let the perfect be the enemy of the good!

We are not testing in a formally prespecified subgroup

However! we do have strong evidence of positive effect...

In a [not perfectly characterized] sub-population.

As Yogi Berra said "Science is more of an art than a science"[5]

[5] This attribution is completely unverified, and may have been made up[6]

[5] It was definitely made up...

Discussion

Rare Disease scenario is hard!

Statistics is meant to support in decision making...

but can never provide “guarantees”...

We always have to make decisions in the absence of perfect info

Discussion: Theratyping

In the CF example...

it is possible that no clinical trial is needed at all

Trikafta is generally fantastically effective[7];

and there may not be other good options

Perhaps observational data alone could be used to evaluate efficacy

[7] It is also *fantastically* expensive, and payers might want formal evidence of benefit to cover $300k/year

Discussion: Beyond CF

CF is a relatively much more "common" rare disease

In more rare disease settings...

60 pts as in my simulations might be a pipe-dream

May not be appropriate there!

(though may be appropriate to similarly combine phase2/3 data)

Discussion: Adaptive Enrichment

There are 2 flavors of adaptive enrichment:

- The one presented here is a bit more aggressive...

Tests H_0: No subgroup (defined by x) benefits from treatment

- Alternative approach is more conservative...

Tests H_0 in pre-specified subgroups (with multiplicity correction)

Both are useful approaches! But appropriate in different scenarios

I think the more aggressive approach is more likely useful here...

- Would love to hear FDA thoughts!

Discussion: Control Arm?

Do we need a concurrent randomized control arm?

In theory, could be modified to use controls from...

- a registry/historical trial

DANGER! DANGER! DANGER!

Theory is easier in theory than in practice

Controls must be comparable to treated patients...

This has to be true as we adapt enrollment criteria

- Might accidentally just adapt to a good prognosis subpopulation

And because I couldn't resist...

Grand-pa and Grand-daughter (same hat[8])

[8] not really the same hat...

Papers

Simon N. and Simon R. Adaptive Enrichment Designs for Clinical Trials (Bisotatistics 2013).

Simon N. and Simon R.Using Bayesian modeling in frequentist adaptive enrichment designs (Biostatistics 2018).

Simon R. and Simon N. Inference for multimarker adaptive enrichment trials (Statistics in Medicine 2017).

Clinical trials in rare diseases: Should we do them differently?

Nigel Stallard

Clinical Trials Unit, Warwick Medical School, University of Warwick, UK

n.stallard@warwick.ac.uk

Acknowledgements

Innovative methodology for small populations research (InSPiRe) project[1]

This project has received funding from the European Union's Seventh Framework Programme for research, technological development and demonstration under grant agreement number FP HEALTH 2013 – 602144

International Rare Diseases Research Consortium (IRDiRC) Task force on Small Population Clinical Trials[2]

[1]Friede et al (2018) *Orphanet Journal of Rare Diseases*, 13: 186.
[2]Day et al (2018) *Orphanet Journal of Rare Diseases*, 13: 195.

What does the guidance say?

Regulation (EC) 141/2000:

> *"patients with [rare] conditions deserve the same quality, safety and efficacy in medicinal products as other patients"*
>
> *"orphan products should therefore be submitted to the normal evaluation process"*

FDA Draft Guidance on Rare Diseases:

> *"The Orphan Drug Act […] does not create a statutory standard […] different from […] common conditions"*

What is actually being done?

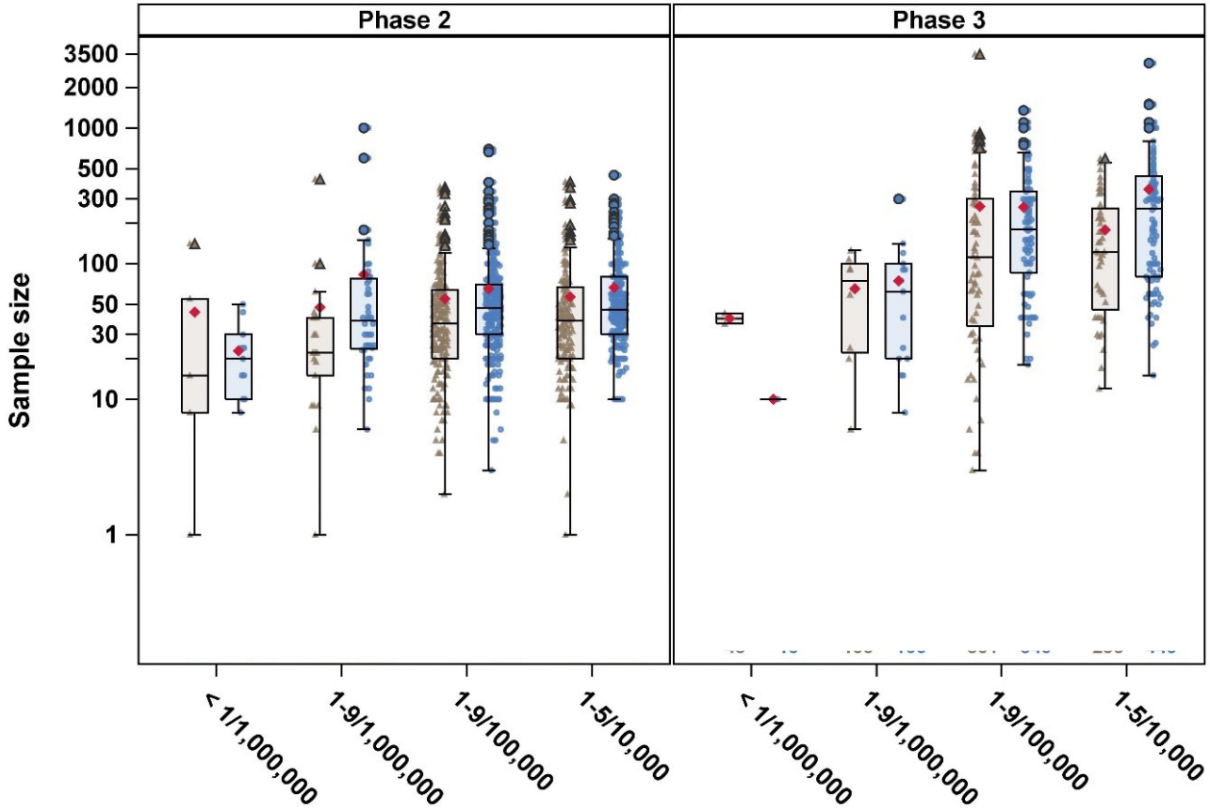
Comparing trials in non-rare and rare diseases[1]

Trial sample size	Non-rare diseases	Rare diseases
0-50	40%	67%
51-100	22%	19%
101-500	30%	13%
500+	8%	1%

[1]Bell and Tudur Smith (2014) *Orphanet Journal of Rare Diseases*, 9: 170.

What is actually being done?

For rare diseases grouped by prevalence[1]



[1]Hee et al (2017) *Orphanet Journal of Rare Diseases*, 12: 44.

What can we do differently?

Challenge of rare diseases research

Decision-making needs high-quality data

Sample sizes are necessarily limited

What might we do differently?

Possibilities include

- get more data
- get more information from same data
- consider changing level of information required

Increase data available - from inside trial

If appropriate to address clinical question:

- do not dichotomise continuous endpoints
- collect baseline covariate data
- collect longitudinal data with long-term follow-up if possible
- collect secondary endpoint data

Increase data available – from outside trial

From outside trial:

- historical controls
  - eg dynamic borrowing methods enable control data to be used when consistent with trial data[1]
  - type I error rates can be inflated[2]

- registry/EHR data
  - eg to generate synthetic controls[3]
  - eg to develop models for *in-silico* trials[4]

[1]Viele et al (2014) *Pharmaceutical Statistics*, 13: 41-54.
[2]Kopp-Schneider et al (2020) *Biometrical Journal*, 62: 361-74.
[3]Bowles et al (2022) medRXiv DOI: 10.1101/2022.12.09.22283281
[4]Musuamba et al (2021) *CPT Pharmacometrics Syst Pharmacol*, 10: 804-25.

Maximise information available from limited data

Use efficient analysis methods

Minimise sample size if possible

- group-sequential designs
- adaptive designs

Use designs that allow patients to receive multiple treatments if possible:

- cross-over, multiple n-of-1 designs

Change level of evidence required

Conventional sample size calculation:

fix type I error rate

choose n to give power to detect specified effect

Why do we control error rates?

concern about consequences of incorrect result

Value of information approach:

Model decision-making and consequences explicitly

Value of information approach to sample size determination

Trade-off between large n: good information on treatments

small n: more patients benefit from result

Future benefits depend on total population size[1,2]

Small populations: optimal α is larger, optimal n is smaller[3,4]

could formalize ad-hoc sample size choice or type I error rate inflation

[1]Cheng et al (2003) *Biometrika*, 90: 923-36.
[2]Stallard et al (2017) *Biometrical Journal*, 59: 609-25.
[3]Abrahmyan et al (2014) *J Gen Intern Med*, 29: 767-73.
[4]Pearce et al (2018) *BMC Med Res Meth*, 18: 20.

Clinical trials in rare diseases: Should we do them differently?

EMA CHMP Guideline: "*No methods exist that are relevant to small studies that are not also applicable to large studies*"

But in small populations must be more efficient, faster, smarter

Need to ensure

- all relevant information is considered
- study design and analysis is as efficient as possible
- decision-making is appropriate

CERSI
Johns Hopkins University

FDA U.S. FOOD & DRUG ADMINISTRATION

Q&A

Kelley Kidwell, PhD
Noah Simon, PhD
Nigel Stallard, MSc, PhD
Gregory Levin, PhD
Associate Director for Statistical Science and Policy
Office of Biostatistics, Center for Drug Evaluation and Research, FDA

CERSI
Johns Hopkins University

FDA U.S. FOOD & DRUG ADMINISTRATION

Break

Upcoming Virtual FDA Workshop

FDA's CDER, CBER, and Duke-Margolis Center for Health Policy Host
Rare Disease Endpoint Advancement Pilot Program Workshop: Novel Endpoints for Rare Disease Drug Development

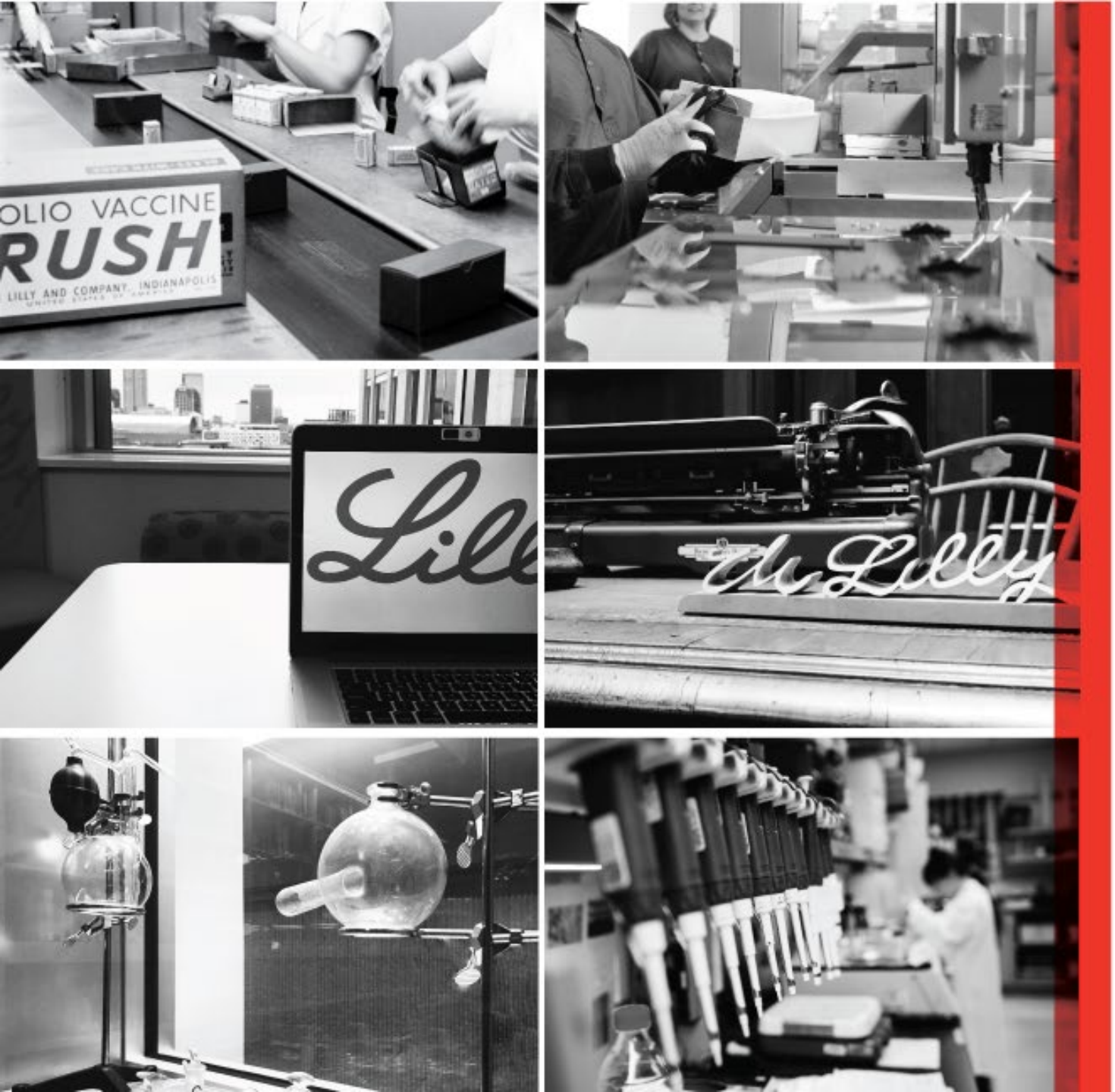
June 7 and 8, 2023; 1-5 pm
Link in the Chat

CERSI
Johns Hopkins University

FDA U.S. FOOD & DRUG ADMINISTRATION

Session 2:
Analysis Methods in Small Populations

Moderator: Michael Rosenblum, PhD
Professor of Biostatistics
Johns Hopkins Bloomberg School of Public Health



Bayesian Approaches and Master Protocols in Rare Disease Drug Development

Karen L. Price, PhD
Associate Vice President, Eli Lilly & Co

03 May 2023

Acknowledgements

- Forrest Williamson, PhD
- Zach Thomas, PhD
- Michael Sonksen, PhD
- Richard Payne, PhD
- Will Landau, PhD

Outline

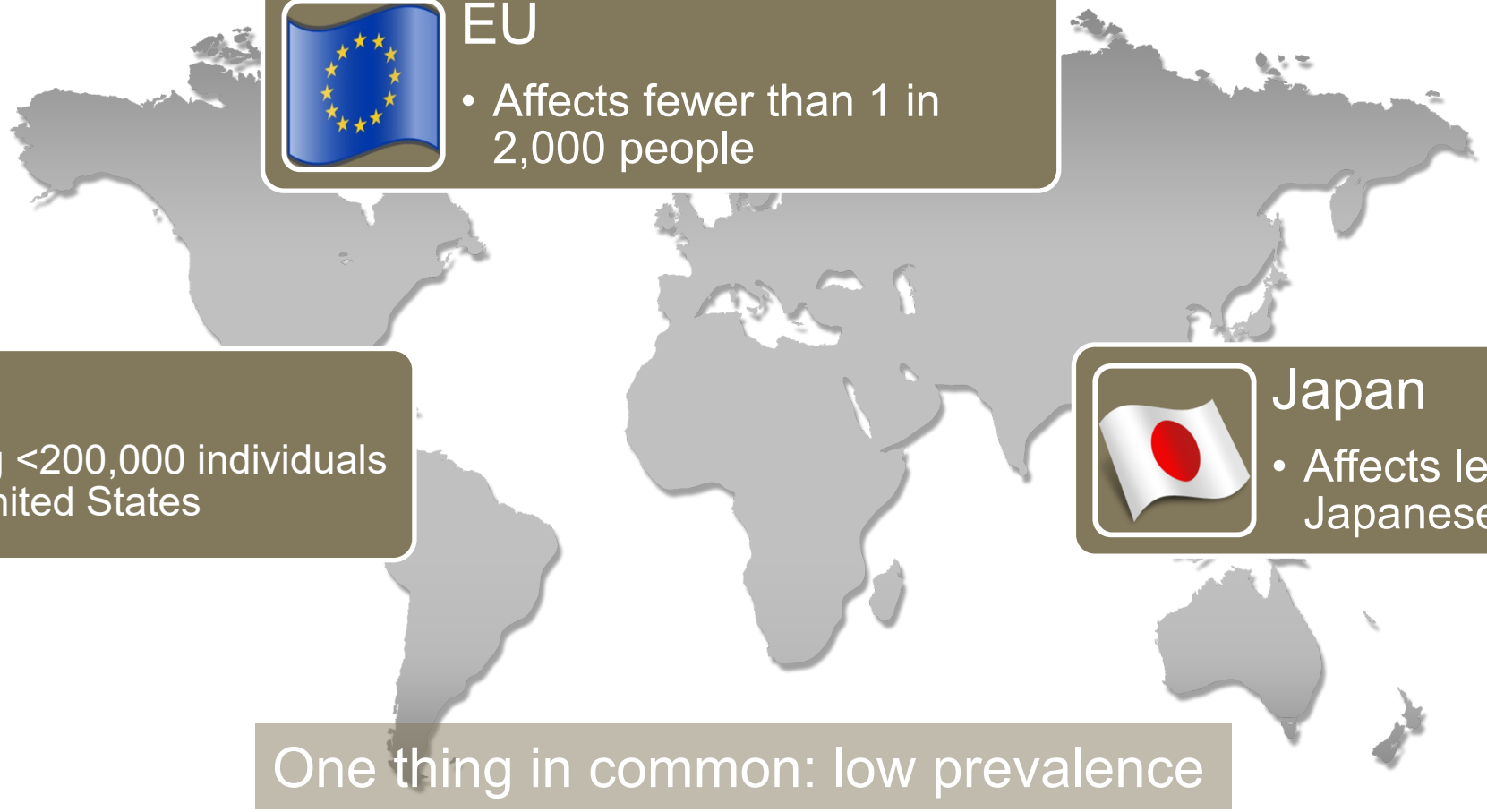
Overview of rare & pediatric diseases

The Bayesian Framework

Examples of Bayesian Application

Master Protocols

What Makes a Disease Rare?



EU

- Affects fewer than 1 in 2,000 people

USA

- Affecting <200,000 individuals in the United States

Japan

- Affects less than 50,000 Japanese people

One thing in common: low prevalence

Rare Diseases in Children

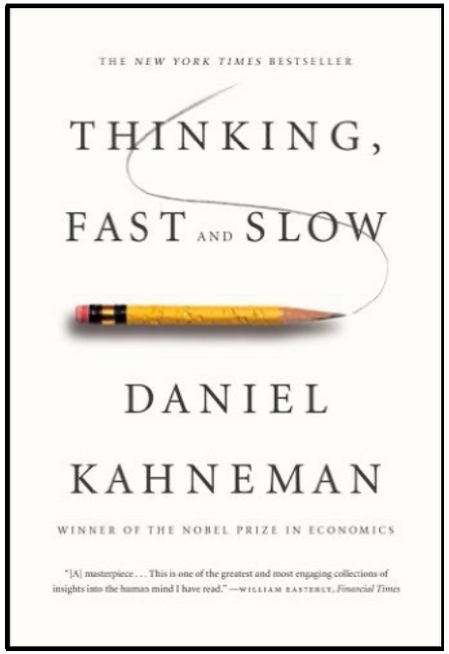
- Rare diseases affect approximately 30 million Americans
  - 20 million of those are children
  - <1% of diseases have FDA approved treatment
  - Numbers are higher in Europe, with similar number of treatments available
- 50%-75% of all rare diseases begin in childhood

THE BAYESIAN FRAMEWORK

Importance of Bayesian Thinking

Humans Struggle with Prediction and Uncertainty

April 2013

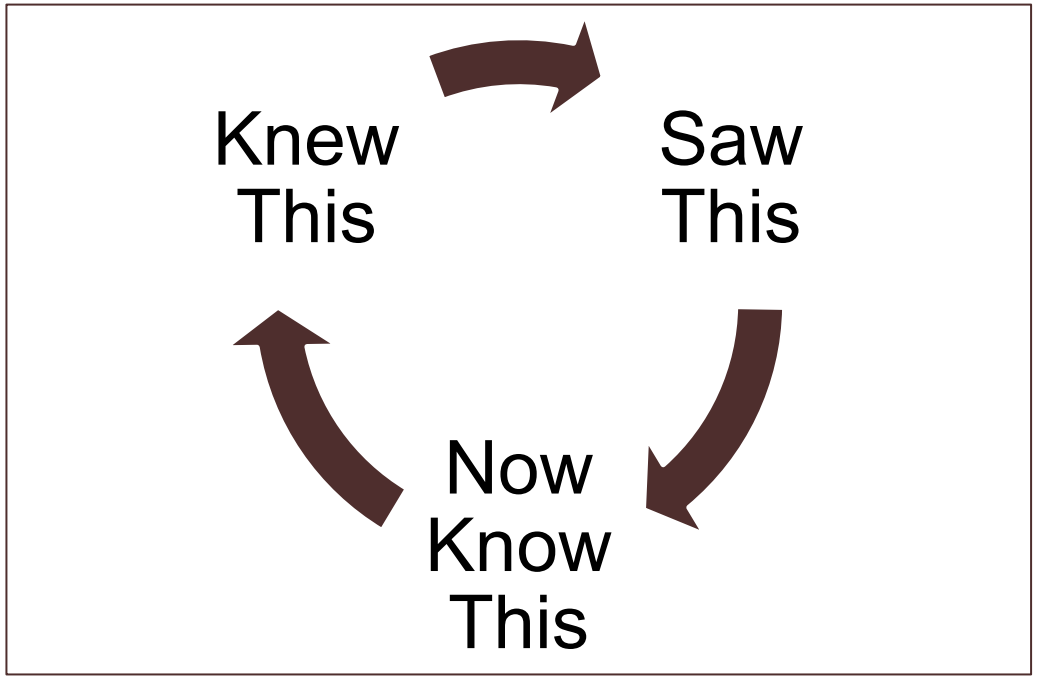


Daniel Kahneman, 2002 Nobel Prize

"The instinctual shortcut that we take when we have 'too much information' is to engage with it selectively, picking out the parts we like and ignoring the remainder..." Nate Lewis

"Our subjective judgments were biased: we were far too willing to believe research findings based on inadequate evidence …" Daniel Kahneman

Learning Requires Formal Process With Regular Updating and Synthesis of Data (i.e., Bayes)



Why Bayes: A working philosophy

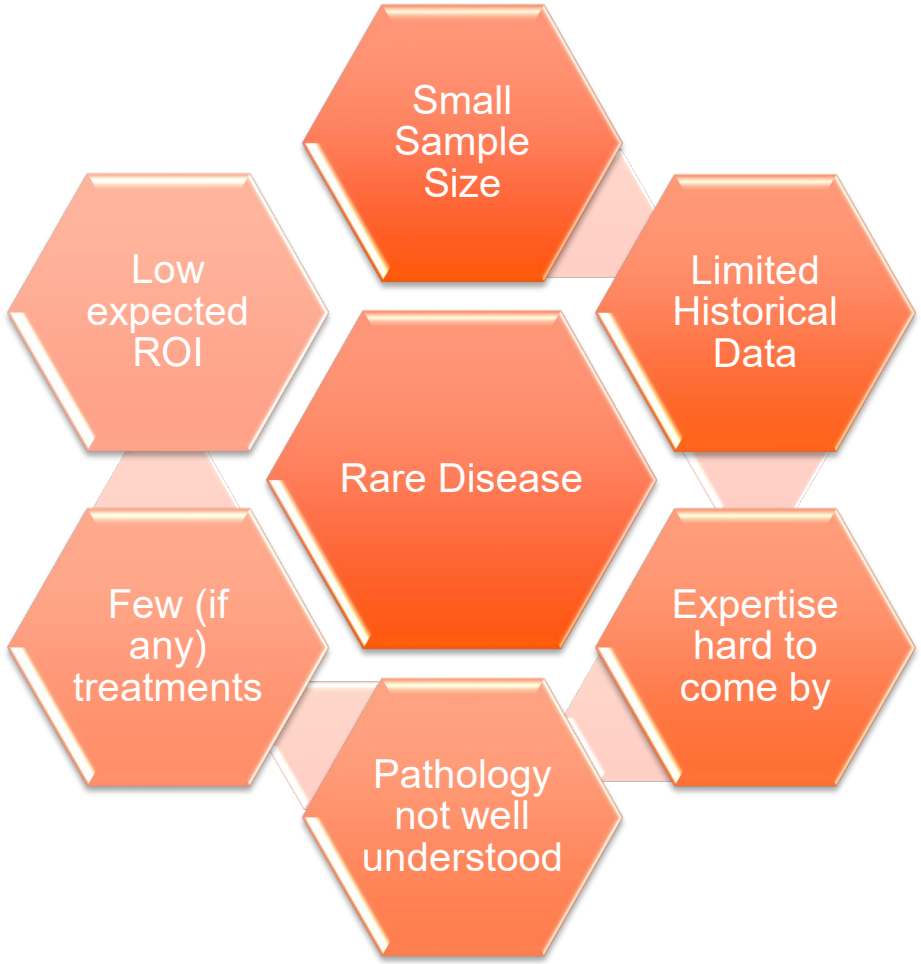
- Bayesian methods lend themselves well to iterative updating of *the science*
  - *'Today's posterior is tomorrow's prior'*

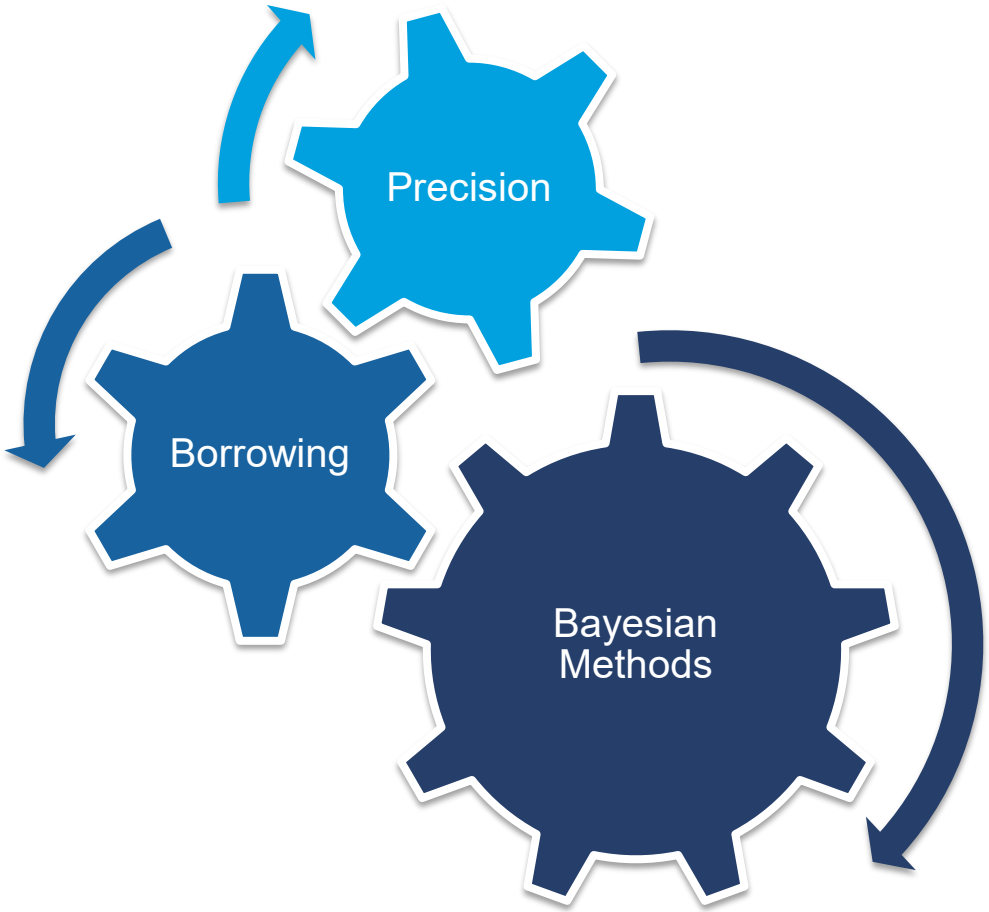
  Dennis Lindley, 1972
  - *"When the facts change, I change my mind. What do you do, sir?"*

  attributed to economist John Maynard Keynes
- Bayes facilitate rigorous integration of what we **know already** (i.e. via informative priors) within analyses of new data designed to shed light on what we **don't know**
  - Strives for **transparent** integration of data from diverse sources to inform decision-making
- Allows straightforward statements of probability and uncertainty
- Bayesian design can reduce sample size/study duration
- Flexible hierarchical modeling with computational conveniences

Rev. Thomas Bayes?

Challenges Necessitate Innovation





BAYESIAN APPLICATION

Borrowing Approaches

- Borrowing can be on control arm and/or treatment arm(s)
- Static vs Dynamic
  - Static
    - Pooling
    - Single arm trials
    - Power priors
  - Dynamic
    - Hierarchical modeling
    - Mixture priors
    - Commensurate priors
- Static vs dynamic can differ for control/treatment

Appeal of dynamic borrowing:
- Borrows more when current data are similar to historical data
- Protects against over-borrowing

See, e.g., Viele, et al., 2014.

Example Potential Data Sources

- Expert/caregiver opinion
- Natural history studies
- Summary level data (RCTs, observational)
- Individual-level patient data
  - Internal to Sponsor or at FDA (or other regulators)
  - Patient registries
  - Observational studies
- PK/PD modeling
- Pre-clinical data

Need to assess relevance of historical data to new data: similar indications, patient population, time since data collection, relevance of endpoints, timepoints, etc. (exchangeability)

Role of Opinion

- Large literature on this topic
- Elicit distributions of belief about key efficacy / safety endpoints
  - There are formal, well-tested protocols
  - May be used as portion of prior or down-weighted
- Elicit distributions about belief in relationships between endpoints, doses, populations, etc.
- Can use to inform about relevance of historical information
- Examples available (see, e.g., MYPAN)

General Comments about Borrowing

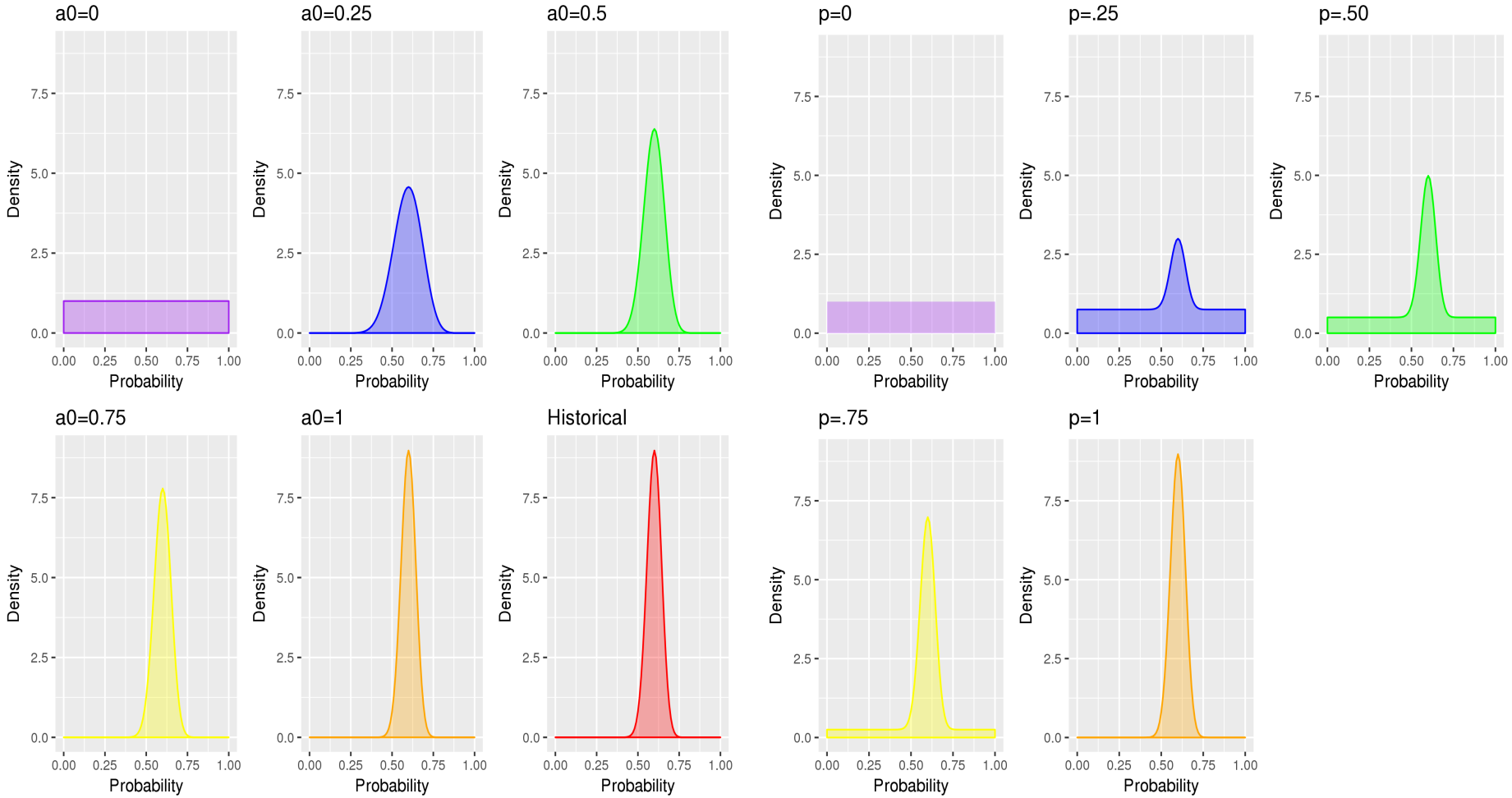
- How much to borrow?
  - ✓ What data is eligible to be included in the prior
  - ✓ Currently need to simulate operating characteristics
  - ✓ Consider “prior effective sample size” and “prior probability of success”
  - ✓ Should assess prior to posterior sensitivity
- May borrow different amounts for different treatments, based on medical need, etc.
- Note, borrowing may ‘dampen’ the effect in current trial (so borrowing does not always favor Sponsor)

Suggestions available in CDRH/CBER Bayesian Guidance document

Hypothetical Example: Borrowing historical control

- Previous data is available on the control group.
  - Specifically, a trial with 120 subjects and 72 responses.
  - Thus the historical rate is 60%.
- This historical information is kept constant throughout the simulation.
- The sample sizes for the current study are 70 for the controls and 140 for the new treatment.

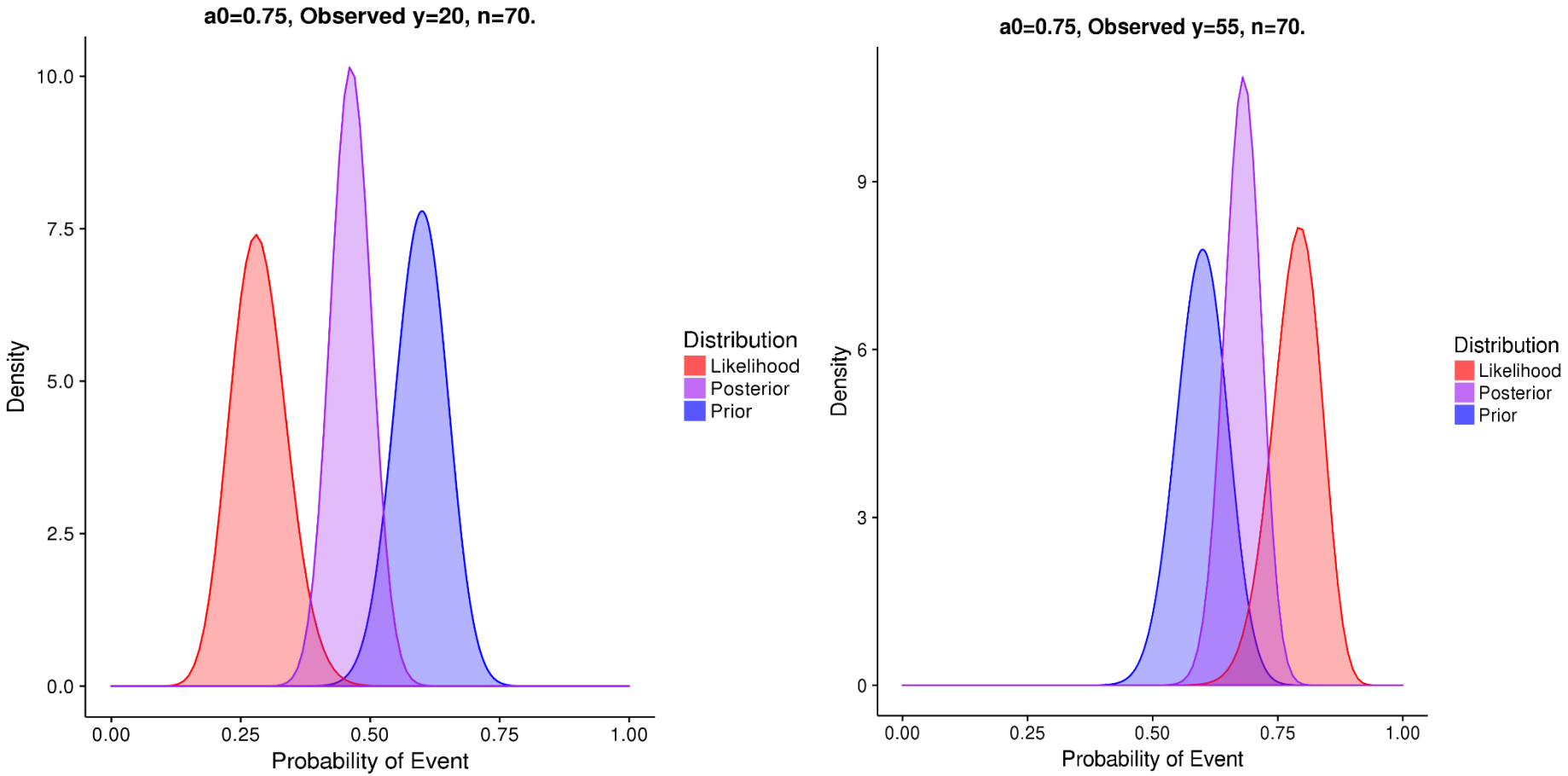
Hypothetical Example: Power Prior vs Mixture Priors



Power prior with various α_0 values

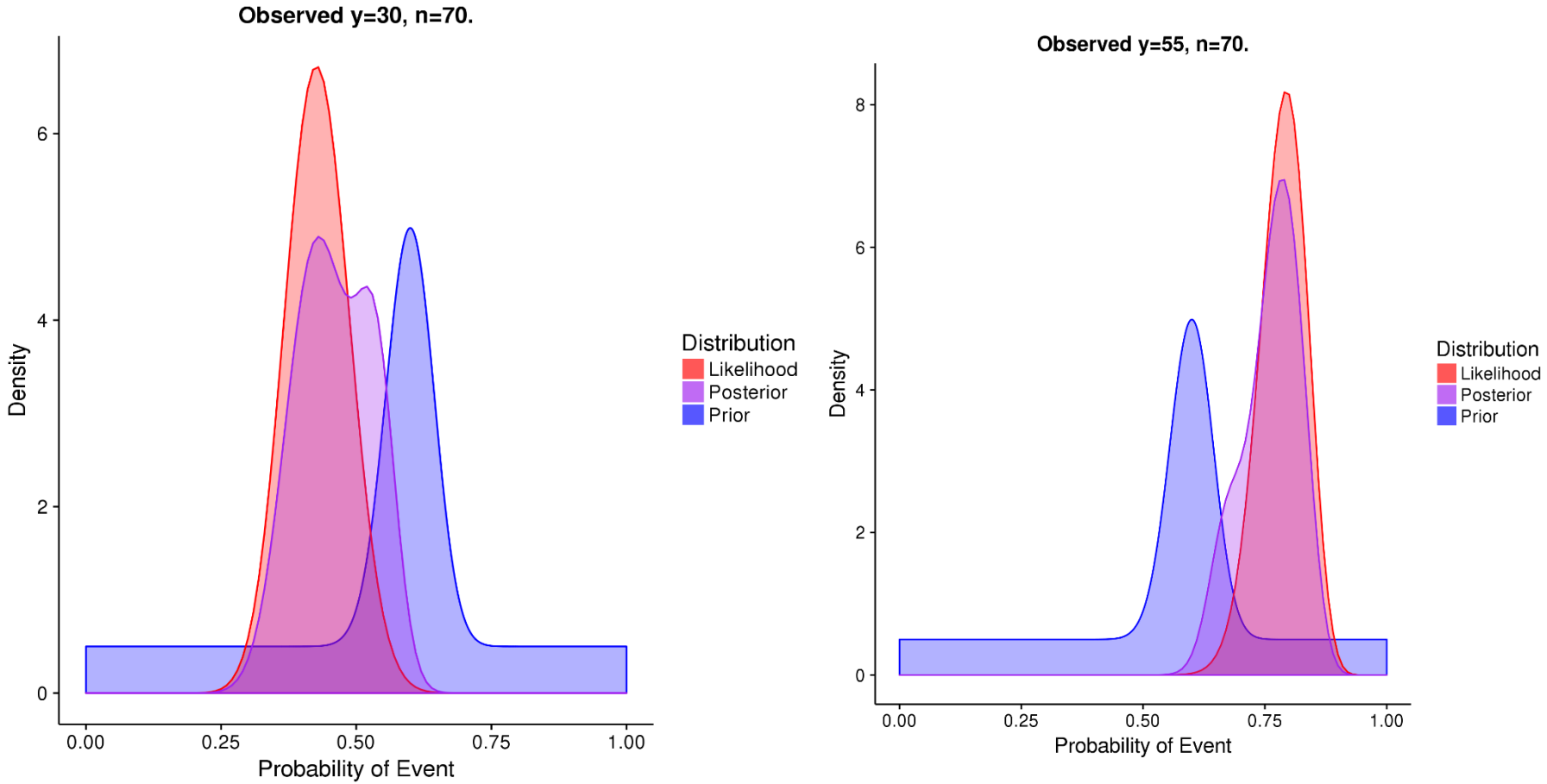
Mixture priors with beta(72, 48) and beta(1,1) at various mixing proportions

Impact of Borrowing on Results



Plots of example posterior distributions for control arm, based on different trial outcomes, using power prior (α_0 = .75)

Impact of Borrowing on Results, cont.



Plots of example posterior distributions for control arm, based on different trial outcomes, using mixture prior (p = .5)

Threshold-crossing approach

Rationale

Open-label study

May include active “reference” arm

- Not powered to test treatment vs. reference
- Only subset of participants are randomized

Control not feasible or unethical

Historical evidence used to establish what a *meaningful response* would be

Ex. Bayesian Decision Rule

Critical Success Factor (CSF): The posterior probability of the treatment response rate exceeding 57% will be calculated, and the study objective will be successfully met (i.e. positive study) if this probability is at least 80%.

- Parameter of interest is the posterior mean response of treated participants at week 24
- 57% is the effect of interest
- 80% is the posterior probability threshold

CSF: Pr(treatment response rate > 0.57) ≥ 80%

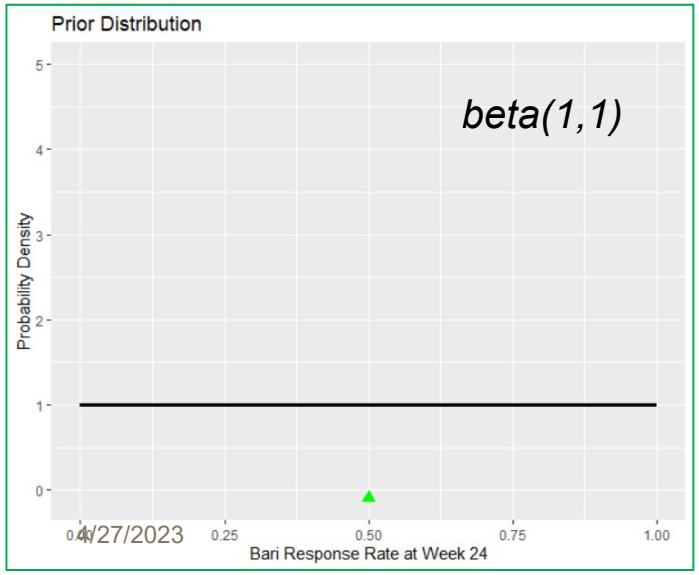
Note: the CSF is only on the novel treatment arm; if there is an active-control arm in the study it is not being used in the primary analysis.

Bayesian Critical Success Factor

Imagine a new study that includes n=30 patients assigned (randomized) to a novel treatment for Juvenile Idiopathic Arthritis.

Primary endpoint is treatment response rate at the end of a treatment period (ex: 24 weeks).

- Note: primary is only on the treatment arm, not treatment vs. control/reference

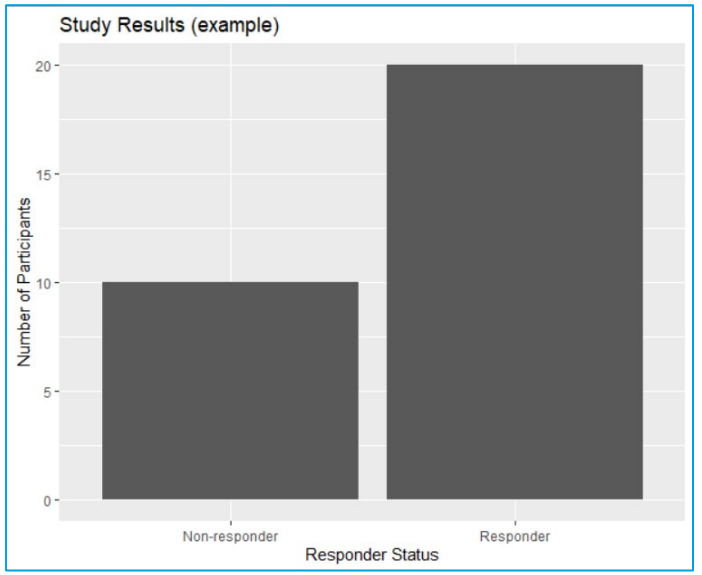


A priori, we believe the response rate can be between 0 and 1 with all values equally likely.

Prior mean response rate = 0.5

Imagine the study is conducted and we observe 20 responders out of 30 on treatment at week 24.

Observed mean response rate = 20/30 = 0.6667



Parameter of interest is mean treatment response rate at week 24

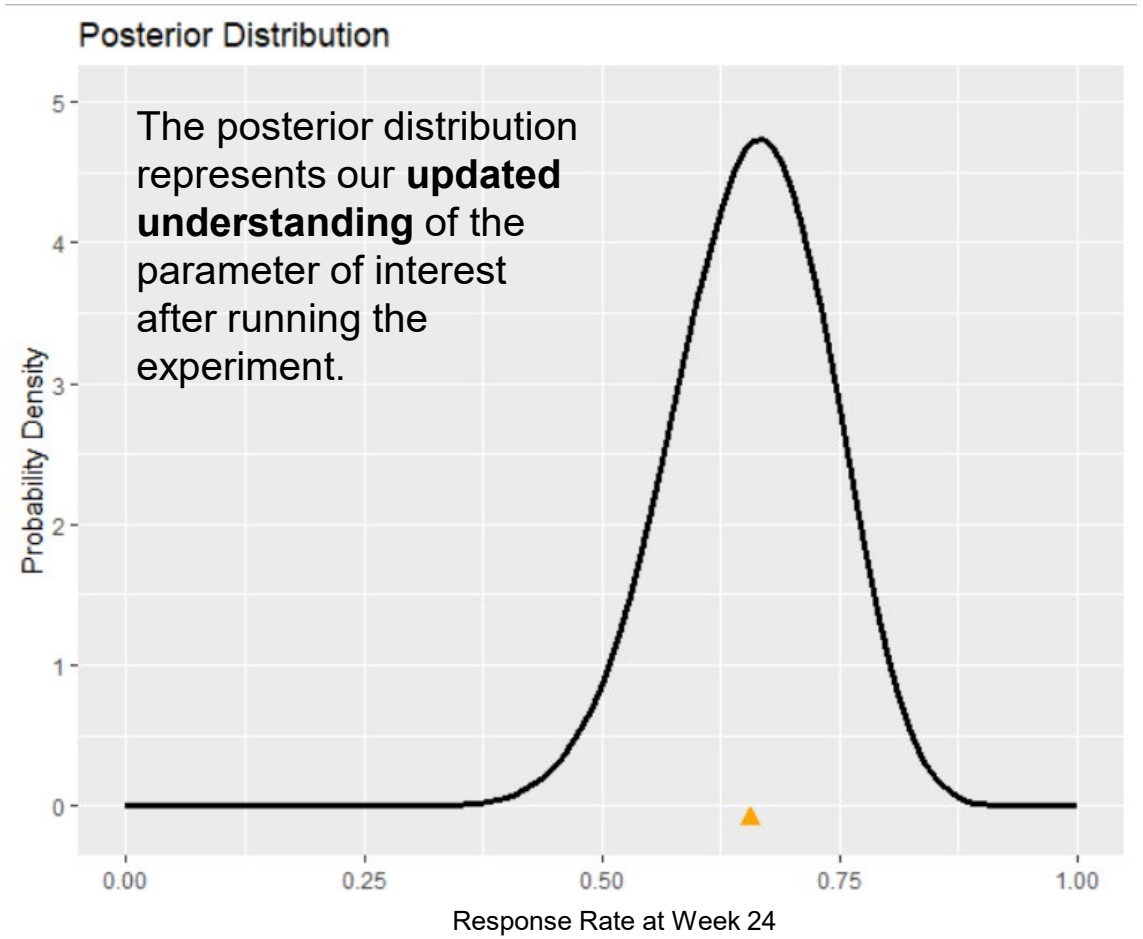
Obtain the posterior distribution

- $beta(21,11)$

Locate the Effect of Interest on the Posterior

Compute the Posterior Probability

Apply the Decision Rule



Posterior mean response rate = 0.6562

Locate the Effect of Interest on the Posterior

Obtain the posterior distribution

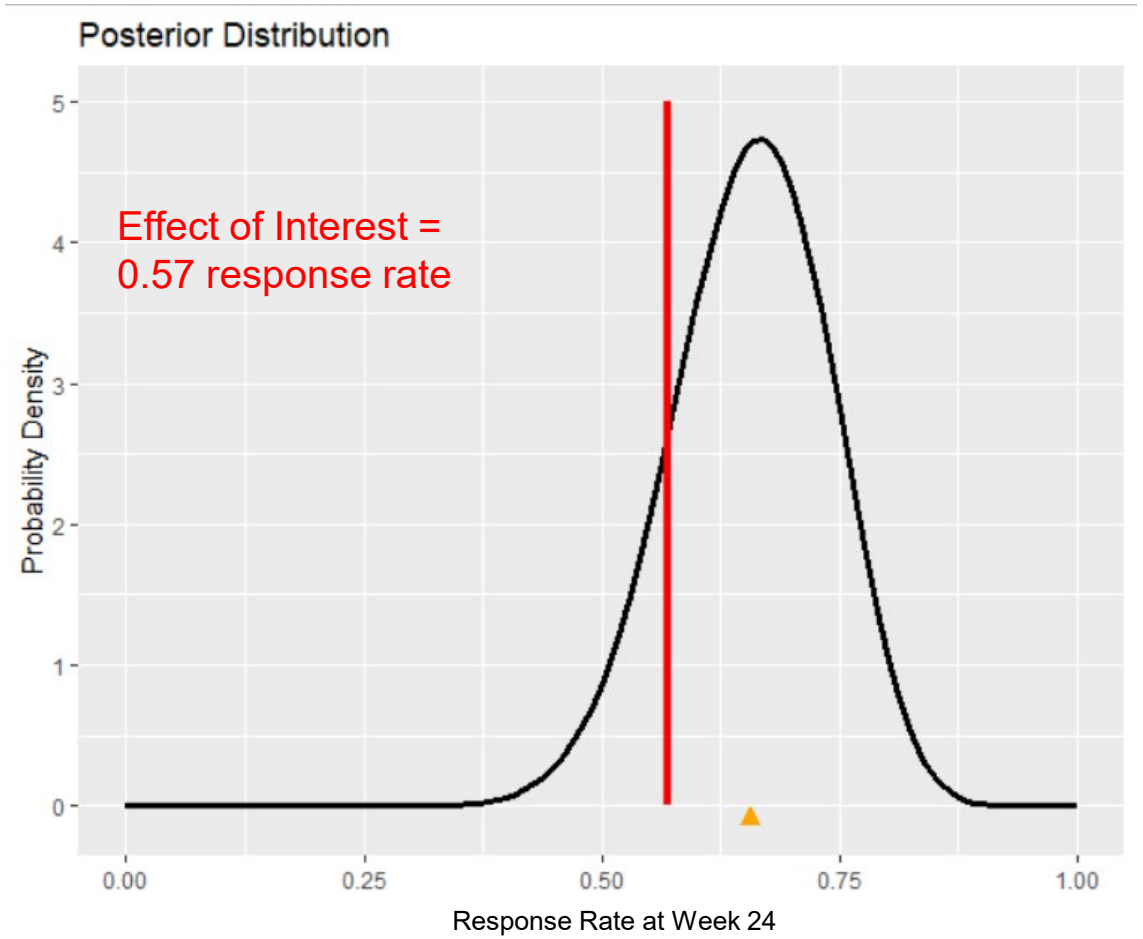
- $beta(21,11)$

Locate the Effect of Interest on the Posterior

- EOI = 0.57

Compute the Posterior Probability

Apply the Decision Rule



Posterior mean response rate = 0.6562

Compute the Posterior Probability

Obtain the posterior distribution

- $beta(21,11)$

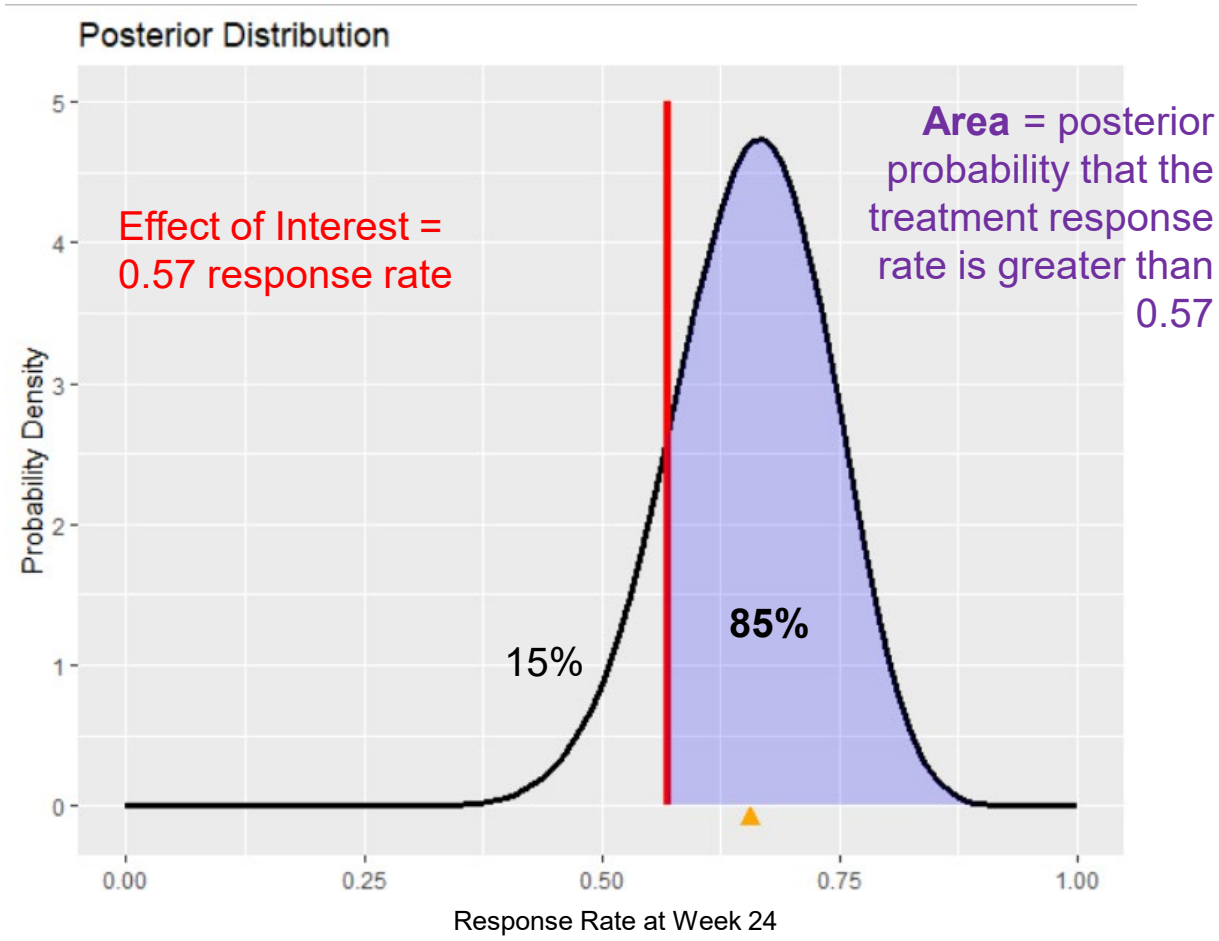
Locate the Effect of Interest on the Posterior

- EOI = 0.57

Compute the Posterior Probability

- Pr(treatment > 0.57) = 85%

Apply the Decision Rule



Posterior mean response rate = 0.6562

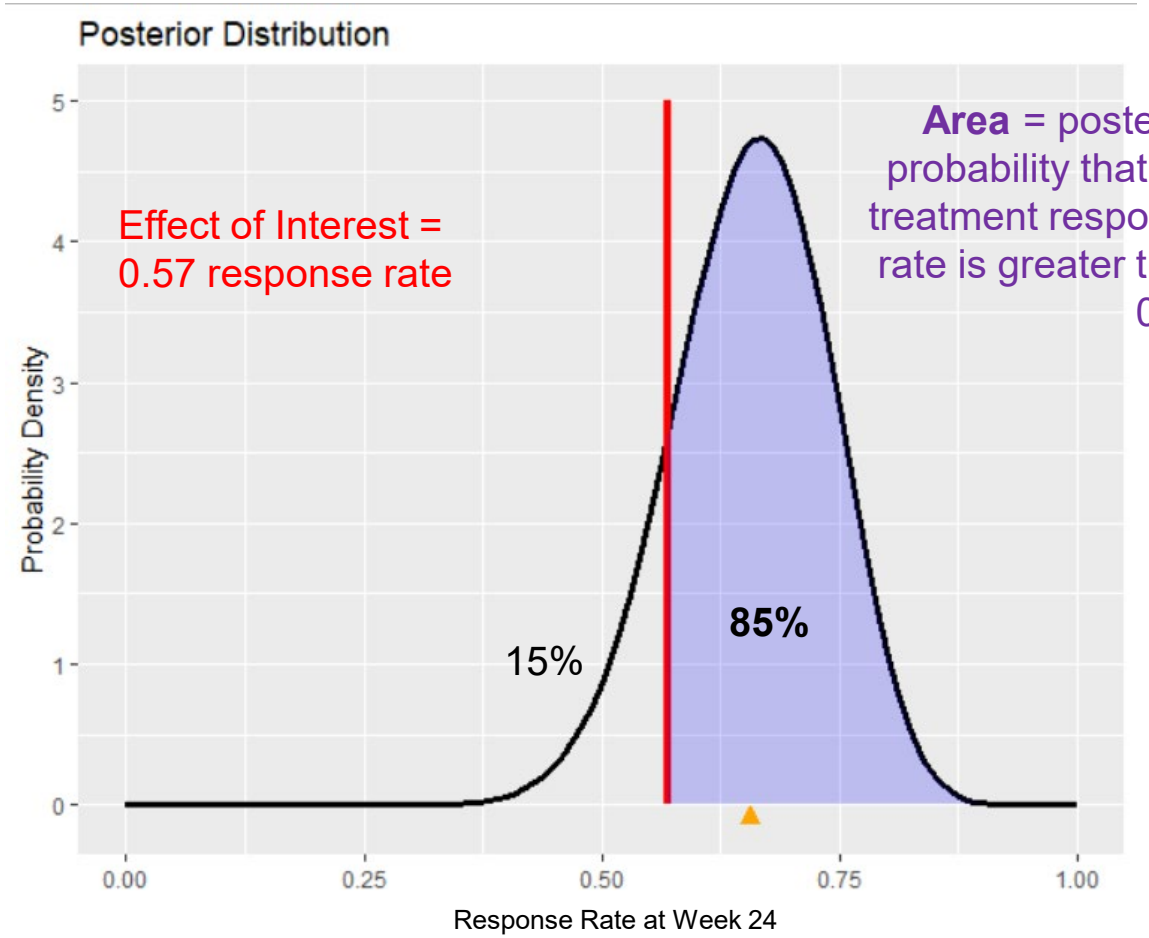
Apply the Decision Rule

Obtain the posterior distribution
- $beta(21,11)$

Locate the Effect of Interest on the Posterior
- EOI = 0.57

Compute the Posterior Probability
- Pr(treatment > 0.57) = 85%

Apply the Decision Rule
- Posterior probability is greater than 80%, so the **success criterion has been met**



Posterior mean response rate = 0.6562

Bayesian Methods Enable Continual Learning

- Consider a trial with 4 interims (n = 20, 50, 75, 90) with a final sample size of 100.
- The trial will be a success at the end of the trial if Pr(response > 0.5) > 0.96
- Should the trial stop early for futility at the following interims?

Probability of Being Successful at N = 100

Interim N	Observed Response	Predictive Probability
20	12 / 20 = 0.60	0.54
50	28 / 50 = 0.56	0.30
75	41 / 75 = 0.55	0.09
90	49 / 90 = 0.54	<.01

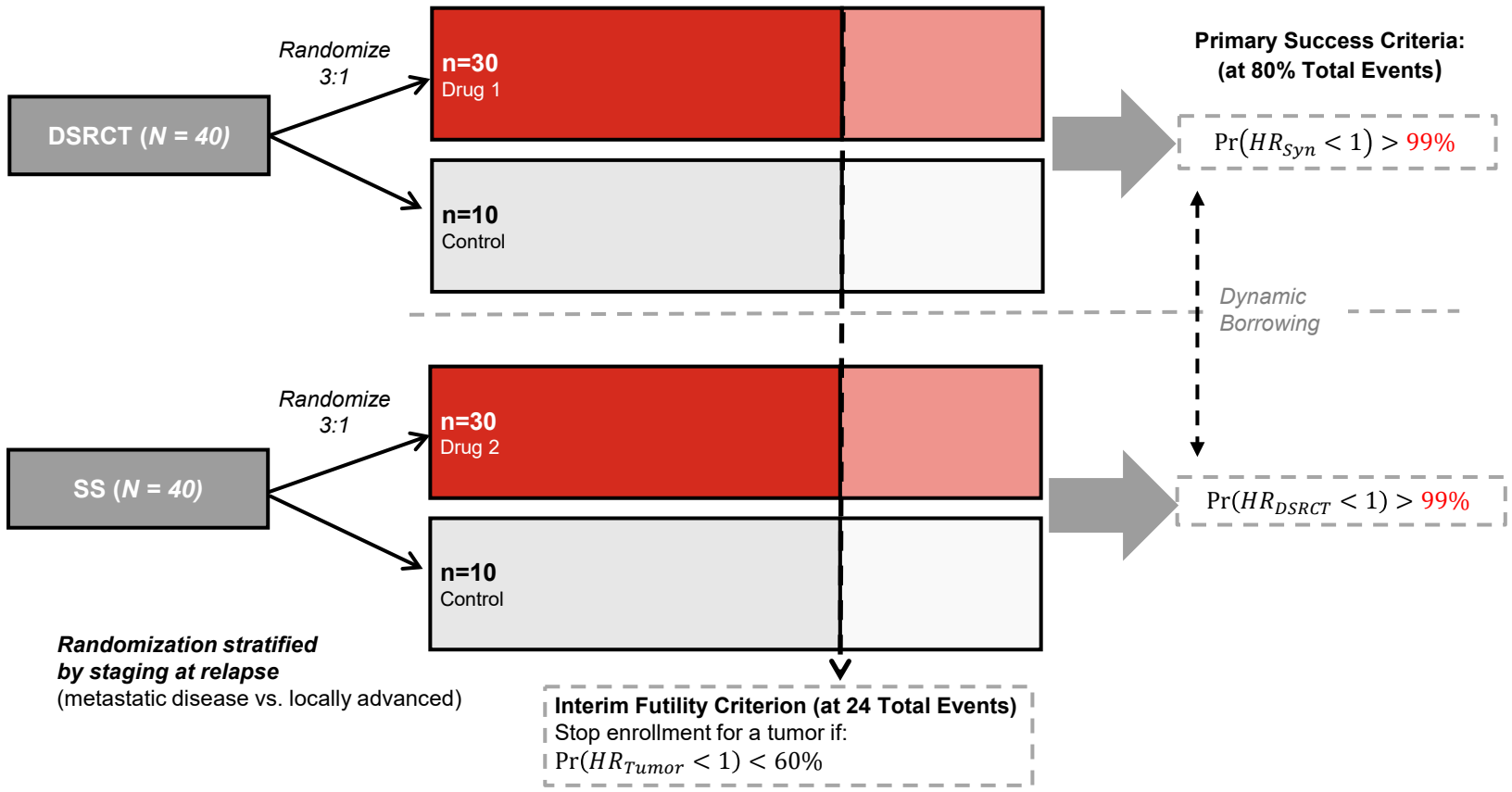
Reference: Saville, B. R., Connor, J. T., Ayers, G. D., & Alvarez, J. (2014). The utility of Bayesian predictive probabilities for interim monitoring of clinical trials. *Clinical Trials*, *11*(4), 485-493.

MASTER PROTOCOLS

Basket-type trial in oncology

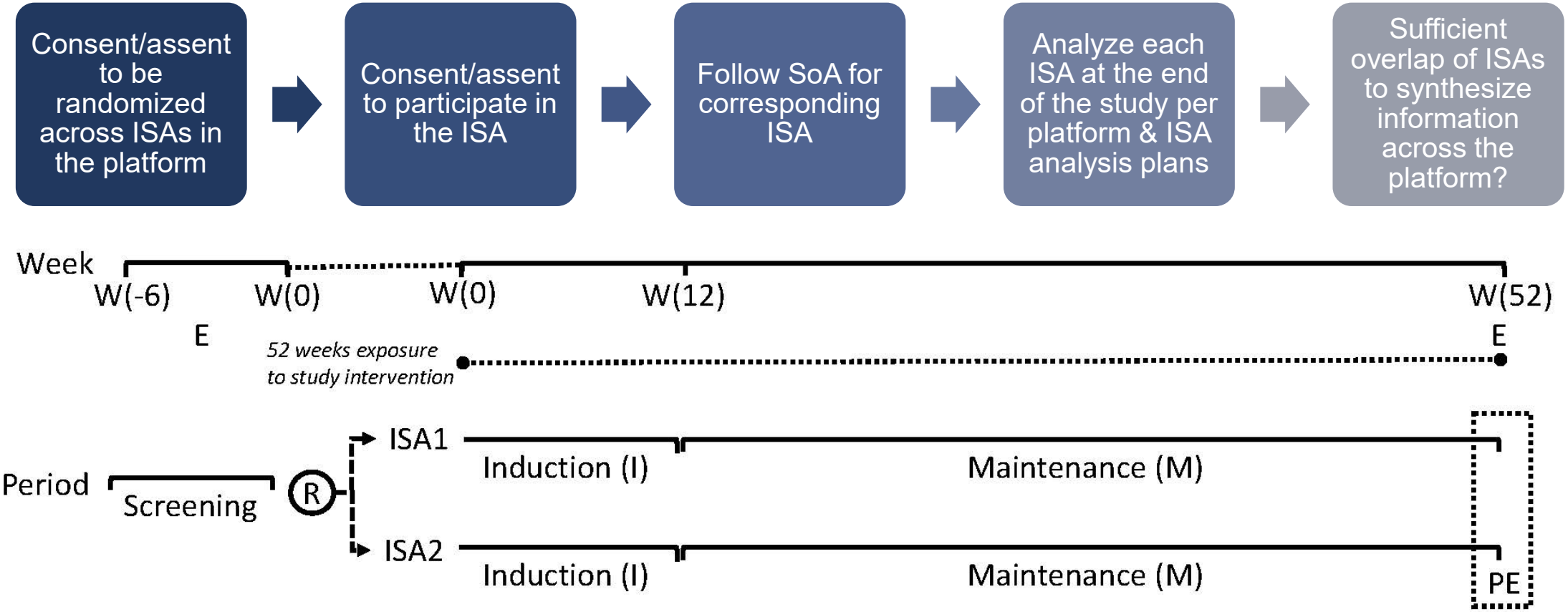
Bayesian Analysis of Primary Endpoint of Progression-Free Survival (PFS)

- Joint Bayesian hierarchical model will be fit to the PFS data from both tumors
- **Likelihood:** parametric (Weibull) survival model with proportional hazards assumption
- **Priors:**
  1) Control arm PFS: **power priors** constructed from propensity-**matched** (individual patient) **real world database** created for this study
  2) Random-effects meta-analytic prior on the two PFS HRs
- 'Dynamic' borrowing on effect-size across tumor
- Individual conclusions for each tumor



DSRCT: Desmoplastic small round cell tumor; SS: Synovial sarcoma

Platform type trial in pediatric IBD



Abbreviations: E=endoscopy; ISA=intervention-specific appendix; PE=primary endpoint; R=randomization between open ISAs; W=Weeks.

Discussion

- Bayesian design and analysis can facilitate rigorous incorporation of relevant scientific context/data in settings of potentially limited sample size
- Incorporation of this context is prespecified by transparent model/prior assumptions and studied via simulation at the design stage
- Can result in increase in power while maintaining low type 1 error
- Bayesian methods enable efficient continual learning
- Master protocols (and other innovative designs) enhance learning
- Collaboration between sponsor and regulatory statisticians/others is critical
- We need **experience** with these designs to continue advancement of statistical methods and operational elements

Bayesian Adaptive Designs and Information Borrowing for Efficient and Accurate Statistical Inference in Rare Diseases

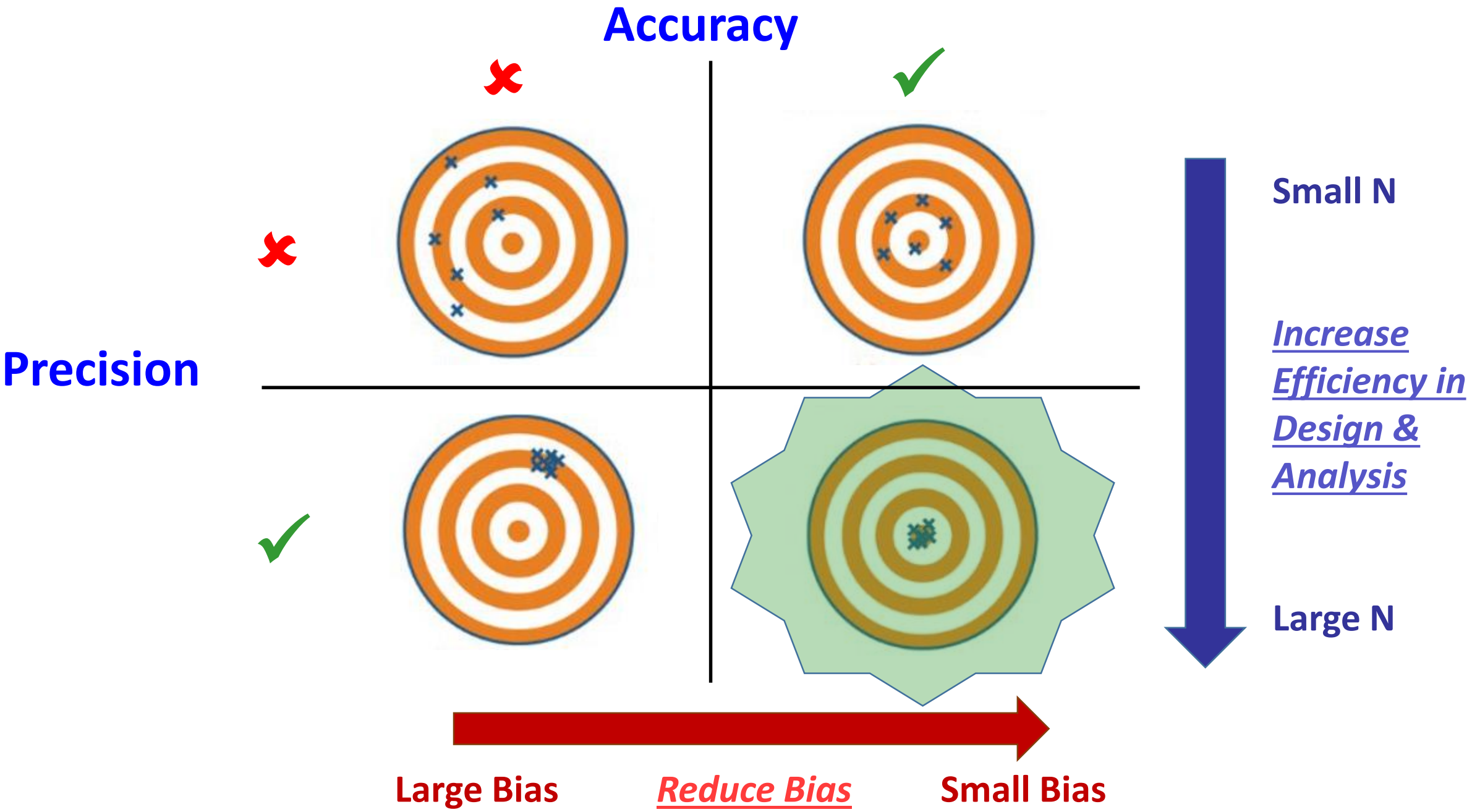
J. Jack Lee, Ph.D., D.D.S.
Professor of Biostatistics

THE UNIVERSITY OF TEXAS
MD Anderson Cancer Center
Making Cancer History®

Outline

- Statistical Challenges and Solutions in Drug Development for Rare Diseases
- Bayesian Statistical Inference
- Clinical Trial Design and Analysis Considerations
  - Bayesian adaptive designs
    - Model-Assisted Designs
  - Adaptive platform designs
  - Bayesian hierarchical models for basket trials
- Concluding Remarks

Goals for Statistical Inference: Accurate and Precise



Statistical Challenges and Solutions in Drug Development for Rare Diseases

- Randomized controlled trials are gold standard.
  - Required large N. Not feasible.

→ **Novel Adaptive Designs**
- **Take all comers**
- **Adaptive randomization**
- **Frequent interim analysis**
- **Easy enroll and conduct**

- Single-arm trials are subject to bias.
  - No comparators. Hard to make a robust inference.

→ **Borrow Information**
- **Concurrent controls**
- **Historical controls**

- Disease registries and EMR data are available.
  - Large N. Heterogeneous groups with mixed data quality.
  - Real-world data

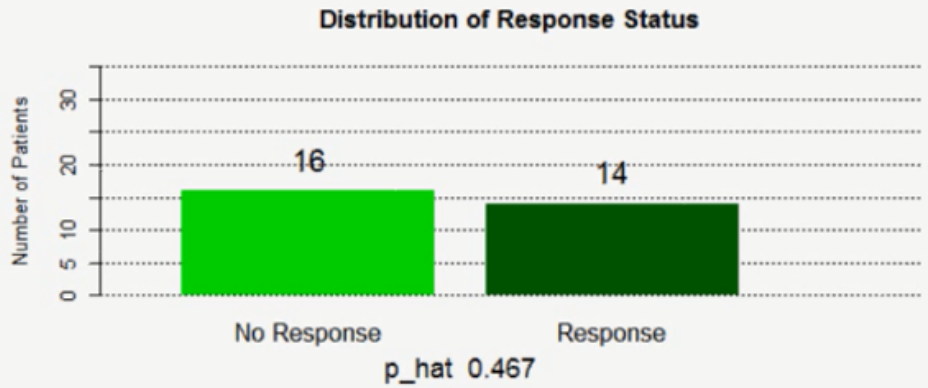
→ **Combine Information**
- **Synthetic controls**
- **Propensity score matching**
- **Network meta-analysis**

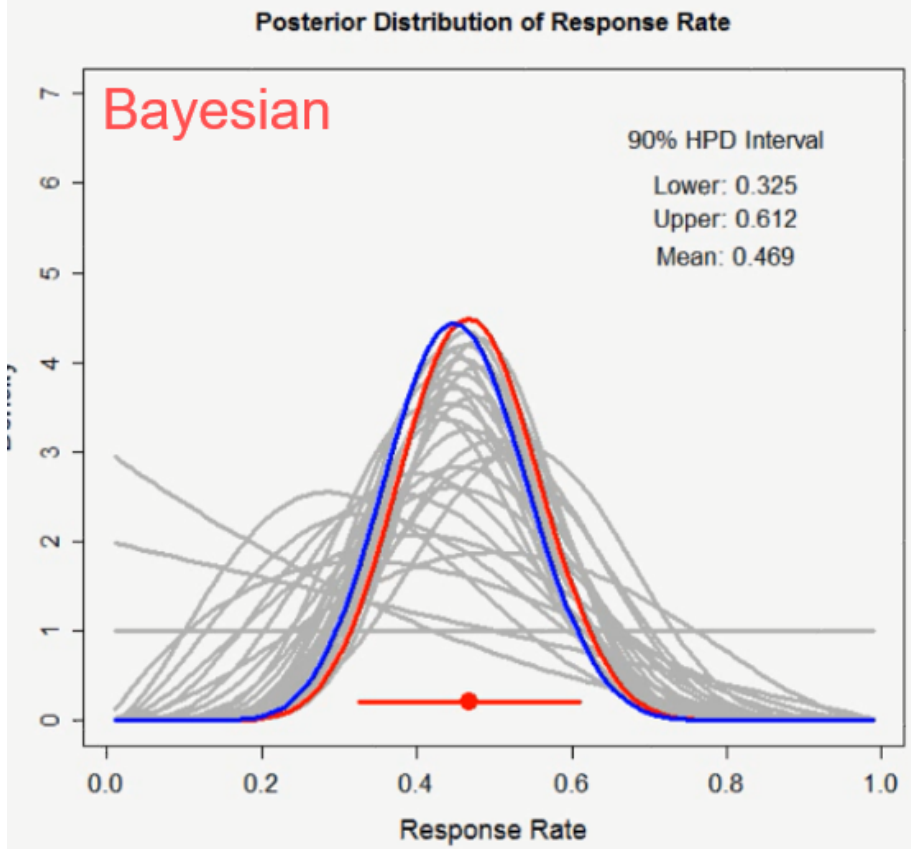
Estimate the Response Rate of A New Drug

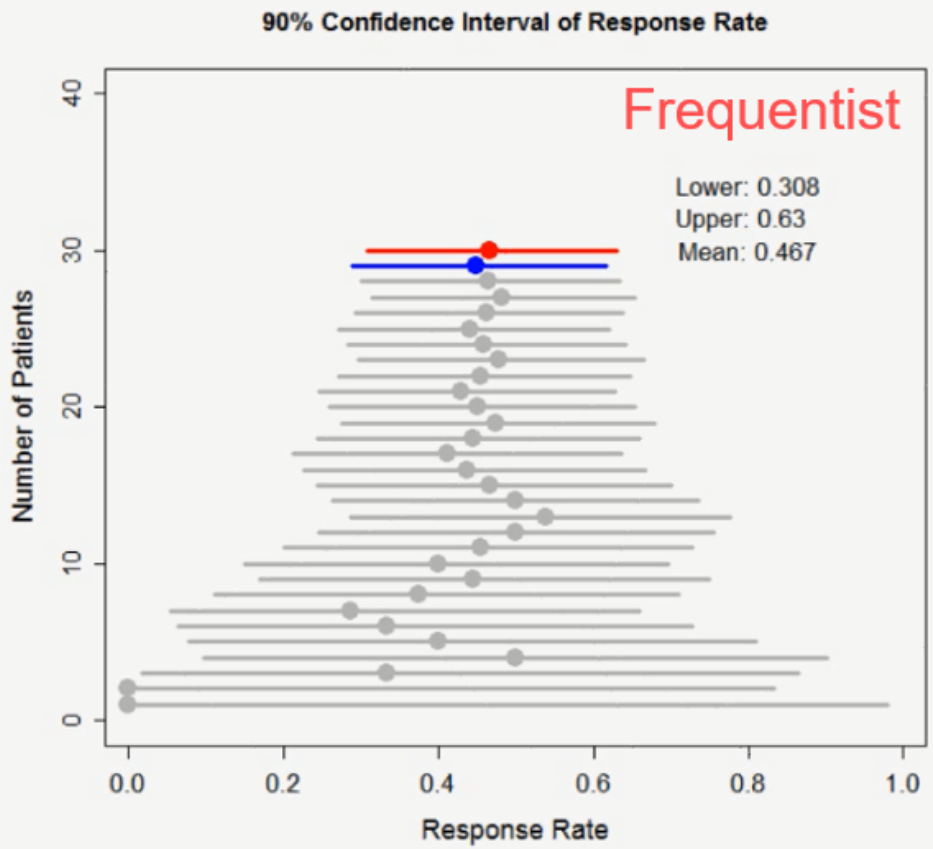
Response Status

10 :	0	0	1	1	0	0	0	1	1	0
20 :	1	1	1	0	0	0	0	1	1	0
30 :	0	1	1	0	0	1	1	0	0	1

Estimated Response Rate 0.467

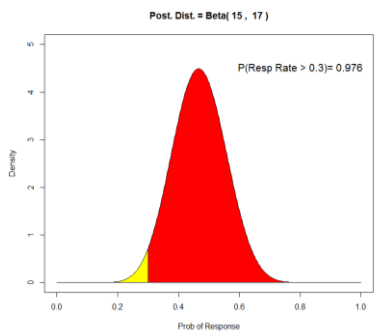


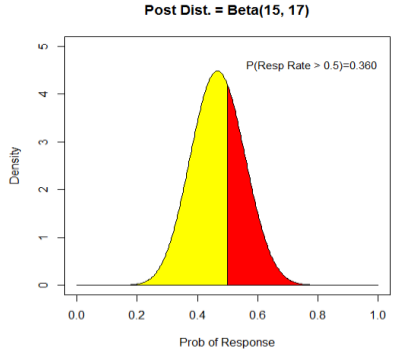


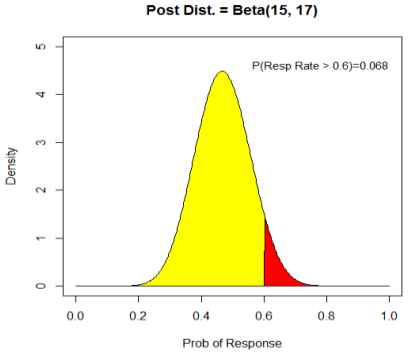


Bayesian Posterior Probability

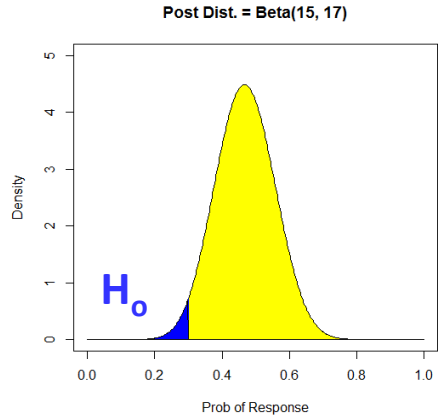
- All information pertinent to the parameter of interest is contained in the posterior distribution: θ ~ Beta(15, 17)
- What is the Probability($\theta > 0.3$)? 0.976
- What is the Probability($\theta > 0.5$)? 0.360
- What is the Probability($\theta > 0.6$)? 0.068

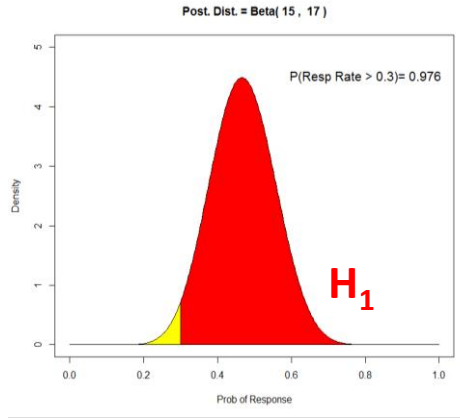






- Testing $H_o : p \leq 0.3$ versus $H_1 : p > 0.3$
- Prior(H_1) = 0.7, Prior(H_o) = 0.3, Prior $Odds_{10}$= 0.7/0.3= 2.33
- Post(H_1|Data)=0.976, Post(H_o|Data)=0.024, Post $Odds_{10}$= 0.976/0.024=40.67
- Bayes $Factor_{10}$ = 17.4. Odds of H_1 vs. H_o true is 17 times stronger compared to the prior odds





Bayesian Paradigm – A Superior Way for Making Statistical Inference

- Advantages of Bayesian Method. It can
  - Model unknown parameters with statistical distributions.
  - Conform to the likelihood principle.
  - Properly address various levels of uncertainty.
  - Use all available information – prior, current, (future); within and outside of the trial via dynamic borrowing to increase efficiency.
  - Allow more frequent monitoring and decision making.
  - Incorporate subjective utility in decision making.
- Be aware
  - Are data and model compatible? Inherent bias due to data heterogeneity.
  - Prior specification /w sensitivity analysis by varying priors.

Clinical Trials: Current Status and Enhancements

- Current Status and Limitations
  - One drug, one study population, one trial at a time.
  - Discrete-phase drug development
    - Phase I → Phase II → Phase III
  - Equal randomization
  - Infrequent interim monitoring
  - Limited use of all available information
    - No borrowing from historical data (external, outside of the trial)
    - No borrowing across subgroups (internal, within the trial)
    - No borrowing across similar trials (external, outside of the trial)

Master Protocol: Seamless phases, Umbrella, basket, Platform trials

→ Adaptive randomization

→ More interim analyses: Early stopping for toxicity, futility, efficacy

→ Bayesian modeling with informative priors

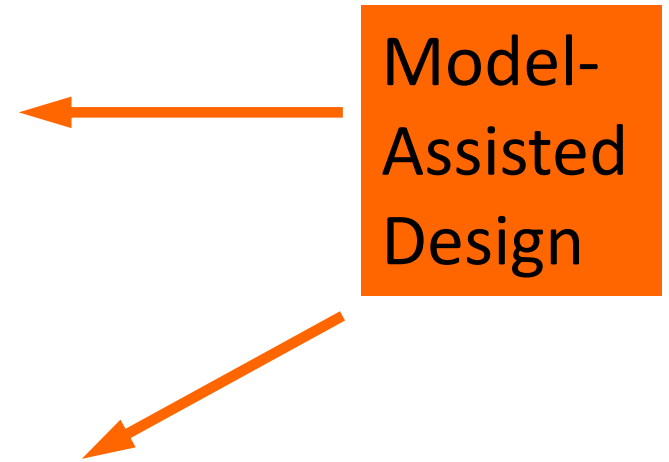
→ Bayesian hierarchical model, Cluster hierarchical model

→ Network meta-analysis

Bayesian Adaptive Designs

Trials that use interim data to guide the study conduct

- Adaptive dose finding
  - Bayesian Optimal INterval (BOIN) Design and iBOIN Design
    - Allow incorporating historical data as informative prior
- Adaptive stopping via posterior or predictive probability
  - Early stopping for toxicity, futility, and/or efficacy
- Adaptive Phase II design with complex endpoints
  - Bayesian Optimal Phase 2 (BOP2) Design
    - Allow informative prior
  - 2-Arm BOP2 Design
    - Allow informative prior
- Adaptive decision making
  - Dropping bad treatments, Add new treatments
  - Umbrella trials, Basket trials, Platform trials



Berry SM, Carlin BP, Lee JJ, and Mueller P. Bayesian Adaptive Methods for Clinical Trials. CRC Press: Boca Raton, FL, 2010.
Yuan Y, Lee JJ and Hilsenbeck SG. Model-Assisted Designs for Early Phase Clinical Trials: Simplicity Meets Superiority. JCO PO 2019

Adaptive Trial Design Shiny Applications

(30+ online programs freely available)

https://trialdesign.org

Novel Designs for Phase I Trials

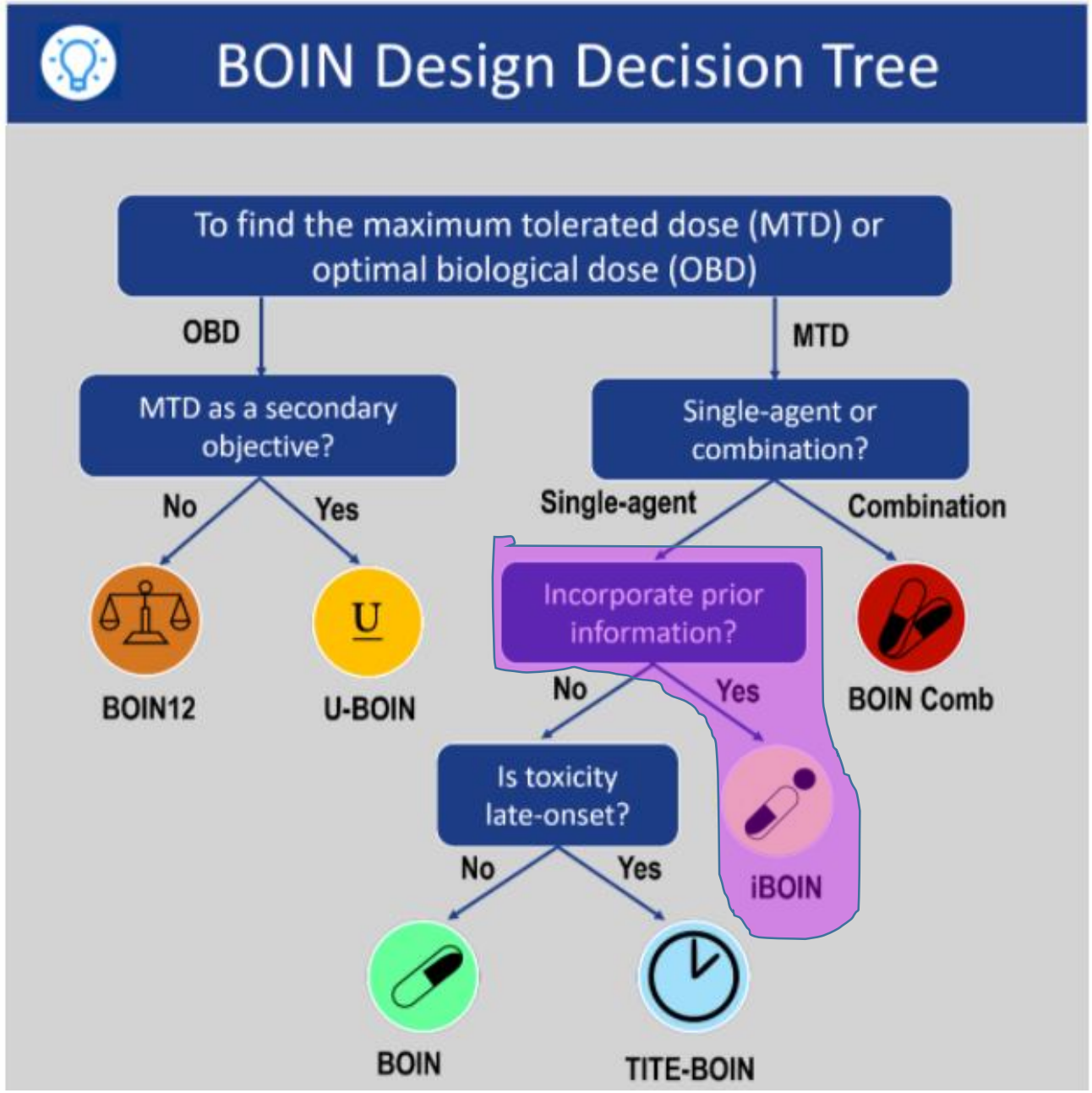
Bayesian Optimal Interval (BOIN) Design

How to choose a design?

Single Agent	Late-onset	Combination	Optimal Biological Dose (OBD)	
				
BOIN/iBOIN Launch Download	**TITE-BOIN** Launch Download	**BOIN Comb** Launch Download	**U-BOIN** Launch	**BOIN12 / TITE-BOIN12** Launch
Find MTD for single-agent trials	**Find MTD in trials with late-onset toxicity or fast accrual**	**Find MTD or MTD contour for combination trials**	**A two-stage design to find OBD for targeted and immune therapy**	**A single-stage design to find OBD for targeted and immune therapies**
BOIN is a novel model-assisted phase-1 trial design that is as easy to implement as the 3+3 design,but yields superior performance compared to more complicated model-based designs, such as CRM.	Time-to-Event BOIN (TITE-BOIN) allows for real-time dose assignment for new patients while some enrolled patients' toxicity data are still pending, thereby significantly shortening the trial duration. It is as easy to implement as the rolling 6 design, but yields much better performance.	BOIN Comb handles combinations of two drugs, each with multiple dose levels. It is as easy to implement as the 3+3 design, but yields superior perfomance compared to more complicated model-based designs.	U-BOIN is a utility-based seamless Bayesian phase I/II trial design to find the optimal biological dose (OBD) for targeted and immune therapies. It allows physicians to incorporate the risk-benefit trade-off to more realistically reflect the clinical practice.	BOIN12 is a simple and flexible Bayesian optimal interval phase I/II (BOIN12) trial design to find the OBD that optimizes the risk-benefit tradeoff. It makes the decision of dose escalation and de-escalation by simultaneously taking account of efficacy and toxicity, and adaptively allocates patients to the dose that optimizes the toxicity-efficacy tradeoff.

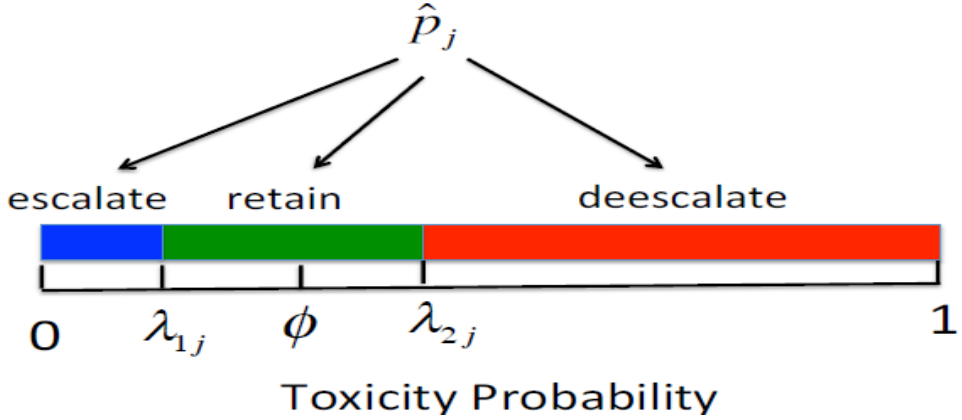
Yuan Y, Lin R, Lee JJ. Model-Assisted Bayesian Designs for Dose Finding and Optimization: Methods and Applications, Chapman & Hall/CRC 2022

How to Choose A Design?



Bayesian Optimal Interval (BOIN) Design

- With the target probability of toxicity ϕ, an interval design makes decision of dose escalation, stay, or de-escalation by comparing the estimated probability of toxicity $\hat{p}_j$ at dose j with a pre-specified toxicity interval.



- The interval boundaries λ_{1j} and λ_{2j} are selected to minimize the decision error of dosing.
- iBOIN Design can incorporate informative prior based on historical data.
- It is long-overdue to abandon the 3+3 design.

Liu S, Yuan Y. Bayesian optimal interval designs for phase I clinical trials. *Appl. Statist.* 64: 507–523, 2015

Zhou, Y., Lee, J. J., Wang, S., Bailey, S., & Yuan, Y. Incorporating historical information to improve phase I clinical trial designs. Pharmaceutical Statistics.;1–18, 2021.

Zhou, Y., Lin, R., Kuo, Y. W., Lee, J. J., & Yuan, Y. BOIN Suite: A Software Platform to Design and Implement Novel Early-Phase Clinical Trials. JCO Clinical Cancer Informatics, 5, 91-101, 2021.

Target Toxicity Rate $\phi = 0.4$

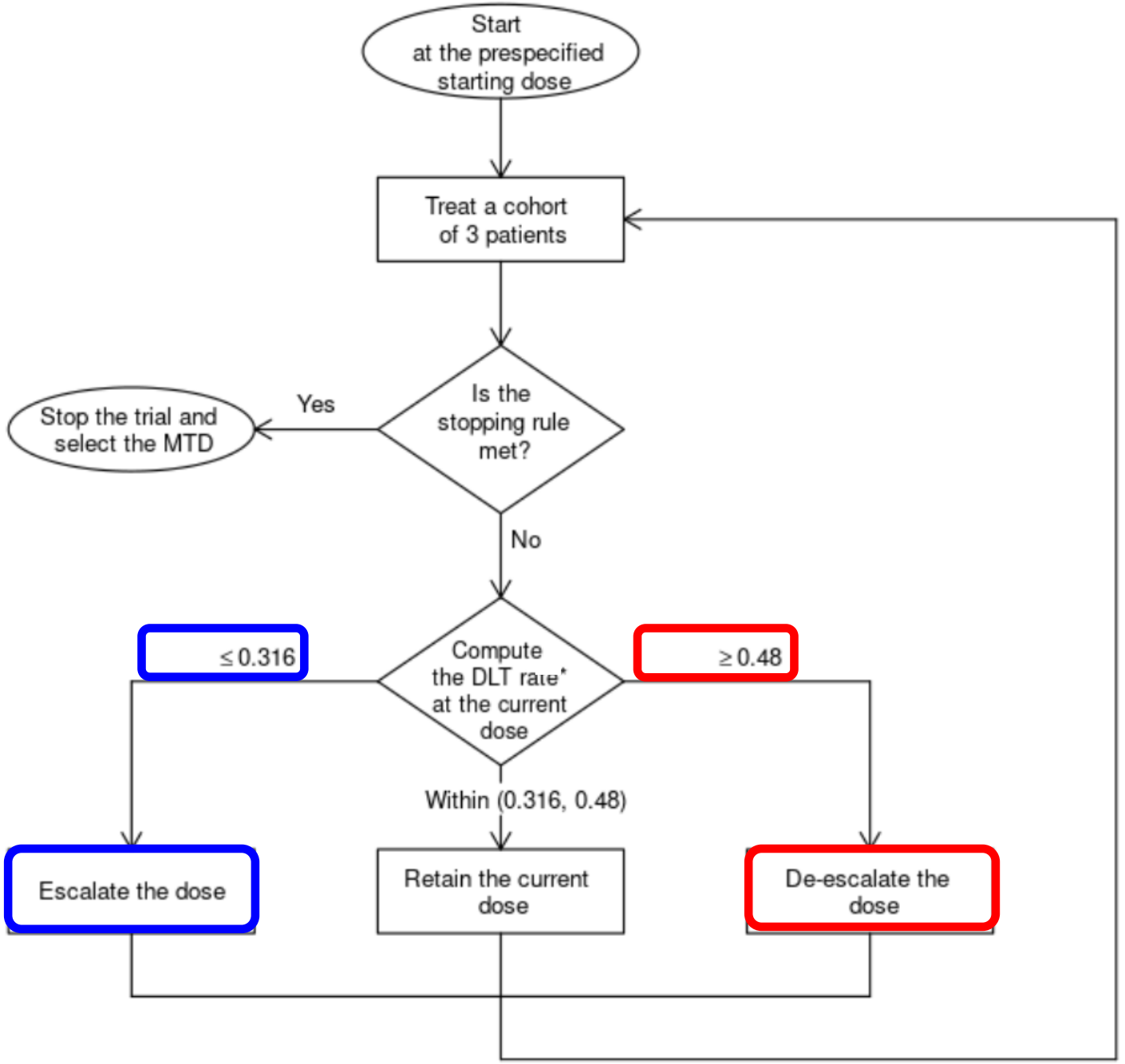
Prior Specification

Enter prior toxicity probability and effective sample size (ESS) at each dose level:

	D1	D2	D3	D4	D5
Pr(toxicity)	0.06	0.14	0.25	0.38	0.50
ESS	10.00	5.00	2.00	2.00	2.00

☐ Check the box to use robust prior

Save Input | Get Decision Table



BOIN: MTD Selection

Trial Name (Optional):

Target Toxicity Probability ϕ :

0.4

Number of doses

5

Please enter the trial data:

Dose level	Number of patients treated	Number of patients with dose limiting toxicity
1	3	0
2	4	1
3	6	2
4	4	2
5	1	1

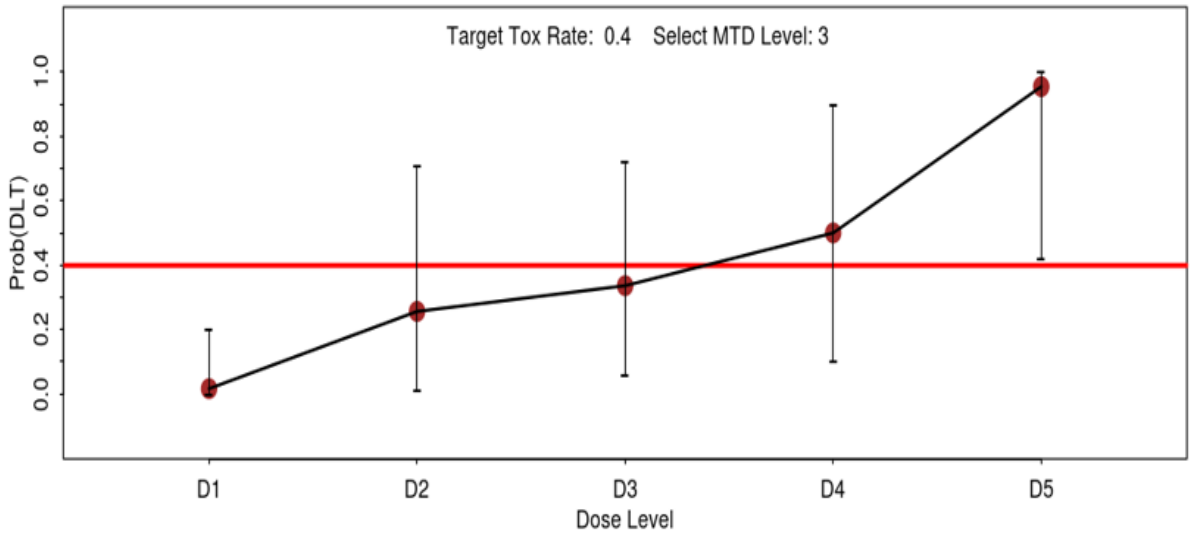
Estimate the MTD

Download MTD Report (PDF)

MTD Selection Result

```
The MTD is dose level  3

Dose    Posterior DLT            95%
Level     Estimate        Credible Interval   Pr(toxicity>0.4|data)
  1         0.02          ( 0.00 ,  0.20 )            0.01
  2         0.26          ( 0.01 ,  0.71 )            0.22
  3         0.34          ( 0.06 ,  0.72 )            0.34
  4         0.50          ( 0.10 ,  0.90 )            0.65
  5         0.95          ( 0.42 ,  1.00 )            0.98
NOTE: no estimate is provided for the doses at which no patient was treated.
```



Novel Designs for Phase II Trials

BOP2 Suite

BOP2

Launch

Make optimal go/no-go interim decisions

The Bayesian optimal phase II (BOP2) design is a flexible Bayesian design that allows any arbitrary number of interim analyses and handles both binary and complicated (e.g., nested and co-primary) efficacy endpoints. Unlike most Bayesian phase II designs, the BOP2 design controls type I error rate and is optimal in the sense that it optimizes power given a fixed sample size, or minimizes the expected sample size if the regimen has low activity (i.e., under the null hypothesis) given a fixed type II error rate.

BOP2-DC

Launch

Make optimal go/consider/no-go decisions

The BOP2-DC design incorporates both statistical significance and clinical significance into decision making. It allows for go/consider/no-go decisions, rather than a binary go/no-go decision. BOP2-DC is optimized to maximize the probability of a go decision when the treatment is effective. The design is highly flexible and accommodates various types of endpoints, including binary, categorical, time-to-event, single and multiple endpoints.

2-Arm BOP2

Launch

BOP2 design for two-arm randomized trials

The 2-arm BOP2 design is an extension of BOP2 design to make go/no-go decisions for two-arm randomized trials. Similar to BOP2, it is flexible, allowing any arbitrary number of interim analyses and accommodating both binary and co-primary endpoints; and it is efficient by maximizing the power of the trial.

TOP

Launch

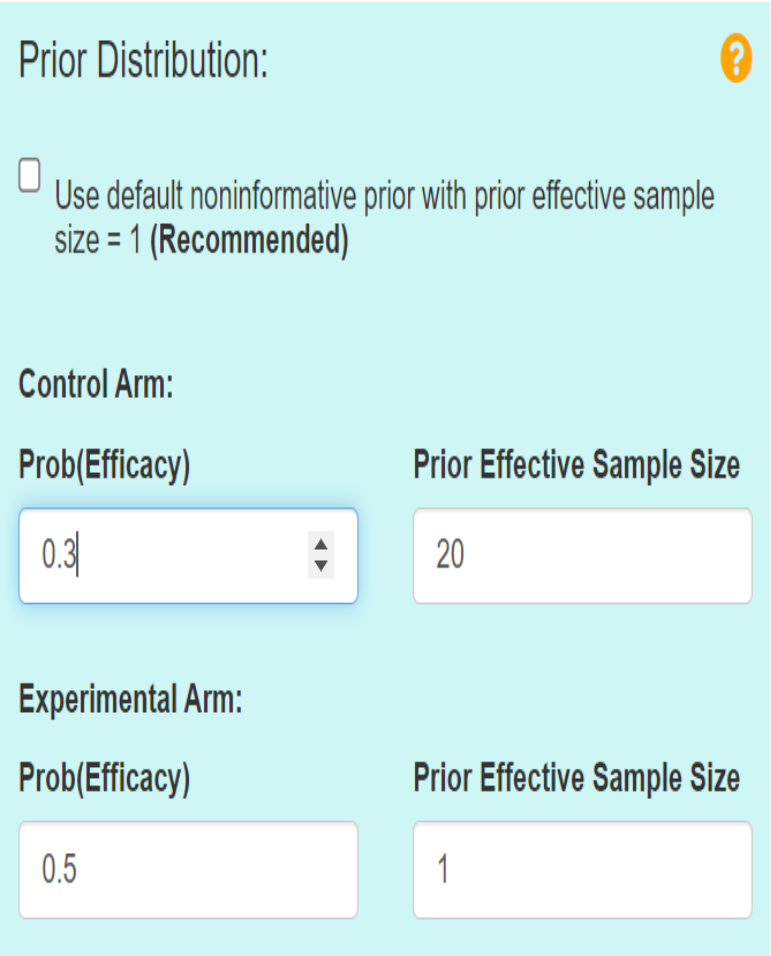
Extension of BOP2 allowing real-time go/no-go decisions

The time-to-event Bayesian Optimal Phase II (TOP) design is a flexible and efficient design for phase II clinical trials. It allows real-time 'go/no-go' interim decision making when some patients' outcomes are still pending. The TOP design maximizes statistical power for detecting effective treatments with well-controlled type I errors. The TOP design is very easy to implement as its decision rules can be tabulated prior to the trial conduct.

Yuan Y, Lin R, Lee JJ. Model-Assisted Bayesian Designs for Dose Finding and Optimization: Methods and Applications, Chapman & Hall/CRC 2022

BOP2: A Bayesian Optimal Design for Phase 2 Clinical Trials with Simple & Complex Endpoints

- Provides a unified framework for phase II trials with simple and complex efficacy and toxicity endpoints.
- Explicitly controls the type I (and II) error rates.
- Is optimal by
  - (i) maximizing power, given a fixed N and type I error; or
  - (ii) minimizing the $E(N|H_0)$, given fixed type I and II error rates.
- 2-Arm BOP2 Design allows comparison between two arms.
  - Allows incorporating informative prior based on historical data



Zhou H, Lee JJ, Yuan Y. BOP2: Bayesian optimal design for phase II clinical trials with simple and complex endpoints. Stat Med. 2017.
Zhao, Y., Yang, B., Lee, J. J., Wang, L., & Yuan, Y. Bayesian Optimal Phase II Design for Randomized Clinical Trials. Statistics in Biopharmaceutical Research, 1-10, 2022.

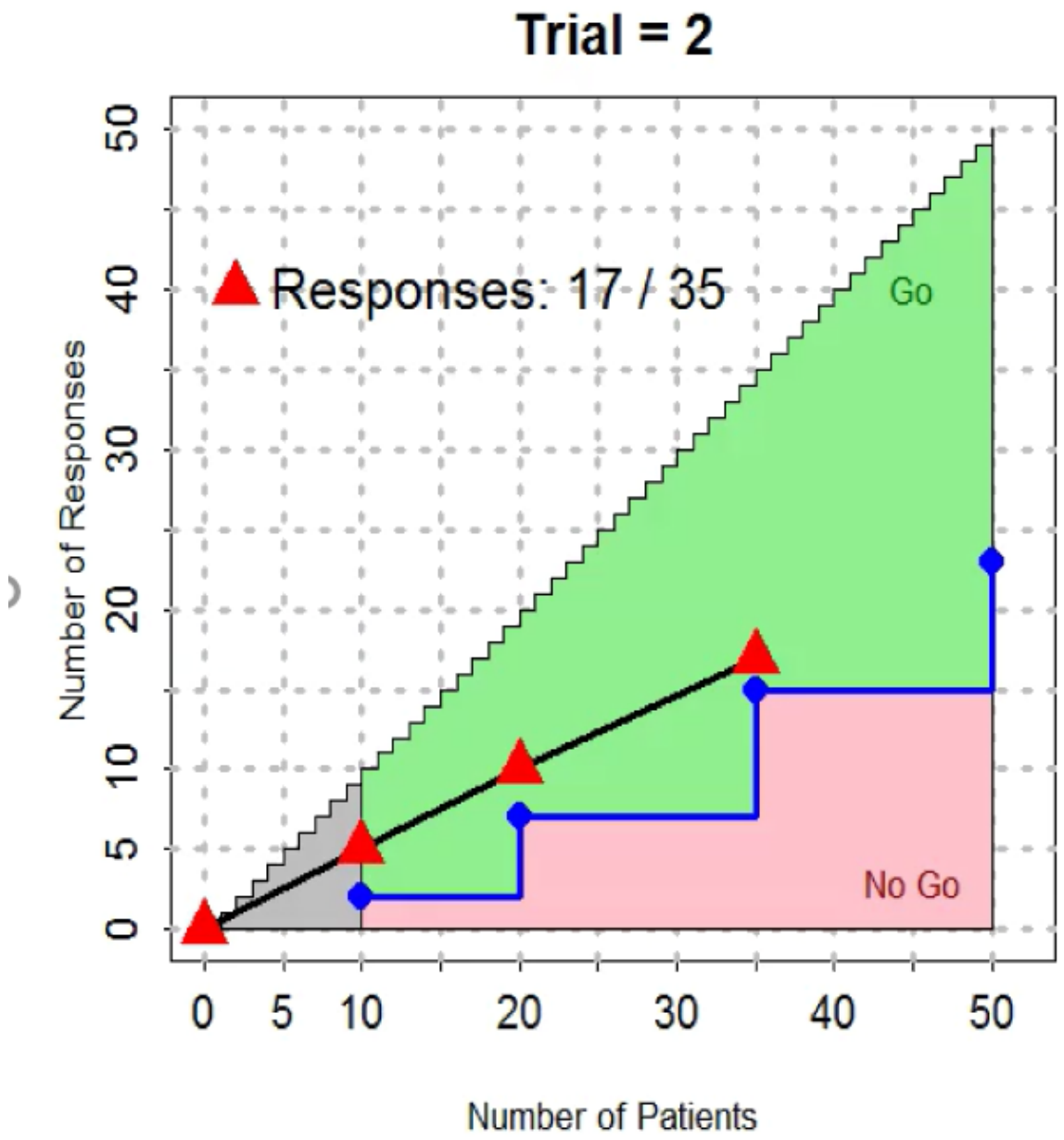
Stopping Boundaries for BOP2 Design

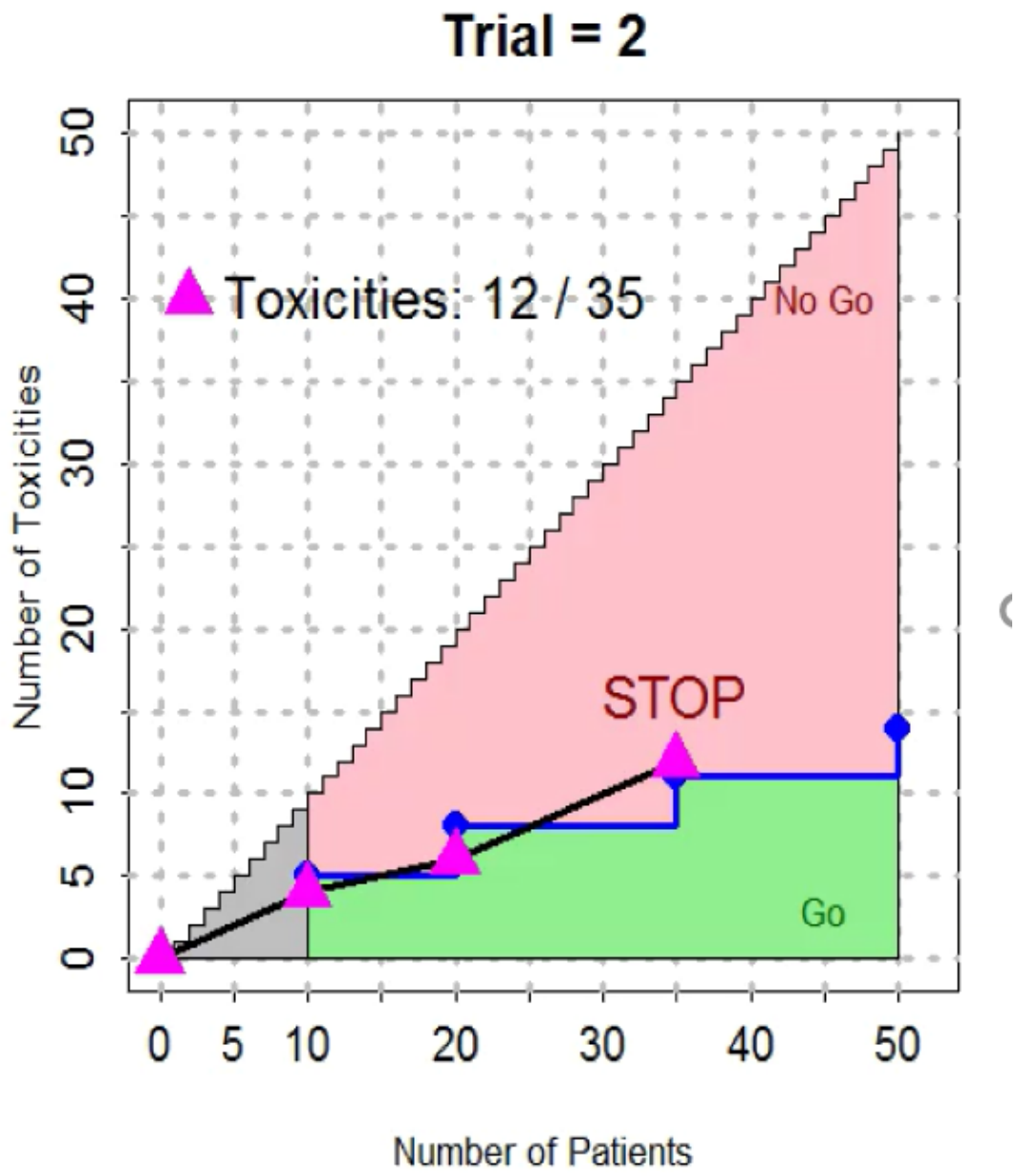
Trial		Stop the trial if	Number of patients treated						
			10	15	20	25	30	35	40
Example 1		# of OR $\leq$	1	2	4	5	7	9	10
Example 2	and	# of CR $\leq$	0	1	3	4	5	7	9
		# of CR/PR $\leq$	2	3	5	8	10	13	16
Example 3	and	# of OR $\leq$	0	1	2	3	4	5	7
		# of PFS6 $\leq$	1	2	4	5	7	9	12
Example 4	or	# of OR $\leq$	2	5	7	10	13	16	19
		# of Toxicities $\geq$	5	6	8	9	10	11	12

OR: objective response

1. $H_0 : Pr(OR) = 0.20; H_1 : Pr(OR) = 0.4$
2. $H_0 : Pr(CR) = 0.15, Pr(CR/PR) = 0.3; H_1 : Pr(CR) = 0.25, Pr(CR/PR) = 0.50.$
3. $H_0 : Pr(OR) = 0.1, Pr(PFS6m) = 0.2; H_1 : Pr(OR) = 0.3, Pr(PFS6m) = 0.35.$
4. $H_0 : Pr(OR) = 0.45, Pr(Toxicity) = 0.30; H_1 : Pr(OR) = 0.60, Pr(Toxicity) = 0.20.$

BOP2 Design with Response and Toxicity Endpoints



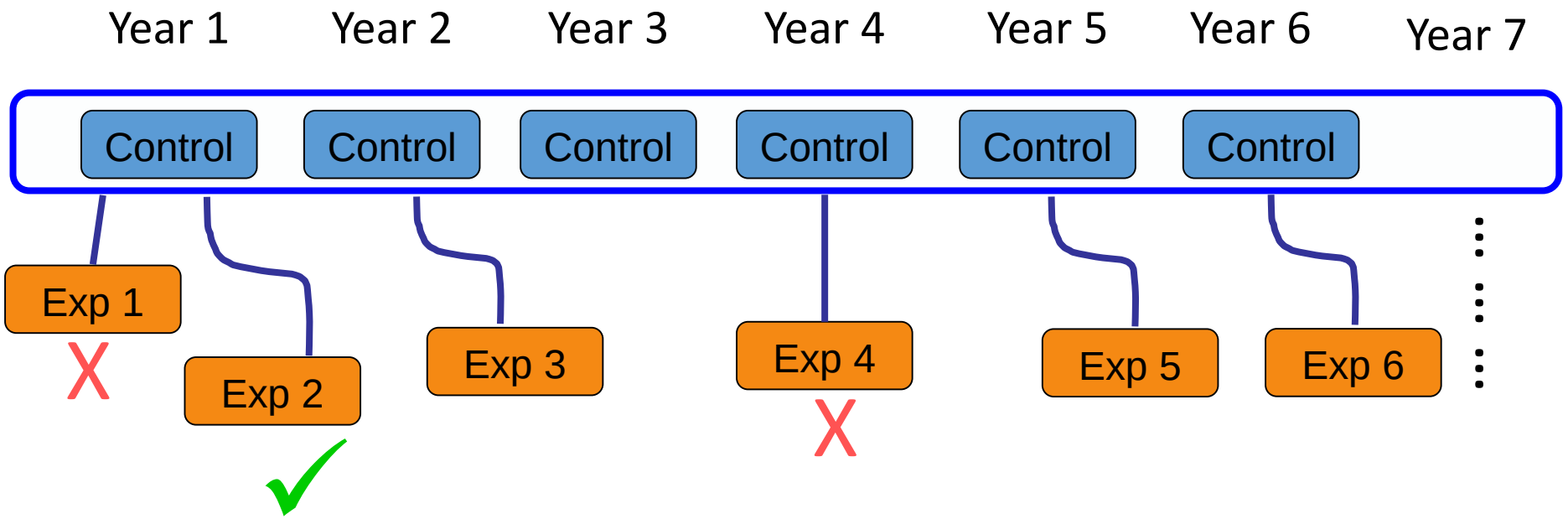


Platform Design with Adaptive Enrichment in Randomized Phase II Trials

- Start with one control and multiple experimental arms or age or histological subgroups
- Continuous toxicity monitoring
  - Drop subgroups when excessive toxicity is found
- Apply equal randomization (ER) or adaptive randomization (AR)
  - Adaptive enrichment via AR
- Calculate the predictive probability or posterior probability of each subgroup being better than the control
  - Sufficiently low: Drop the subgroup
  - Sufficiently high: Graduate the subgroup
  - Otherwise, continue patient enrollment until reach N_{max}
- A perpetual, drug screening platform
  - Write a protocol with the "backbone" infrastructure
  - Add new treatments whenever needed
  - Amend the protocol by adding or enriching subgroups showing promising results

Hobbs BP, Chen N, Lee JJ. Controlled multi-arm platform design using predictive probability. Stat Methods Med Res 27(1):65-78, 2018
Zhu H, Piao J, Lee JJ, Hu F, Zhang L. Response adaptive randomization procedures in seamless phase II/III clinical trials. J Biopharm Stat 30(1):3-17, 2019

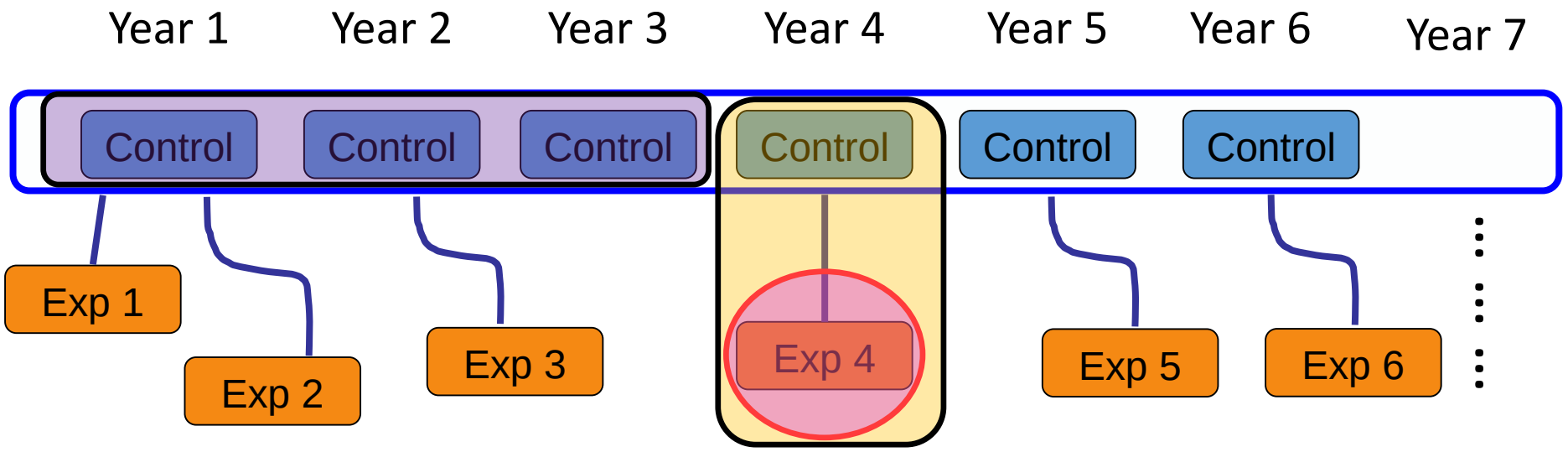
Adaptive Platform Design



Control: Backbone of the Platform

Experimental Treatments: Modules

Adaptive Platform Design



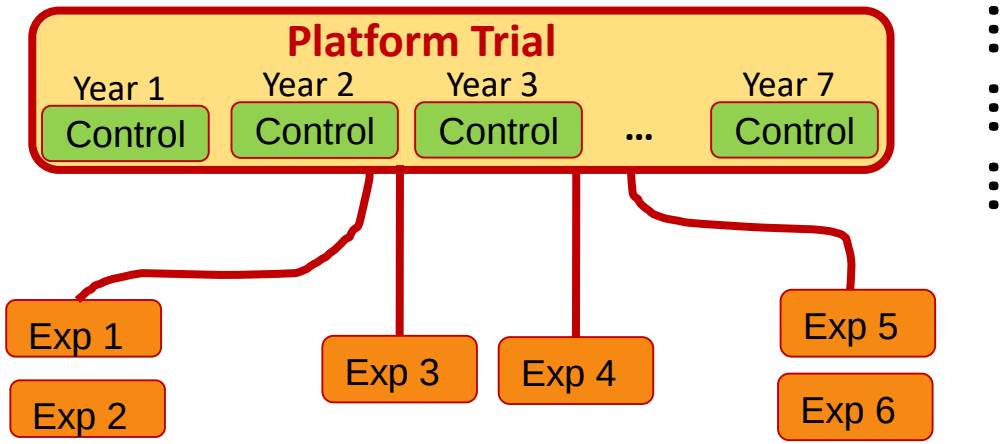
Historical Control + **Concurrent Control**

- Large Sample Size
- More heterogeneous
- Population drift over time

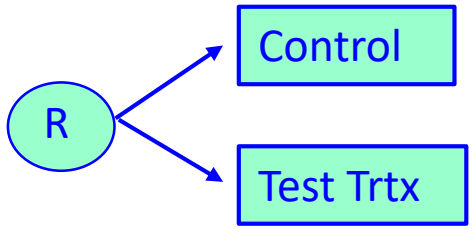
- Small Sample Size
- More homogeneous
- No population drift

Novel Designs with Master Protocols

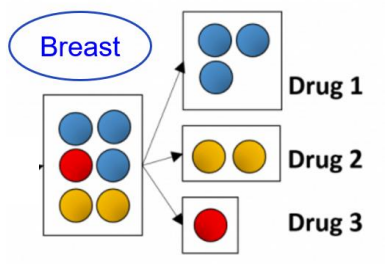
Exploratory: Learning and Signal Seeking



Confirmatory

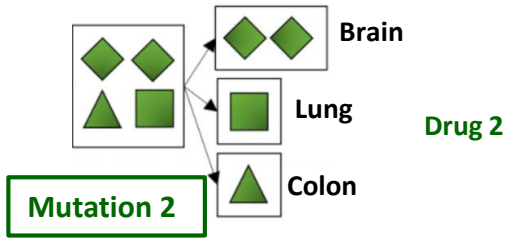


Umbrella Trial



Focused, Small, Phase III Trial in Selected patient population

Basket Trial



Bayesian Hierarchical Model for Synthesizing Information for Subgroups in Basket Trials

- Clinical Trials often have subgroups
  - Different histology subtypes or age or region subgroups
- Bayesian hierarchical model can borrow information across subgroups
  - More borrowing when subgroups are more alike and less borrowing when subgroups are more different. (nice!)
- Bayesian Classification and Information Sharing (BaCIS) allows *smart borrowing* which borrows across "similar" subgroups and does not borrow across "dissimilar" ones.
- Bayesian cluster hierarchical model (BCHM) forms clusters first. Subgroups within the same cluster are exchangeable but not exchangeable across clusters.
  - Borrow information within each cluster
- Bayesian hierarchical model can synthesize multi-sources real-world data

Two Goals for Bayesian Hierarchical Model in Borrowing Information

- Accuracy
  - Identify subgroups in which drug works
  - Identify subgroups in which drug do not work

- Efficiency
  - Use smaller sample size to achieve the accurate inference, i.e.,. making the correct “go” or “no go” decision

Kaizer AM, Koopmeiners JS, Hobbs BP. Bayesian hierarchical modeling based on multisource exchangeability. Biostatistics 1;19(2):169-184, 2018.
Chen and Lee, Bayesian hierarchical classification and information sharing for clinical trials with subgroups and binary outcomes, Biometrical Journal 2019.
Chen and Lee, Bayesian cluster hierarchical model for subgroup borrowing in the design and analysis of basket trials with binary endpoints, *Statistical Methods in Medical Research* 2020

Borrowing Across Subgroups

Parameters for BHM

Hyper Prior for overall mean, mu:

μ_0 | τ_0

0 | 0.000001

Hyper Prior for precision of mu, tau:

Γ_a | Γ_b

0.01 | 0.1

Prior distribution for ρ_i :Beta(a_0 ,b_0)

a_0 | b_0

0.001 | 0.001

Credible Interval (1- α) x 100%: α

0.2

WinBUGS Settings

Burnin | Iter

1000 | 10000

Sample Information

Number of Groups

5

Number of Success for Each Group *(separate by a comma)*

1,2,3,7,8

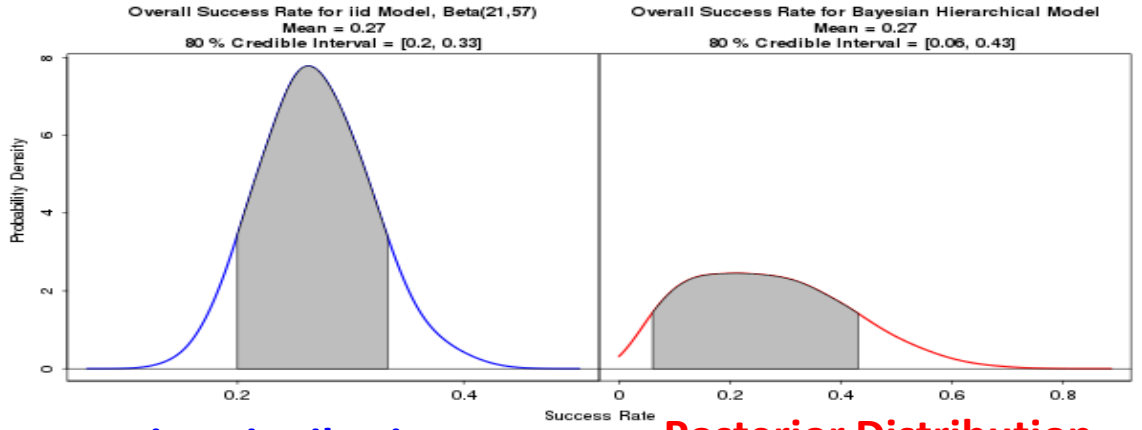
Sample Size for Each Group *(separate by a comma)*

15,18,10,15,20

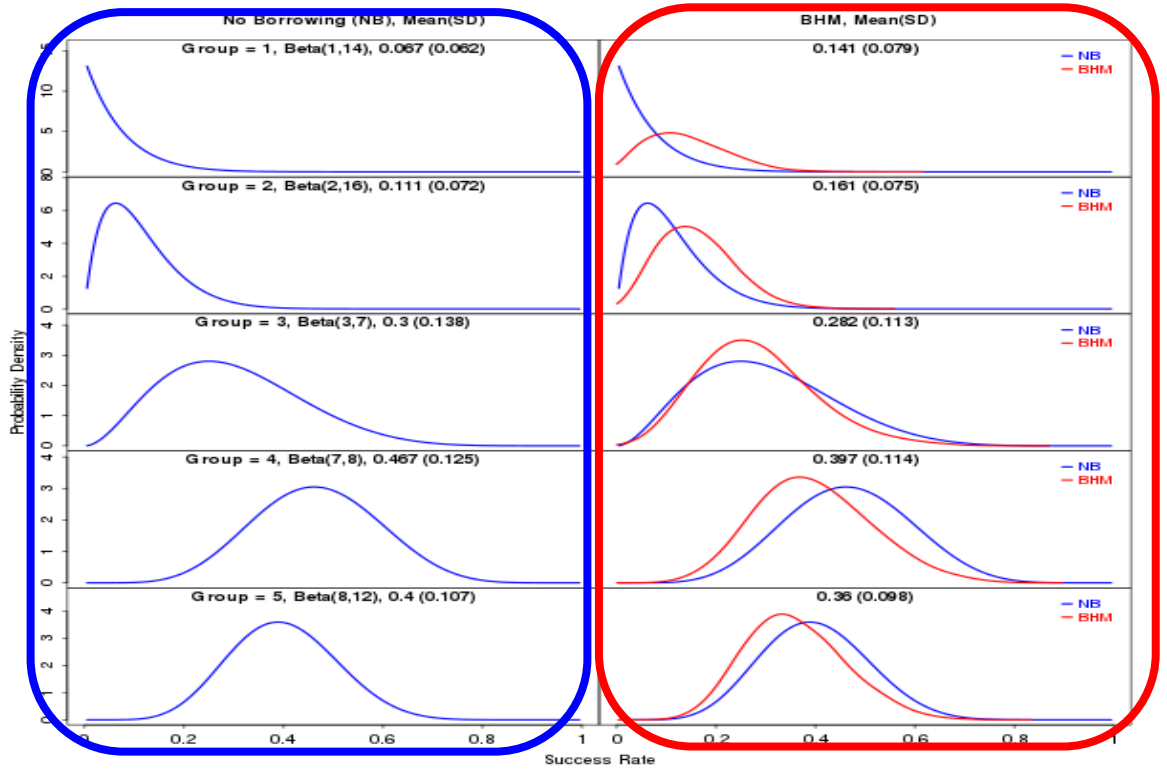
Calculate

Edit | Start Over

BHM - Binomial Data | Support Document



Prior Distribution | **Posterior Distribution**

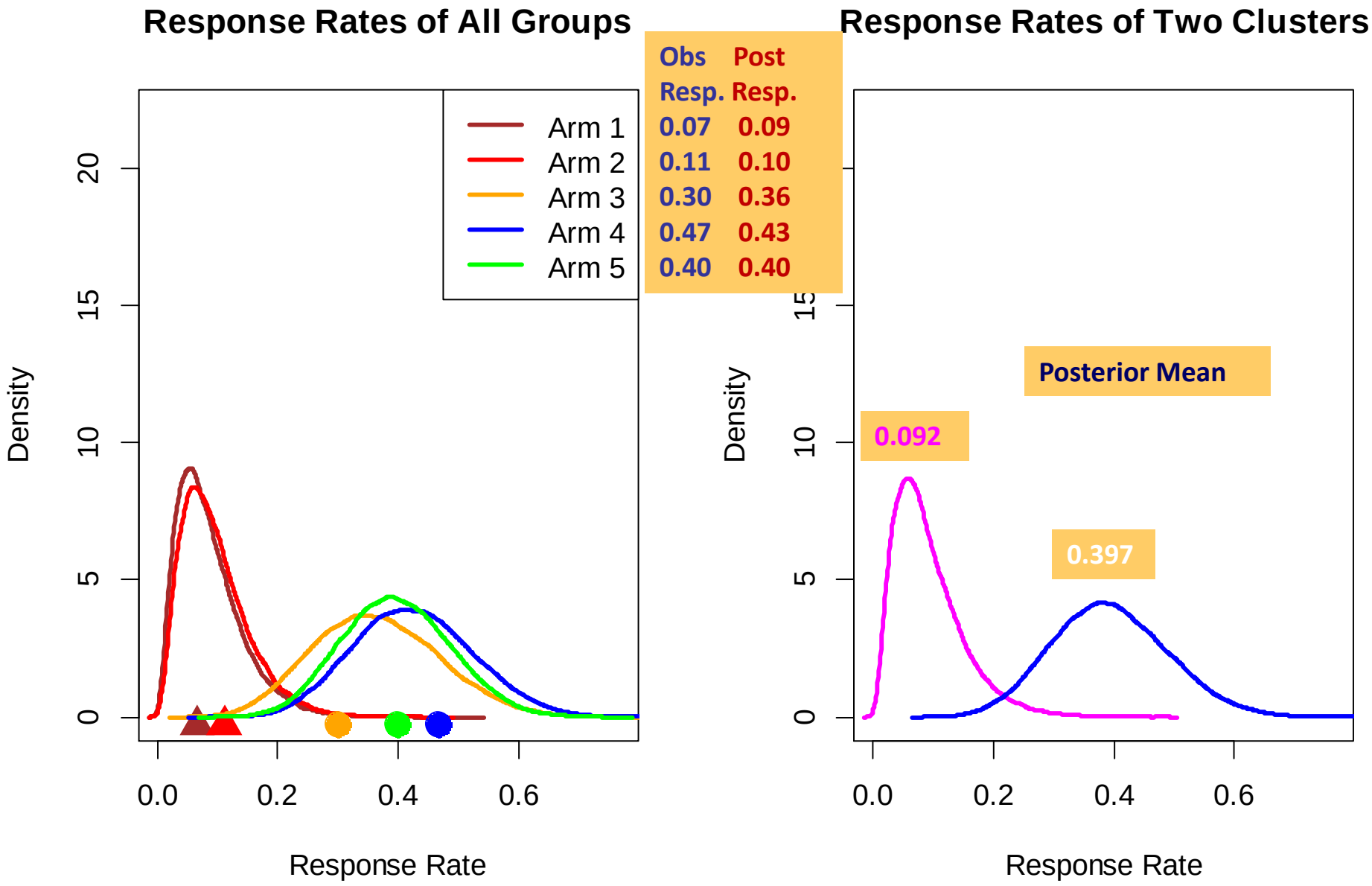


Response Rates in 5 Age Subgroups: 1/15, 2/18, 3/10, 7/15, 8/20

Age
12-18
6-11
2-5
1-2
<1

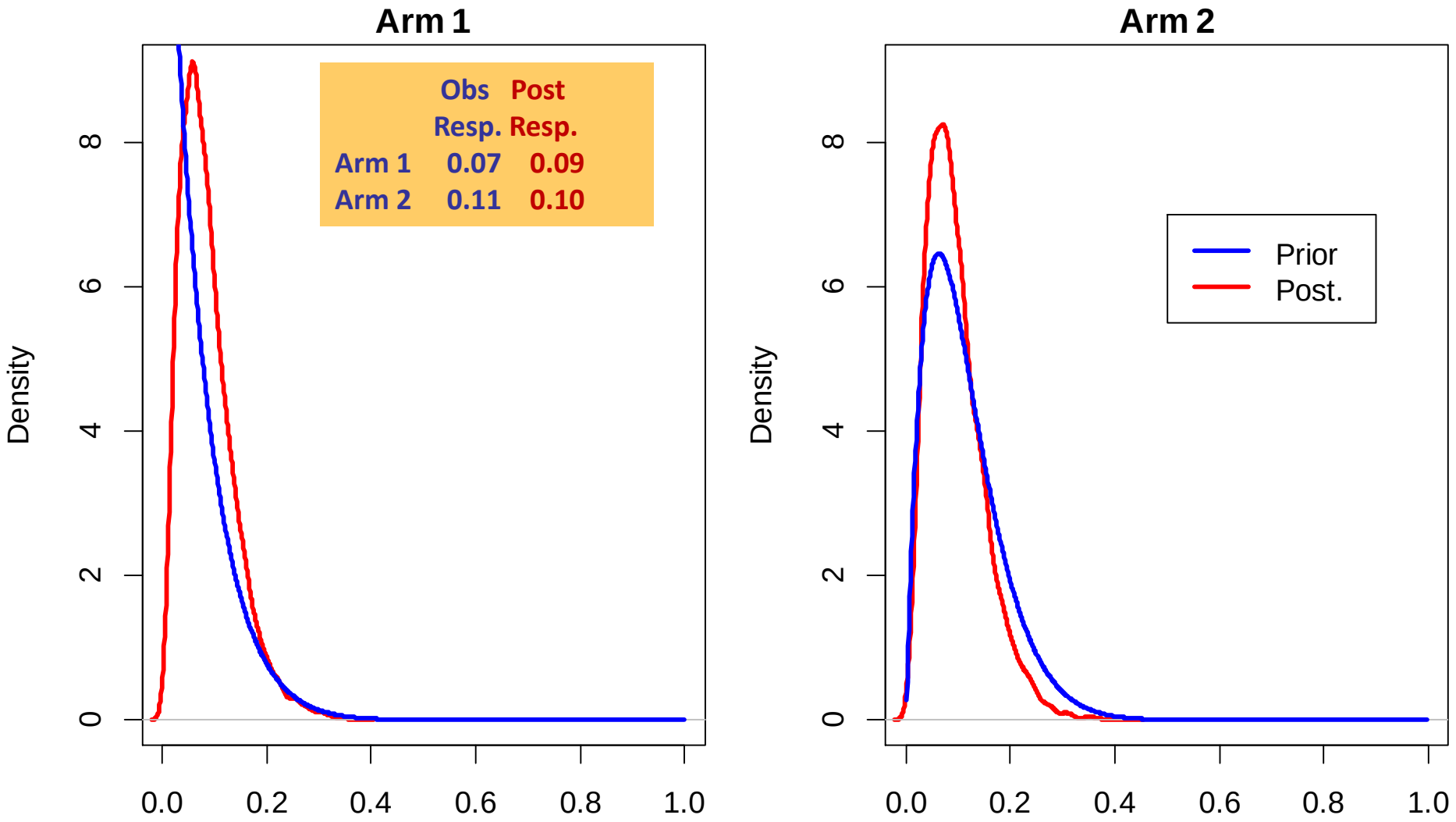
26

Posterior Distributions of Response Rates



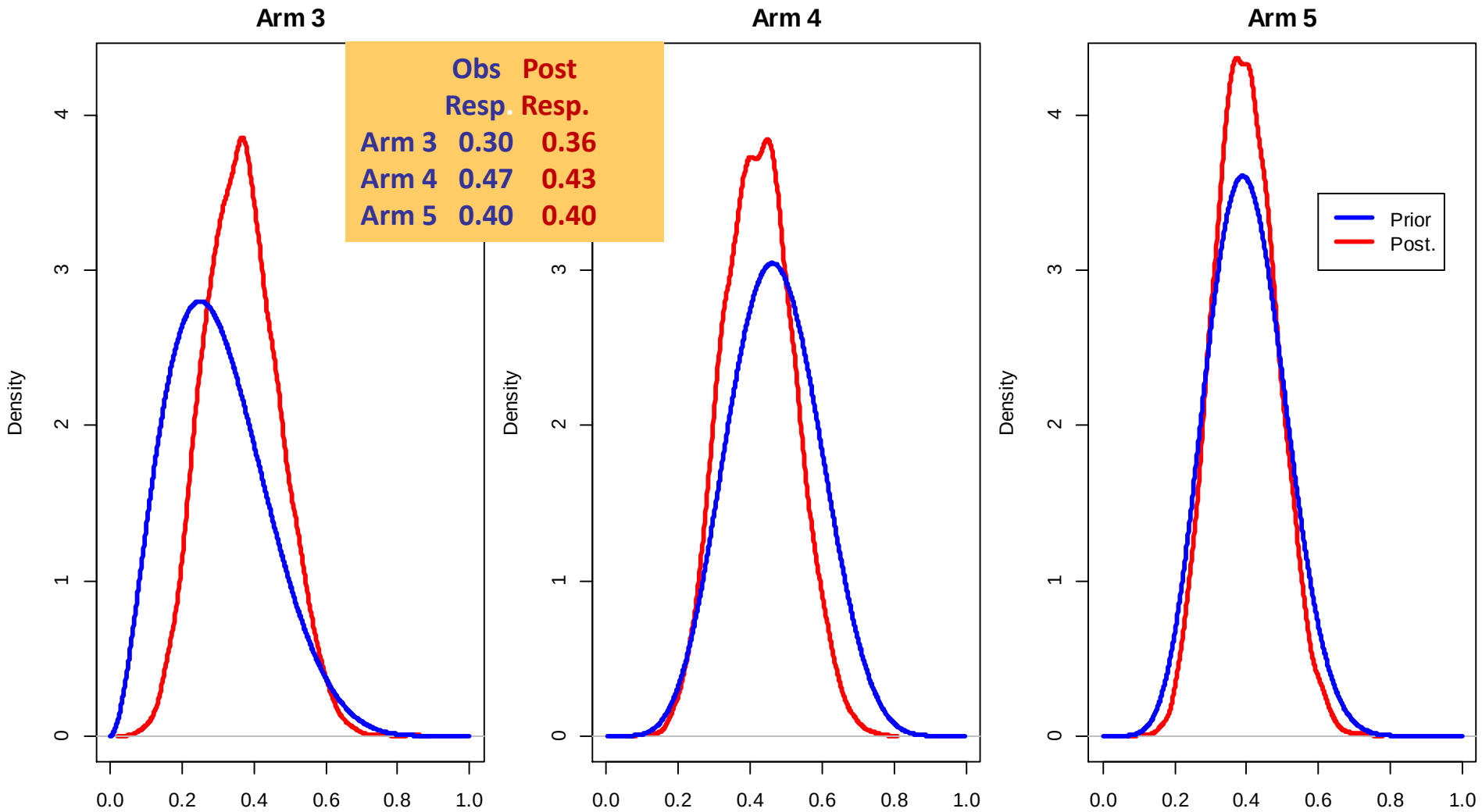
5 Arms with outcomes (1/15, 2/18, 3/10, 7/15, 8/20). $\emptyset_1$=0.1, $\emptyset_2$=0.3, τ_2=0.001, τ_4=0.1, α=5, β=1.

Cluster 1



5 Arms with outcomes (**1/15, 2/18**, 3/10, 7/15, 8/20)

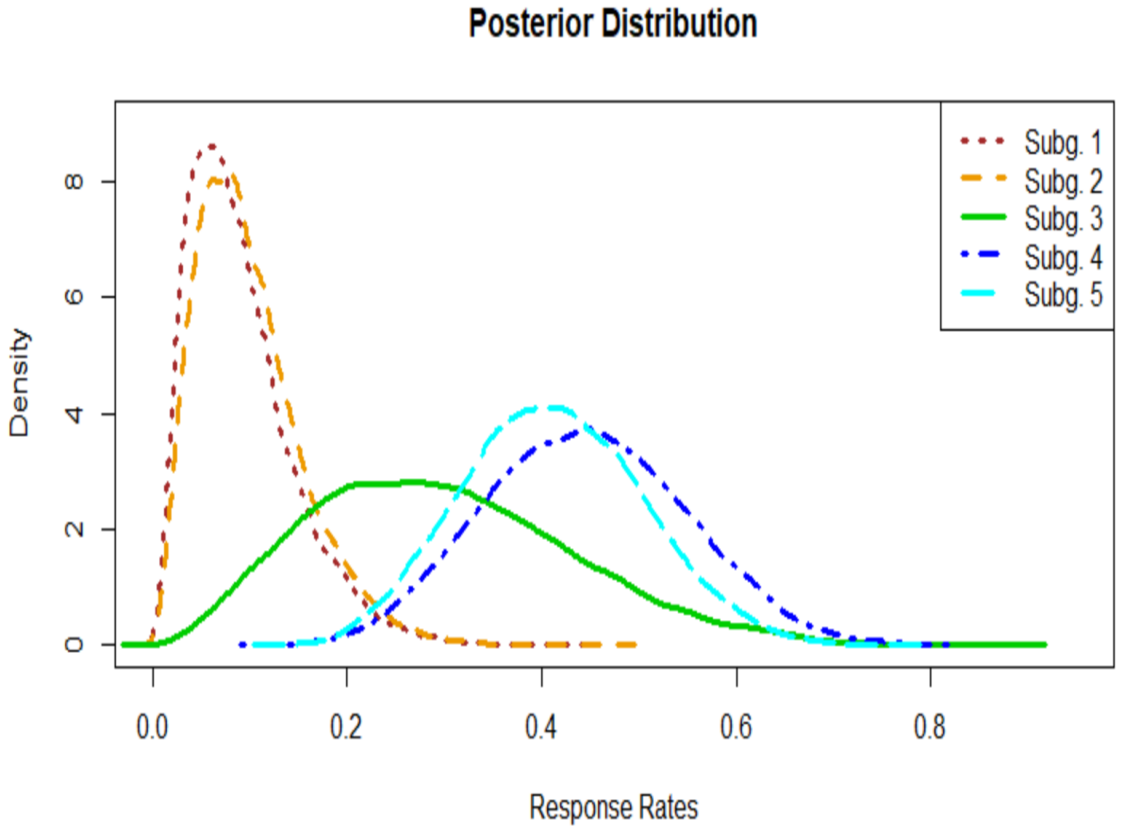
Cluster 2

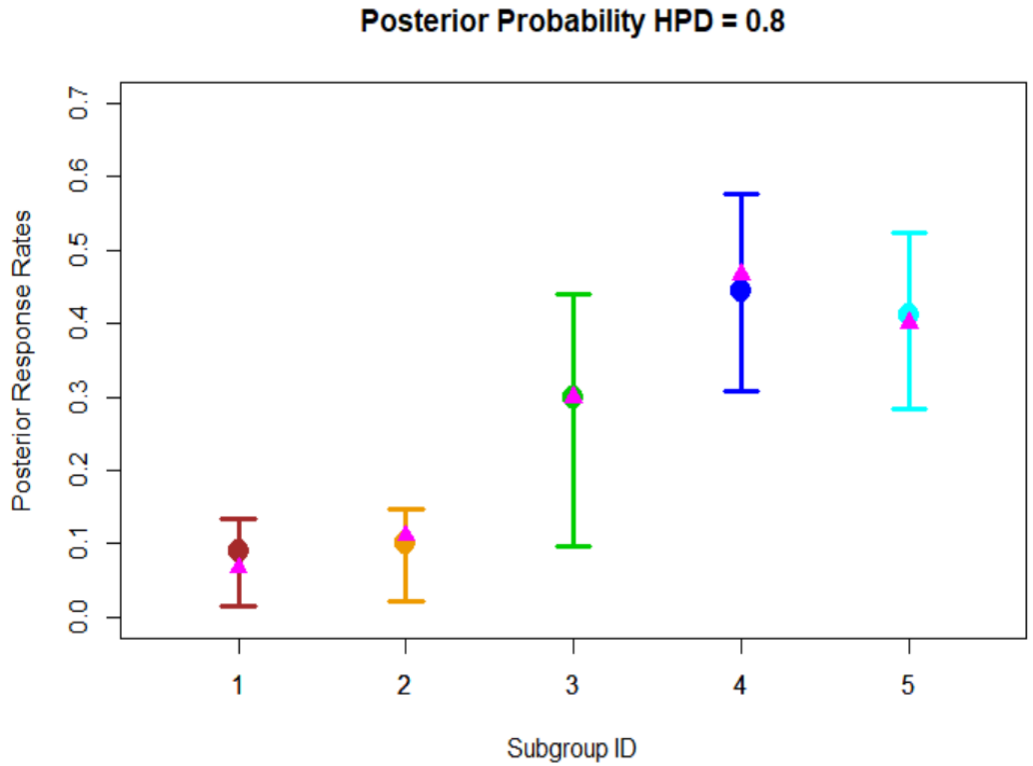


5 Arms with outcomes (1/15, 2/18, **3/10, 7/15, 8/20**)

29

R-Package: BCHM
BCHM Output with alpha = 1e-10, 3 Clusters





Circle: Posterior Mean; Triangle: Observed Mean

Concluding Remarks

- Statistics can help in extracting signals from the noise in the data
  - Avoid bias
  - Reduce variability / Increase efficiency
- There is no free lunch. But there are some lunch specials.
- Bayesian paradigm takes the "*we learn as we go*" approach and is particularly useful in rare diseases
  - Flexible and adaptive
  - Continuous learning
  - Naturally and easily to incorporate and synthesize all relevant information
- Bayesian adaptive designs are efficient and robust in drug development
- All signals found need to be validated in prospective trials.

Working closely with statisticians from beginning to end and applying rigorous statistical methods to maximize the success of every project.

FDA/CDER and John Hopkins University CERSI Workshop Addressing Challenges in the Design and Analysis of Rare Disease Clinical Trials: Considerations and Tools

Leveraging Longitudinal Data in Design and Analyses

Rima Izem, PhD
May 3rd, 2023

Take home

- Multiple randomized study designs in rare diseases leverage longitudinal data (repeated measures) and within-subject comparison (self-control) to establish efficacy or safety

- Multiple observational study designs can also leverage longitudinal data (repeated measures) and within-subject comparison but need to control for multiple sources of bias (e.g., confounding and selection biases)

- Introducing within-subject design and analyses methods in comparative studies has the potential advantages of increasing analyses units, reducing outcome variability, and reducing confounding compared to between-subject comparisons

Outline

1. Review of randomized designs in rare diseases leveraging longitudinal data collection (repeated measures) of outcomes over time
2. Introducing observational study methods leveraging longitudinal collection
3. Design and analyses considerations

Randomized vs. observational

Randomized studies control for **all confounding and** selection through **randomization** and planning

Epidemiological studies control for **multiple sources of bias** using target RCT emulation

- Parallel Arms, factorial designs ⟷ - Cohort study, use of external control to a single arm, case-control
- Crossover, and N-1 ⟷ - Self-control case series or case-crossover
- Sequential randomization studies ⟷ - Sequential control for confounding

Source: Izem R, McCarter R. Randomized and non-randomized designs for causal inference with longitudinal data in rare disorders. Orphanet J Rare Dis. 2021 Nov 23;16(1):491.

Randomized longitudinal studies

There are more choices for randomization than parallel control and including within-subject comparisons has several advantages

Review of randomized designs

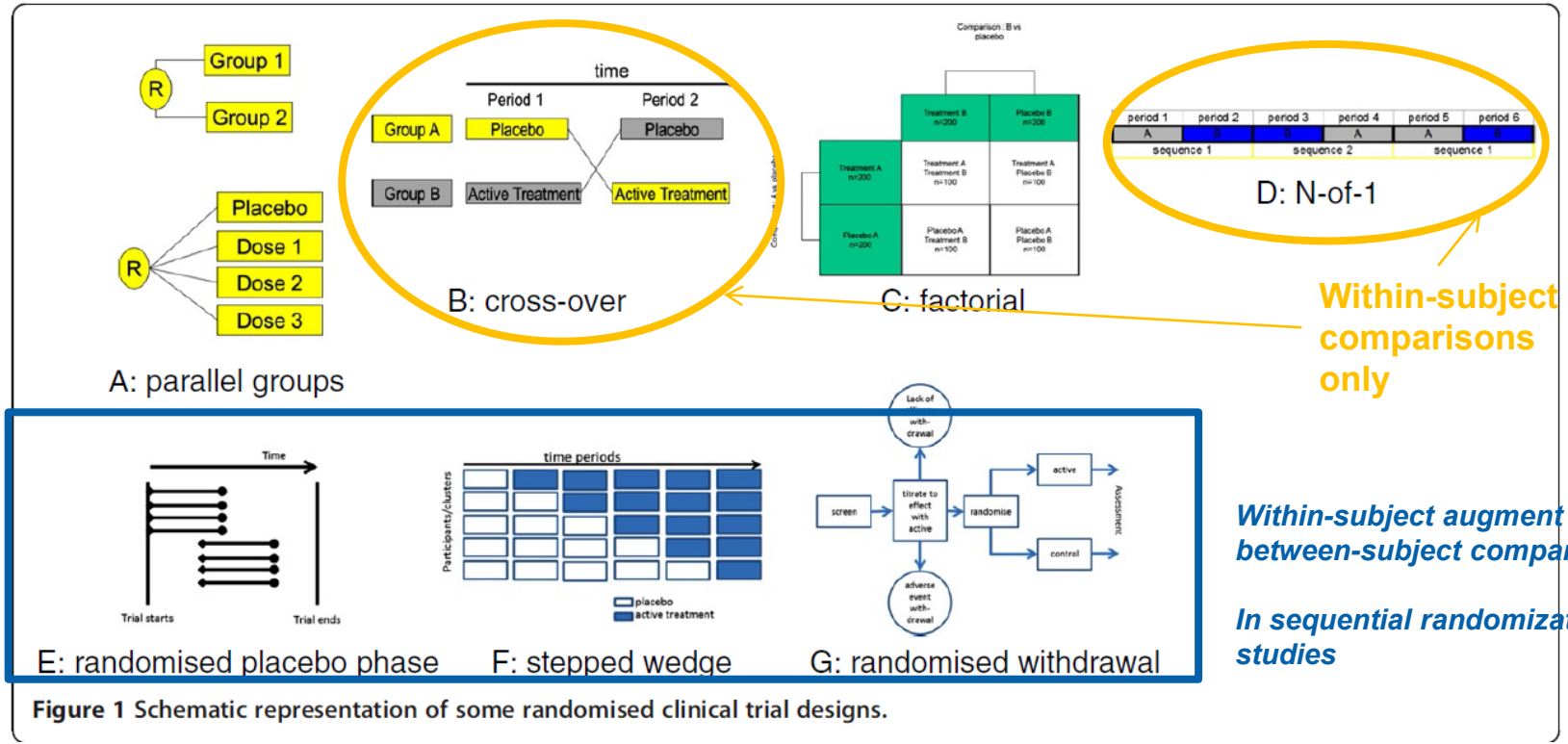


Figure 1 Schematic representation of some randomised clinical trial designs.

Source: Cornu et al (2013). Experimental designs for small randomised clinical trials: an algorithm for choice." Orphanet journal of rare diseases 8 (2013): 1-12.

N-1 randomized study example
within-subject comparison only

- Rare disorder: urea cycle disorder (UCD)
- Study subject OTCD female > 45 years of age, did not take L-arginine for a few months prior to study
- Trial over a 6-week period, 3 paired weeks (L-arginine and placebo pairs), blinded to treatment physician and patient

Table 1
Weekly efficacy indicators comparing placebo and L-arginine treatments

Efficacy indicator	Pair 1				Pair 2				Pair 3				Mean		Paired *t*-test	
	Placebo		Active		Placebo		Active		Placebo		Active		Placebo	Active	0-Tail	1-Tail
	Day				Day				Day							
	5	6	5	6	5	6	5	6	5	6	5	6	5 + 6	5 + 6		
Questionnaire score	3.1	1.5	5.5	5.8	5.4	4	5.7	4.9	2.7	3.2	5	5.1	3.3	5.3	0.0162*	
Plasma arginine μmol/L	102		156		60		148		91		108		84	138	0.122	0.061
Plasma glutamine μmol/L	716		611		628		539		690		471		678	540	0.078	0.039*

Plasma arginine reference range; 34–118 μmol/L.
Plasma glutamine reference range; 385–862 μmol/L.
* Significant at $P < 0.05$.

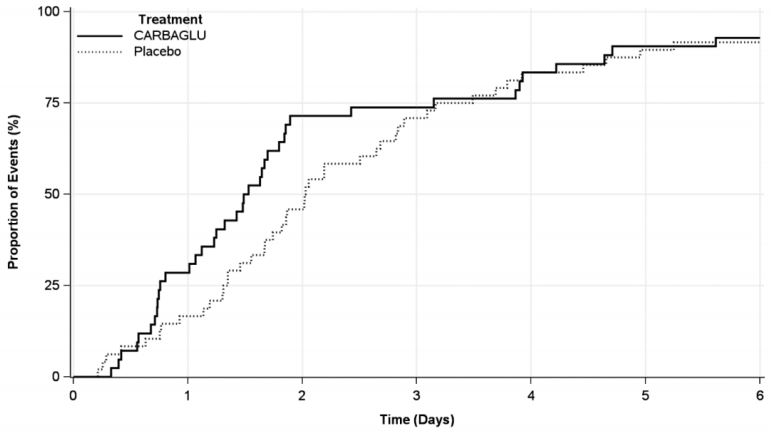
Source: Hackett A, Gillard J, Wilcken B: n of 1 trial for an ornithine transcarbamylase deficiency carrier. Mol Genet Metab 2008, 94:157–161.

Useful reference: Senn, S., Sample size considerations for n-of-1 trials. Stat Methods Med Res, 2019. 28(2): p. 372-383.

Randomized treatment periods example
within and between-subject comparison

“The efficacy evaluation was based on 90 hyperammonemic episodes (42 treated with CARBAGLU and 48 with placebo) in 24 patients (12 male and 12 female) with PA (n = 15) or MMA (n = 9) [...] The primary endpoint was the time from the first dose of drug to the earlier of plasma ammonia level ≤ 50 micromol/L (normal range) or hospital discharge [...]Throughout the first three days of treatment, a higher proportion of CARBAGLU-treated episodes reached the primary endpoint compared to placebo-treated episodes (Figure 2)”

Figure 2: Episodes Reaching the Earlier of Plasma Ammonia Level ≤ 50 micromol/L or Hospital Discharge in Patients with PA or MMA Treated with CARBAGLU or Placebo for up to 7 days



Source: CARBAGLU Label (2021)

Summary and further considerations
Randomized longitudinal designs

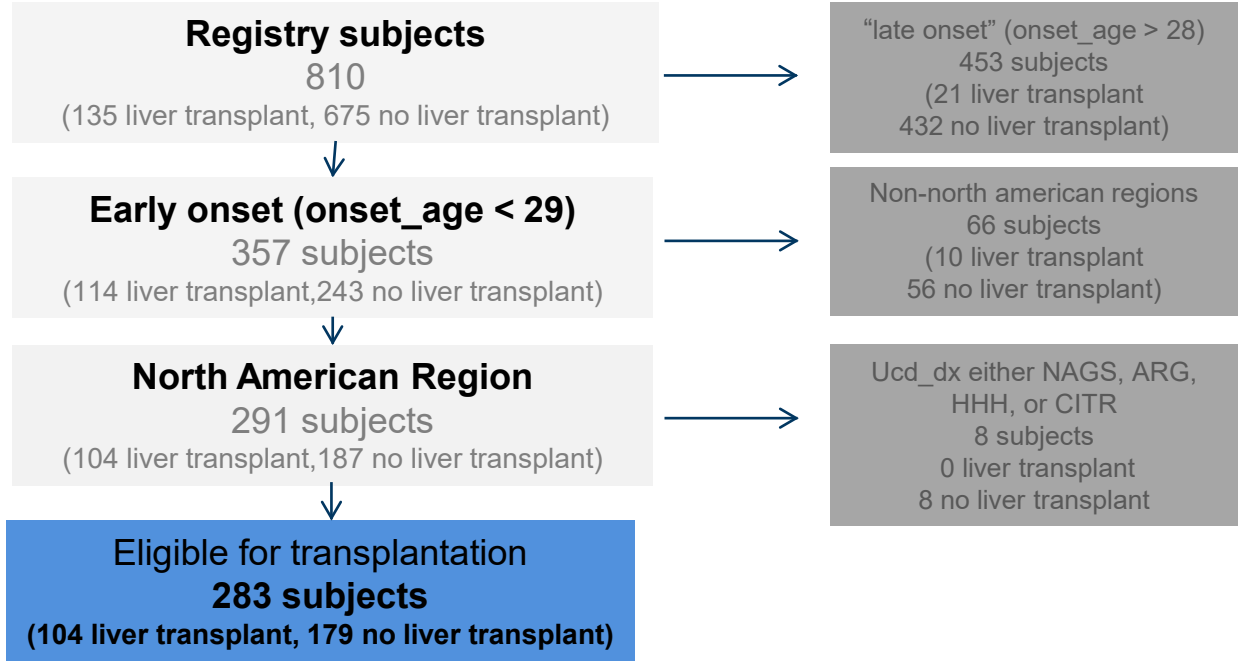
- Multiple randomized longitudinal designs using within-subject comparisons exist
  - Those include: N-of-1 design, sequential randomization (stepped wedged, early withdrawal, delayed therapy)
- Main advantages of longitudinal designs using within-subject comparisons (self-control)
  - Unit of analysis is a [subject x time] unit
  - Within-subject outcome variability typically < between-subject variability
  - Longitudinal data inform natural history considerations
- Important considerations: Timing
  - Example: duration long enough to observe change in outcome? Short enough to assume exchangeability of time periods?

Observational longitudinal designs

A best practice is to design the study to mimic a target randomized study

Cohort study evaluating an intervention

Example: Flow chart from registry to Cohort in UCD study evaluating effectiveness of liver transplantation



Emulating a target randomized trial to decrease biases (e.g., selection, confounding) reduced the analytical cohort size by at least 65%

See Batshaw ML, et al (2014) A longitudinal study of urea cycle disorders. Mol Genet Metab 113(1-2):127–130

Cohort study evaluating an intervention

Example: UCD study evaluating effectiveness of liver transplantation (continued)

- Emulating a target trial was sufficient to eliminate potential sources of bias for some outcomes

  (e.g., Liver transplant was curative in managing hyperammonemia)

- Additional control for confounding (with propensity matching/weighting or risk set matching) and selection bias was necessary and limited the inference population to the comparable groups.

  -> Inference for quality of life and survival limited to “common support group” of medically managed and transplanted.

More details on design and results: Ah Mew N, McCarter R, Izem R, et al. (2020). Comparing Treatment Options for Urea Cycle Disorders. Patient-Centered Outcomes Research Institute (PCORI). https://doi.org/10.25302/12.20.CER.150227816

Single arm with external control
A *challenging* approach, warranted?

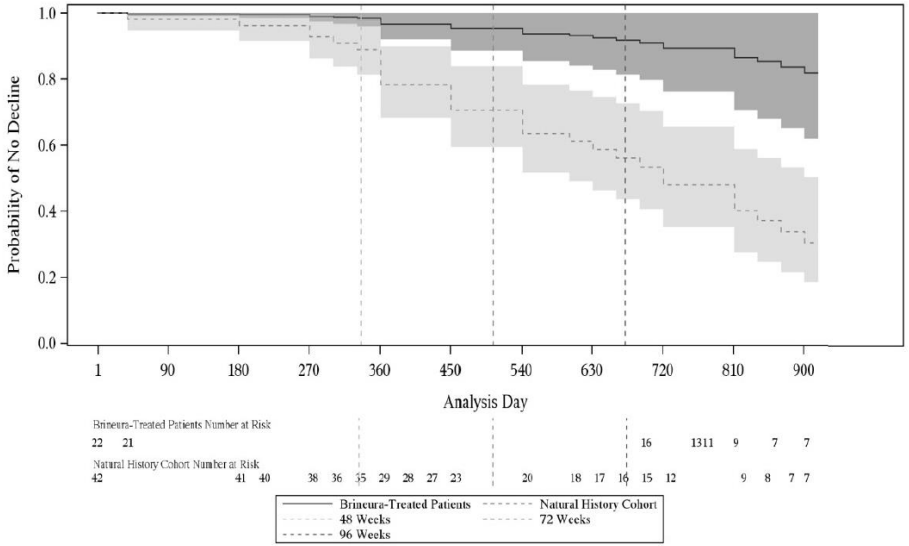
Comparison of findings of a single arm study to an external comparator requires fitness-for-purpose evaluation of database and methods. This includes demonstrating

- Comparability of cohort extracted from the external database and single arm (population, treatment, outcome, frequency of assessments, start and end of follow-up)
- Adequacy of control for confounding: capture of potential confounders in single arm and external control & use of adjustment methods for measured confounding
- Pre-specifying adequately the analytical methods

A review with multiple case studies: Izem R et al (2022). Real-World Data as External Controls: Practical Experience from Notable Marketing Applications of New Therapies. Ther Innov Regul Sci. 2022 Jun 8.

Single arm with external control
Example: Cerliponase Alfa

Figure 7. Estimated Time to Unreversed (Sustained) 2-Category Decline or Unreversed Score of Zero in Motor Domain for Symptomatic Pediatric Patients in the Brineura Single-Arm Clinical Study with Extension and for Patients in a Natural History Cohort (Based on the Cox Proportional Hazards Model Adjusting for Covariates)



To mitigate confounding bias:

- identified potential confounders (e.g., age at diagnosis, baseline motor score and genotype)
- adjusted for confounding using multiple methods (e.g., regression, exact matching)

To mitigate selection bias: emulated a concurrent control with external control (choice of eligibility, inclusion/exclusion, start of follow-up)

Source: Brineura label

RCT vs. observational study

Randomized studies control for **all confounding and** selection through **randomization** and planning

Epidemiological studies control for **multiple sources of bias** using target RCT emulation

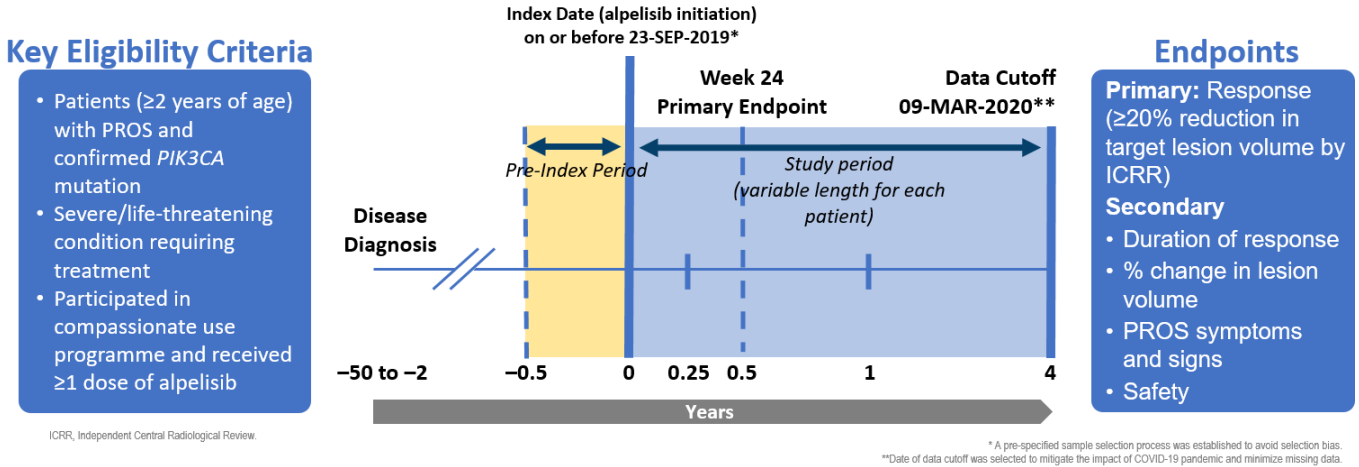
- Parallel Arms, factorial designs ⟷ - Cohort study, use of external control to a single arm, case-control
- Crossover, and N-1 ⟷ - Self-control case series or case-crossover
- Sequential randomization studies ⟷ - Sequential control for confounding

Source: Izem R, McCarter R. Randomized and non-randomized designs for causal inference with longitudinal data in rare disorders. Orphanet J Rare Dis. 2021 Nov 23;16(1):491.

Questions in leveraging longitudinal data

- Can unit of analysis be subject x time rather than subject?
- Can duration of look-back between diagnosis and intervention be exploited?
- Can sequential or time-varying confounding control methods help with assessment?
- Can a self-controlled study (e.g., case crossover or self-controlled case series study) answer the causal inference?

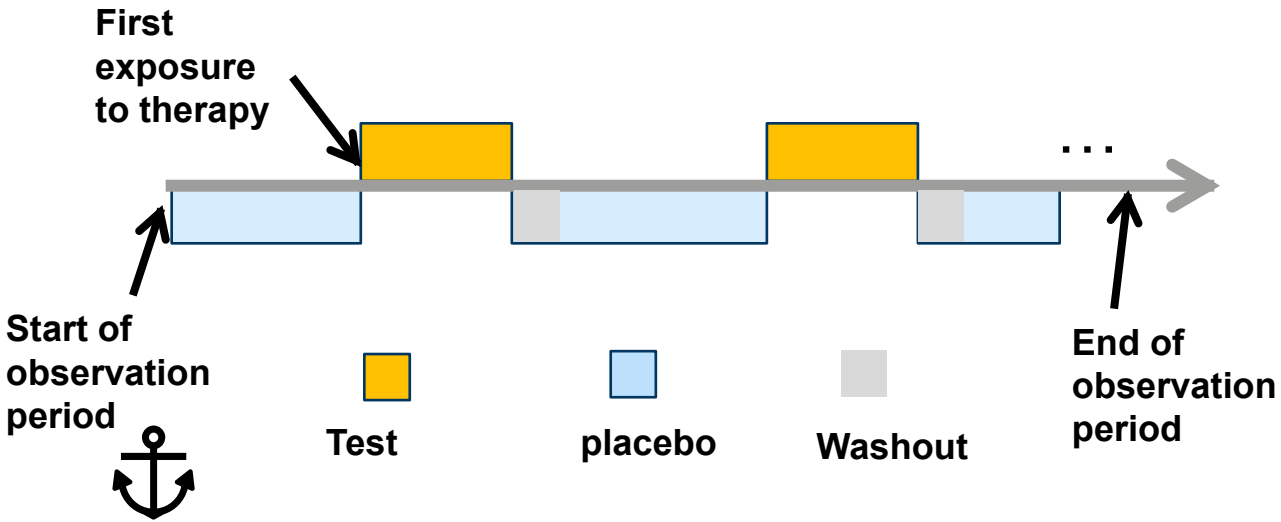
Pre-post self controlled comparison
Alpelisib case study



Select results: By week 24, 37% responders (≥ 20% reduction from baseline in the sum of target lesion volume) for a median length of exposure of 18.1 months

Source: O'Connell P, Ridolfi A, Fretault N (2023) Case study using RWD in the context of a pivotal trial for regulatory approval in a rare disease, Journal of Biopharmaceutical Statistics
More details on design: Canaud, G., et al (2021). LBA23 EPIK-P1: Retrospective Chart Review Study of Patients (Pts) with PIK3CA-Related Overgrowth Spectrum (PROS) Who Have Received Alpelisib (ALP) as Part of a Compassionate Use Programme. *Annals of Oncology* 32:S1297.

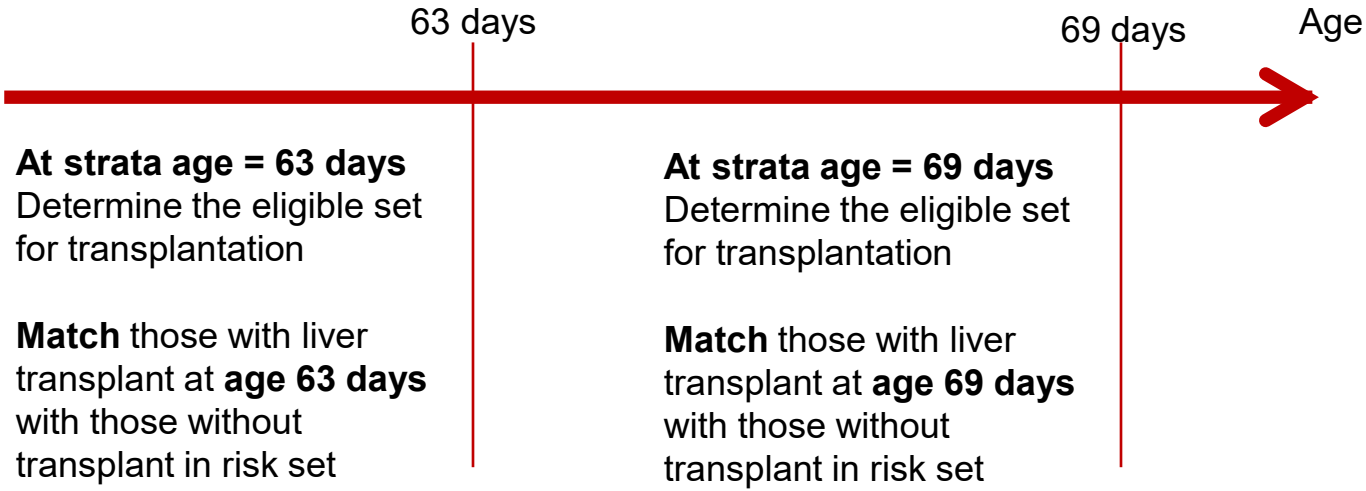
self-controlled case-series
generalizing the pre-post comparison



More details on design and analyses: Whitaker, H.J., et al., *Tutorial in biostatistics: the self-controlled case series method.* Stat Med, 2006. **25**(10): p. 1768-97.

Sequential cohort entry

Example: UCD study evaluating liver transplantation



Ref for risk set matching: Li, Y.F.P., K.J. Propert, and P.R. Rosenbaum, Balanced risk set matching. Journal of the American Statistical Association, 2001. 96(455): p. 870-882.

Summary and further considerations
longitudinal non-randomized comparisons

- All non-randomized comparisons need to minimize or mitigate confounding and selection bias for a valid inference
  - Emulating a target randomized study can improve the design, several analytical methods can further adjust or quantify the impact of for potential sources of bias
- Advantages of longitudinal designs using within-subject comparison (self-control) compared to cohort design/external control comparison
  - Unit of analysis is a [person x time] unit
  - Within-subject outcome variability typically < between-subject variability
  - Bias due to confounding is typically lessened for within-subject comparisons
  - Comparability of time 0 and duration of follow-up built-in by design

Summary and further considerations
longitudinal non-randomized comparisons (contd)

- Important considerations: Timing
  - Example: duration long enough to observe change in outcome? Short enough to assume exchangeability of time periods? Are some sources of bias (such as confounding) time-varying?
- Analytical considerations – use of paired test or hierarchical models for adjustments. For example,
  - Within subject correlation is adjusted in analysis (e.g., paired t-test or McNemar's test)
  - Time-varying confounding (e.g.; age-dependent) is adjusted in analysis
  - Anchor in case-series is first exposure; anchor in case-control is first event; control periods can be before and/or after anchor.

Take home

- Multiple randomized study designs in rare diseases leverage longitudinal data (repeated measures) and within-subject comparison (self-control) to establish efficacy or safety

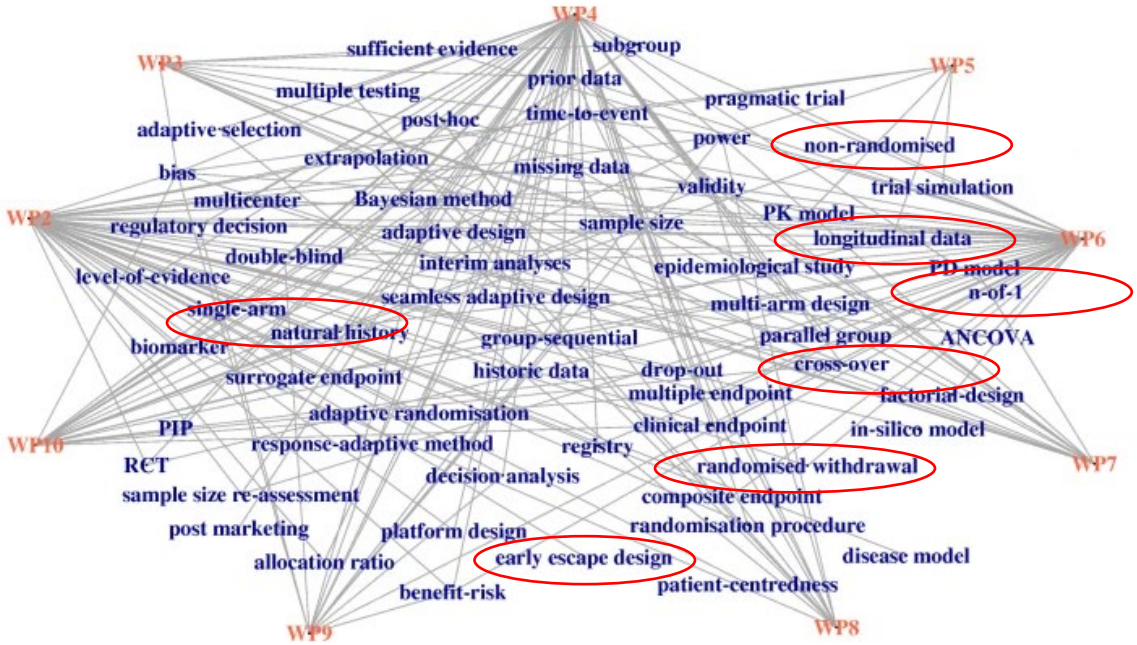
- Multiple observational study designs can also leverage longitudinal data (repeated measures) and within-subject comparison but need to control for potential confounding and selection biases

- Introducing within-subject design and analyses methods in comparative studies has the potential advantages of increasing analyses units, reducing outcome variability, and reducing confounding compared to between-subject comparisons

Acknowledgements

- Children's National Research Institute: Robert McCarter, Urea Cycle Disorders Network, and grant funding from PCORI and NIH
- Berkeley forum: rare diseases working group
- ASA Biopharm pediatric subgroup
- Novartis: teams in Pediatric Center of Excellence at Novartis and analytics team in Alpelisib

Fig. 1



IDeAl-net-1 relating IRDiRC task force report design and analysis topics to IDeAl's work package output

Some keywords in this presentation

Thank you

rima.izem@novartis.com

CERSI
Johns Hopkins University

FDA U.S. FOOD & DRUG ADMINISTRATION

Q&A

Karen Price, PhD
J. Jack Lee, PhD, MS, DDS
Rima Izem, PhD
Frank Harrell, PhD

Expert Biostatistics Advisor to Center for Drug Evaluation and Research, FDA
Professor of Biostatistics, Vanderbilt University

Frank Harrell's Panelist Remarks

Available At:

https://hbiostat.org/talks/cder-jhu.html

CERSI
Johns Hopkins University

FDA U.S. FOOD & DRUG ADMINISTRATION

Concluding Remarks

Dionne Price, PhD

Deputy Director

Office of Biostatistics, Office of Translational Sciences

Center for Drug Evaluation and Research, FDA