Statistical considerations for drug development in rare disease

BASS 2025

November 5, 2025

Emily Morris
Mathematical Statistician
Office of Biostatistics, CDER, FDA

Disclaimer

This presentation reflects the views of the author and should not be construed to represent FDA's views or policies.

Challenges and the role of the statistician

DRUG DEVELOPMENT IN RARE DISEASE

Context

- Very rare disease (prevalence is much smaller than 200,000 in the US)
- Slowly progressive disease
- Uncertainty about natural history of disease
- Heterogeneous disease presentation

Examples: Niemann-Pick Disease Type C (NPC), Fabry disease, Hunter syndrome (MPS II)

Example: Hunter Syndrome

Mucopolysaccharidosis type II (MPS II) or Hunter Syndrome

- X-linked lysosomal storage disorder
- Sugar molecules called GAGs build up within the lysosomes and may cause breathing problems, heart disease, joint abnormalities, seizures, neurological decline
- 'Mild' and 'severe' subtypes; though patients typically fall on a continuum from attenuated to severe
- Onset typically between 2-4 years old and life expectancy is 10-20 years
- Prevalence is 1 in 100,000-170,000 male births

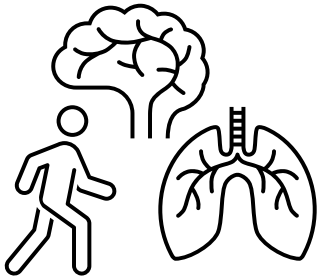
Sources: Hunter Syndrome (MPS II): Symptoms & Causes and Mucopolysaccharidosis type II (Hunter syndrome): a clinical review and recommendations for treatment in the era of enzyme replacement therapy - PMC

Goal

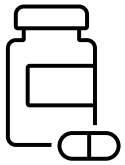
- Consider designing the pivotal study used to demonstrate evidence of efficacy. We aim to have a conclusive answer at the end of the trial whether a drug has a meaningful effect or not.

- More specifically we care about:

Who benefits?



What is impacted?



Optimal dose

Challenges

- Small number of patients

  → Limited ability to increase sample size

Challenges

- Small number of patients

  → Limited ability to increase sample size

- Heterogeneous disease

  → How to select a sensitive endpoint?

  → What should the target population be?

Challenges

- Small number of patients
  - → Limited ability to increase sample size
- Heterogeneous disease
  - → How to select a sensitive endpoint?
  - → What should the target population be?
- Unknown natural history
  - → How heterogenous is the disease?
  - → How quickly does the disease progress?
  - → What endpoints impact patients the most?

How can statistics help?

- Statisticians are critical for ensuring that studies are adequately powered under reasonable assumptions
- Simulations can help inform
  - Endpoint (compare continuous vs categorical/dichotomized)
  - What analysis methods can maximize power?
  - What sample size is needed to detect a reasonable treatment effect size?

Challenge: Small Sample Size

Consider using a cross-over design

- How long do each of the treatment periods need to be to detect a treatment effect?
- How long of a washout period is needed?
- Is functional unblinding a concern?
- Is it reasonable for patients to go off treatment and restart later?

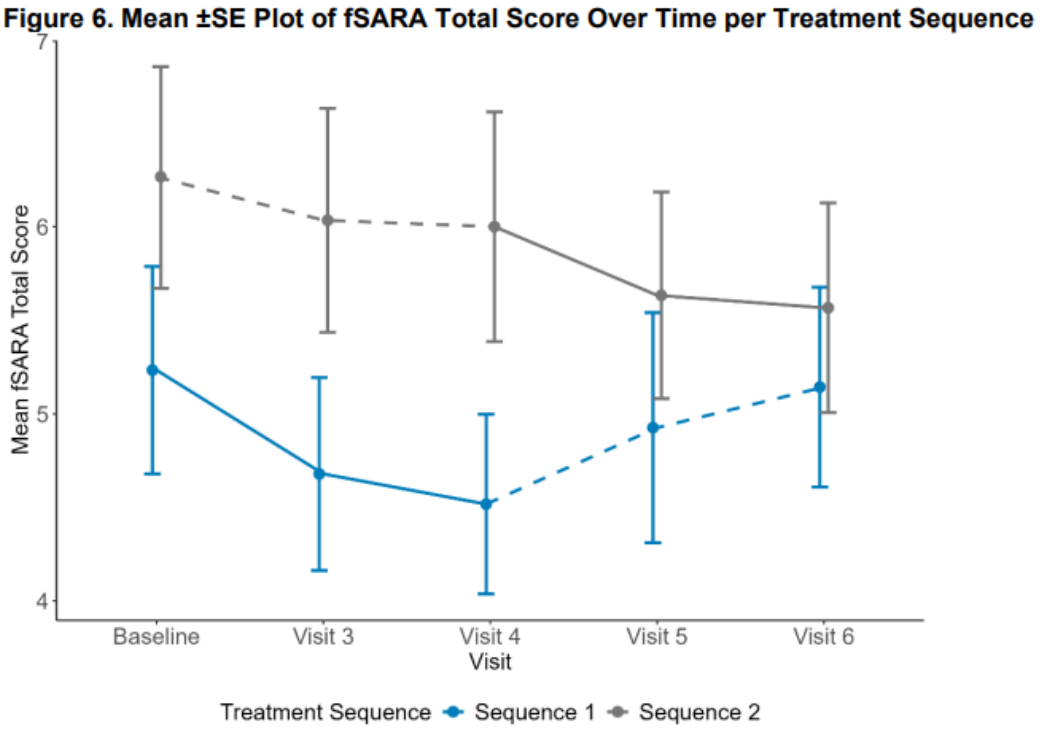
Example: NPC

Aqneursa (levacetylleucine)

- Design: randomized, double-blind, placebo-controlled cross-over design of 60 subjects for the treatment of Niemann-Pick disease type C
- Primary endpoint: modified Scale for Assessment and Rating of Ataxia (SARA)
- Treatment duration: 12 weeks per treatment
- Target population: mild neurological symptoms at baseline

Example: NPC

- Modified SARA measures gait, sitting, stance, and speech disturbance (scored 0 to 16)
- 30 patients per sequence, but we have data from both placebo and test product from all 60 subjects



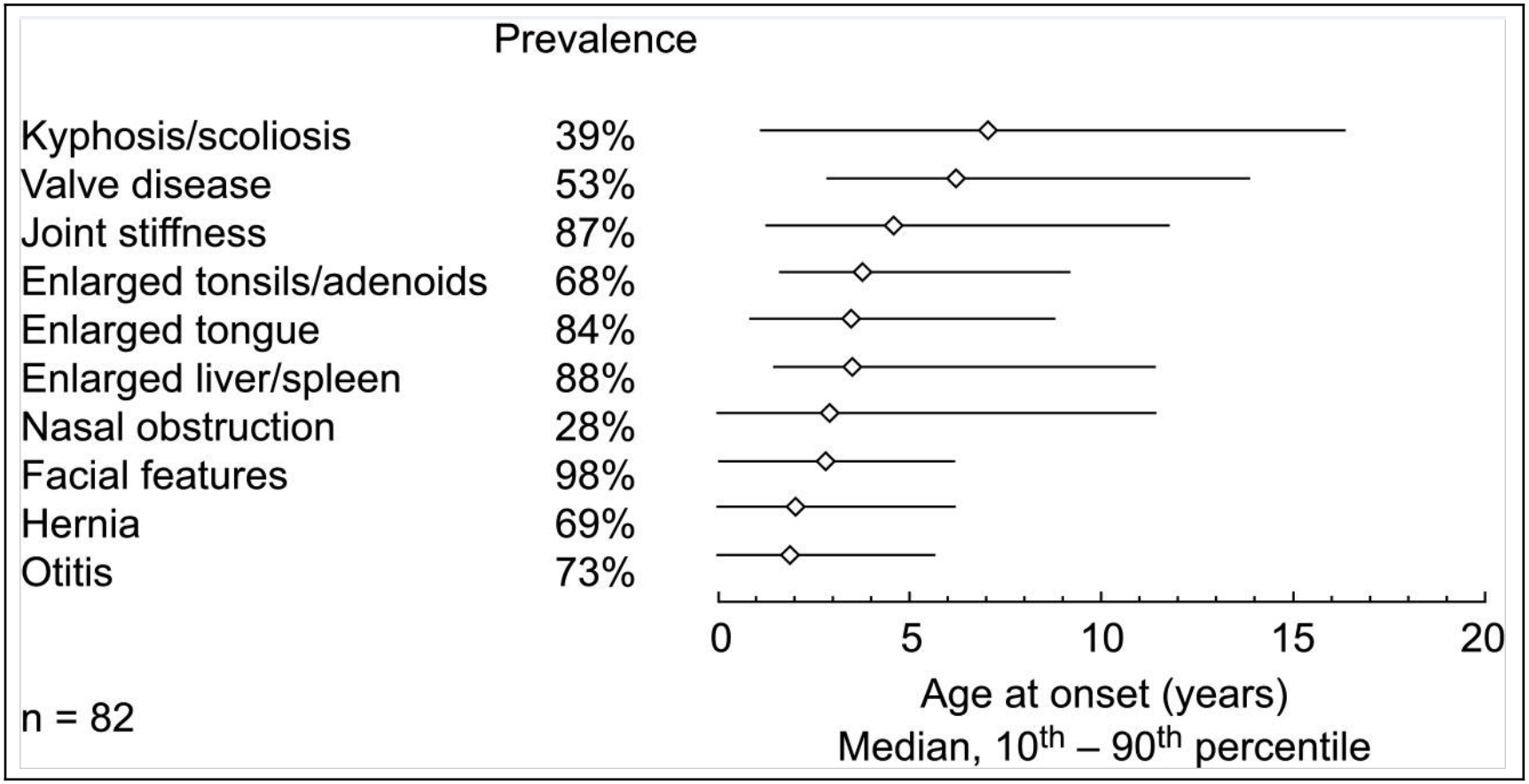
Sequence 1 = received levacetylleucine in period 1 and placebo in period 2
Sequence 2 = received placebo in period 1 and levacetylleucine in period 2

Source: https://www.accessdata.fda.gov/drugsatfda_docs/nda/2024/219132Orig1s000IntegratedR.pdf

Challenge: Heterogeneous Disease

- Consider using multiple endpoints
  - Would a global test increase power?
- Consider a design where different populations have different endpoints
  - How to most efficiently combine into one test?
- Incorporate longitudinal data
  - How much is gained by including multiple timepoints?

Example: Hunter Syndrome



Source: Wraith, J. Edmond, et al. "Mucopolysaccharidosis type II (Hunter syndrome): a clinical review and recommendations for treatment in the era of enzyme replacement therapy." European journal of pediatrics 167.3 (2008): 267-277.

Example: Hunter Syndrome

Elaprase (idursulfase)

- Design: randomized, double-blind, placebo-controlled trial
- Primary endpoints: change from baseline to week 53 in 6-minute walk test distance walked and %-predicted forced vital capacity (FVC)
- Primary analysis: O'Brien rank-sum global test*
- Target population: %-predicted FVC <80%, age ranged from 5 to 31 years old

*O'Brien, Peter C. "Procedures for comparing samples with multiple endpoints." *Biometrics* (1984): 1079-1087.
Source: https://www.accessdata.fda.gov/drugsatfda_docs/label/2018/125151s197lbl.pdf

Example: Hunter Syndrome

O'Brien global rank-sum test procedure:

1. Rank each endpoint across treatment groups
2. Sum across ranks within subjects
3. Test for difference in ranks between treatment groups using ANCOVA with baseline age group and baseline disease score as covariates

Challenge: Unknown Natural History

Consider using an adaptive design

- Using accrued information to adapt features may help overcome knowledge gaps at the design phase

ADAPTIVE DESIGNS FOR RARE DISEASE TRIALS

ICH E20 Draft Guidance

- Adaptive design

  "a clinical trial design that allows for **prospectively planned modifications** to one or more aspects of the trial based on interim analysis of accumulating data from participants in the trial"

- **Pros:** Helpful when uncertainty may remain at the design stage, can minimize exposure to an inferior product by stopping early
- **Cons:** added complexity, risk for trial integrity

Source: ICH E20 draft guidance from June 2025

Adaptations

Options:

- Sample size
- Population selection
- Treatment/dose selection
- Treatment duration

Sample Size Adaptation

Conduct an interim analysis to determine whether the final sample size should be increased

- Comparative or non-comparative options
- Common approach: promising zone using conditional power (CP)

Helpful when there is uncertainty at the design stage about estimates of treatment effect that impact sample size estimation (effect size, variability of outcome, placebo response rate, etc.)

Sample Size Adaptation: Conditional Power

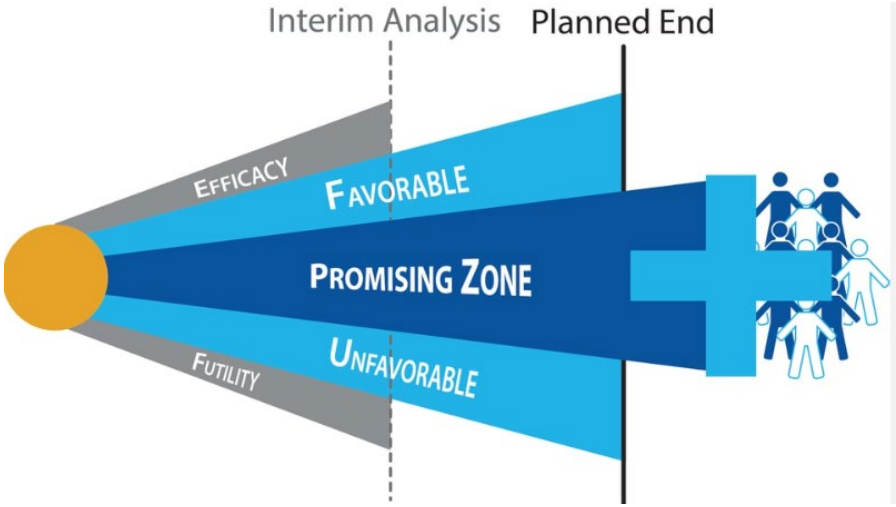
$$\text{Conditional power} = \Pr(Z_2 \geq z_\alpha \mid Z_1 = z_1)$$

The probability of rejecting the null hypothesis at the final analysis given the test statistic (z_1) at the interim analysis, assuming that the treatment effect estimated at the interim analysis is the true effect

Sample Size Adaptation: Conditional Power

Interim analysis decisions:

- If conditional power is promising then increase sample size to achieve sufficient conditional power
- If CP is favorable or unfavorable then continue to planned sample size
- May also stop for efficacy or futility



Source: Cytel article "Operational and regulatory considerations in a promising zone trial" https://www.cytel.com/blog/regulatory-and-operational-considerations-promising-zone

Population Selection

Conduct interim analysis to determine whether the trial should be limited to a specific subpopulation.

Helpful where there is uncertainty about which patients are likely to benefit from treatment:

- If only effective in a subpopulation but tested in a broader population, this may lead to an underpowered study
- If only tested in a subpopulation but effective in a broader population, this would unnecessarily restrict the indication

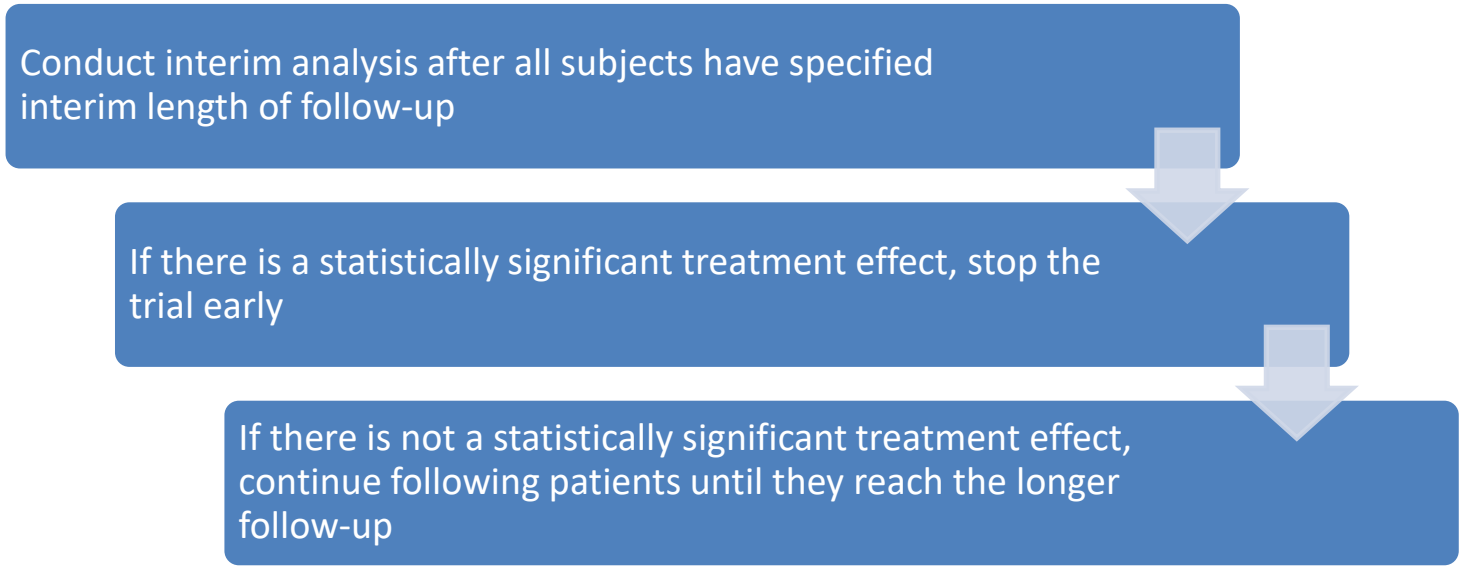
Treatment/Dose Selection

Compare multiple test products or multiple doses of one product to a control at the interim analysis with the intent of selecting the most promising product for the remainder of the trial

Helpful when there is uncertainty about the benefit-risk profile of more than one treatment

Adaptive Treatment Duration

Conduct an interim analysis to evaluate whether the length of follow-up will be sensitive to detecting a treatment effect



Helpful when there is uncertainty about how quickly the disease progresses with or without treatment

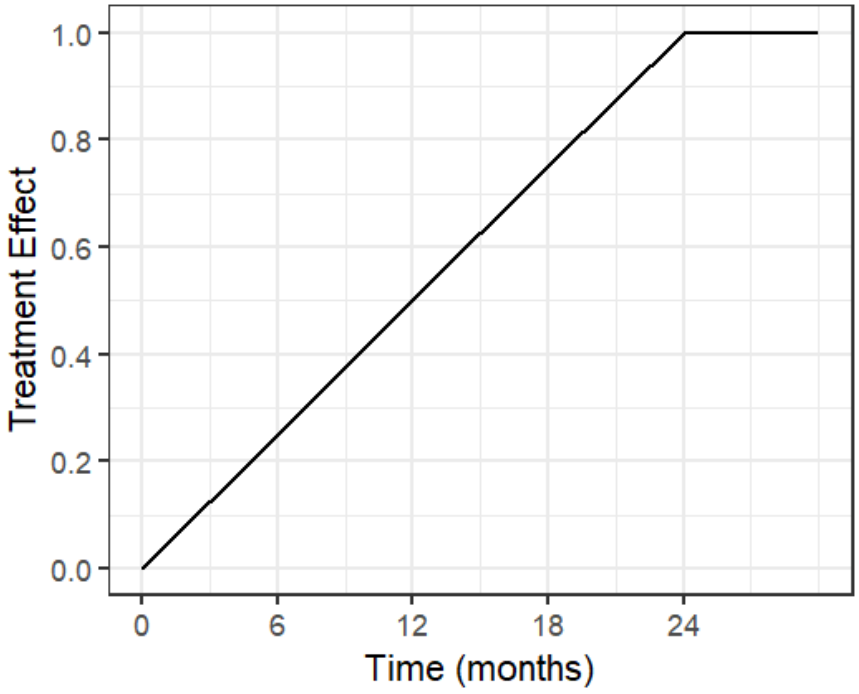
Adaptations

Options:

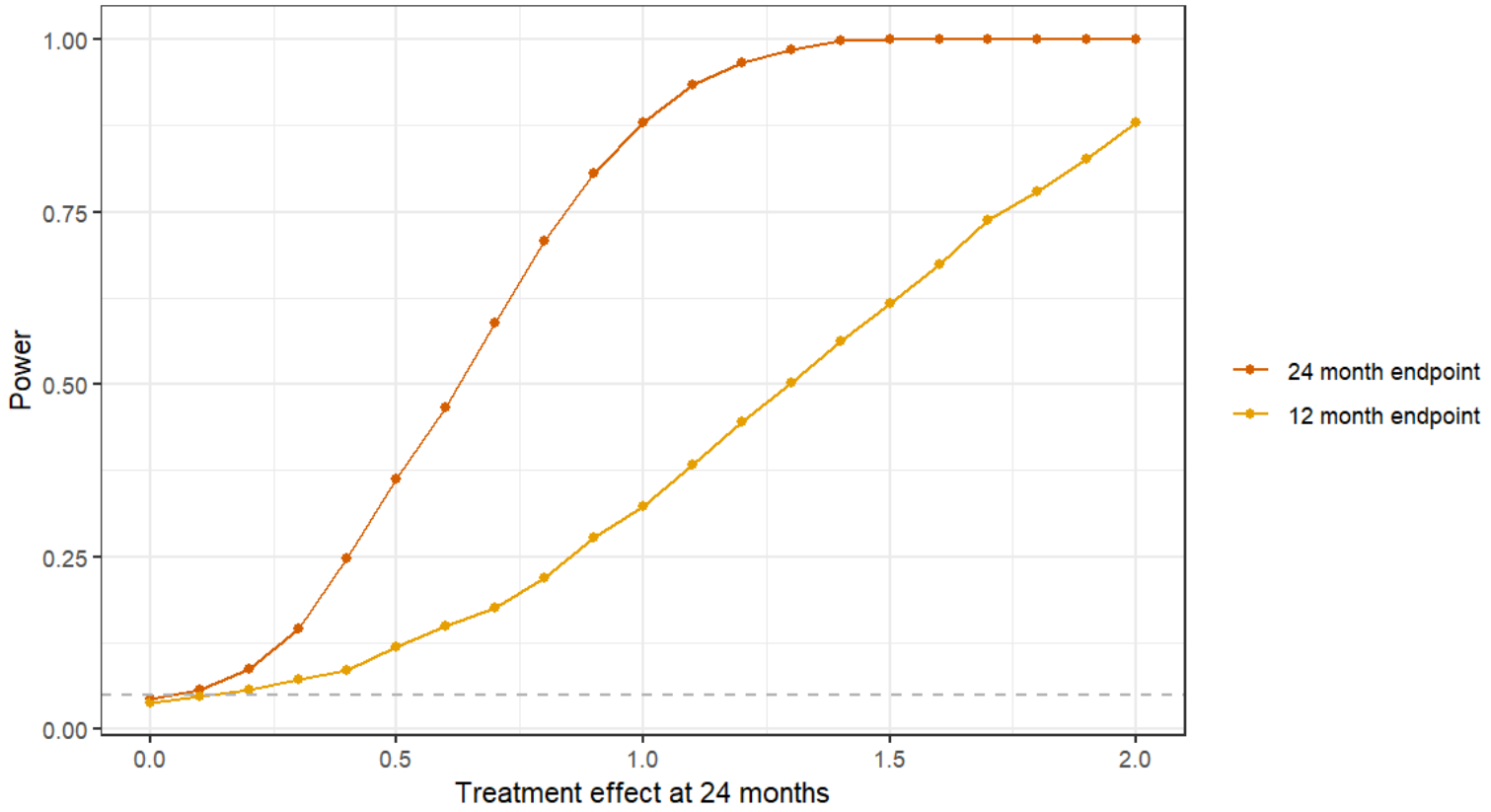
- Sample size ←may be difficult to increase
- Population selection ←may not have enough subjects to enrich for a subpopulation
- Treatment selection ←may not apply
- Treatment Duration

Example Setting

- Sample size = 40
- 1:1 randomization to placebo or study drug
- Endpoint: continuous
- Assume a linearly increasing treatment effect over time followed by plateau
- Treatment duration?



Fixed time at 1 or 2 years



If the treatment effect is 1.0 after 2 years of treatment, the 1-year trial will only have power of 33% and the 2-year trial will have power of 88%

Adaptive Duration (AD)

Interim analysis: 1-year outcome

Final analysis: 2-year outcome

- Outcomes are different but likely correlated
- Multiplicity issue

How to account for multiplicity?

Group Sequential Design (GSD)

GSD: Conduct one or multiple interim analyses after a prespecified proportion of subjects have final endpoint (information fraction)

- Tests the same endpoint
- Allows for a more efficient use of time and resources by potentially ending the trial sooner and reducing patient exposure to inferior treatments

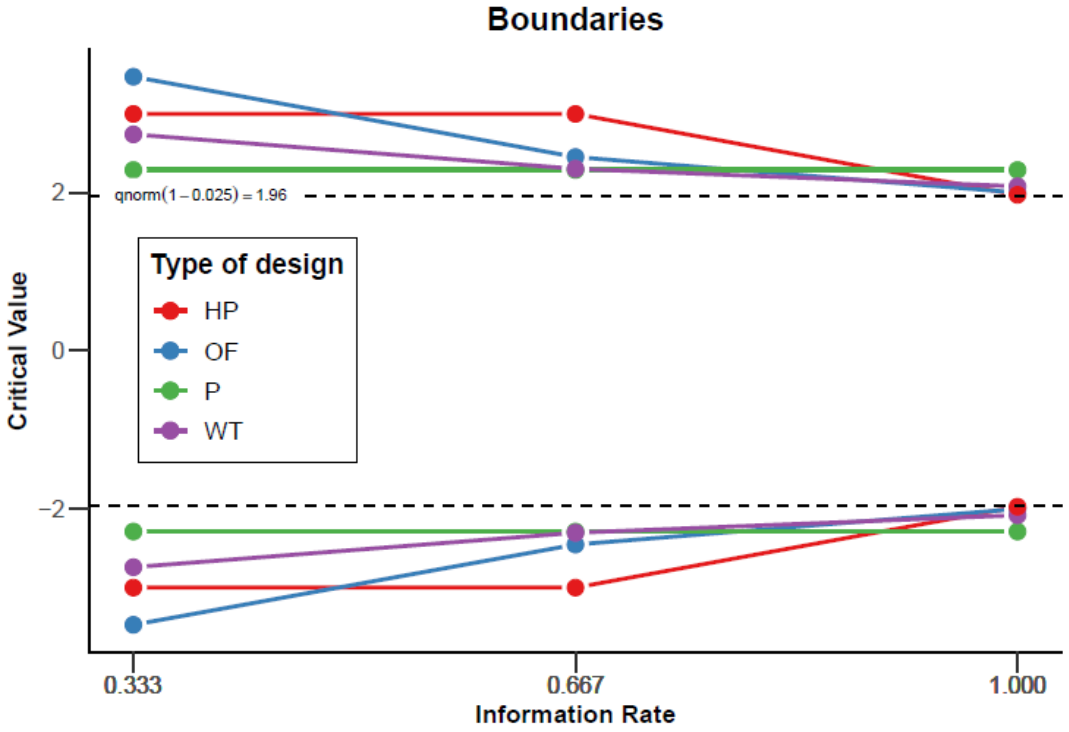
Well established approaches to adjust for multiplicity in GSD

- O'Brien Fleming (spends small amount of alpha at the interim analysis)
- Pocock (spends alpha evenly at each analysis)

Group Sequential Design

Different approaches to select the stopping boundary for efficacy

- O'Brien-Fleming is a common choice which requires very persuasive early results to stop
- Pocock boundary has higher probability of early stopping
- Alpha-spending is a more flexible approach (flexible timing and frequency of interim analyses)



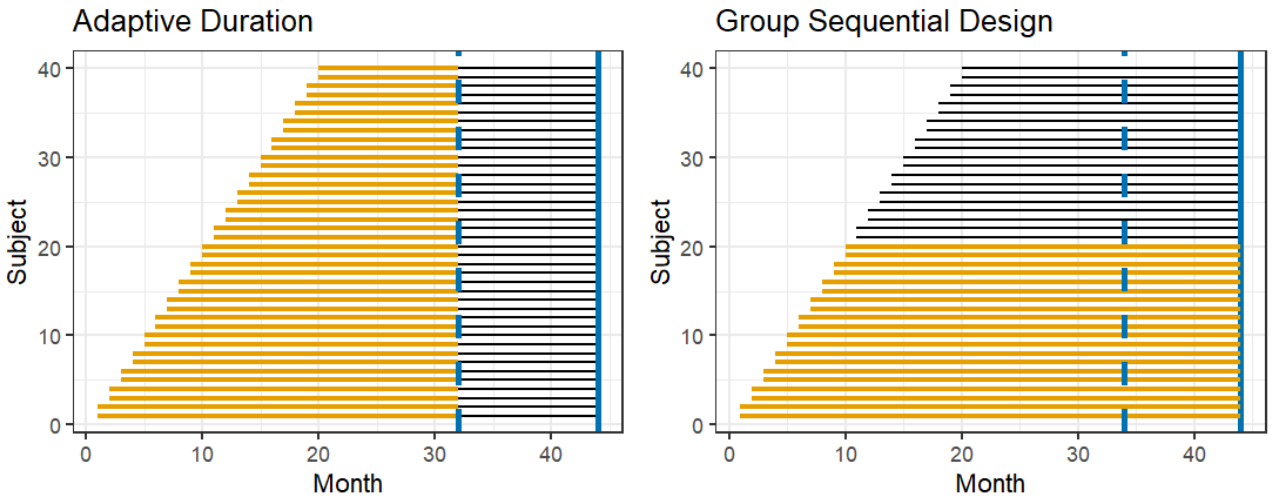
Source: Lakens, Daniel, Friedrich Pahlke, and Gernot Wassmer. "Group sequential designs: A tutorial." (2021).

GSD Boundaries

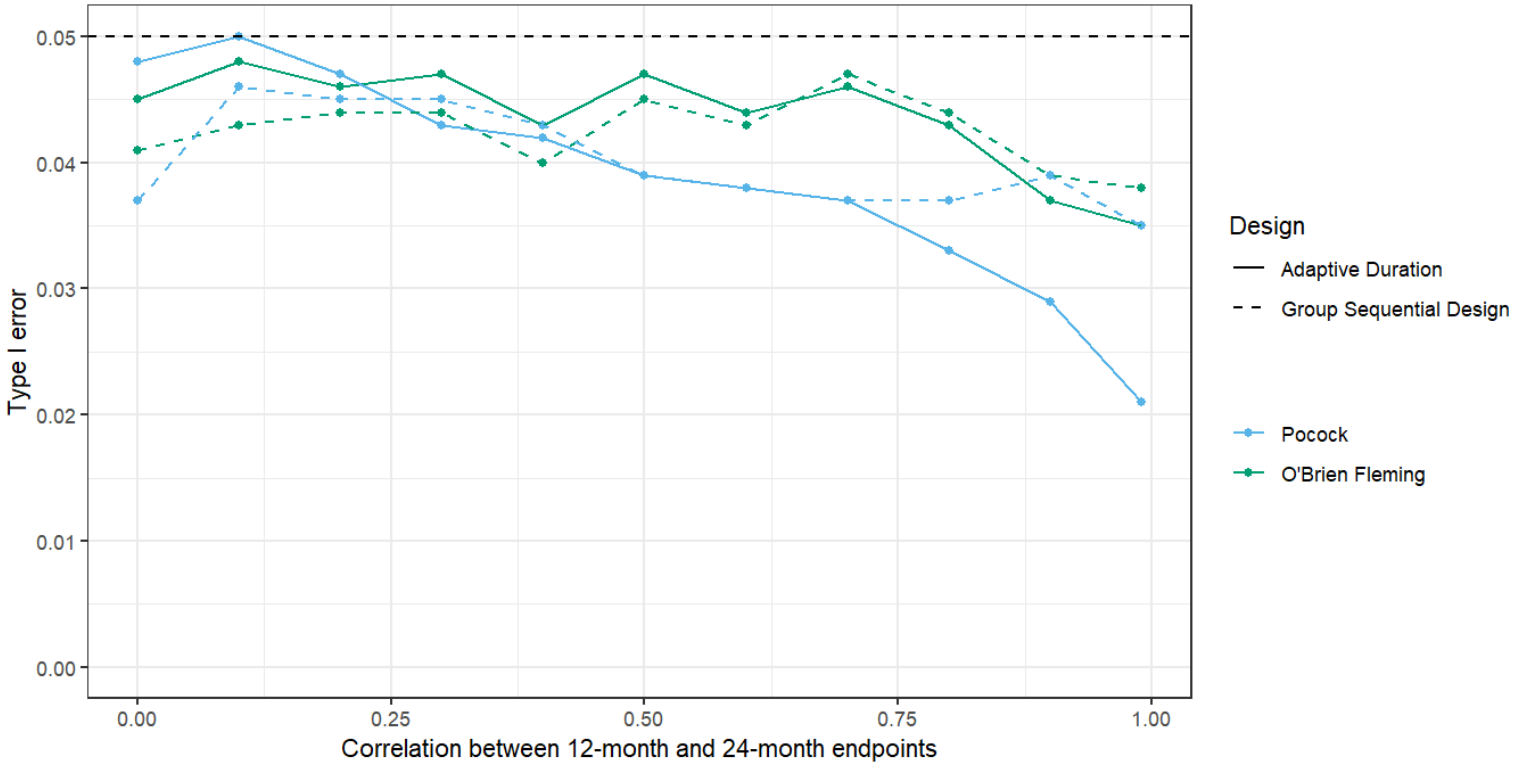
Can we apply similar rules to adjust alpha at the interim analysis when adapting the duration?

- Information fraction is now based on length of follow-up per subject for all subjects

- Enrollment rate = 2 subjects/month
- Interim analysis after 12 months
- Final analysis after 24 months



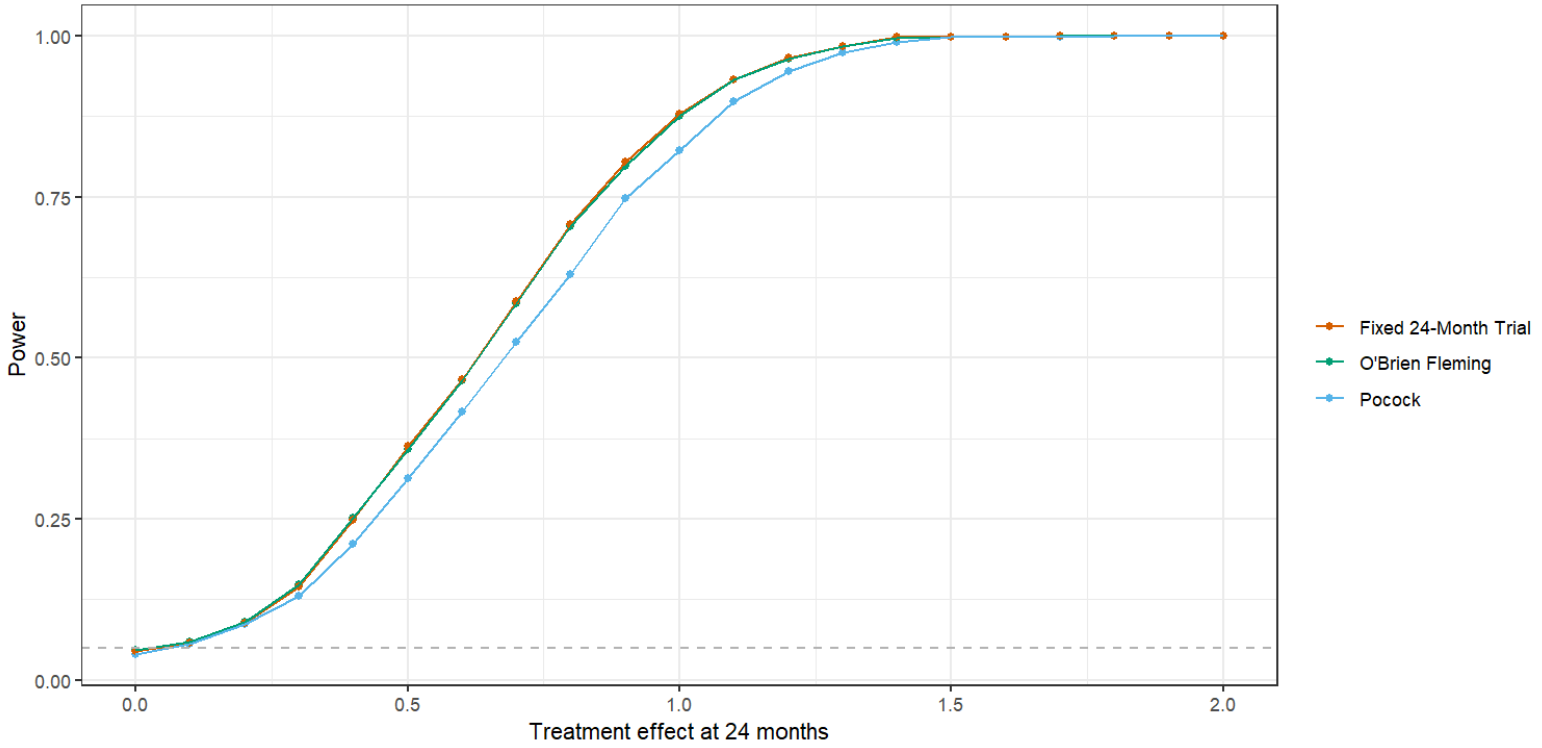
Type I error vs correlation



- Appears to control type I error for adaptive duration design in this setting
- Simulations are needed for the specific context to demonstrate that type I error is not inflated; may need to consider other types of endpoints, different sample size, etc.

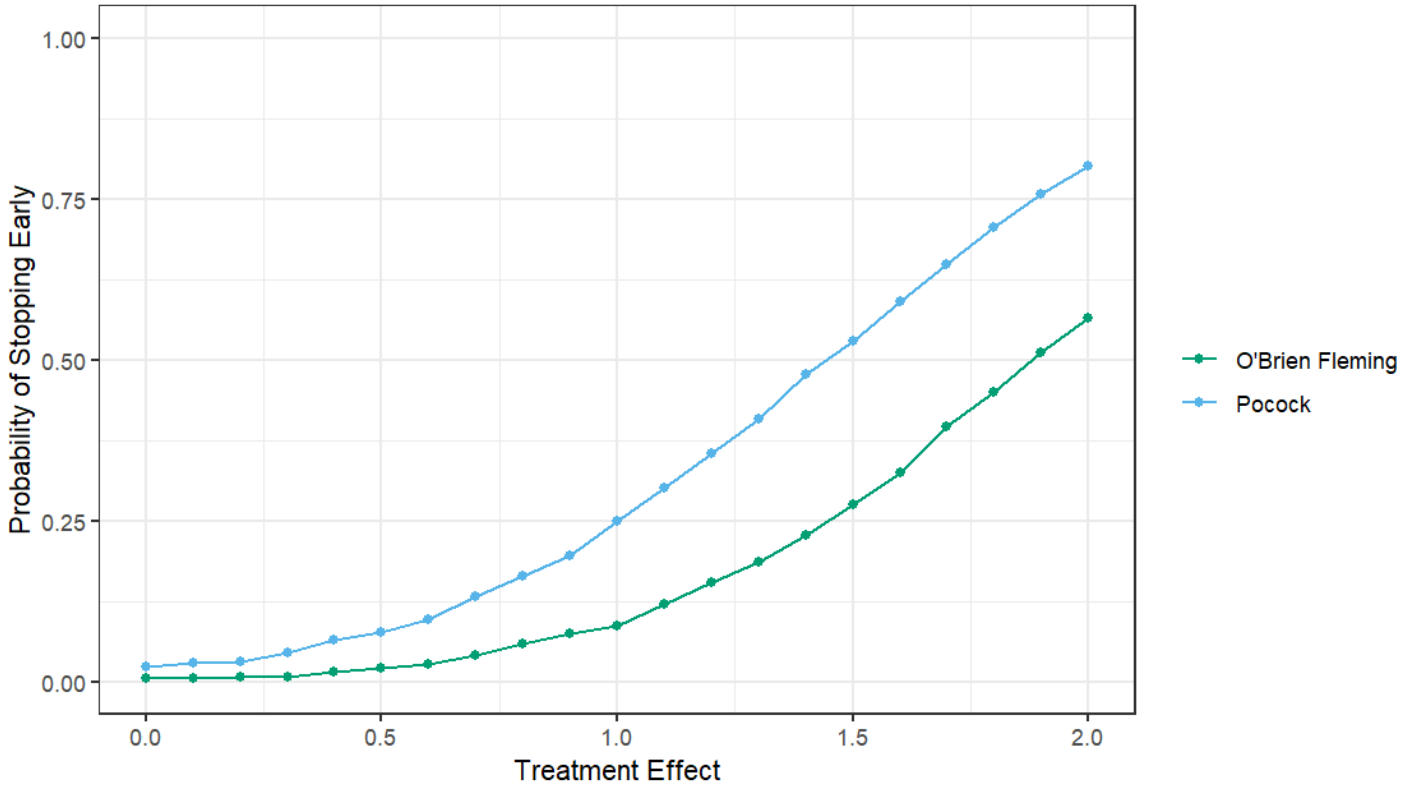
Power vs treatment effect

- Power of O'Brien Fleming approach is similar to that of the fixed 24-month trial
- Pocock approach has slightly lower power

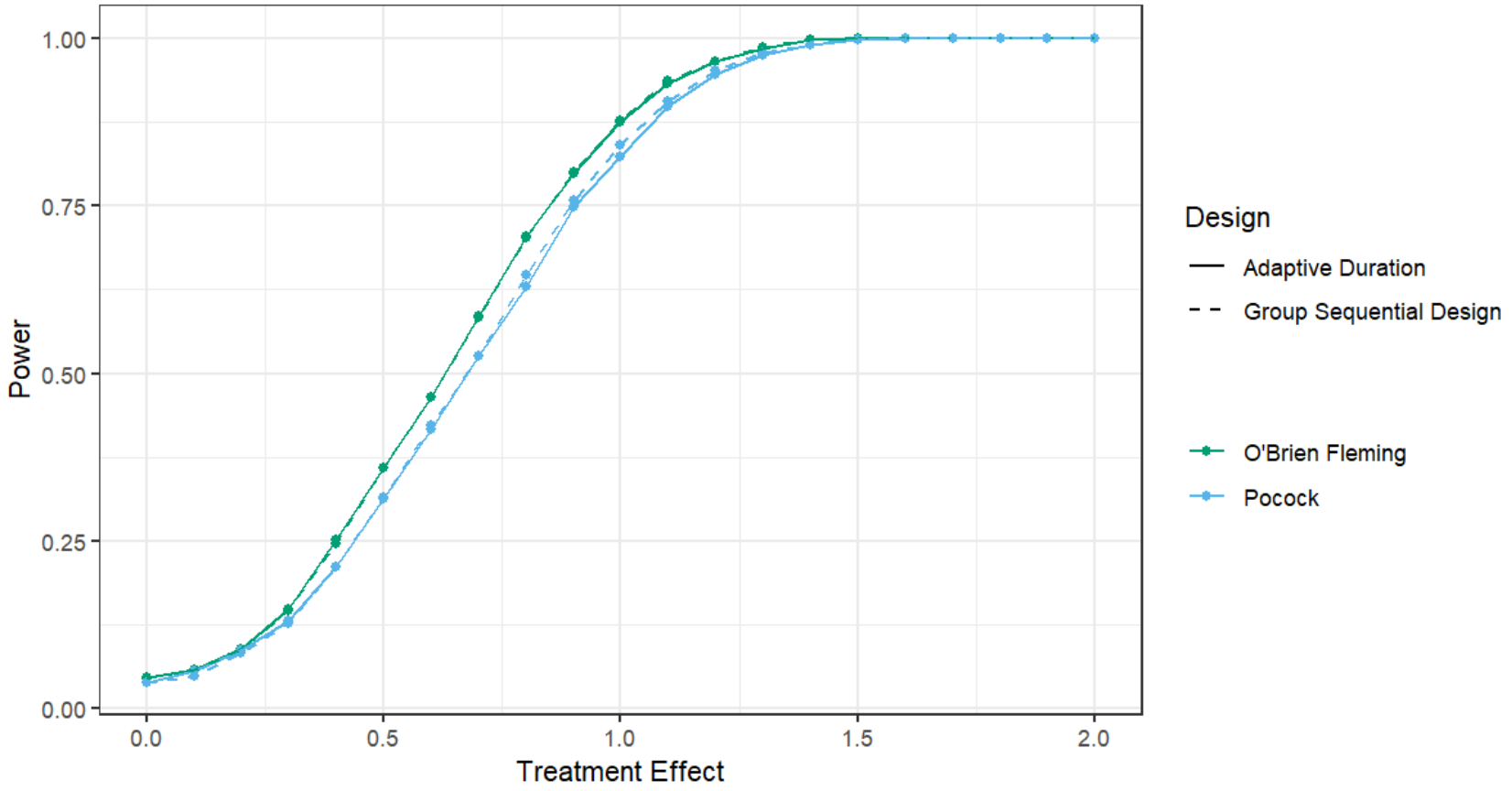


Probability of stopping early

- Pocock has much higher probability of stopping early compared to O'Brien Fleming approach (25% when the treatment effect is 1.0 at 2 years compared to 9%)



Compare power of AD to GSD



Simulation Takeaways

- Operating characteristics
  - Adaptive duration with adjustments used in group sequential design appears to have type I error controlled for this setting
  - Power for O'Brien Fleming AD design similar to that of the fixed duration trial at the maximum length
  - Power for GSD and AD are similar, though GSD may be slightly more powerful

Considerations

- Characteristics that may change these conclusions
  - Correlation between earlier and later timepoints
  - Different analysis methods
  - Combining features to increase power such as multiple endpoints or using longitudinal data at each analysis
- Factors to consider with this design
  - Ability to conduct the longest fixed duration trial vs adding an interim analysis
  - Weigh the desire to stop early with overall trial power
  - Enrollment rate

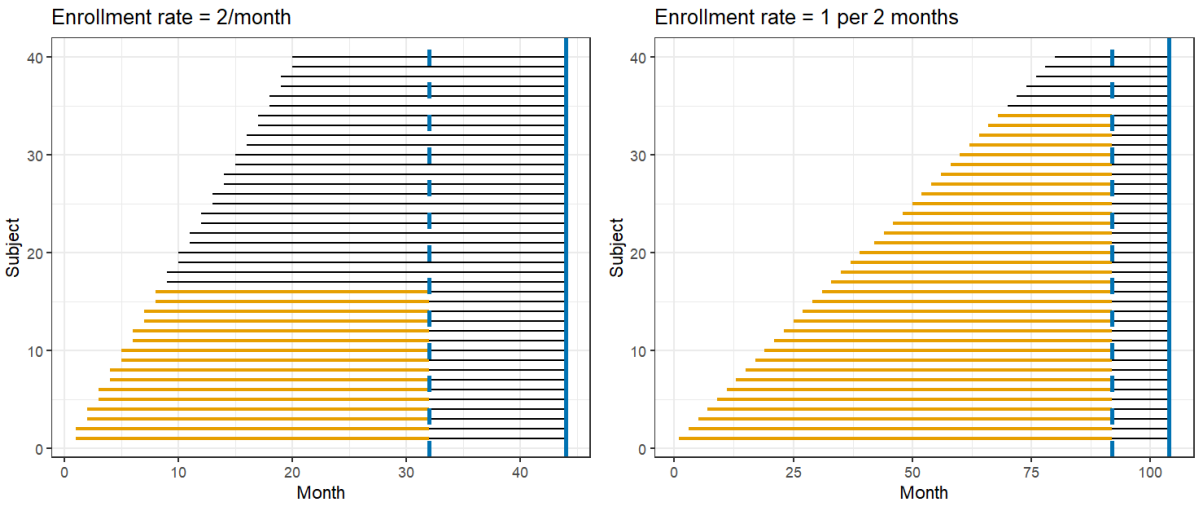
Enrollment Rate

Enrollment rate = 2 subjects per month

- Interim analysis occurs 32 months after first subject is enrolled
- 40% of subjects have complete data at the interim analysis

Enrollment rate = 1 subject every 2 months

- Interim analysis occurs 92 months after first subject is enrolled
- 85% of subjects have complete data at the interim analysis



Additional Considerations for Adaptive Designs (from ICH E20)

- Maintaining trial integrity will be more challenging
  - Knowledge of adaptation may influence behavior
  - Ensure there is a clear rule for adaptations and that IDMC can make adaptation recommendations
- Ensure sufficient information for full benefit-risk analysis
- Estimates may be highly variable if interim analysis is too early or based on too little information and lead to an erroneous decision
- Limit complexity of adaptations in confirmatory trial

Conclusion

- Conduct simulations to evaluate operating characteristics of any non-standard design
- Consider aspects that could be adapted where there is an information gap at the design phase
- Case-by-case considerations
  - Will be more complex if you use multiple endpoints or have multiple interim analyses

Sources

- ICH E20 Draft Guidance Adaptive Designs for Clinical Trials
- O'Brien, Peter C. "Procedures for comparing samples with multiple endpoints." *Biometrics* (1984): 1079-1087.
- Shives E, Gurmu Y, Lee W, Morris E, Wang Y. Novel Clinical Trial Design With Stratum-Specific Endpoints and Global Test Methods for Rare Diseases With Heterogeneous Clinical Manifestations. Stat Med. 2025 Aug;44(18-19):e70206. doi: 10.1002/sim.70206. PMID: 40772797.
- Wraith, J. Edmond, et al. "Mucopolysaccharidosis type II (Hunter syndrome): a clinical review and recommendations for treatment in the era of enzyme replacement therapy." *European journal of pediatrics* 167.3 (2008): 267-277.
- MPS II Cleveland Clinic: Hunter Syndrome (MPS II): Symptoms & Causes
- Chen, YH Joshua, David L. DeMets, and K. K. Gordon Lan. "Increasing the sample size when the unblinded interim result is promising." Statistics in medicine 23.7 (2004): 1023-1038.
- Mehta, Cyrus R., and Stuart J. Pocock. "Adaptive increase in sample size when interim results are promising: a practical guide with examples." Statistics in medicine 30.28 (2011): 3267-3284.

Thanks

Joint work with:

Rebecca Chiu, Greg Levin, Emily Nguyen, and Yan Wang

Questions?

Thank you for listening!


FDA
U.S. FOOD & DRUG
ADMINISTRATION