Conditional Power: The Good, The Bad (and The Ugly?)

Applications to Interim Analyses and Adaptation in Clinical Trials

Keaven M. Anderson and Yujie Zhao

August 5, 2025

Quote

(If you use conditional power,) if you are not careful you will make mistakes.

–Jason Liao (Incyte)

Acknowledgement

Thanks to Yihui Xie for formatting help.

Abstract

Conditional power has been used to define futility boundaries, for sample size re-estimation, and for decision making at in interim point in a clinical trial. The general question is “What is the probability the trial will succeed given what we observe today?” Useful answers to this question depend on how well the treatment effect for the rest of the trial can be approximated. Related quantities of predictive probability of success and, from the beginning of the trial, probability of success according to a prior distribution (average power) will be computed and discussed. The gsDesign R package and its Shiny interface will be discussed. We will discuss how to do computations that are easily interpretable and usable for customers of the quantities derived. We will also caution against problematic uses and interpretation of conditional power.

Keywords: conditional power, interim analysis, group sequential design, adaptive design

Overview

- Brownian motion approach to conditional power (CP) Proschan, Lan, and Wittes (2006).
- Mapping between design characteristics.
- CP at interim depends on future effect size.
- Interim effect size estimate for CP can be misleading.
    - 2 examples.
- Quick comments on conditional power sample size re-estimation

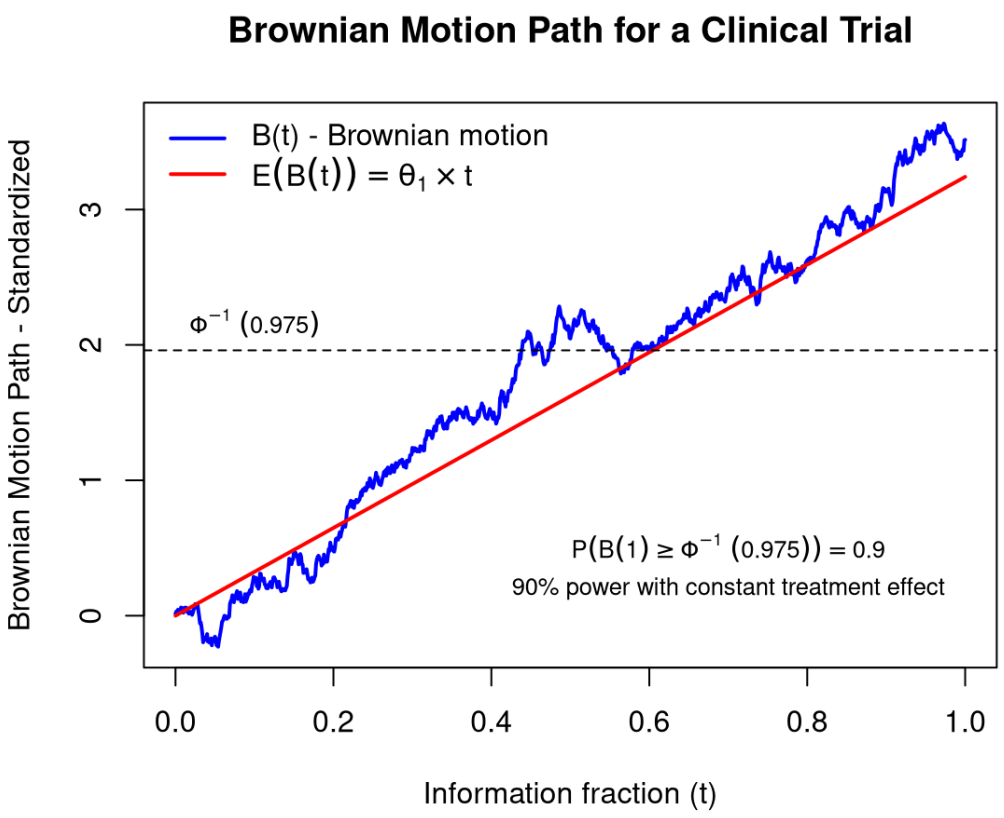
Sample Size Re-estimation Using Conditional Power

Source: gettyimages.com

- Much literature on this! (e.g., talk with Cyrus Mehta for a more positive view)
- Logistics can make this problematic
  - Fast enrollment
  - Long endpoint follow-up
- Interim treatment effect estimates can be unreliable
  - Lack of homogeneity
  - Randomness
- I am not a fan; this is in my “(and the Ugly?)” category

Brownian Motion and Conditional Power

Brownian Motion



- If you test continuously during a trial, asymptotically, results are like a Brownian motion process.
- Standardized treatment effect

$$\theta_1 = \Phi^{-1}(0.975) + \Phi^{-1}(0.9)$$

- Assuming a constant treatment effect, as observations (events for time-to-event endpoint), the B(t)-values vary at random about a trend line.

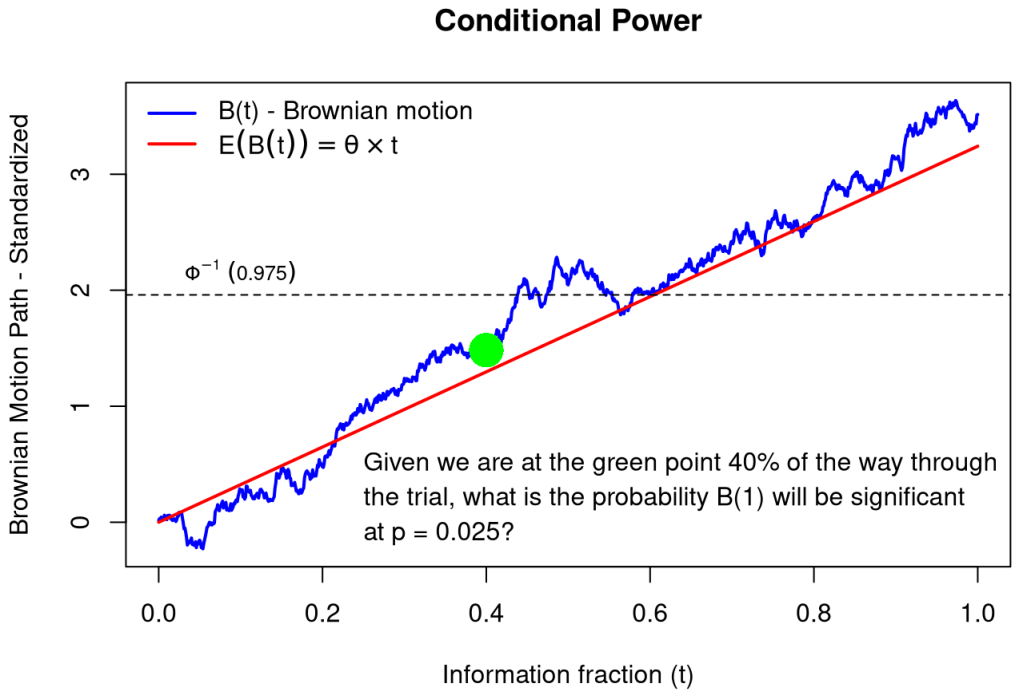
$$B(t) \sim \text{Normal}(\mu = \theta_1 \times t,\ \sigma^2 = t).$$

$$B(1) - B(t) \sim \text{Normal}(\mu = \theta_1 \times (1-t),\ \sigma^2 = 1-t).$$

- $B(t)$ and $1 - B(t)$ are **independent increments**
- Standard normal random variable:

$$Z(t) = B(t)/\sqrt{t}$$

Conditional power



- Effect size at interim:

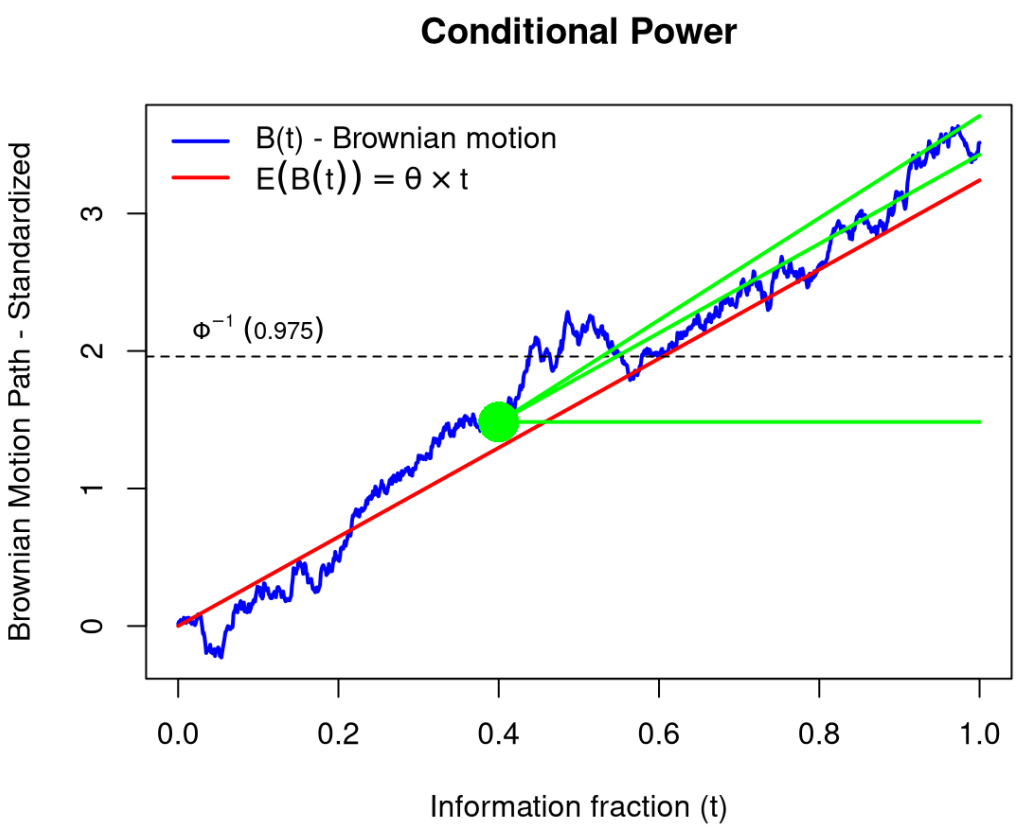
$$\hat{\theta}(t) = \frac{B(t)}{t}$$

$$\hat{\theta}(t) \sim \text{Normal}(\mu = \theta \times t,\ \sigma^2 = 1/t).$$

- Note: Variance of estimated effect size is big for small t.
- Conditional power at information fraction t

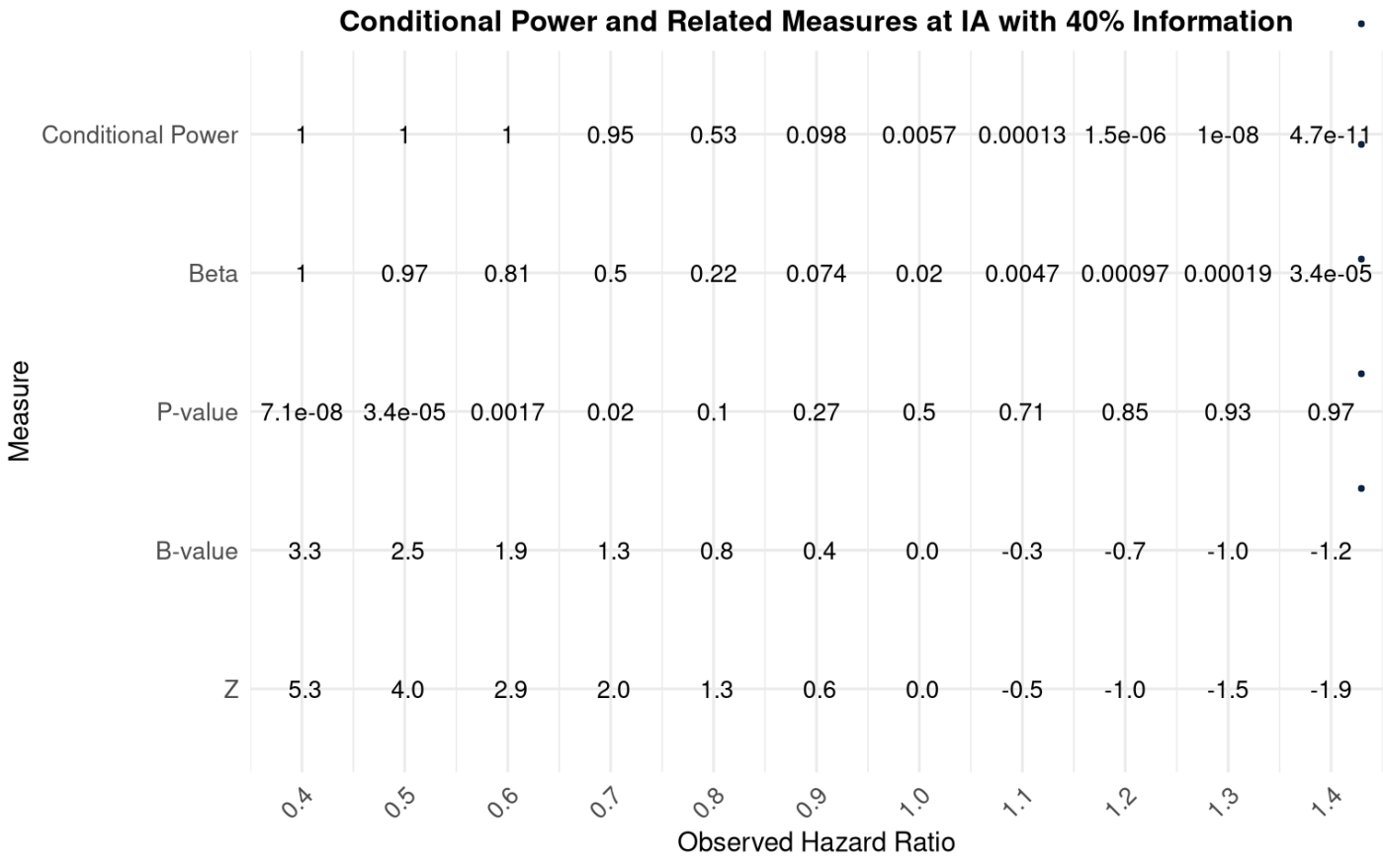
$$\text{CP}(t, b(t), \theta, \alpha) = 1 - \Phi\left(\frac{\Phi^{-1}(1-\alpha) - b(t) - \theta \times (1-t)}{\sqrt{1-t}}\right)$$

Conditional Power



Conditional Power at 40% of Information	
Interim Analysis	
Estimated Effect Size	Conditional Power
$\hat{\theta}(0.4)$	0.99
θ_1	0.97
0 (conditional error)	0.27

Mapping Conditional Power to Other Quantities



- Assume 90% power for HR = 0.7, 1-sided $\alpha = 0.025$
- Schoenfeld (1983) approximation suggests 331 events.
- Conditional power (at 40%) maps simply to other characteristics.
- Compute all characteristics of interest for bounds!
- Consider using β-spending or interim treatment effect (gsDesign2) for futility bound

Design - Futility

Require trend; γ = -2			
Analysis	Value	Efficacy	Futility
IA 1: 40%	Z	7.052	0.003
N: 340	p (1-sided)	0.000	0.499
Events: 135	~HR at bound	0.297	1.000
Month: 12.1	Spending	0.000	0.019
	B-value	4.450	0.002
	CP	1.000	0.006
	CP H1	1.000	0.511

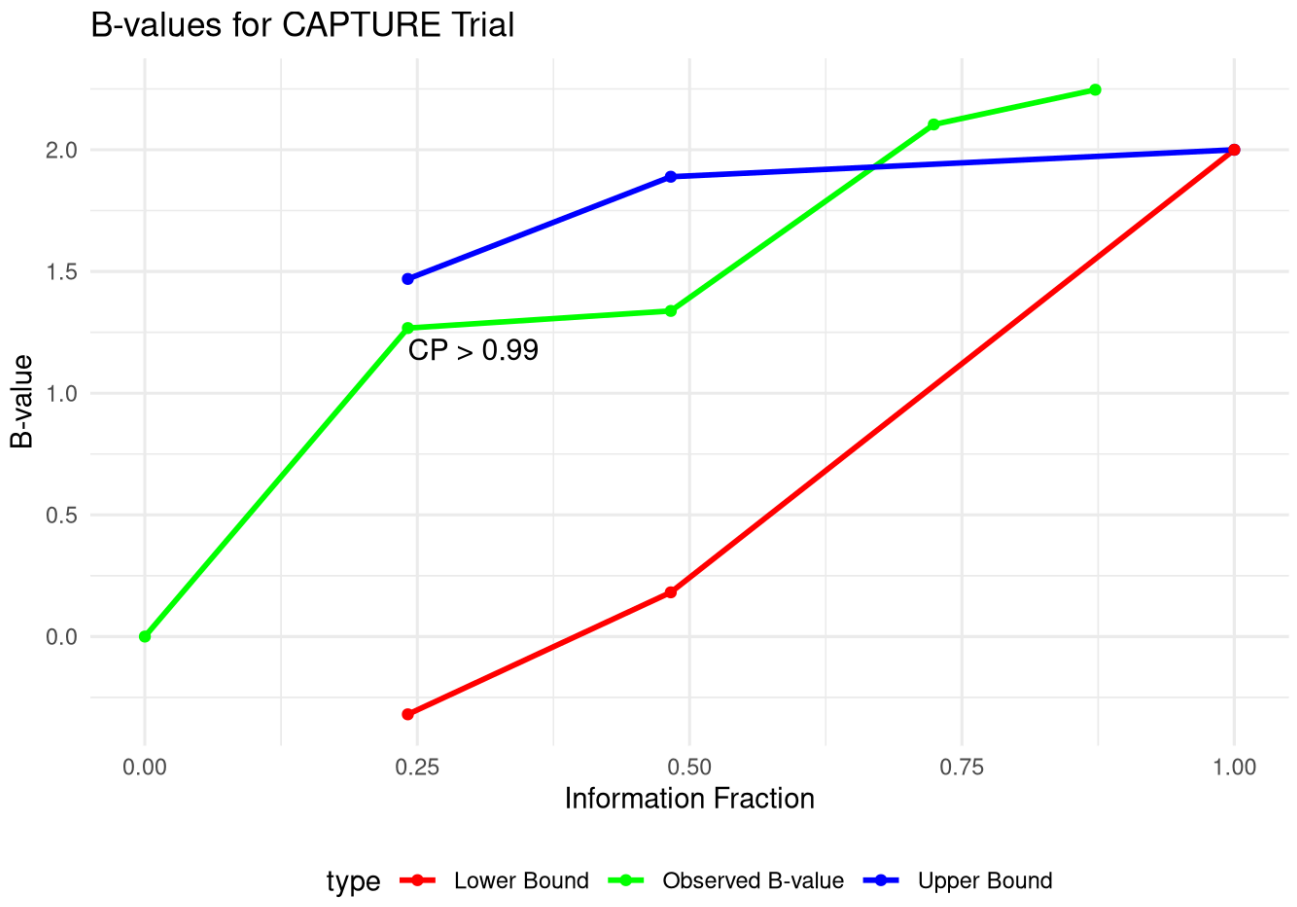
Require HR < 0.89; γ = 2.2; CP = 15%			
Analysis	Value	Efficacy	Futility
IA 1: 40%	Z	7.048	0.738
N: 398	p (1-sided)	0.000	0.230
Events: 158	~HR at bound	0.326	0.889
Month: 12.1	Spending	0.000	0.066
	B-value	4.452	0.466
	CP	1.000	0.153
	CP H1	1.000	0.797

Require HR < 0.95; γ = -0.45			
Analysis	Value	Efficacy	Futility
IA 1: 40%	Z	7.049	0.297
N: 352	p (1-sided)	0.000	0.383
Events: 140	~HR at bound	0.304	0.951
Month: 12.1	Spending	0.000	0.035
	B-value	4.452	0.188
	CP	1.000	0.027
	CP H1	1.000	0.622

- Sample size and Type II error increase with increasing CP bound - assuming 90% power.
- Alternative to increasing sample size is to decrease power.
- Bound with CP > 15% may suggest dropping power to 85%.

Trials are Often Not Homogeneous

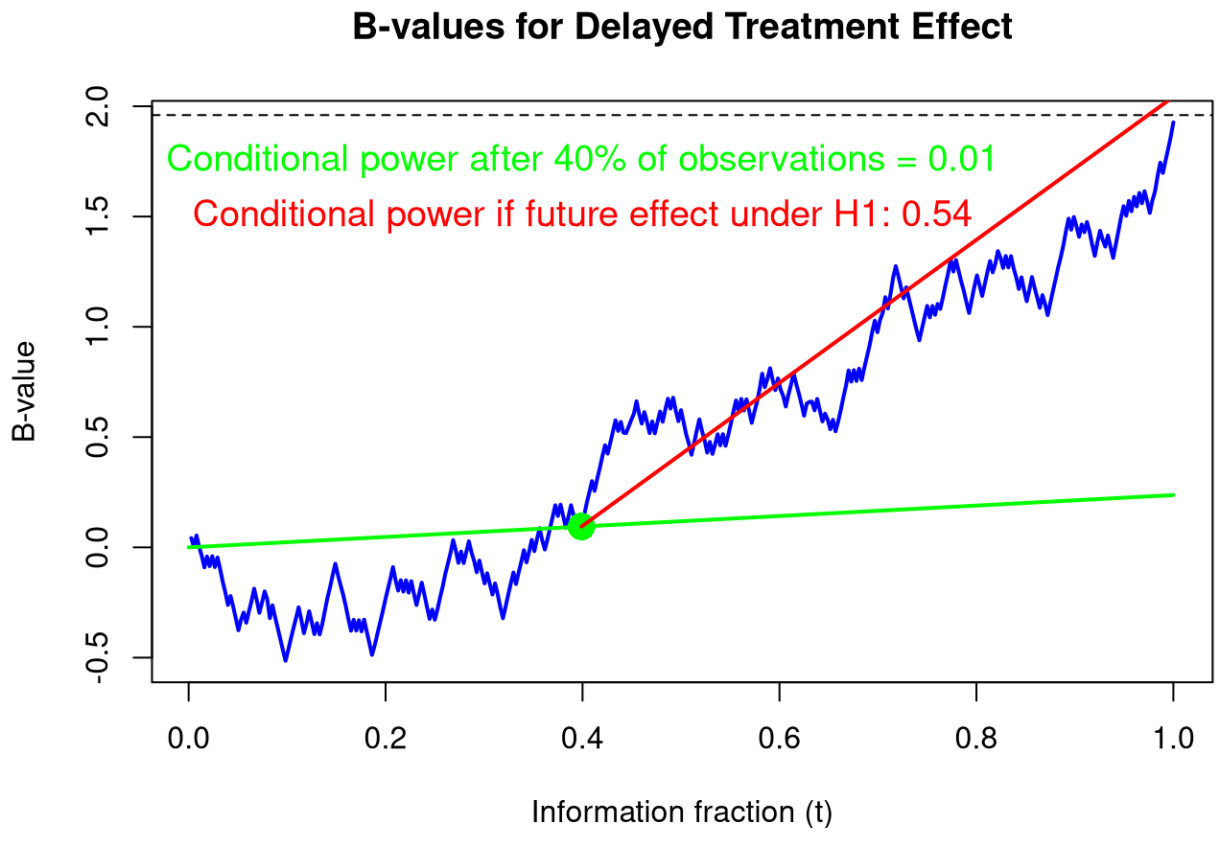
- The CAPTURE Investigators (1997) trial used a group sequential design with 3 analyses planned after 25%, 50% and 100% of the information fraction.
- Here we provide bounds and data that are similar to the trial.
- Is effect size not homogeneous over time?
- Country participation broadened after IA1 patients enrolled.



Delayed Treatment Effect

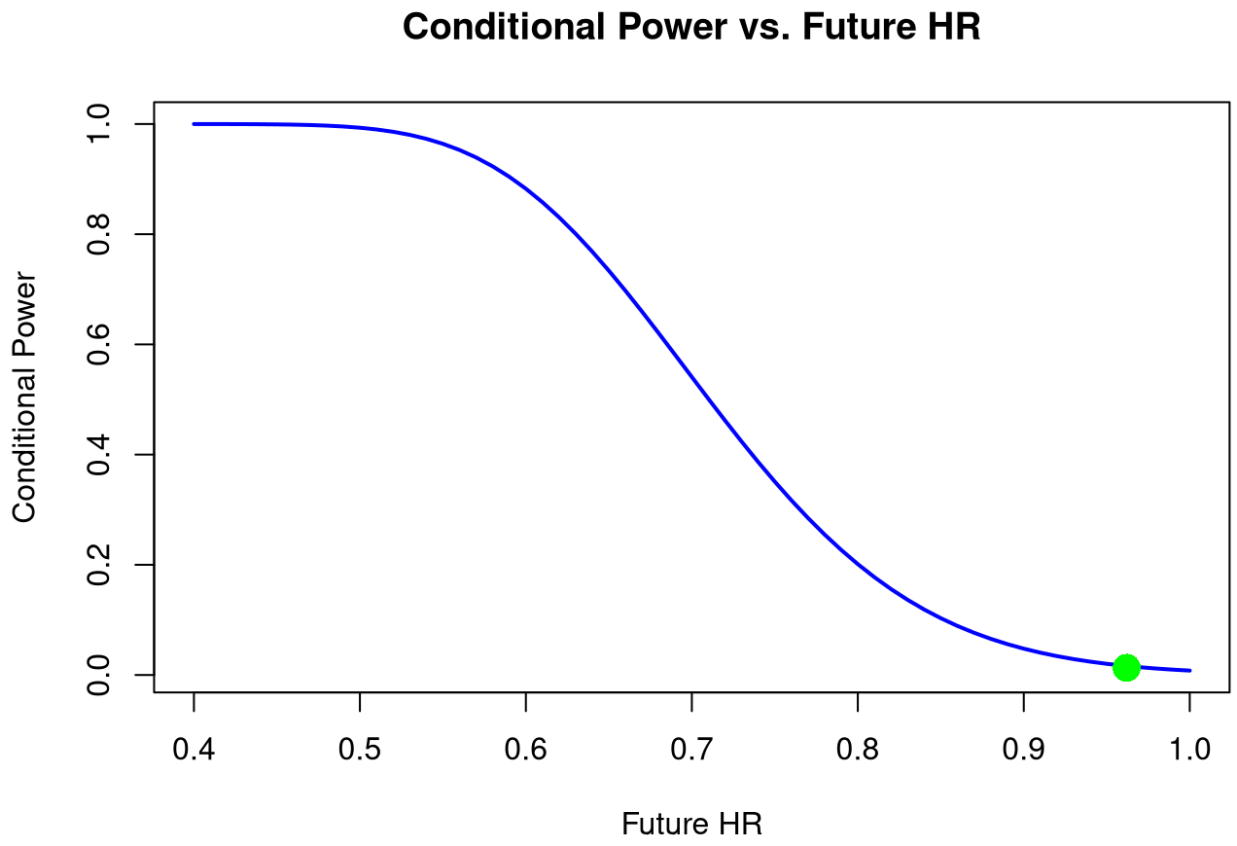
Delayed effect

- This example assumes:
  - HR = 1 for 4 months, HR = 0.6, thereafter.
  - 12 month expected trial enrollment.
  - 36 month expected trial duration.
- Slope (effect size) increases over time.
- Conditional power based on interim effect size is now deceptive!



Conditional Power by Future HR

- Assumes b = 0.09 and t = 0.4.
- IA $\widehat{\text{hr}} = 0.98$.



Futility Bound Accounting for Delayed Effect

Bound	Z	~HR at bound	Nominal p	Alternate hypothesis	Null hypothesis
Analysis: 1 Time: 12.2 N: 628 Events: 178 AHR: 0.87 Information fraction: 0.4					
Futility	-0.62	1.0971	0.7317	0.0636	0.2683
Analysis: 2 Time: 36 N: 628 Events: 446 AHR: 0.71 Information fraction: 1					
Efficacy	1.96	0.8306	0.0250	0.9022	0.0249

- Futility bound accounting for delayed effect.
- Still have β-spending of about 6% at IA, as before.
- However, now ~HR at bound is ~1.1 rather than previous ~0.9 under proportional hazards.
- Now using `gsDesign::gs_design_ahr()` to enable delayed treatment effect.
- Conditional power at bound TBD

Cautions and Conclusions

The Conversation

- Q: What is the conditional power?
- A: It depends! I will give you a range of plausible values you can think about!

Cautions and Conclusions

- Conditional power attempts to answer a natural question: “What is the probability the trial will succeed given what we observe today?”
- Trials are not often homogeneous, so look at CP across a range of possible future effect sizes.
- For futility, conditional power at bound is just one operating characteristics.
    - No need to select bound based on CP.
    - However, CP is a good characteristic to compute.

Thank you

Email: keaven_anderson@merck.com **Website:** keaven.github.io

Session Information

```
## R version 4.3.1 (2023-06-16)
## Platform: x86_64-pc-linux-gnu (64-bit)
## Running under: Red Hat Enterprise Linux 9.6 (Plow)
##
## Locale:
##   LC_CTYPE=en_US.UTF-8       LC_NUMERIC=C
##   LC_TIME=en_US.UTF-8        LC_COLLATE=en_US.UTF-8
##   LC_MONETARY=en_US.UTF-8    LC_MESSAGES=en_US.UTF-8
##   LC_PAPER=en_US.UTF-8       LC_NAME=C
##   LC_ADDRESS=C               LC_TELEPHONE=C
##   LC_MEASUREMENT=en_US.UTF-8 LC_IDENTIFICATION=C
##
## Package version:
##   base64enc_0.1.3     bigD_0.3.1          bitops_1.0.9
##   bslib_0.9.0         cachem_1.1.0        cli_3.6.5
##   codetools_0.2.19    commonmark_1.9.5    compiler_4.3.1
##   corpcor_1.6.10      cpp11_0.5.2         curl_6.4.0
##   data.table_1.17.6   digest_0.6.37       doFuture_1.0.1
##   dplyr_1.1.4         evaluate_1.0.4      farver_2.1.2
##   fastmap_1.2.0       fontawesome_0.5.3   foreach_1.5.2
##   fs_1.6.6            future_1.34.0       future.apply_1.11.3
##   generics_0.1.4      ggplot2_3.5.2       globals_0.16.3
##   glue_1.8.0          graphics_4.3.1      grDevices_4.3.1
##   grid_4.3.1          gsDesign_3.6.9      gsDesign2_1.1.5.1
```

References

Proschan, Michael A, KK Gordan Lan, and Janet Turk Wittes. 2006. *Statistical Monitoring of Clinical Trials: A Unified Approach*. Springer.

Schoenfeld, David A. 1983. “Sample-Size Formula for the Proportional-Hazards Regression Model.” *Biometrics*, 499–503.

The CAPTURE Investigators. 1997. “Randomized Placebo-Controlled Trial of Abciximab Before and During Coronary Intervention in Refractory Angina: The CAPTURE Study.” *Lancet* 349: 1429–35.