Some Considerations in Late-Stage Oncology Trial Design

Keaven M. Anderson[1]

[1]Merck Research Laboratories keaven_anderson@merck.com

April, 2014

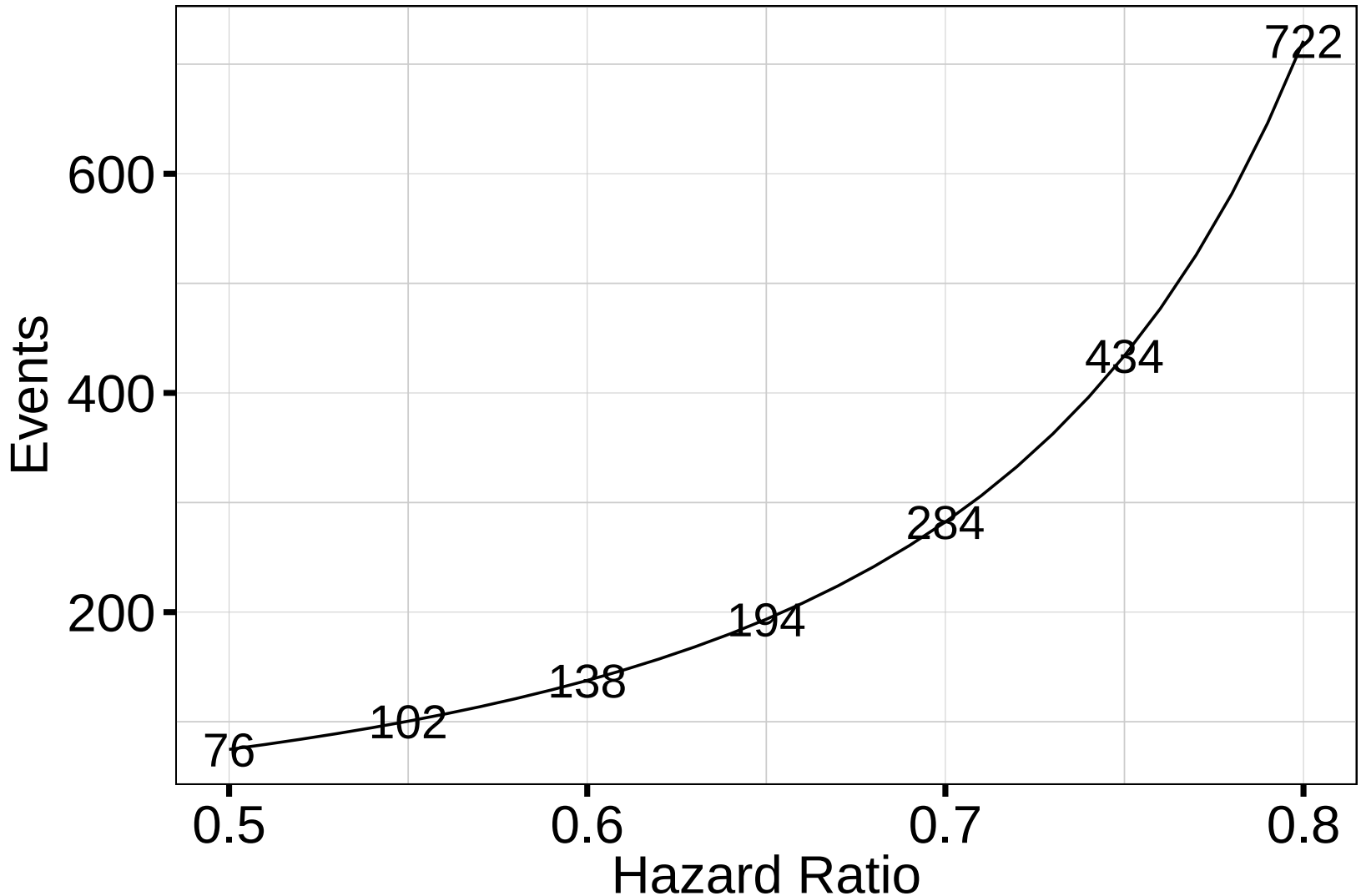
Disclaimer and Background

- While the content of this presentation reflects discussions of studies with regulators, it represents thoughts of the author and is intended to promote discussion
- Examples represent ongoing work and input is welcome
- Numeric examples generated using gsDesign R package available at CRAN; web interface available at `http://spark.rstudio.com/gsDesign/prod`
- Probability of success is related to other works, such as Chuang-Stein [2006], Stallard et al. [2005]

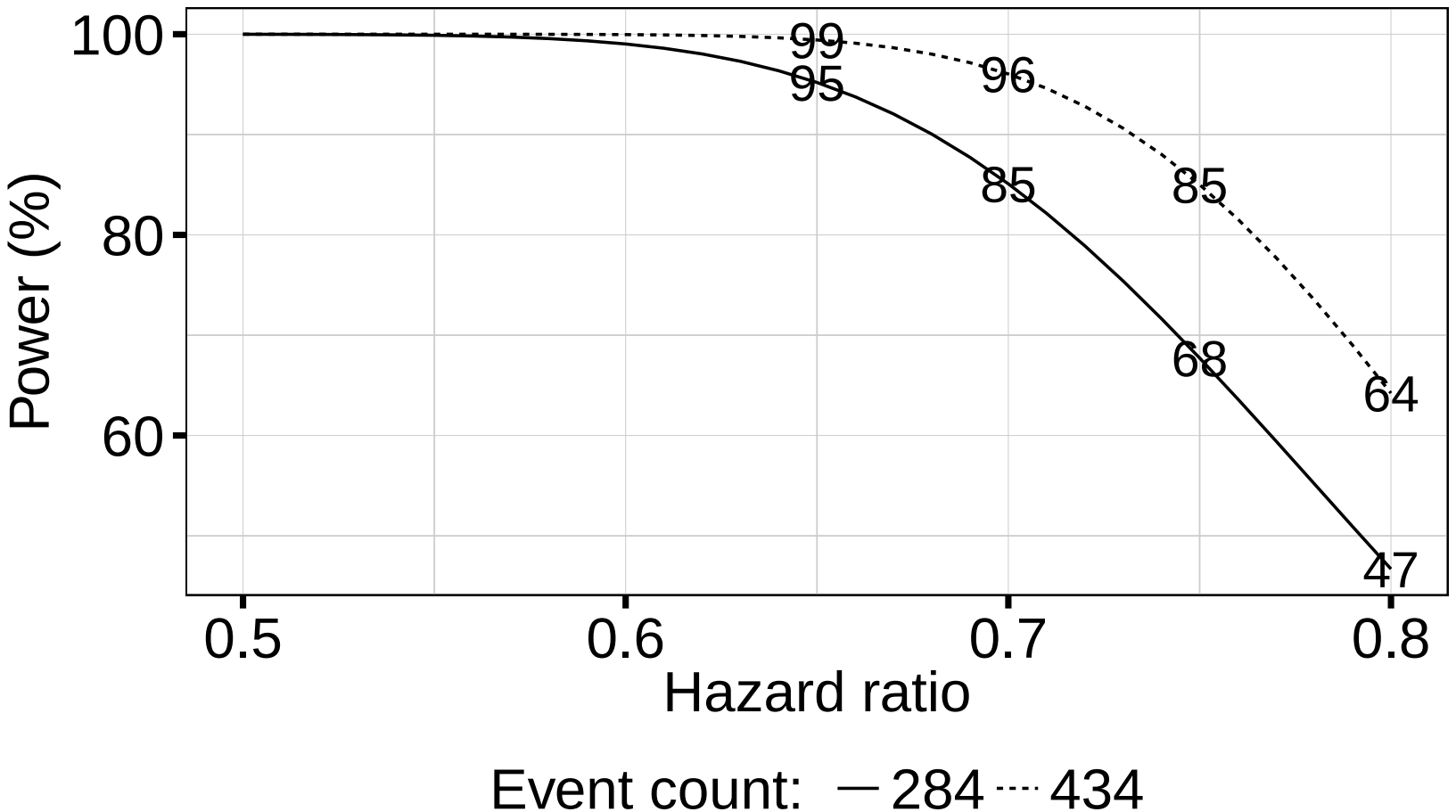
Topics of interest

- Phase 2 and Phase 3 sample size: probability of success considerations
- When should a trial stop early?
  - Setting appropriate endpoints, bounds and timing for analyses
  - Ethical and bias issues
  - Heterogeneity and missing data
- Designing interim analysis plans with two endpoints
- Designing interim analysis plans with biomarker subgroups (limited)
- Accelerated approval and full approval at different stages of a trial

Events Required for 85% Power by Hazard Ratio



Power by underlying hazard ratio for event counts of 284, 434



OS Example; Source: Barclays Capital/Huang, April 4, 2014

"**Amgen: T-Vec (Phase III/Melanoma) Narrowly Missed OS; Commercial Opportunity Limited for Monotherapy...** Amgen (AMGN) announced ... the primary overall survival (OS) analysis of the Phase 3 pivotal trial evaluating T-Vec for advanced melanoma ... (p=0.051). The estimated OS hazard ratio and improvement in median OS were similar to previously reported interim OS analysis. The interim analysis showed OS of 23.3 mos vs. 19.0 mos with HR=0.79... Commercial opportunity is limited for T-vec Monotherapy in melanoma...
Real opportunity for T-Vec may exist in combination..."

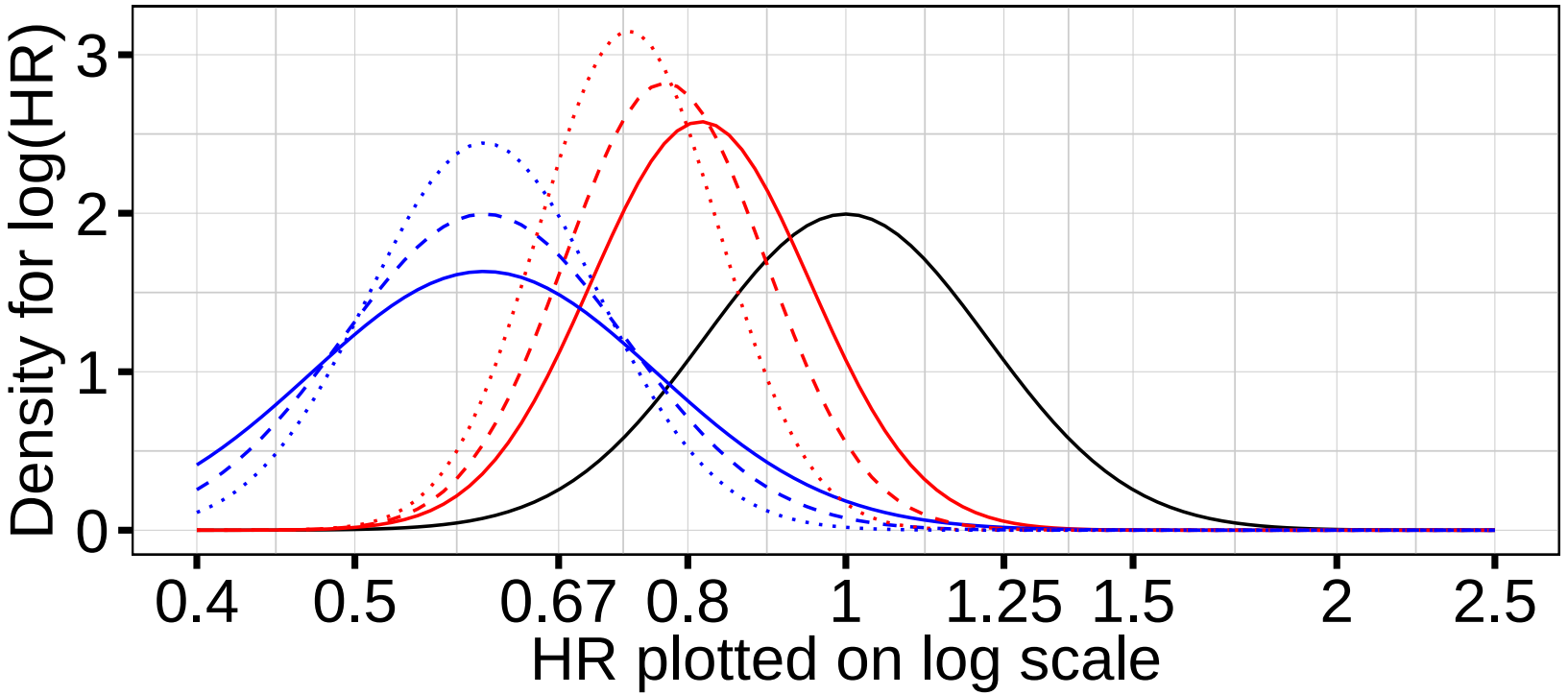
- Trial has positive finding for durable response
- Was greater investment justified, or was spreading investment across other opportunities better?

Phase 2 Decision Making Incorporating Prior Distribution

- We assume trial is first indication for a new drug and mechanism of action
- After a positive Phase 2 study, many Phase 3 studies fail
- A 'negative' prior can adjust for over-optimism from Phase 2
  - Where a drug or mechanism already has a related positive Phase 3, prior information is more positive
- Phase 3 probability of success (POS) takes into account Phase 2 results and Phase 3 design
  - This can be considered in sizing Phase 2 and in setting interim Phase 2 bounds; e.g., move to Phase 3 when POS $> 50\%$

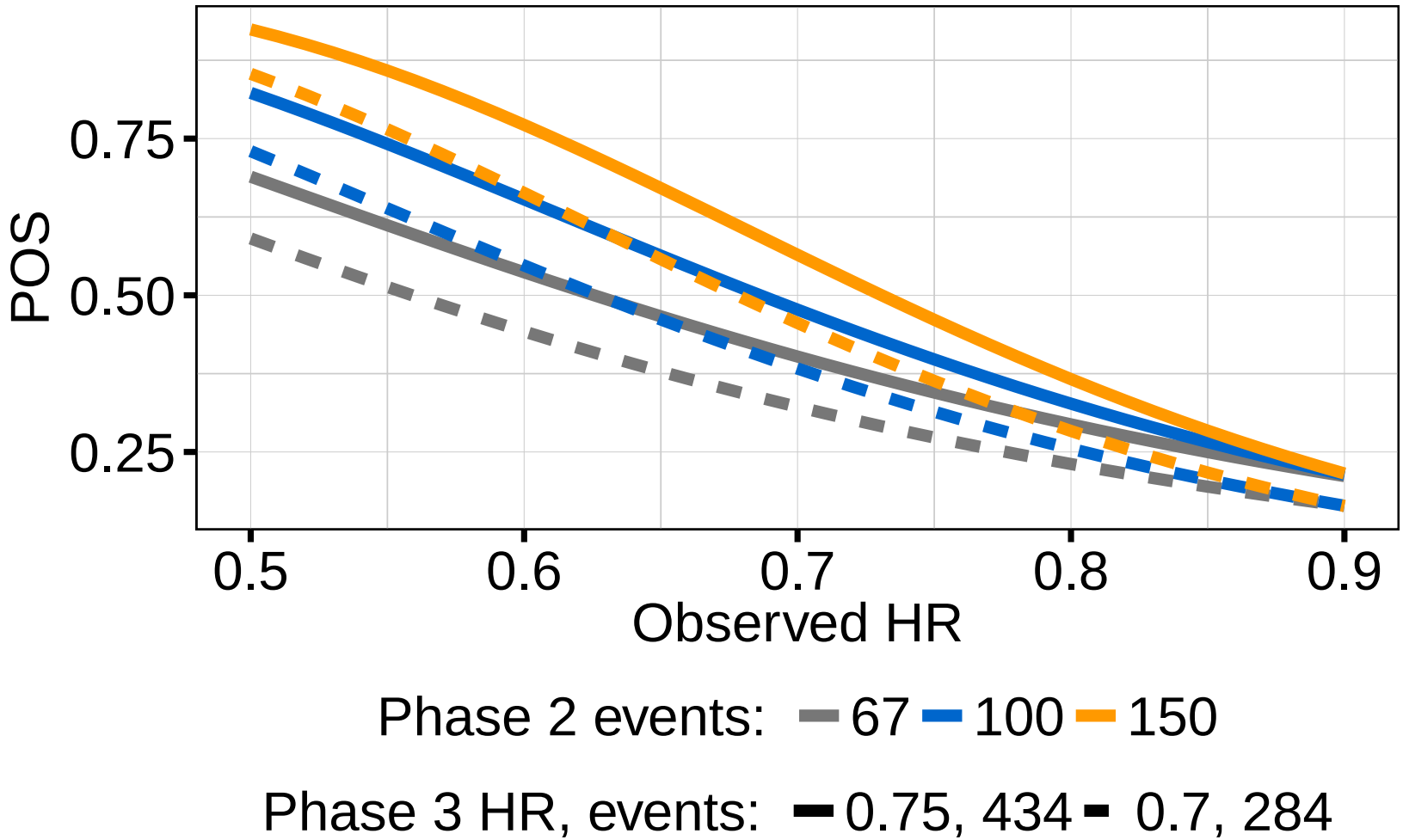
Phase 2 Prior and Posterior Distributions

Prior equivalent to observed HR=1 with 100 events; shrinks observed HR towards 0



Prior (black) and posterior (red) distributions before and after Phase 2 at 67 (solid), 100 (dashed) and 150 (dotted) events. Blue lines show assumed result of HR=.6 along with posterior

POS Based on Negative Prior



Phase 3 events provide 85% power for given hazard ratio

Approximate 1-sided P-value Required to Reach 50% POS

Phase 2			Phase 3	
Events	Obs HR	p-value	Events	HR
67	0.560	0.009	284	0.70
67	0.625	0.027	434	0.75
100	0.625	0.009	284	0.70
100	0.685	0.029	434	0.75
150	0.680	0.009	284	0.70
150	0.730	0.027	434	0.75

- Phase 2 p-values are 1-sided
- Phase 3 events provide 85% power for Phase 3 HR
- Do 'larger' Phase 2 trials offer better chance for early approval with strong results?

Designing a trial with PFS and OS

- Progression is often a reasonable 'surrogate endpoint' (and often not) for overall survival (OS) in oncology
- For drugs with a large expected benefit, may be reasonable to study progression free survival (PFS) during the course of the study prior to a definitive evaluation of OS.
- Treatment effect usually expected to be bigger for PFS and endpoints occur sooner
- Example 2-arm considerations
  - PFS: 5 months median control; power trial for HR=.5
  - OS: 12 months median control; power trial for HR=.7
  - Assumed enrollment over 2 years (9 month ramp-up), 18 months minimum OS follow-up, 85% power, 2.5% alpha
  - Single PFS efficacy analysis planned; OS interim efficacy and futility at that time (could do earlier PFS futility analysis)

Matching up Final PFS and Interim OS Analyses

- Matching up of anticipated PFS analysis and OS IA done using gsDesign; can be done 'manually' with EAST
- Relative timing of PFS events may end up differing from planned OS interim
  - Need to ensure futility and efficacy bounds are reasonable assuming different interim event count using spending functions
- Type I error controlled for 2 endpoints by allocating 0.5% to PFS, 2% to OS. With a positive PFS finding, full 2.5% can be given to OS.
- PFS powered at 95% to help synchronize PFS analysis with a 'reasonable' timing for OS analysis

OS Design

Analysis	Value	Efficacy	Futility
IA 1: 34%	Z	4.87	0.75
N: 434	p (1-sided)	0.00	0.23
Events: 128	HR at bound	0.42	0.88
Month: 21	P(Cross) if HR=1	0.00	0.77
	P(Cross) if HR=0.7	0.00	0.10
Final	Z	2.05	2.05
N: 522	p (1-sided)	0.02	0.02
Events: 372	HR at bound	0.81	0.81
Month: 42	P(Cross) if HR=1	0.02	0.98
	P(Cross) if HR=0.7	0.85	0.15

Asymmetric design with $\alpha = .02$, 85% power, Hwang-Shih-DeCani spending functions with $\gamma = -16$ (efficacy), $\gamma = 3$ (futility). Boundary crossing probabilities at final analysis are cumulative.

Estimation Bias Associated with Positive Interim Finding

- Regulators have encouraged us to minimize interim analyses
- Is analysis with extreme efficacy after 128 deaths too little to stop trial?
  - Can set up separate OS analysis or delay by requiring higher power/smaller PFS alpha with more events
- A positive interim OS finding would be expected to provide a biased estimate of treatment effect
- Should a bias-adjustment be made for an interim stop for efficacy?
  - For example, Whitehead suggests subtracting expected bias at observed treatment effect (Whitehead [1986], Jennison and Turnbull [2000])

PFS Design

- Timing: 21 months
- Sample size: 433
  - Given delay for IA prep, enrollment likely complete before IA.
- Events: 200
- $\alpha = .005$ (1-sided), Power = 95%
- Approximate hazard ratio for a positive result: 0.69 (approximately 2.2 monthsincrease in median PFS)
- If central PFS, may discontinue PFS review after 'definitive' PFS analysis
- Submit PFS results for preliminary approval without definitive OS? Likely to need OS trend
- Consider crossover of patients given an early approval may compromise definitive OS finding

Biomarker Design

- May consider Freidlin and Simon design to select biomarker at an interim analysis
  - Only independent patient not in interim count toward definitive subgroup
- Suppose +/− biomarker selected before trial start
  - Test independently in Biomarker+ and overall population
  - Is Biomarker− independent subgroup supportive or must results be definitive in this subgroup?
  - Can step down from Biomarker+ testing to overall or Biomarker−
  - If Biomarker− group is large, independent testing makes sense
  - If Biomarker− group is small, it's tempting to
    - test overall population (increased sample size or possible power loss), or
    - not enroll Biomarker− patients

Summary and Discussion

- Many challenges in oncology development
- How do we quickly get highly effective treatments to patients with high POS?
- Appropriate sizing of Phase 2 and Phase 3 is a challenge give competive and cost pressures, as well as pressure for meaningful results
  - Using small sample size in Phase 2 may be short-sighted in many cases
- Can decision theory improve Phase 2 and Phase 3 program planning as opposed to typical powering and sample size?
- Appropriate interim timing and bounds should allow early stopping
- Adding biomarkers to the mix creates a new set of challenges

References

Christy Chuang-Stein. Sample size and the probability of a successful trial. *Pharmaceutical Statistics*, 5:305–309, 2006. doi: 10.1002/pst.232.

Christopher Jennison and Bruce W. Turnbull. *Group Sequential Methods with Applications to Clinical Trials*. Chapman and Hall/CRC, Boca Raton, FL, 2000.

Nigel Stallard, John Whitehead, and Simon Cleal. Decision-making in a phase II clinical trial: a new approach combining Bayesian and frequentist concepts. 4:119–128, 2005. doi: 10.1002/pst.164.

John Whitehead. On the bias of maximum likelihood estimation following a sequential test. *Biometrika*, 73:573–581, 1986.