Adaptive Monitoring: Optimal wait-time to control false discoveries

Jonathan J Chipman, PhD

Population Health Sciences, Division of Biostatistic
University of Utah
Huntsman Cancer Institute

March 23, 2020

Overview

Co-authors Jeffrey D Blume, PhD and Robert A Greevy, Jr., PhD

AM SGPV Adaptive monitoring on Second Generation p-Value (Chipman 2019)

Burn-in To ensure $\alpha < 0.05$ with unlimited sample size

Prematurely Ending Clinical Trial(s)

Towards a Revolution in COPD Health (TORCH) (Calverley 2007)

Primary Aim: Establish whether beta-agonist (salmeterol plus fluticasone propionate) has survival benefit in participants with chronic obstructive pulmonary disease

2007 6112 participants

- HR 0.825 (95% CI: 0.681-1.002, p-adjusted=0.052)
- Awkward Conclusion: primary outcome did not reach statistical significance, yet 'significant benefits in all other outcomes.'

Prematurely Ending Clinical Trial(s)

Towards a Revolution in COPD Health (TORCH) (Calverley 2007)

Primary Aim: Establish whether beta-agonist (salmeterol plus fluticasone propionate) has survival benefit in participants with chronic obstructive pulmonary disease

2007 6112 participants

- HR 0.825 (95% CI: 0.681-1.000, p-adjusted=0.05)
- Awkward Conclusion: primary outcome did not reach statistical significance, yet 'significant benefits in all other outcomes.'

Prematurely Ending Clinical Trial(s)

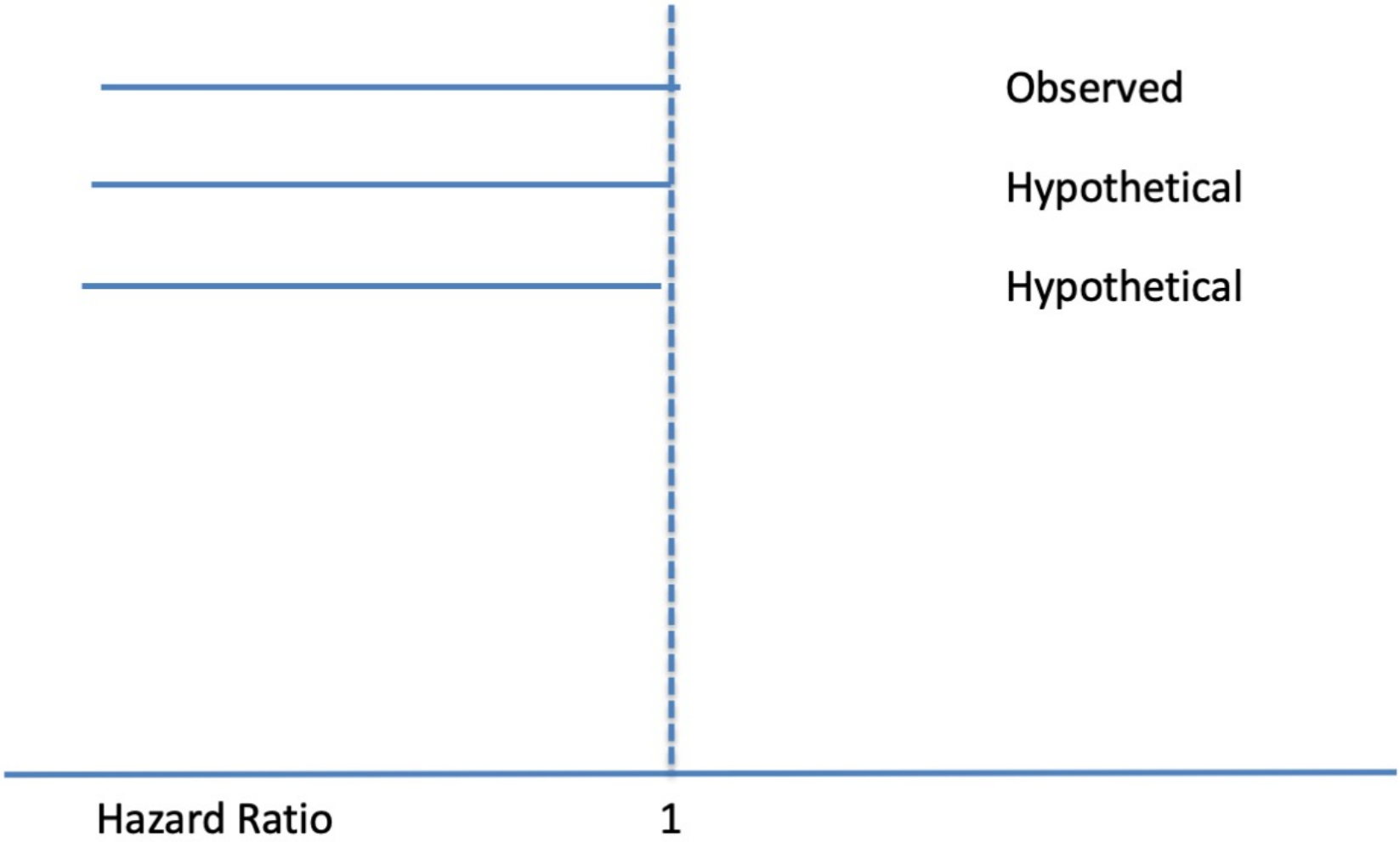
Towards a Revolution in COPD Health (TORCH) (Calverley 2007)

Primary Aim: Establish whether beta-agonist (salmeterol plus fluticasone propionate) has survival benefit in participants with chronic obstructive pulmonary disease

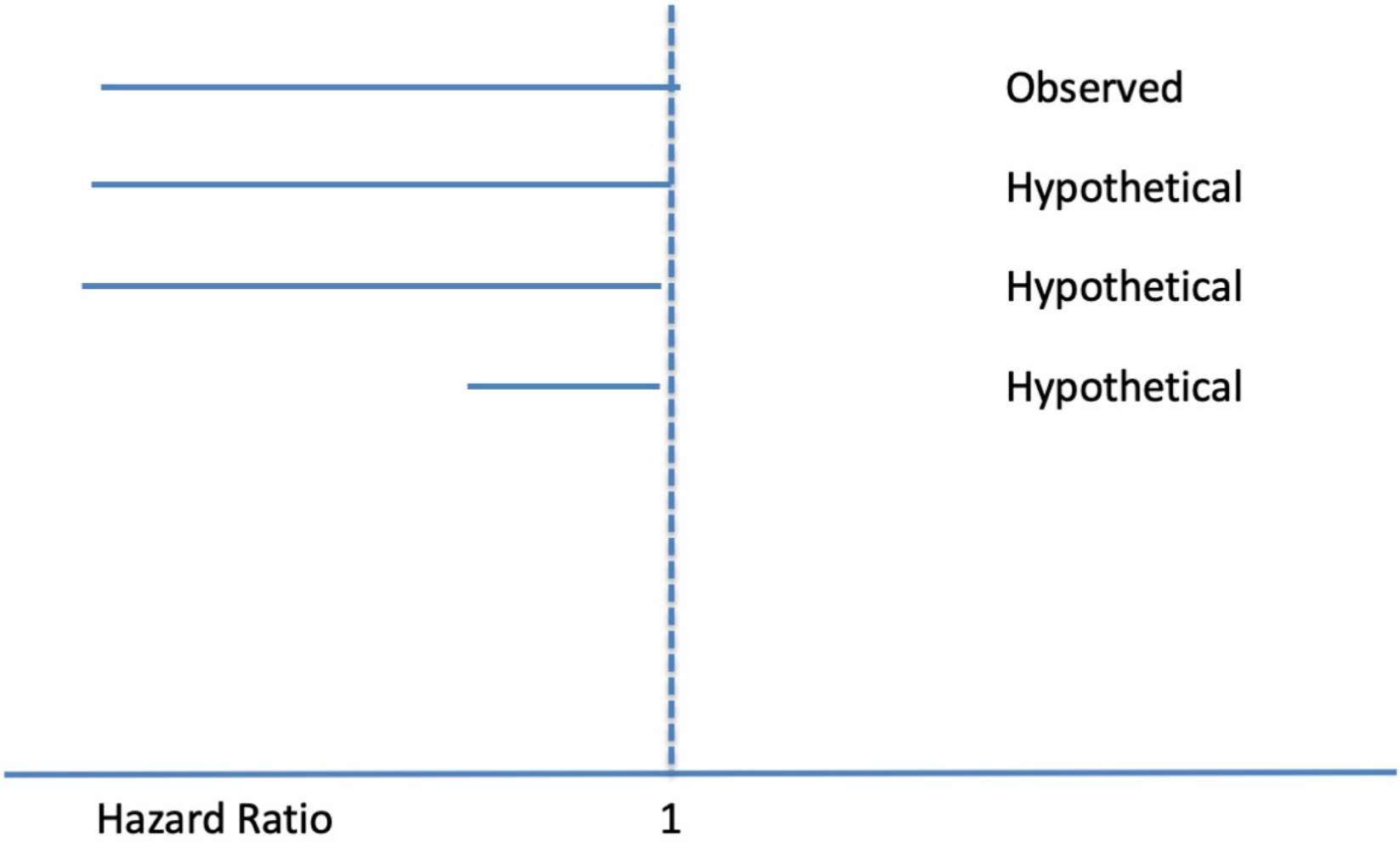
2007 6112 participants

- HR 0.825 (95% CI: 0.681-0.998, p-adjusted=0.0498)
- Awkward Conclusion: primary outcome did not reach statistical significance, yet 'significant benefits in all other outcomes.'

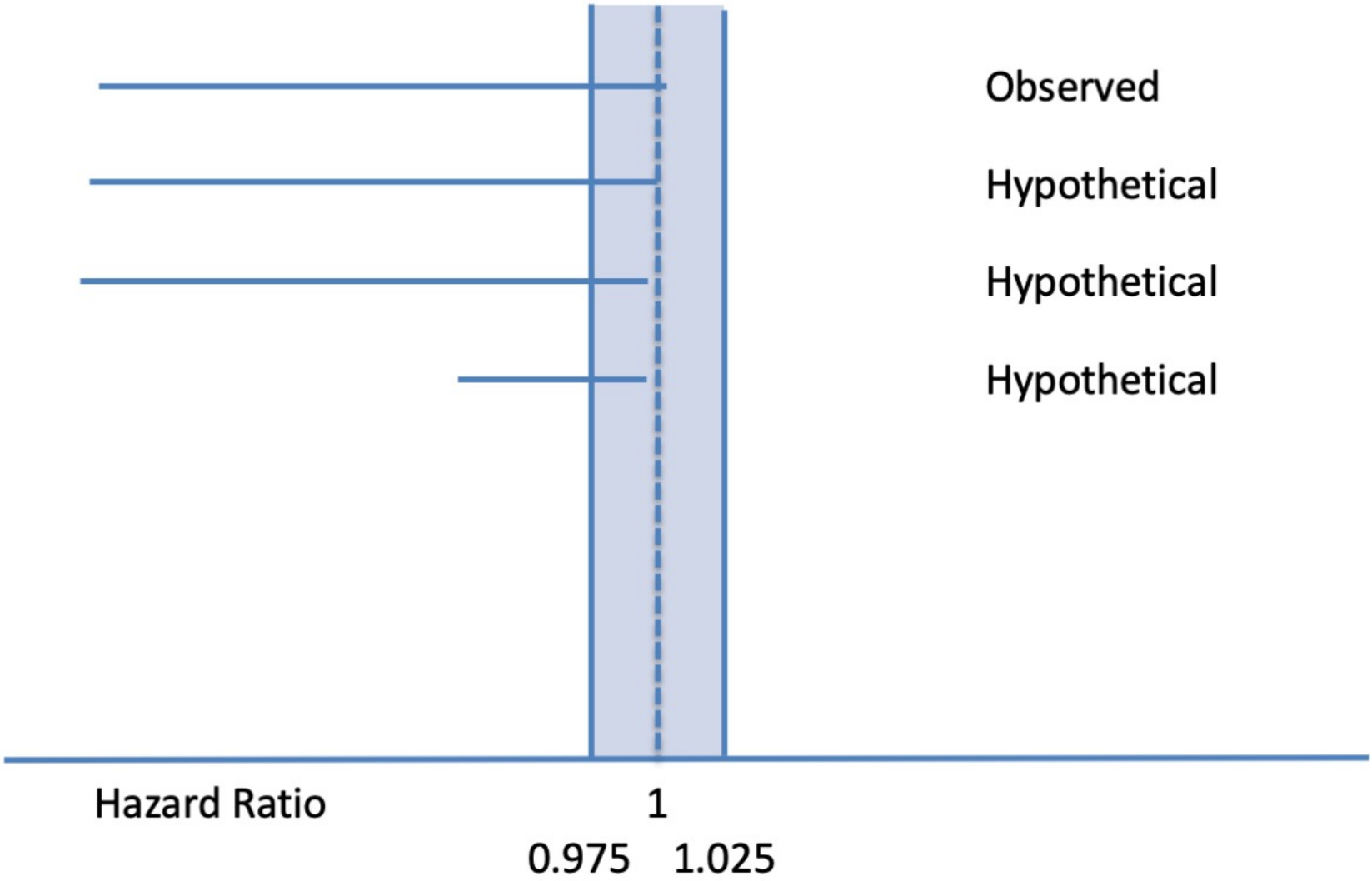
Envisioning treatment effects



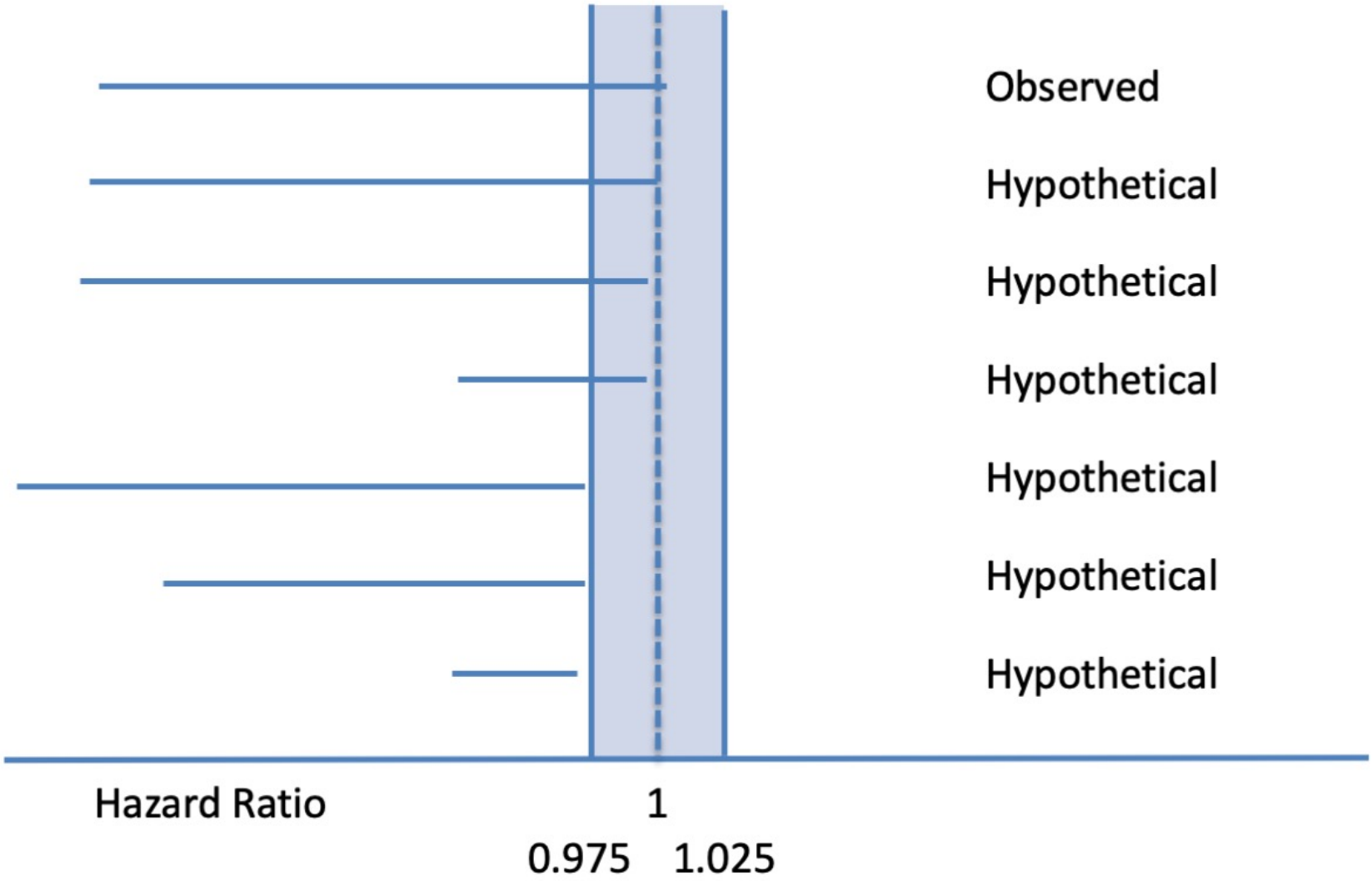
Envisioning treatment effects



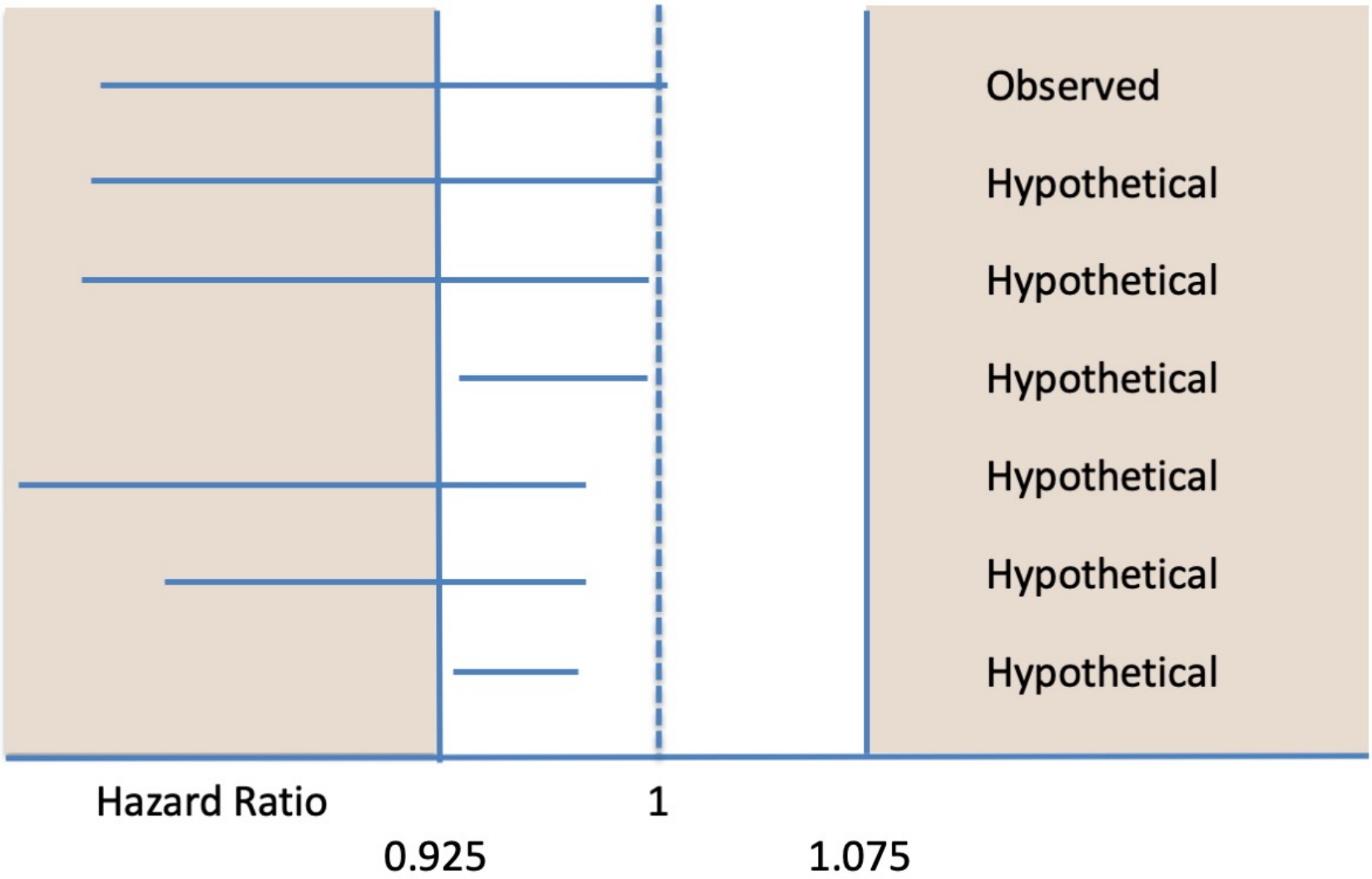
Region of Practically Equivalent Effects (ROPE) (Kruschke 2011)



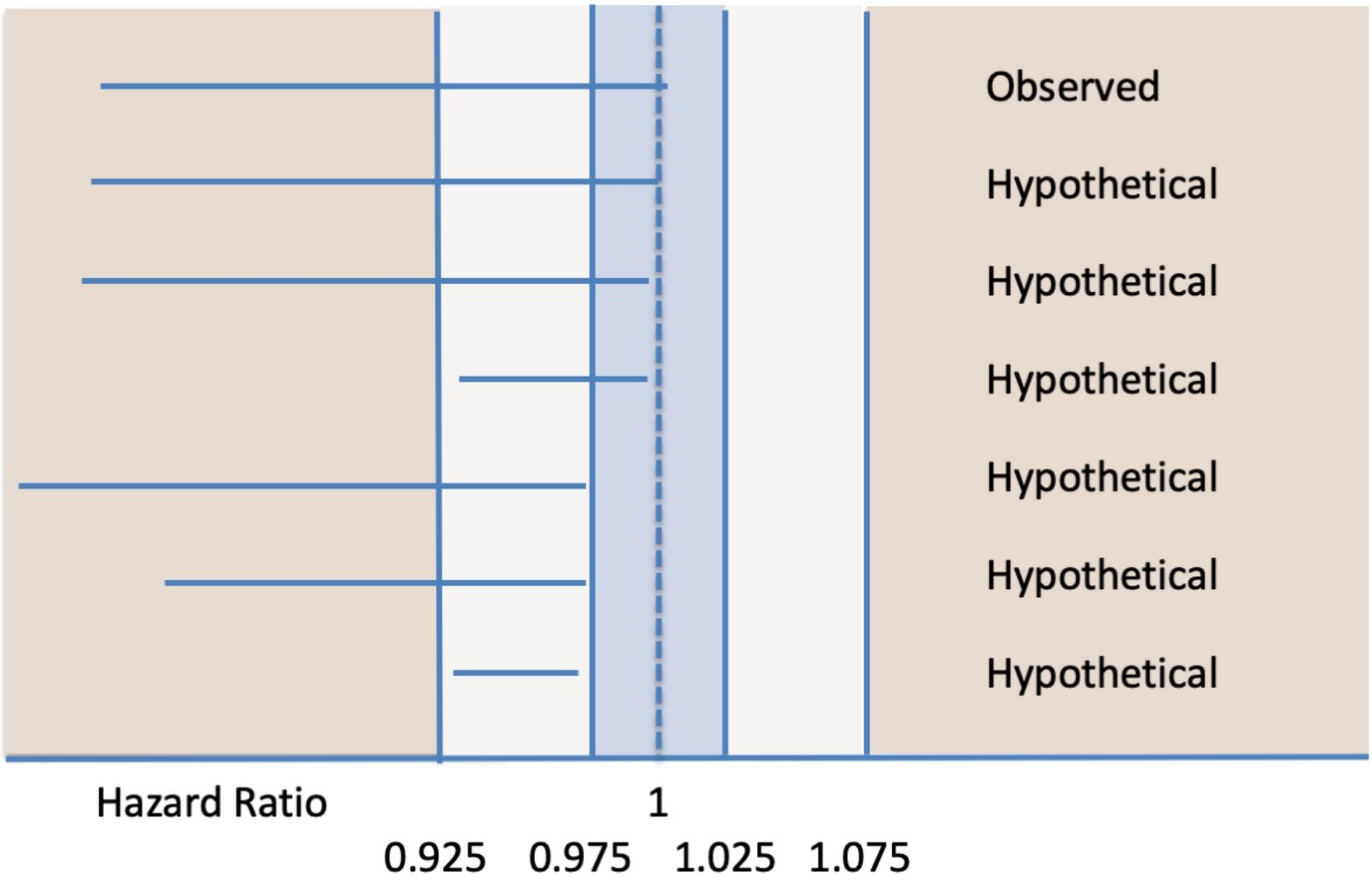
Region of Practically Equivalent Effects (ROPE) (Kruschke 2011)



Region of Meaningful Effects (ROME)

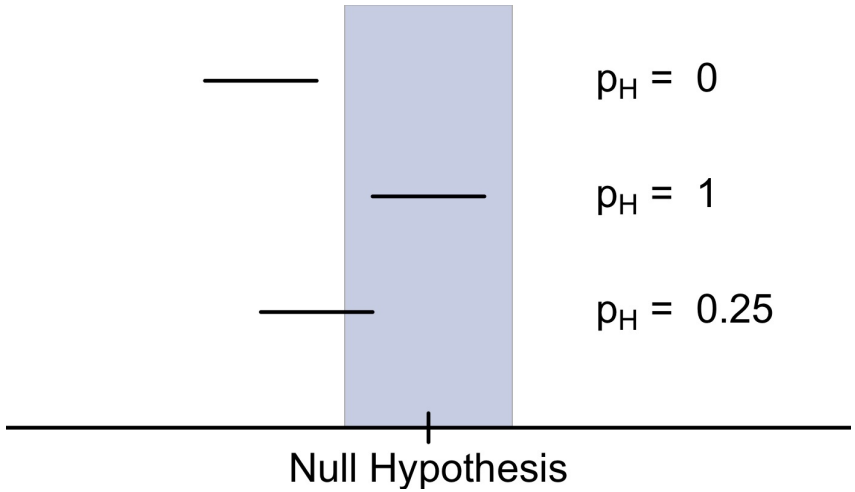


ROPE and ROME



Second Generation p-value (SGPV; Blume et al. 2018, 2019)

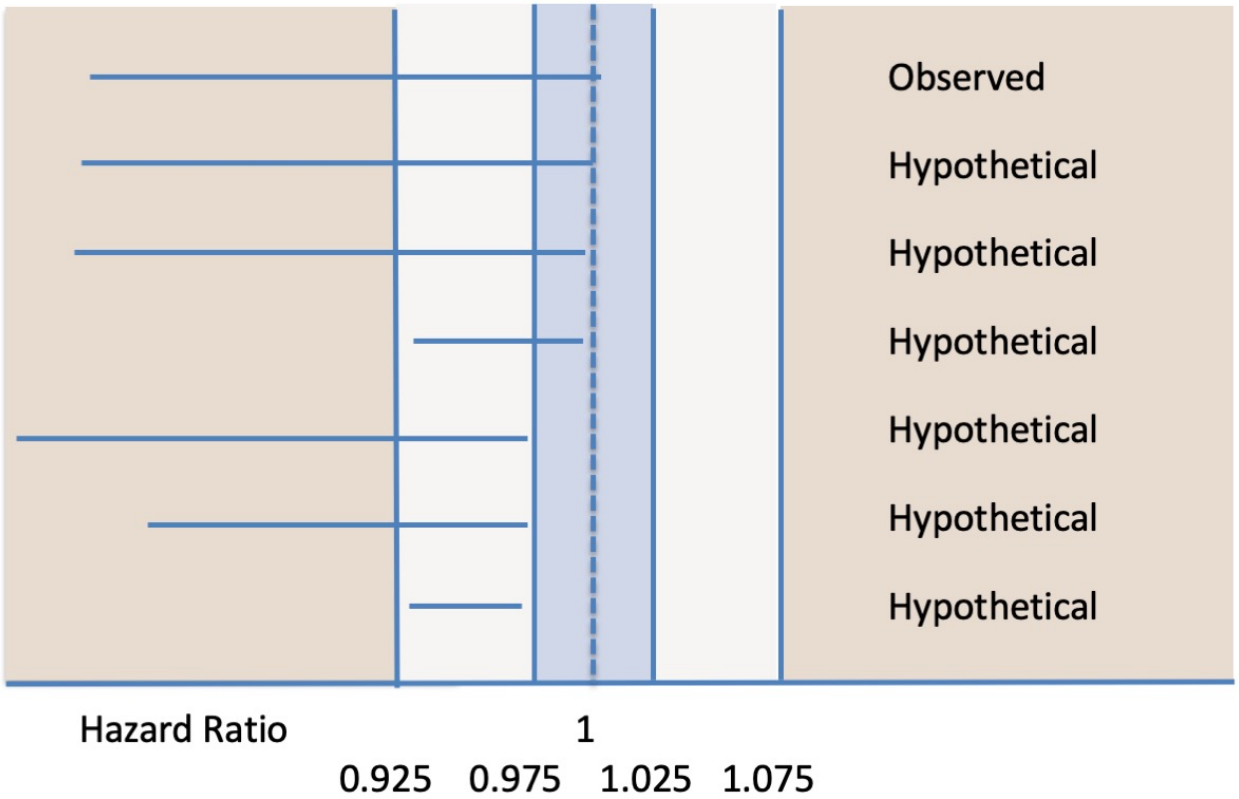
What proportion of interval overlaps with Δ_H?



Interpretation of p_H

- $p_H = 0$: Evidence to rule out hypothesis effects
- $p_H = 1$: Evidence supporing hypothesis effects
- $0 < p_H < 1$: Inconclusive, need more data

TORCH with ROPE and ROME



$0 < p_{ROPE} < 1 ; \ 0 < p_{ROME} < 1$

$0 < p_{ROPE} < 1 ; \ 0 < p_{ROME} < 1$

$0 < p_{ROPE} < 1 ; \ 0 < p_{ROME} < 1$

$0 < p_{ROPE} < 1 ; \quad p_{ROME} = 0$

$p_{ROPE} = 0 ; \ 0 < p_{ROME} < 1$

$p_{ROPE} = 0 ; \ 0 < p_{ROME} < 1$

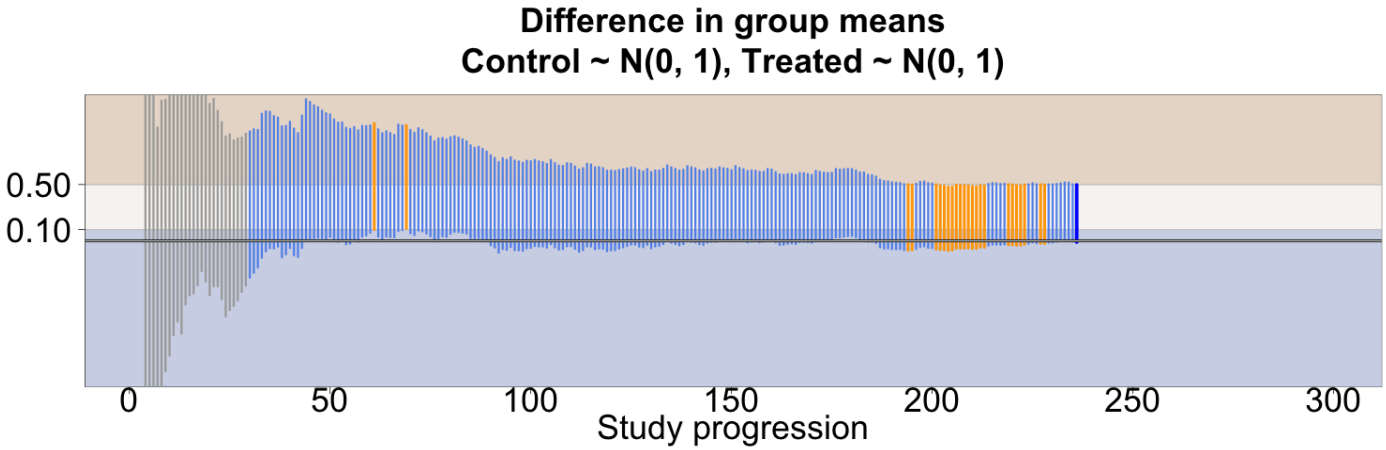
$p_{ROPE} = 0 ; \quad p_{ROME} = 0$

Adaptive Monitoring with SGPV

Wait	Interval width stabilizes
Monitor	Interval and SGPV at desired looks
Alert	$p_{ROPE} = 0.0$: Ruled out ROPE effects
	$p_{ROME} = 0.0$: Ruled out ROME effects
Affirm	Stop if same conclusion k participants later
	End of resources
Report	Only the final interval when stopping

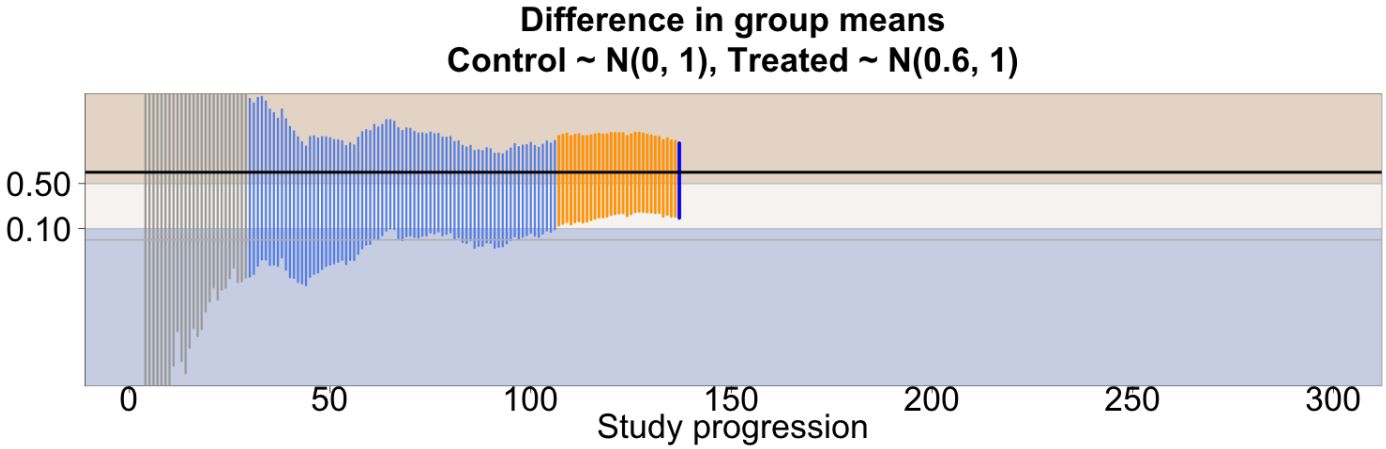
Adaptive Monitoring with SGPV

Wait	Interval width stabilizes
Monitor	Interval and SGPV at desired looks
Alert	$p_{ROPE} = 0.0$: Ruled out ROPE effects
	$p_{ROME} = 0.0$: Ruled out ROME effects
Affirm	Stop if same conclusion k participants later
	End of resources
Report	Only the final interval when stopping



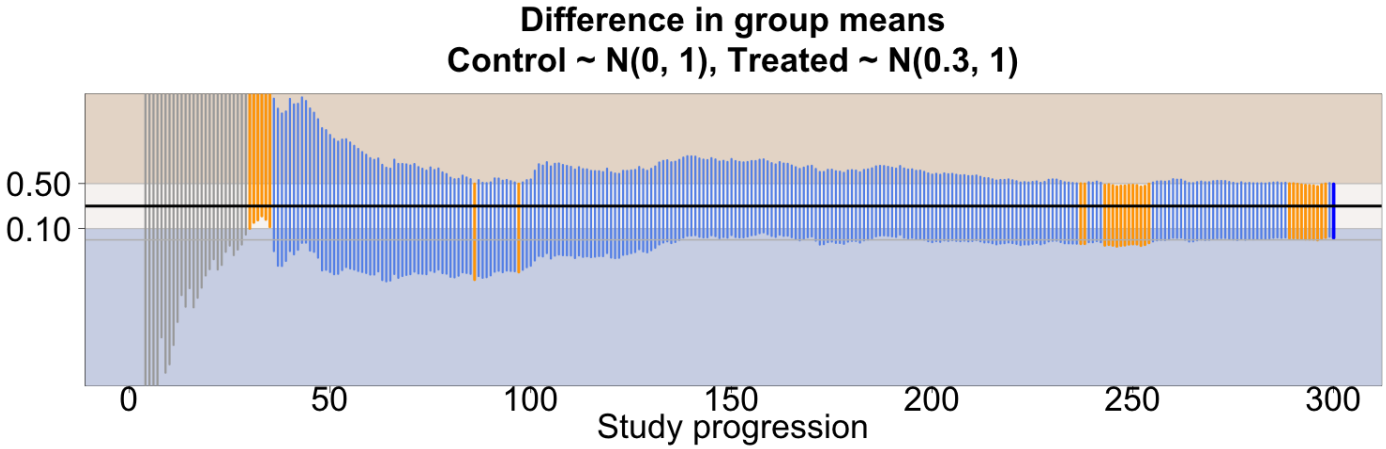
Adaptive Monitoring with SGPV

Wait	Interval width stabilizes
Monitor	Interval and SGPV at desired looks
Alert	$p_{ROPE} = 0.0$: Ruled out ROPE effects
	$p_{ROME} = 0.0$: Ruled out ROME effects
Affirm	Stop if same conclusion k participants later
	End of resources
Report	Only the final interval when stopping



Adaptive Monitoring with SGPV

Wait	Interval width stabilizes
Monitor	Interval and SGPV at desired looks
Alert	$p_{ROPE} = 0.0$: Ruled out ROPE effects
	$p_{ROME} = 0.0$: Ruled out ROME effects
Affirm	Stop if same conclusion k participants later
	End of resources
Report	Only the final interval when stopping



Error probabilities

Error probabilities (such as Type I Error) depend on

- ROPE and ROME
- Wait time
- Affirmation steps (k)

Error probabilities

Error probabilities (such as Type I Error) depend on

- ROPE and ROME (set by science)
- Wait time
- Affirmation steps

Wait time with unlimited sample size

- Measured by inferential width

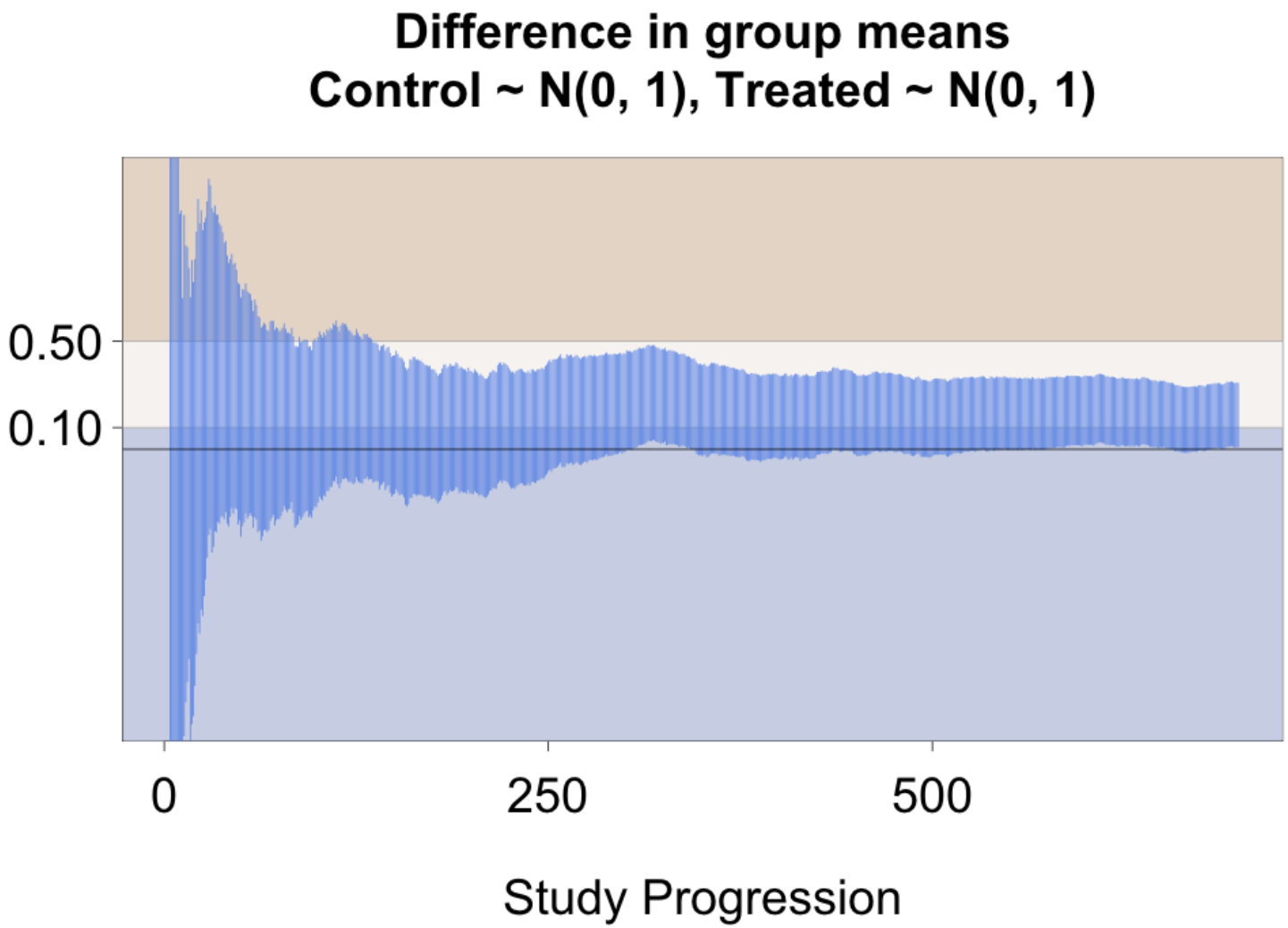
Controlling errors

Simulations with 40K replicates with varying wait times

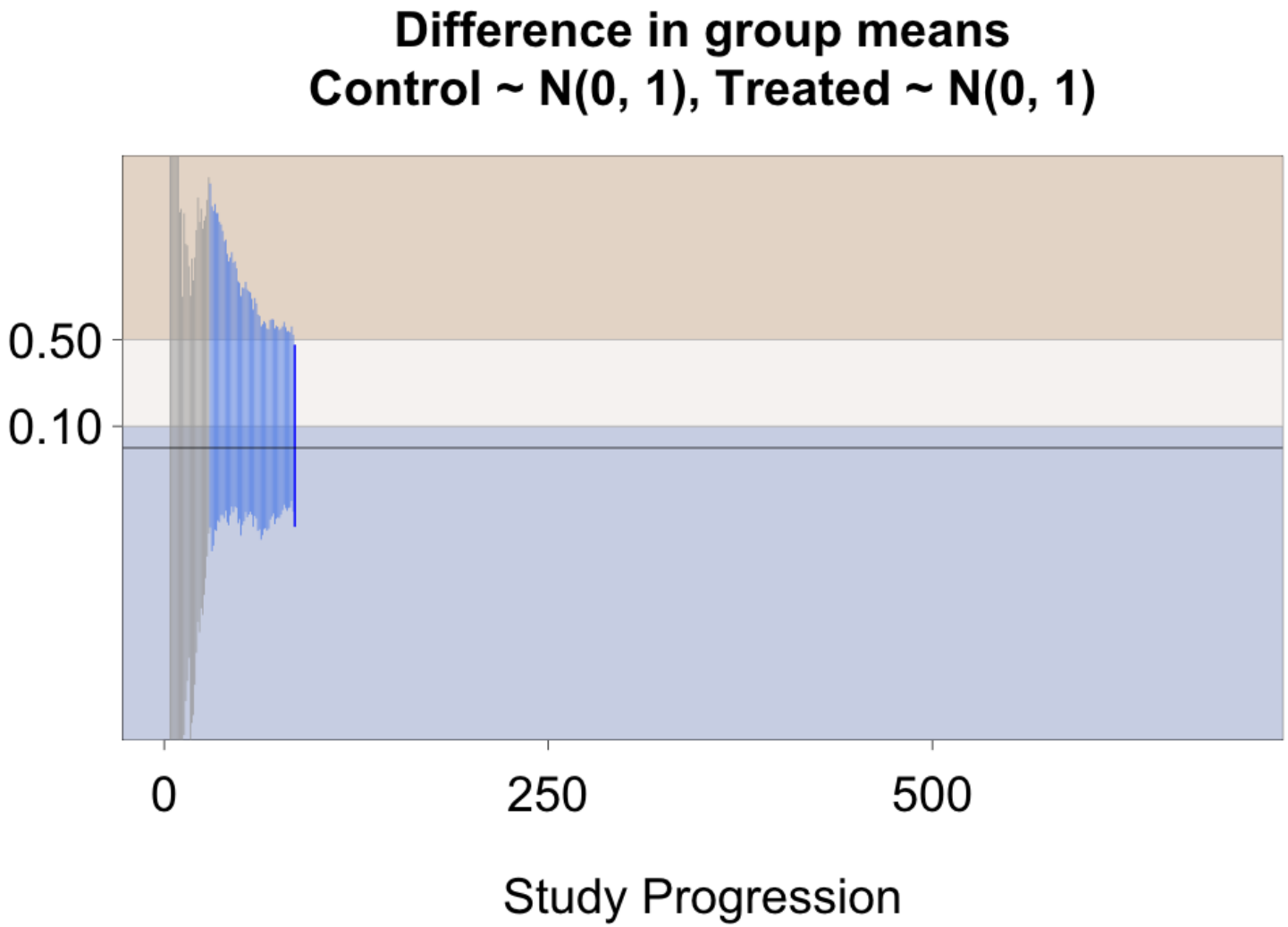
Holding constant:

- Control~N(0,1), Treated~N(0,1)
- ROME and ROPE regions (10 settings)
- $k = 0$

How long to wait before applying monitoring rules: A one-sided example

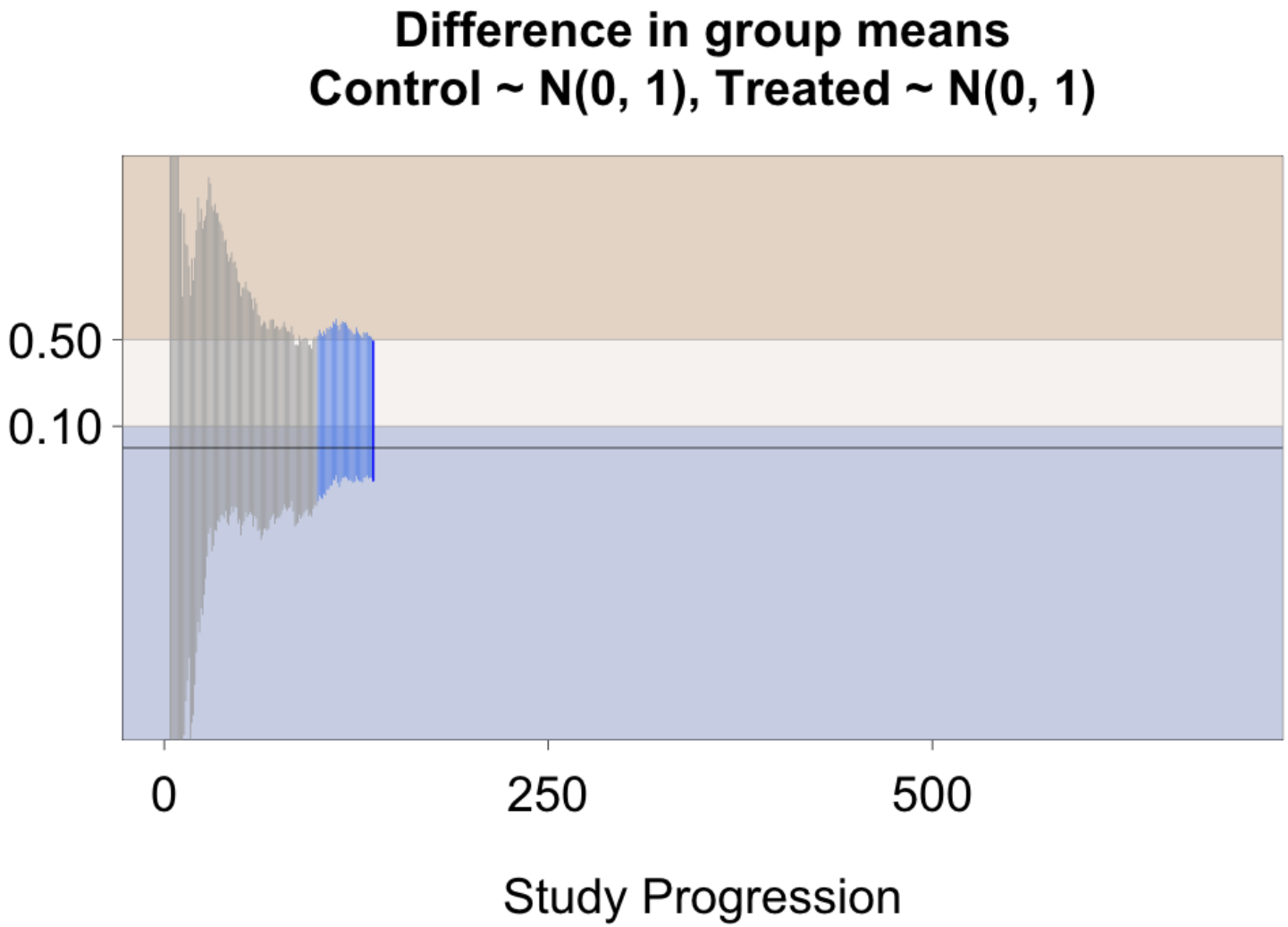


How long to wait before applying monitoring rules: A one-sided example



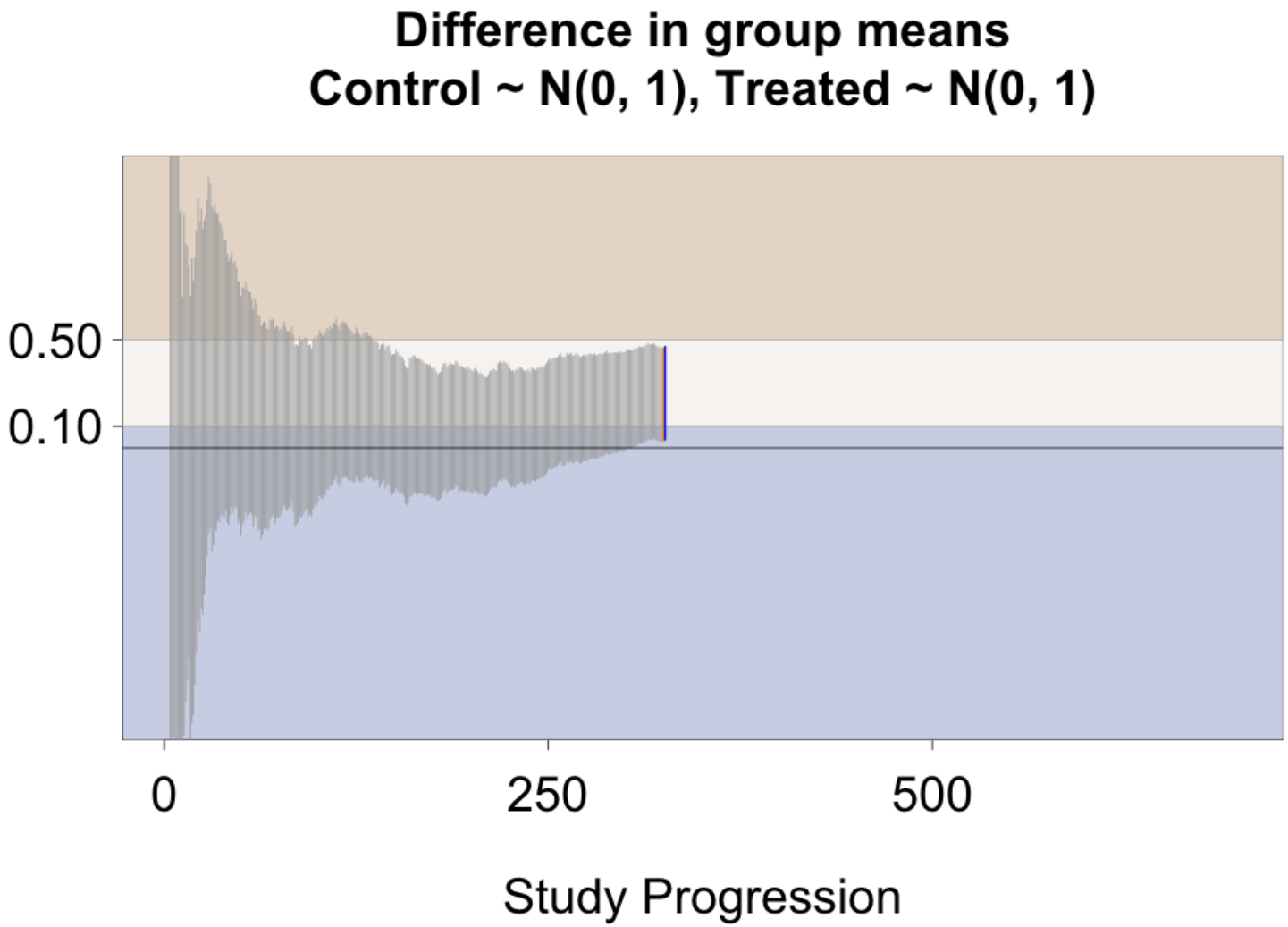
Wait until expected margin if error (1/2 width of CI) is 0.5

How long to wait before applying monitoring rules: A one-sided example



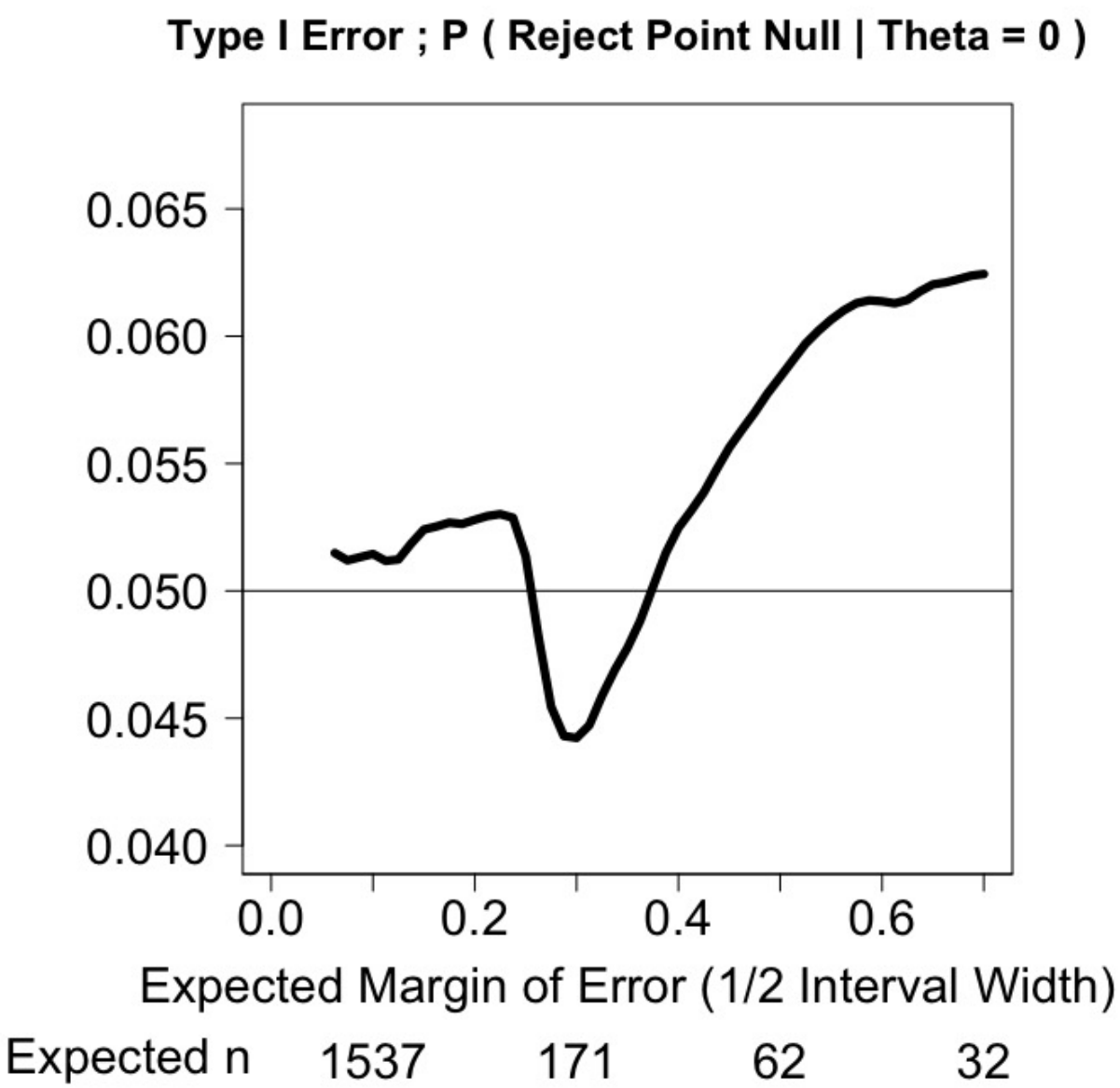
Wait until expected margin if error (1/2 width of CI) is 0.25

How long to wait before applying monitoring rules: A one-sided example

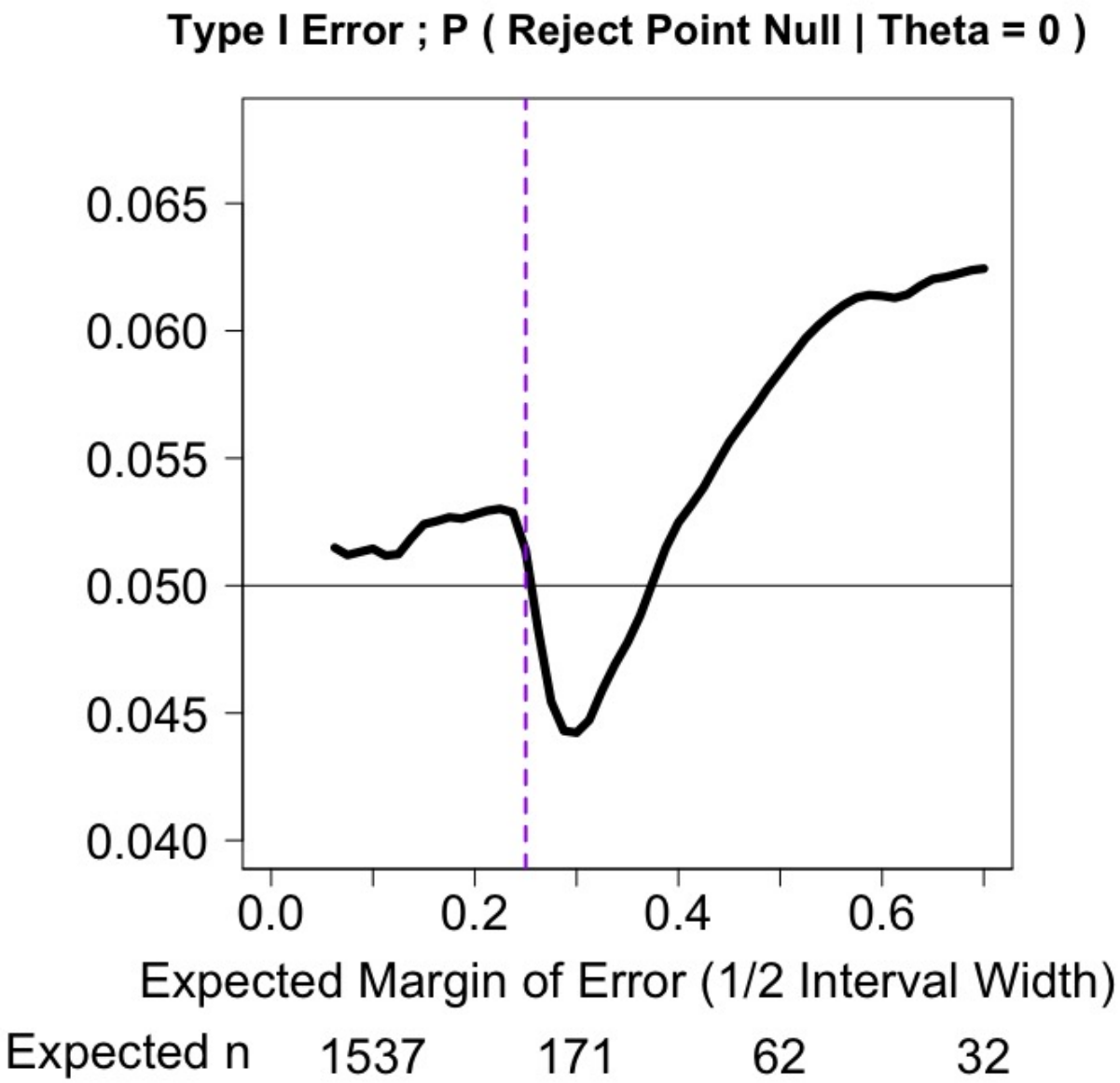


Wait until expected margin if error (1/2 width of CI) is 0.15

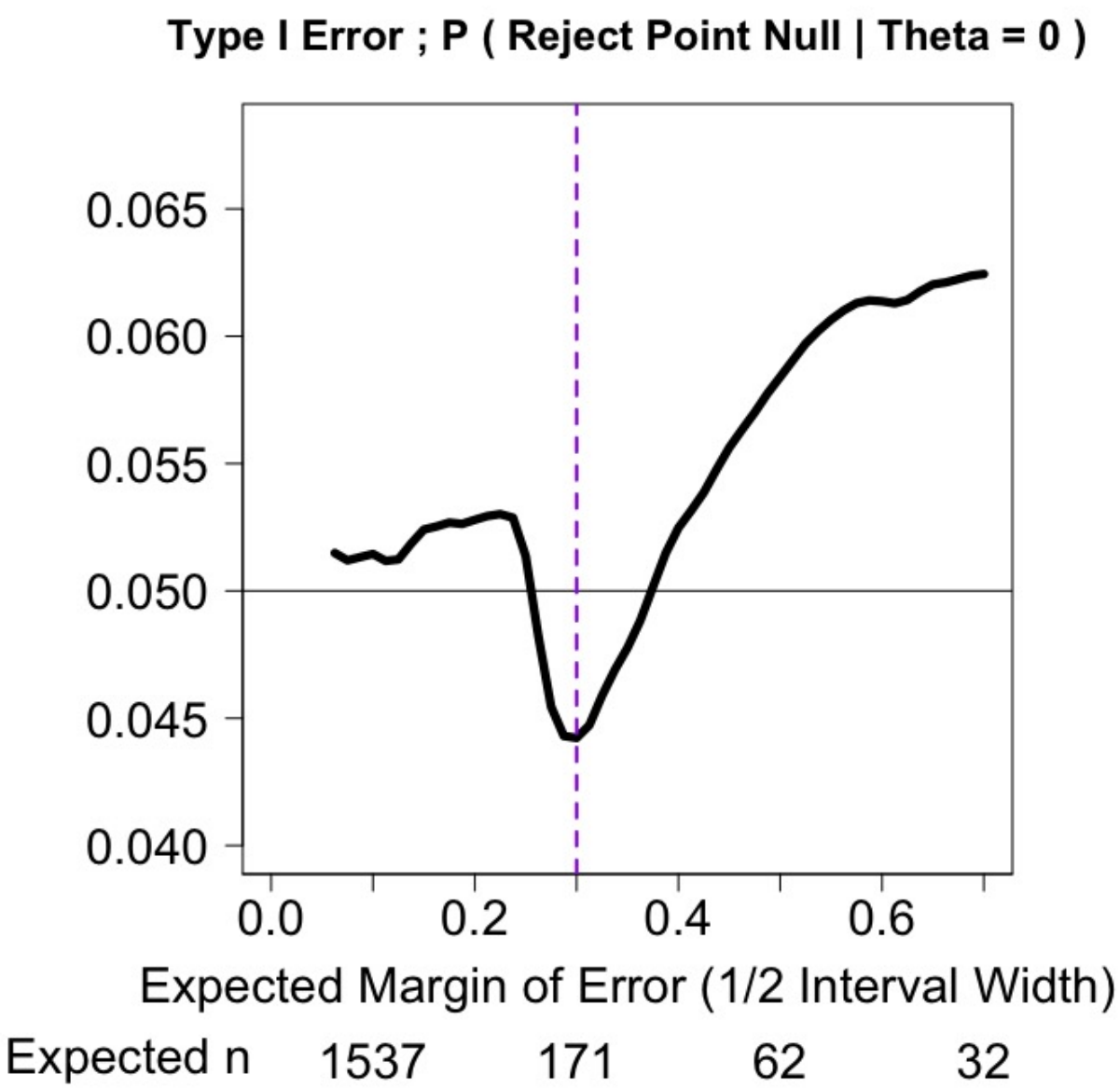
Type I Error


Type I Error ; P (Reject Point Null | Theta = 0)
0.065
0.060
0.055
0.050
0.045
0.040
0.0
0.2
0.4
0.6
Expected Margin of Error (1/2 Interval Width)
Expected n
1537
171
62
32

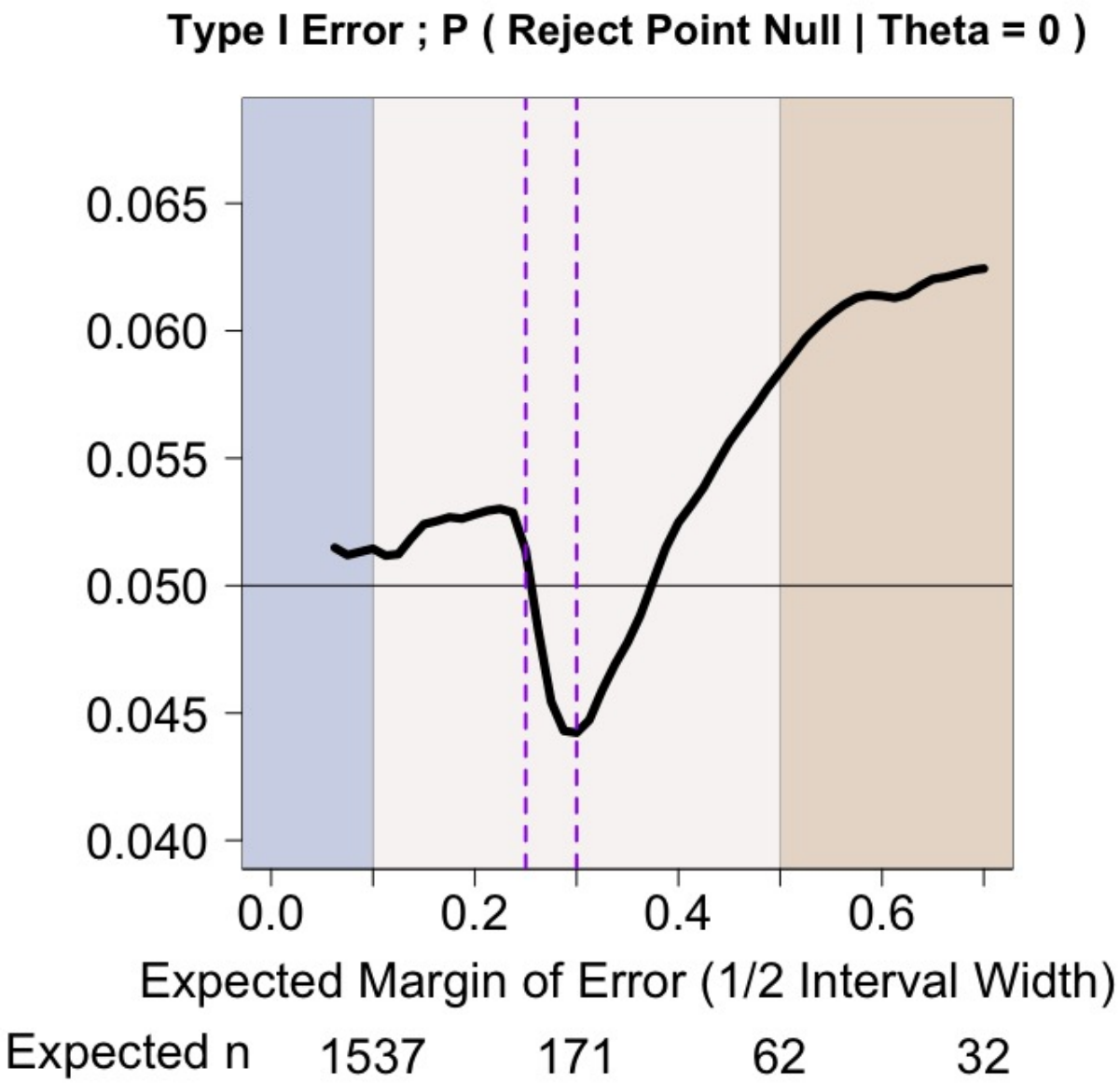
Type I Error


Type I Error ; P (Reject Point Null | Theta = 0)
0.065
0.060
0.055
0.050
0.045
0.040
0.0
0.2
0.4
0.6
Expected Margin of Error (1/2 Interval Width)
Expected n
1537
171
62
32

Type I Error


Type I Error ; P (Reject Point Null | Theta = 0)
0.065
0.060
0.055
0.050
0.045
0.040
0.0
0.2
0.4
0.6
Expected Margin of Error (1/2 Interval Width)
Expected n
1537
171
62
32

Type I Error


Type I Error ; P (Reject Point Null | Theta = 0)
0.065
0.060
0.055
0.050
0.045
0.040
0.0
0.2
0.4
0.6
Expected Margin of Error (1/2 Interval Width)
Expected n
1537
171
62
32

Conclusions

1. AM SGPV novel design to follow studies until ruling out practially null or meaningful effects

2. With immediate stopping (k=0) and unlimited sample size, Type I error can be controlled through the wait time.

Thank you and questions