Sequential Rematching Randomization and Adaptive Monitoring with the Second-Generation p-value to increase the efficiency and efficacy of Randomized Clinical Trials

Jonathan Chipman, MS

Department of Biostatistics
Vanderbilt University

May 10, 2019

RCT Costs and Benefits

2015-2016 Pivotal Trials accepted by FDA **?**

- 138 Trials to approve 59 novel therapeutic drugs
- Enrolled on median 488 patients (IQR: 230, 740)
- Cost per patient on median $41K (IQR: $32K - $82K)

Outline

Three papers to increase study efficiency and efficacy.

- Paper 1: Sequential Rematched Randomization
- Paper 2: Adaptive Monitoring Using the Second Generation p-value
- Paper 3: Estimation of operating characteristics of AM with SGPV

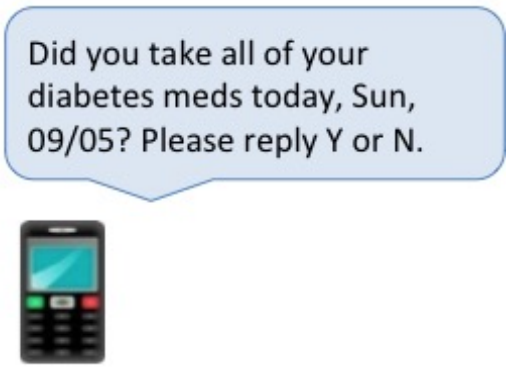
Motivating Example: REACH Trial

REACH Trial

Rapid **E**ducation/Encouragement **A**nd **C**ommunications for **H**ealth

Population:	Adults with Type 2 Diabetes (DM)
Purpose:	Increase glycemic control and adherence to medications
Main Intervention:	Text message-delivered diabetes support for 12 months
Outcome:	Glycemic control (A1c) compared to control
Multi-site enrollment:	512 patients from Vanderbilt and Non-Vanderbilt Clinics
Clinical Relevance:	Change clinical practice – Decrease in A1c of 0.5 or more

REACH



Key Baseline Covariates

Biological Factors

- Baseline A1c*
- Age at baseline
- Race / Ethnicity
- DM Rx type*
- Time since DM dx*

Socio-economic Factors

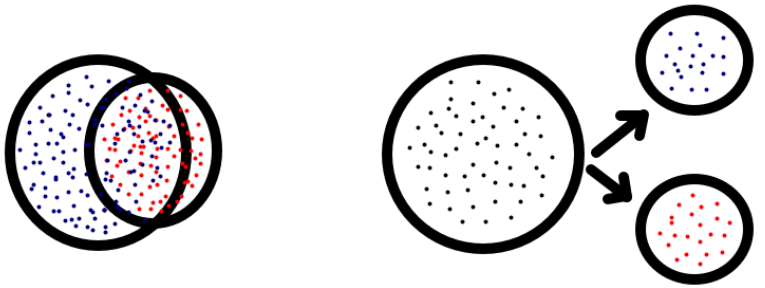
- Yrs of education
- Income level
- Insurance type

* Post-hoc analysis: Greater association with outcome

Stratified Block Randomization

- Limited in number of covariates creating strata
- 4000+ Strata - Stratifying on all covariates*
- 2 * 3 * 3 * 2 = 36 Strata - Stratifying on:
  - 2 Sites
  - 3 Baseline A1c levels
  - 3 DM Rx Types
  - 2 Time since DM Dx levels

Matched Randomization [?]



Observational Study: Bipartite (2 group) matching

Randomization Trial: Non-bipartite matching

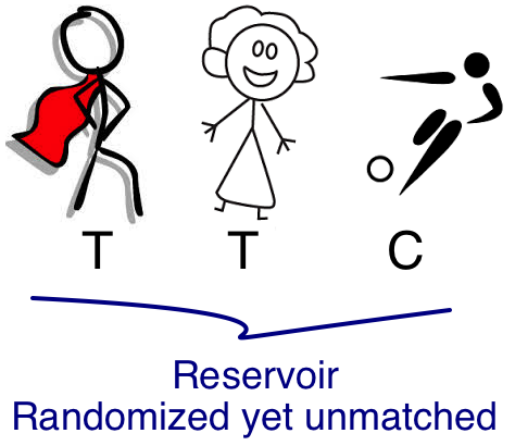
Refined stratification into pairs of patients using distance matrix

- Pairs selected for smallest total distance
- Allows for any number of continuous and categorical covariates
- All participants known before randomization

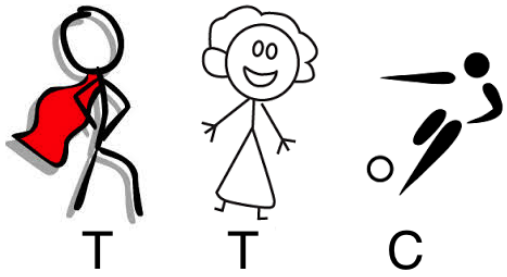
Sequential Matching [?]

Sequential Matching [?]

Sequential Matching [?]

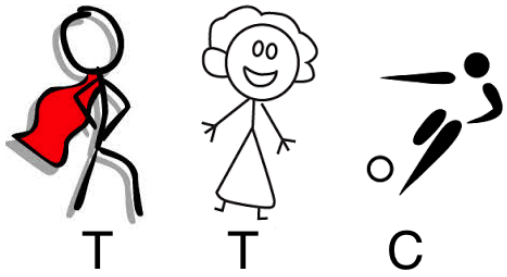


Sequential Matching [?]



If $\text{Distance}_2 < \text{Threshold}$

Sequential Matching [?]



If Distance$_2$ > Threshold

Matching OTF, some additional notes [?]

Pre-specification:

- Initial reservoir size
- Threshold to denote degree of similarity

Fixed Threshold: Match is better than [20%] of random matches

- Mahalanobis Distance of random pairs scales to $F_{(p,n-p)}$
- Assumes normally distributed baseline covariates

Reservoir:

- Not required to deplete
- May result in unequal treatment group sizes

Extension 1: Dynamic Threshold

Dynamic percentile of random match distances

1. Chance of matching now versus later
2. Random match distances (from data, ie not assumed normal)
3. Remove threshold at end to ensure all patients match

Patients enroll in b enrollment blocks
$||U_b||$: Number of unmatched patients at bth enrollment
$||R_b||$: Number remaining patients after bth enrollment

$$Q_b = \frac{||U_b|| - 1}{||U_b|| + ||R_b|| - 1}$$

$$Threshold_b = \begin{cases} F_b^{-1}(Q_b) & ||U_b|| < ||R_b|| \\ \text{best match(es)} & ||U_b|| \geq ||R_b|| \end{cases}$$

F_b is estimated by $\hat{F}_b$ by bootstrap sampling random matches from distance matrix.

Extension 2: Sequential Re-Matching

Sequential Matching

- Formed matches remain through study

Sequential Re-Matching

- Matches are allowed to break if a better match enters
- However, participants keep original treatment assignment
  - New participant: Match to anybody
  - Former enrolled participant: Match to anyone of opposite treatment or new participant

Case study using REACH Data to comparing randomization schemes

Balance of baseline covariates

Efficiency estimating treatment effect

Results will vary based on baseline covariates

- Number of patients
- Number of covariates
- Distribution of covariates
- Adjusted R^2 of covariates

Simulations to evaluate Sequential Rematching Randomization

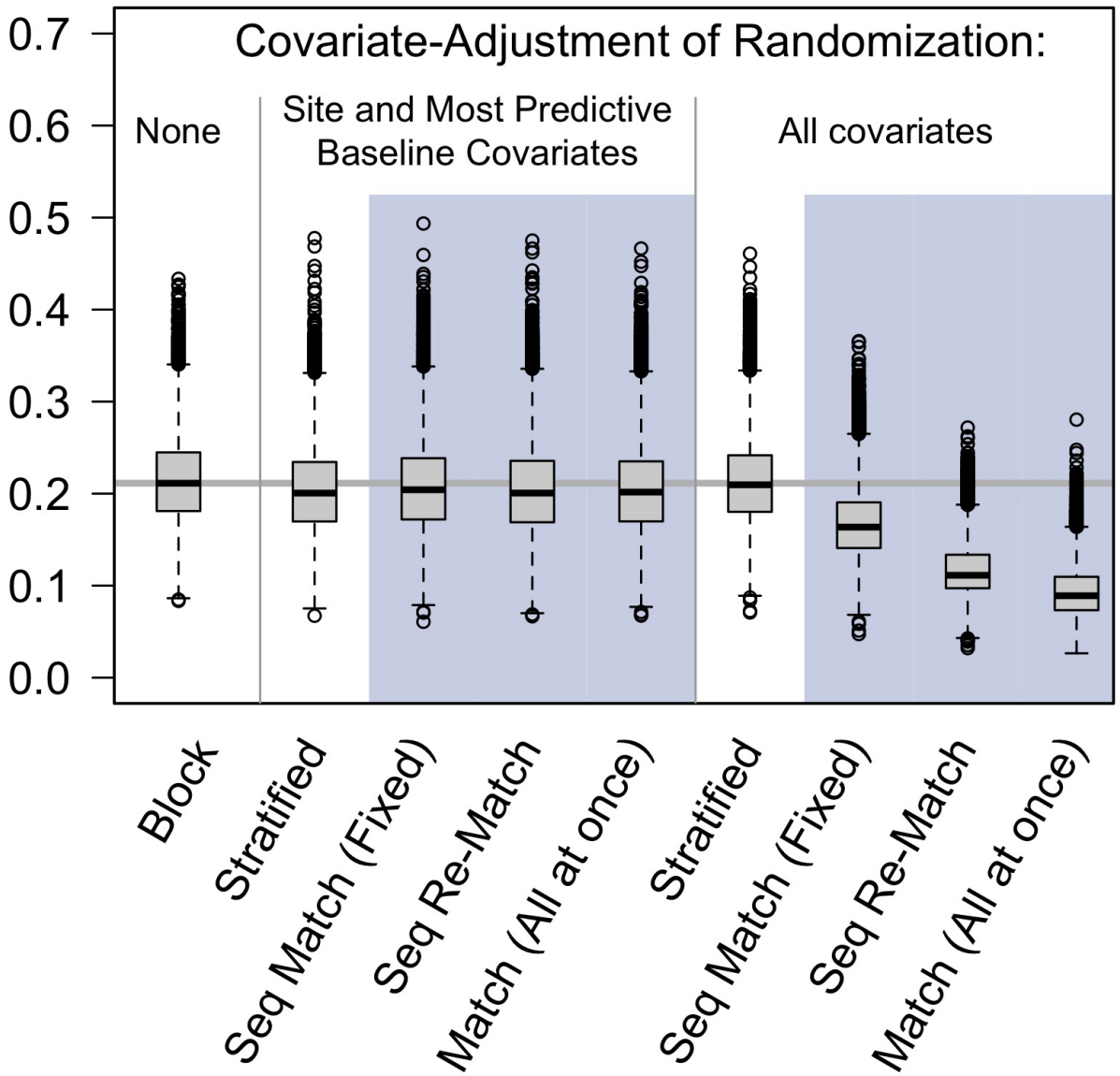
Generate 20K MCMC replicates under potential outcomes framework:

- Bootstrap sample 512 observations from REACH trial
- Outcomes
  - Y(0) = Predicted three month A1c + random residual
  - Y(1) = Y(0) - 0.5
- Generate randomization scheme under
  - Block Randomization
  - Stratified Block Randomization
  - Matched Randomization
- Observe
  - Balance: Average Standardized Mean Difference among all covariates
  - Efficiency: CI Width of 1) fully-adjusted linear model and 2) Permutation distribution

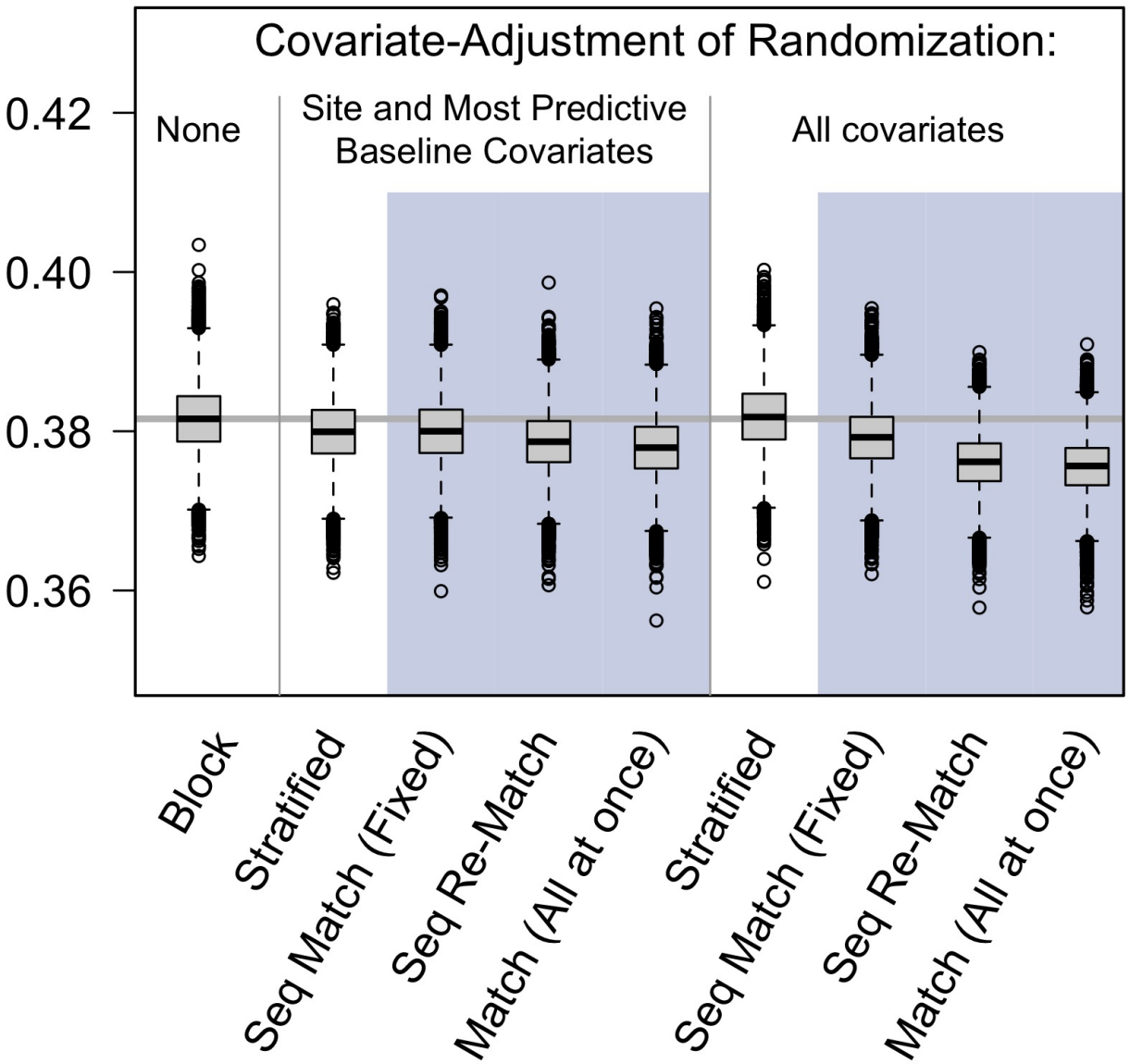
Where Stratified and Matched randomization adjusts for

- Site, Baseline A1c, DM Rx Type, and Time since DM Dx
- All baseline covariates

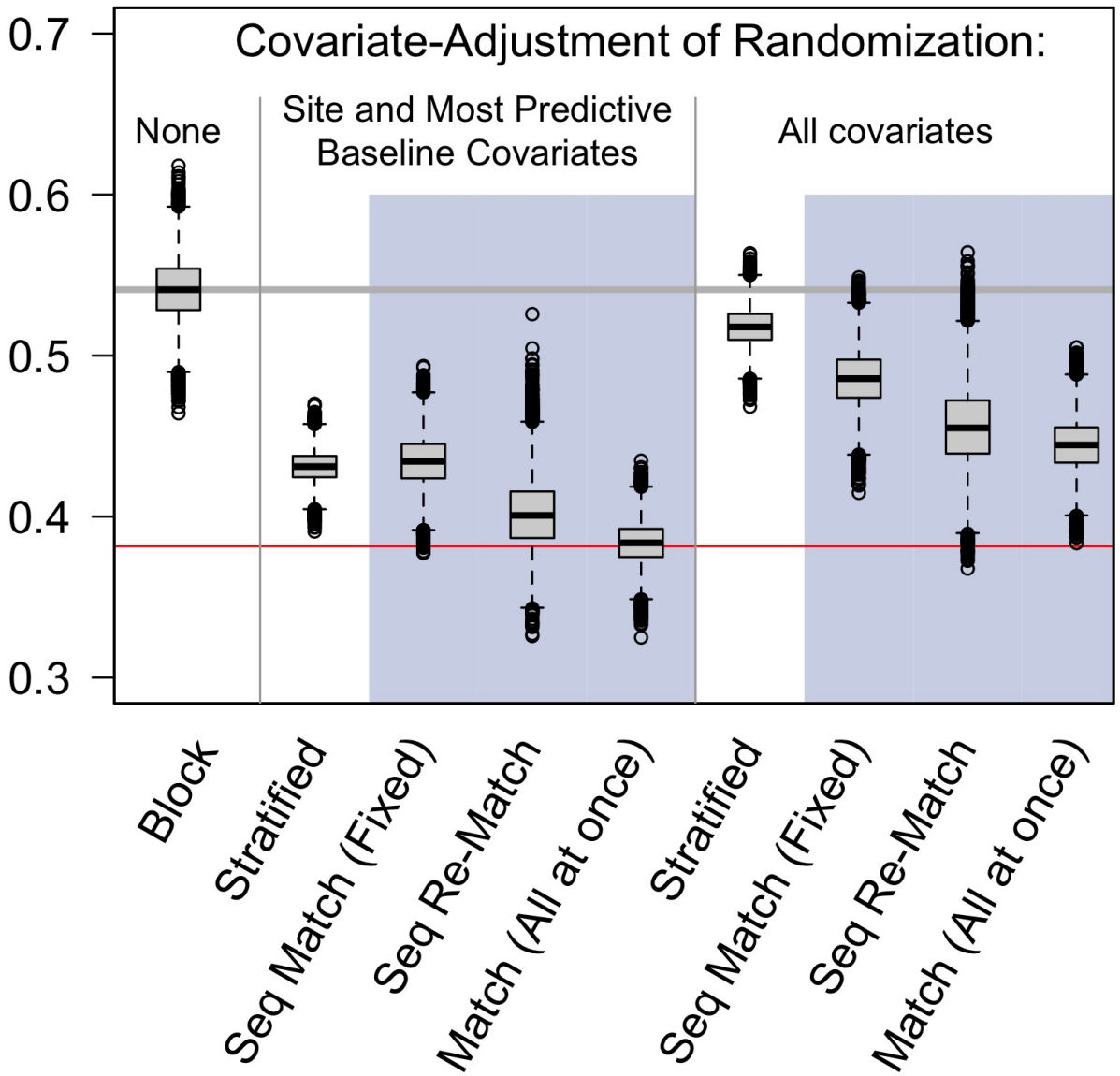
Maximum Standardized Mean Difference (all baseline covariates)



Covariate-Adjusted OLS Confidence Interval Width of estimated treatment effect



Permutation Confidence Interval Width of estimated treatment effect



Recommendations regarding balance and efficiency

Conclusions

- Greatest overall balance by adjusting for all baseline covariates
- Greatest linear gains in efficiency
  - Model efficiency: adjusting randomization to all baseline covariates
  - Permutation efficiency: adjusting randomization to priority covariates
- Permutation efficiency achieves nearly same efficiency as fully-adjusted model with Block Randomization

Practical Recommendations

- Recommend adjusting to all baseline covariates
- Adaptively monitor until reaching a clear clinical conclusion

Paper 2: Adaptive Monitoring Using the Second Generation p-value

Prematurely Ending Clinical Trial(s)

Towards a Revolution in COPD Health (TORCH) (**?**)

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

2007 6112 patients

- 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) (**?**)

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

2007 6112 patients

- 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) (**?**)

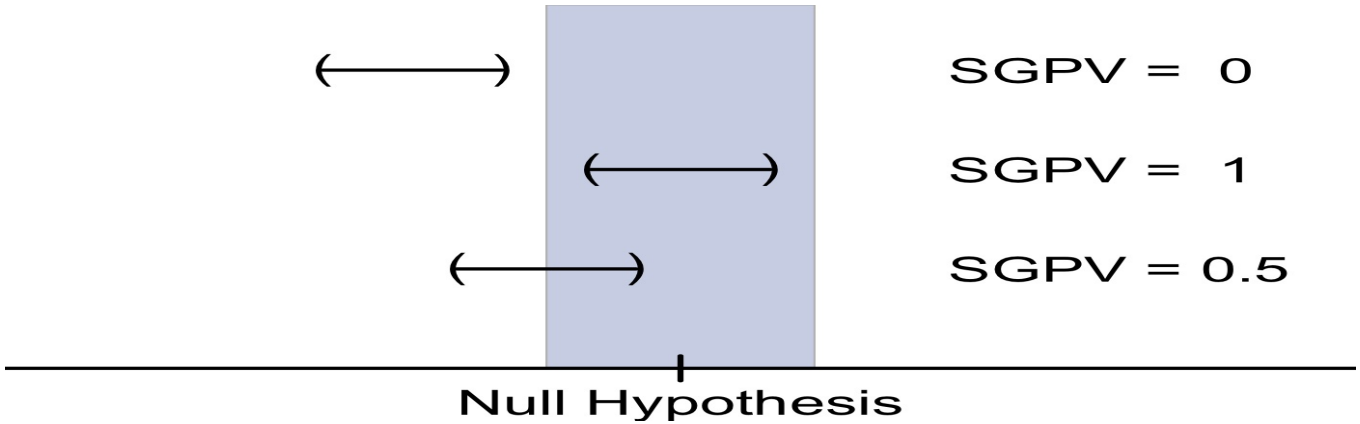
Primary Aim: Establish whether beta-agonist (salmeterol plus fluticasone propionate) has survival benefit in patients with chronic obstructive pulmonary disease

2007 6112 patients

- 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.'

Second Generation p-value (SGPV; ?)

What proportion of interval overlaps with trivial effects?

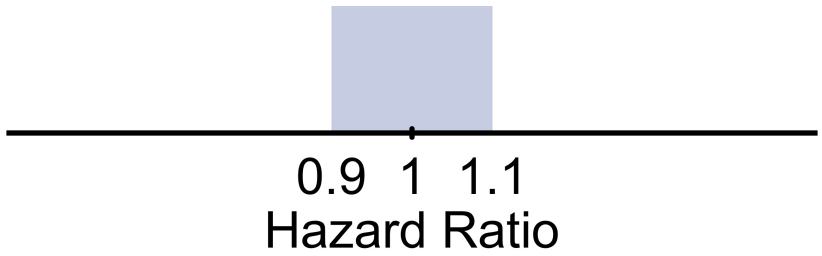


Interpretation of SGPV_T (T for trivial effects)

- $\text{SGPV}_T = 0.0$: Evidence to rule out trivial effects
- $\text{SGPV}_T = 1.0$: Evidence to rule out non-trivial effects
- $\text{SGPV}_T = 0.5$: Inconclusive, need more data

Second Generation p-value (SGPV; ?)

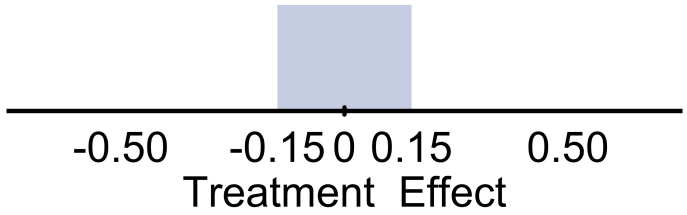
Example of indifferenze zone with TORCH



If HR were 0.825 (95% CI: 0.681-0.899)
Interpretation of $\text{SGPV}_T = 0$: Rule out trivial effects

REACH

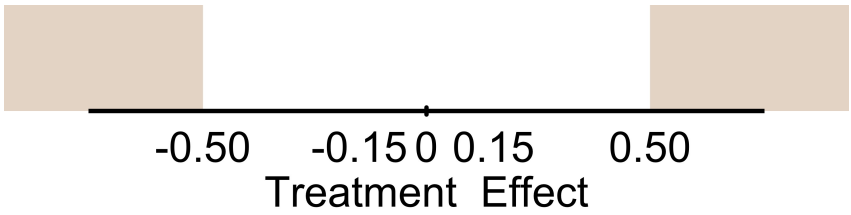
Rapid **E**ducation/Encouragement **A**nd **C**ommunications for **H**ealth



A decrease is HbA1c of 0.15 (ex. 7.5 to 7.35)
is considered essentially no change

REACH

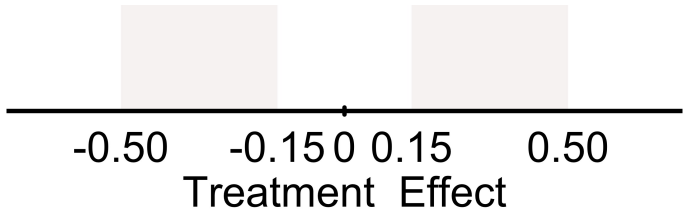
Rapid **E**ducation/Encouragement **A**nd **C**ommunications for **H**ealth



A decrease is HbA1c of 0.5 (ex. 7.5 to 7.0)
would highly suggest adopting intervention

REACH

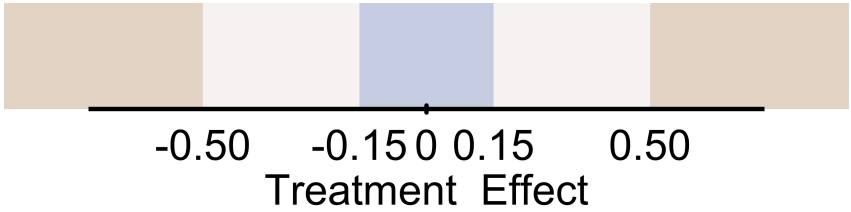
Rapid **E**ducation/Encouragement **A**nd **C**ommunications for **H**ealth



A decrease is HbA1c of between 0.15 and 0.5 is interesting
it may suggest adopting intervention

REACH

Rapid **E**ducation/Encouragement **A**nd **C**ommunications for **H**ealth



SGPV_T to rule out trivial effects
SGPV_{HA} to rule out highly-actionable effects

$\text{SGPV}_T = 0.0$: Ruled out trivial effects

$\text{SGPV}_{HA} = 0.0$: Ruled out highly actionable effects

Adaptive Monitoring with SGPV

Wait	interval width stabilizes
Monitor	interval and SGPV at desired looks
Alert	$\text{SGPV}_T = 0.0$: Ruled out trivial effects
	$\text{SGPV}_{HA} = 0.0$: Ruled out highly actionable effects
Affirm	stop if same conclusion [40] patients later
	end of resources ($n = 512$)
Report	only the final interval when stopping

Adaptive Monitoring with SGPV

Wait	interval width stabilizes
Monitor	interval and SGPV at desired looks
Alert	$\text{SGPV}_T = 0.0$: Ruled out trivial effects
	$\text{SGPV}_{HA} = 0.0$: Ruled out highly actionable effects
Affirm	stop if same conclusion [40] patients later
	end of resources ($n = 512$)
Report	only the final interval when stopping

Adaptive Monitoring with SGPV

Wait	interval width stabilizes
Monitor	interval and SGPV at desired looks
Alert	$\text{SGPV}_T = 0.0$: Ruled out trivial effects
	$\text{SGPV}_{HA} = 0.0$: Ruled out highly actionable effects
Affirm	stop if same conclusion [40] patients later
	end of resources ($n = 512$)
Report	only the final interval when stopping

Adaptive Monitoring with SGPV

Wait	interval width stabilizes
Monitor	interval and SGPV at desired looks
Alert	$\text{SGPV}_T = 0.0$: Ruled out trivial effects
	$\text{SGPV}_{HA} = 0.0$: Ruled out highly actionable effects
Affirm	stop if same conclusion [40] patients later
	end of resources ($n = 512$)
Report	only the final interval when stopping

Adaptive Monitoring with SGPV

Wait	interval width stabilizes
Monitor	interval and SGPV at desired looks
Alert	$\text{SGPV}_T = 0.0$: Ruled out trivial effects
	$\text{SGPV}_{HA} = 0.0$: Ruled out highly actionable effects
Affirm	stop if same conclusion [40] patients later
	end of resources ($n = 512$)
Report	only the final interval when stopping

Adaptive Monitoring with SGPV

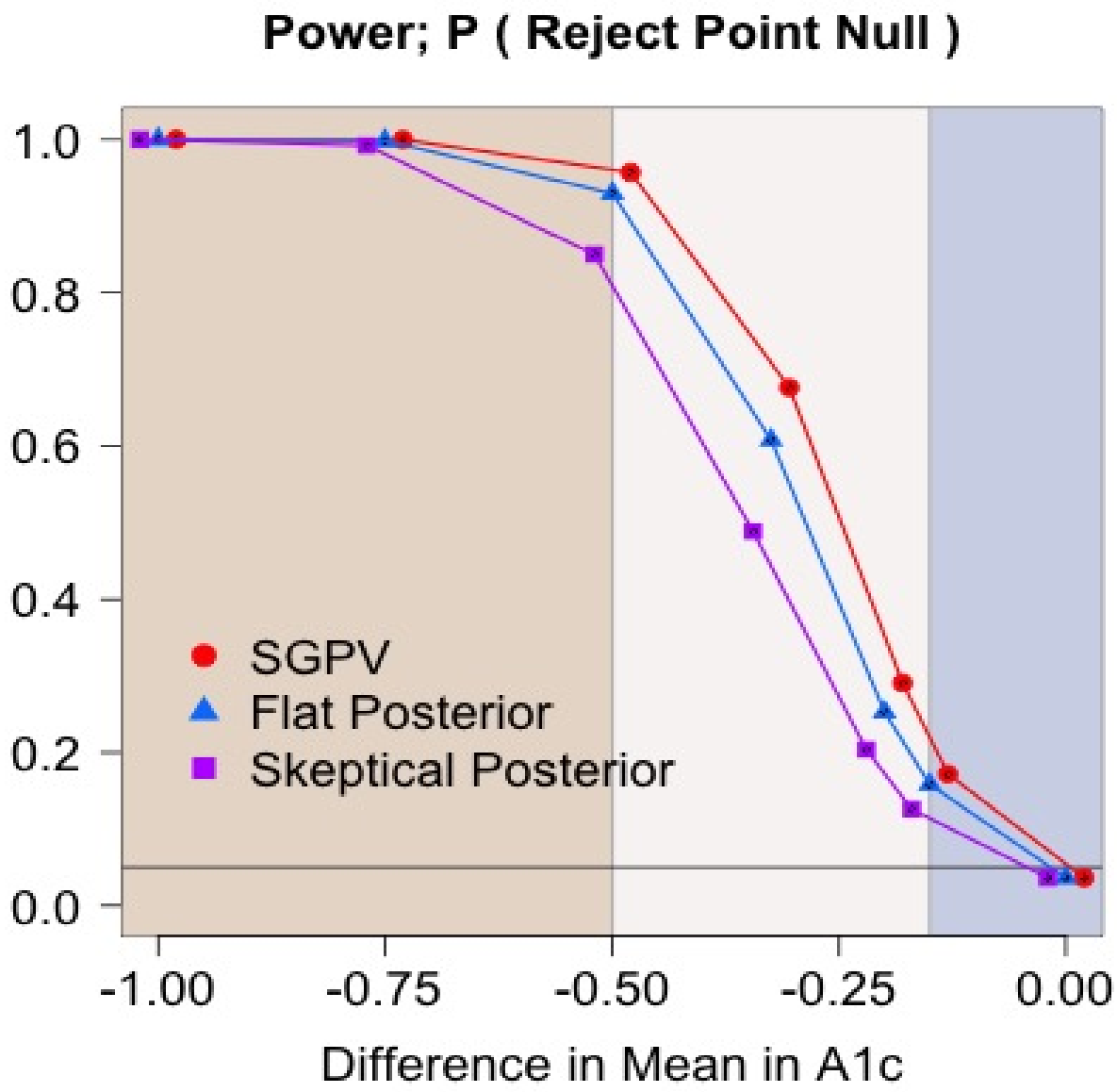
Wait	interval width stabilizes
Monitor	interval and SGPV at desired looks
Alert	$\text{SGPV}_T = 0.0$: Ruled out trivial effects
	$\text{SGPV}_{HA} = 0.0$: Ruled out highly actionable effects
Affirm	stop if same conclusion [40] patients later
	end of resources ($n = 512$)
Report	only the final interval when stopping

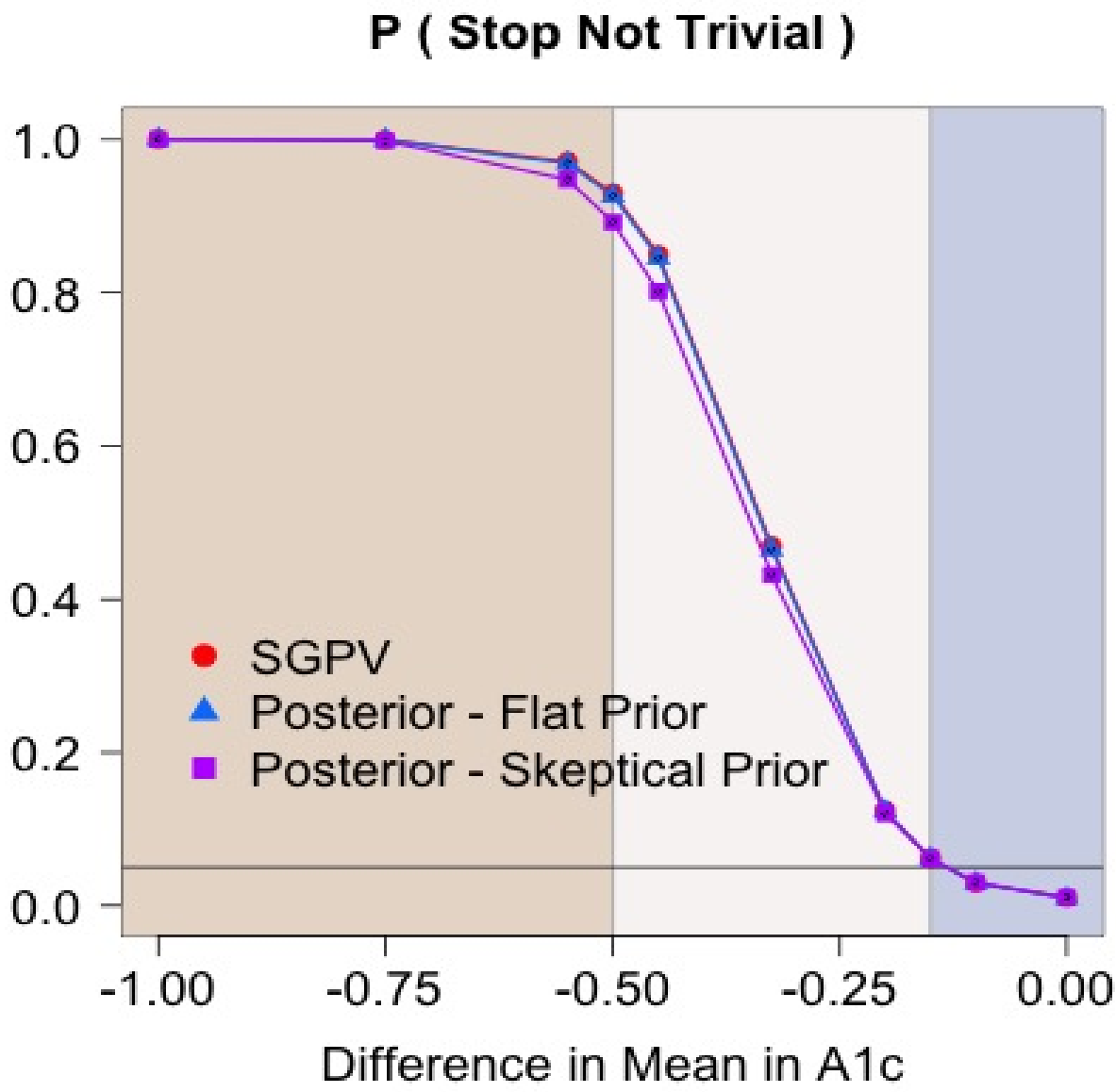
Adaptive Monitoring with SGPV

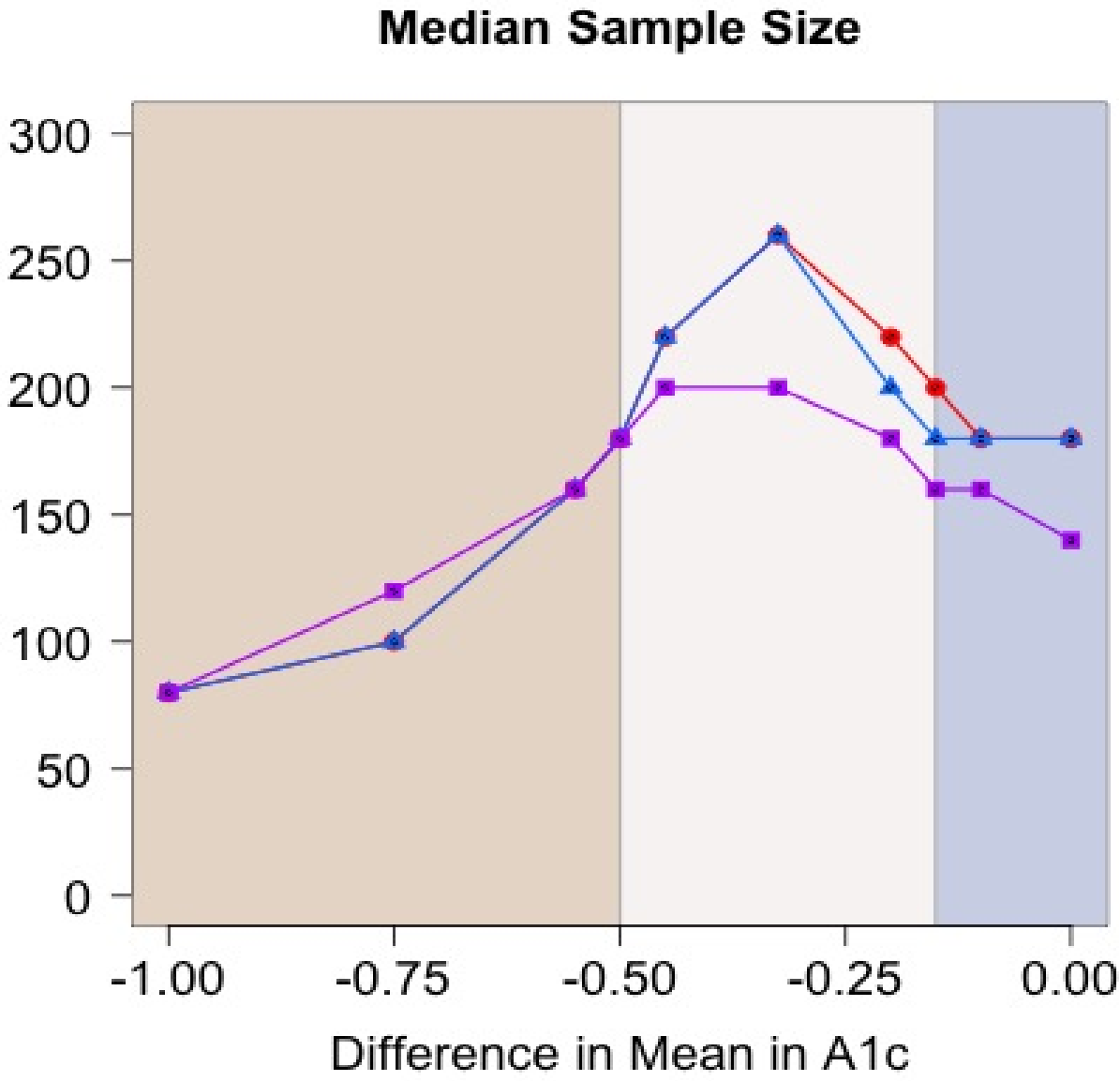
Wait	interval width stabilizes
Monitor	interval and SGPV at desired looks
Alert	$\text{SGPV}_T = 0.0$: Ruled out trivial effects
	$\text{SGPV}_{HA} = 0.0$: Ruled out highly actionable effects
Affirm	stop if same conclusion [40] patients later
	end of resources ($n = 512$)
Report	only the final interval when stopping

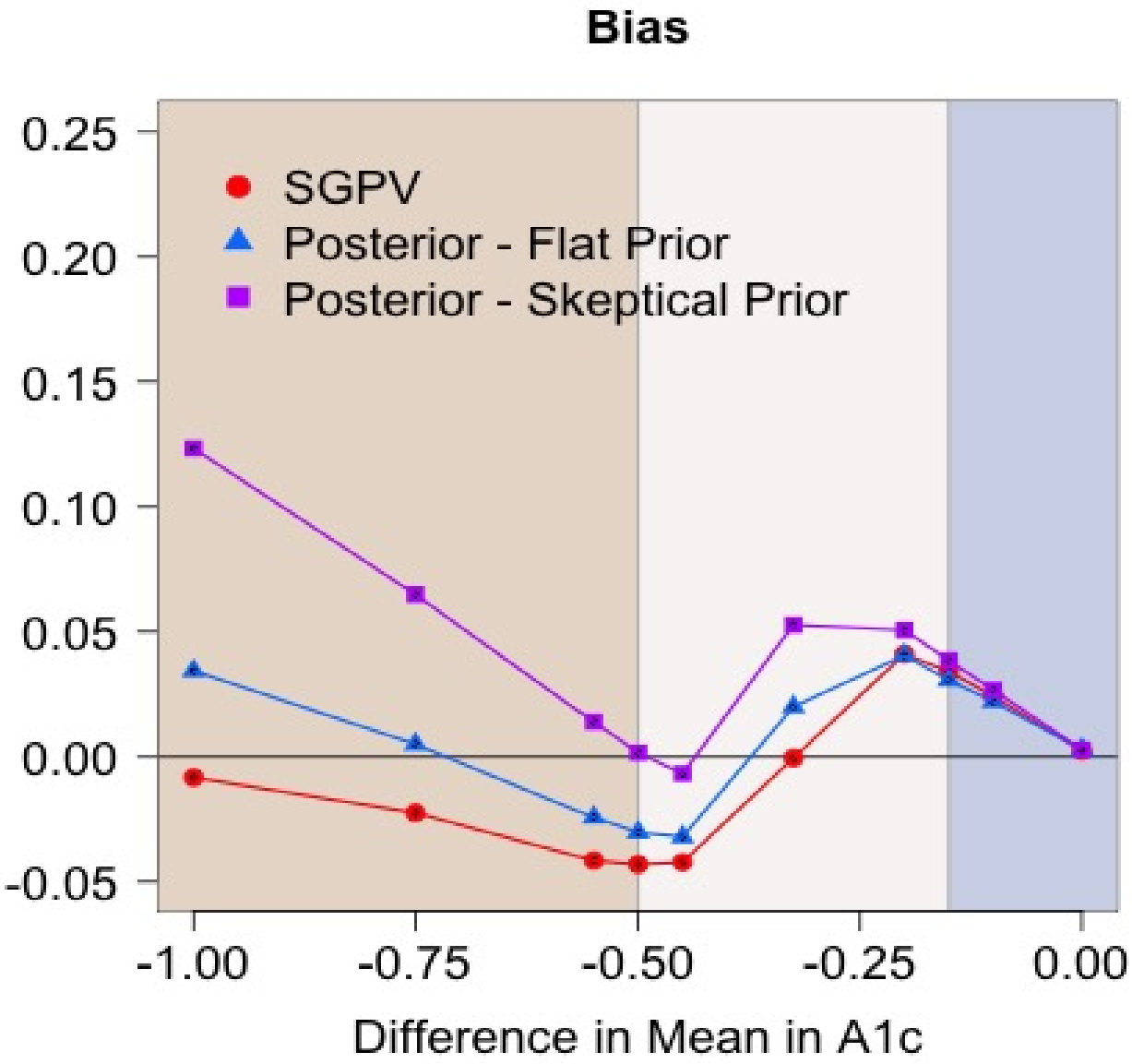
Adaptive Monitoring with SGPV

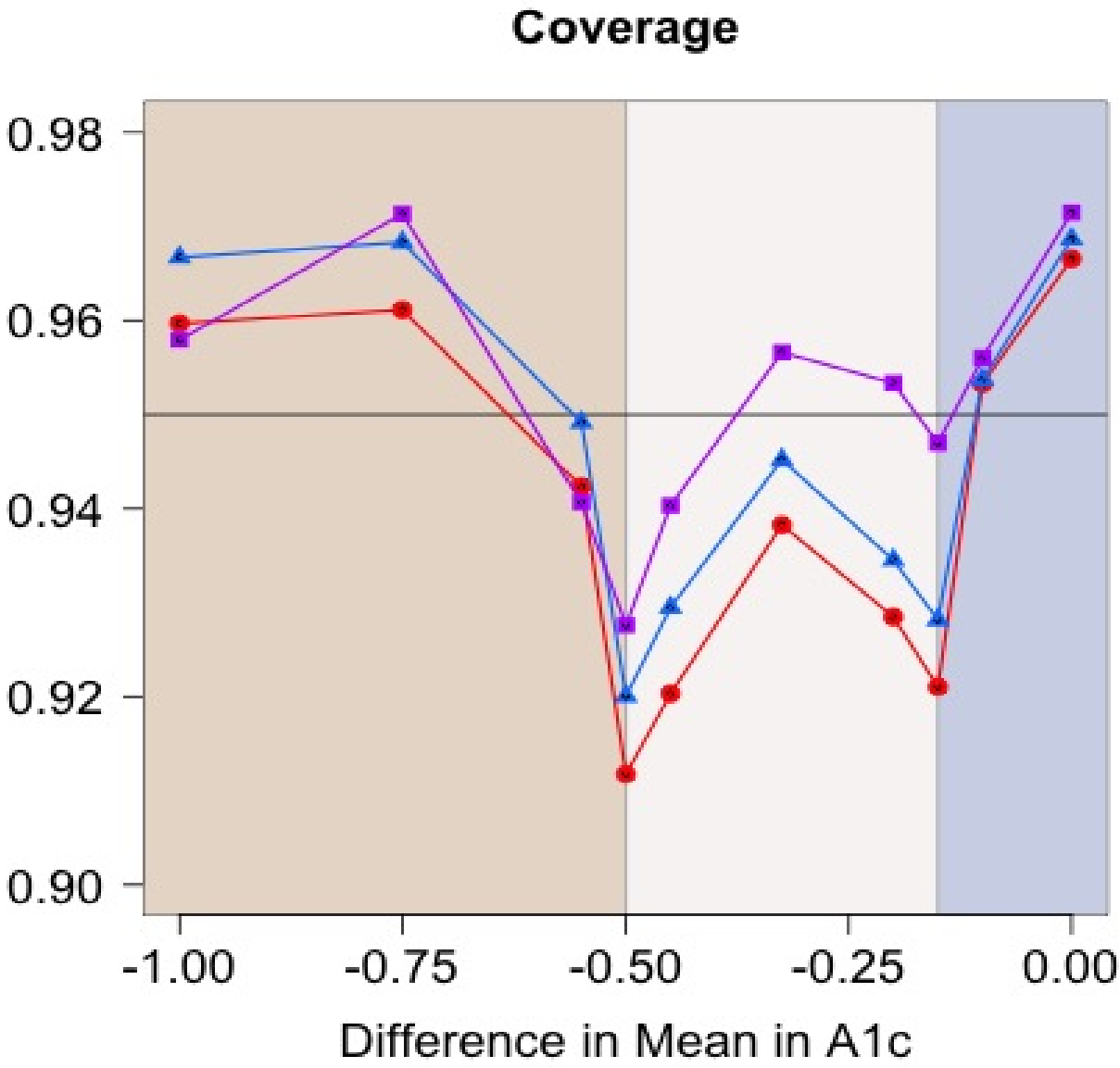
Wait	interval width stabilizes
Monitor	interval and SGPV at desired looks
Alert	$SGPV_T = 0.0$: Ruled out trivial effects
	$SGPV_{HA} = 0.0$: Ruled out meaningful effects
Affirm	stop if same conclusion [40] patients later
	end of resources (n = 512)
Report	only the final interval when stopping


Power; P (Reject Point Null)
1.0
0.8
0.6
0.4
0.2
0.0
SGPV
Flat Posterior
Skeptical Posterior
-1.00
-0.75
-0.50
-0.25
0.00
Difference in Mean in A1c


P (Stop Not Trivial)
1.0
0.8
0.6
0.4
0.2
0.0
SGPV
Posterior - Flat Prior
Posterior - Skeptical Prior
-1.00
-0.75
-0.50
-0.25
0.00
Difference in Mean in A1c


Median Sample Size
300
250
200
150
100
50
0
-1.00
-0.75
-0.50
-0.25
0.00
Difference in Mean in A1c


Bias
SGPV
Posterior - Flat Prior
Posterior - Skeptical Prior
0.25
0.20
0.15
0.10
0.05
0.00
-0.05
-1.00
-0.75
-0.50
-0.25
0.00
Difference in Mean in A1c


Coverage
0.98
0.96
0.94
0.92
0.90
-1.00
-0.75
-0.50
-0.25
0.00
Difference in Mean in A1c

Paper 3: sgpvAM R package and practical guidance to control adaptive monitoring errors

sgpvAM package

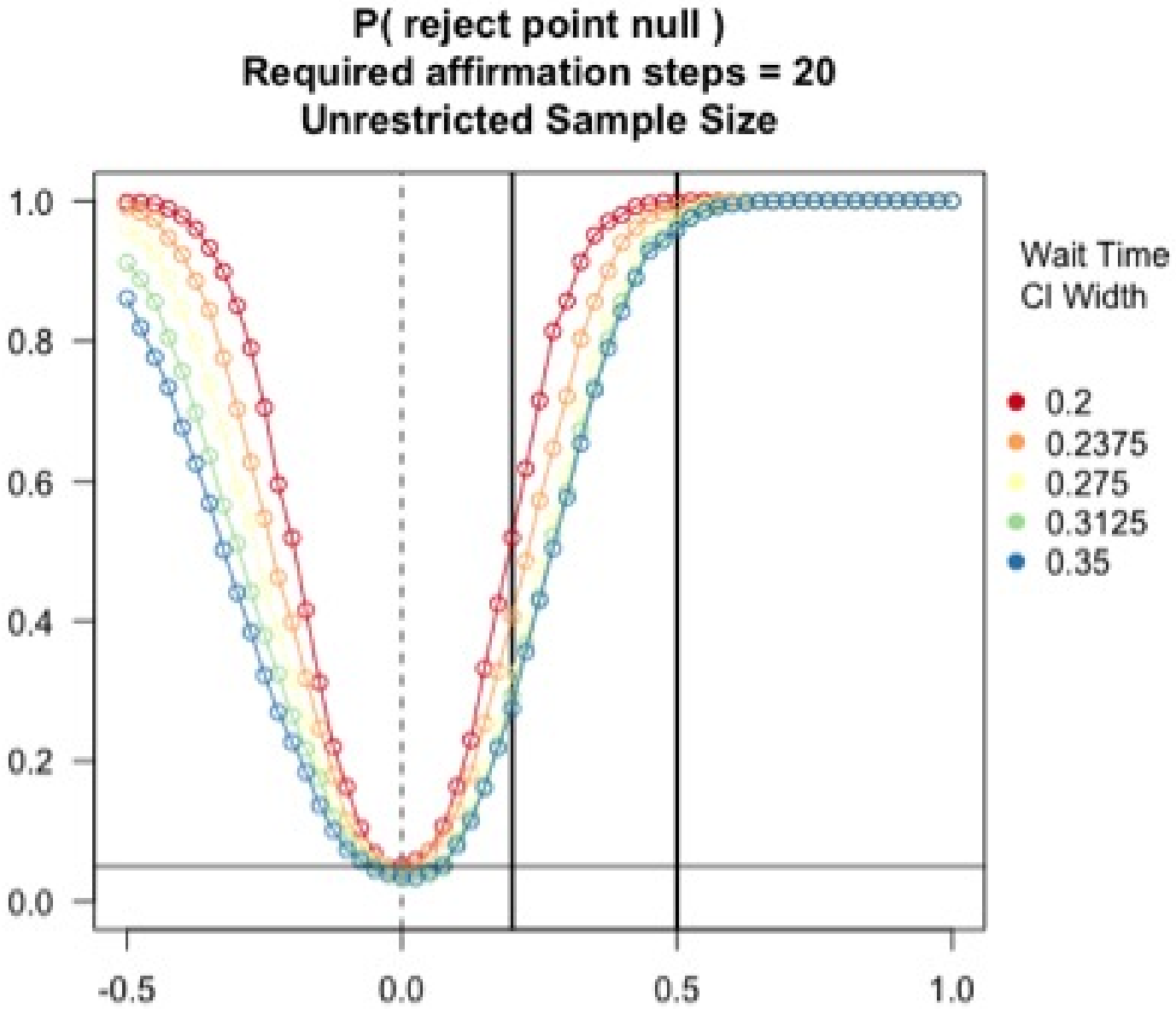
1. How to easily operationalize?
2. Encourage not changing clinical relevance
3. Under fully sequentially monitored study, how long to wait to ensure Type I Error < 0.05?
4. What are operating characteristics when observations are not observed immediately? When limited in sample size?

sgpvAM package

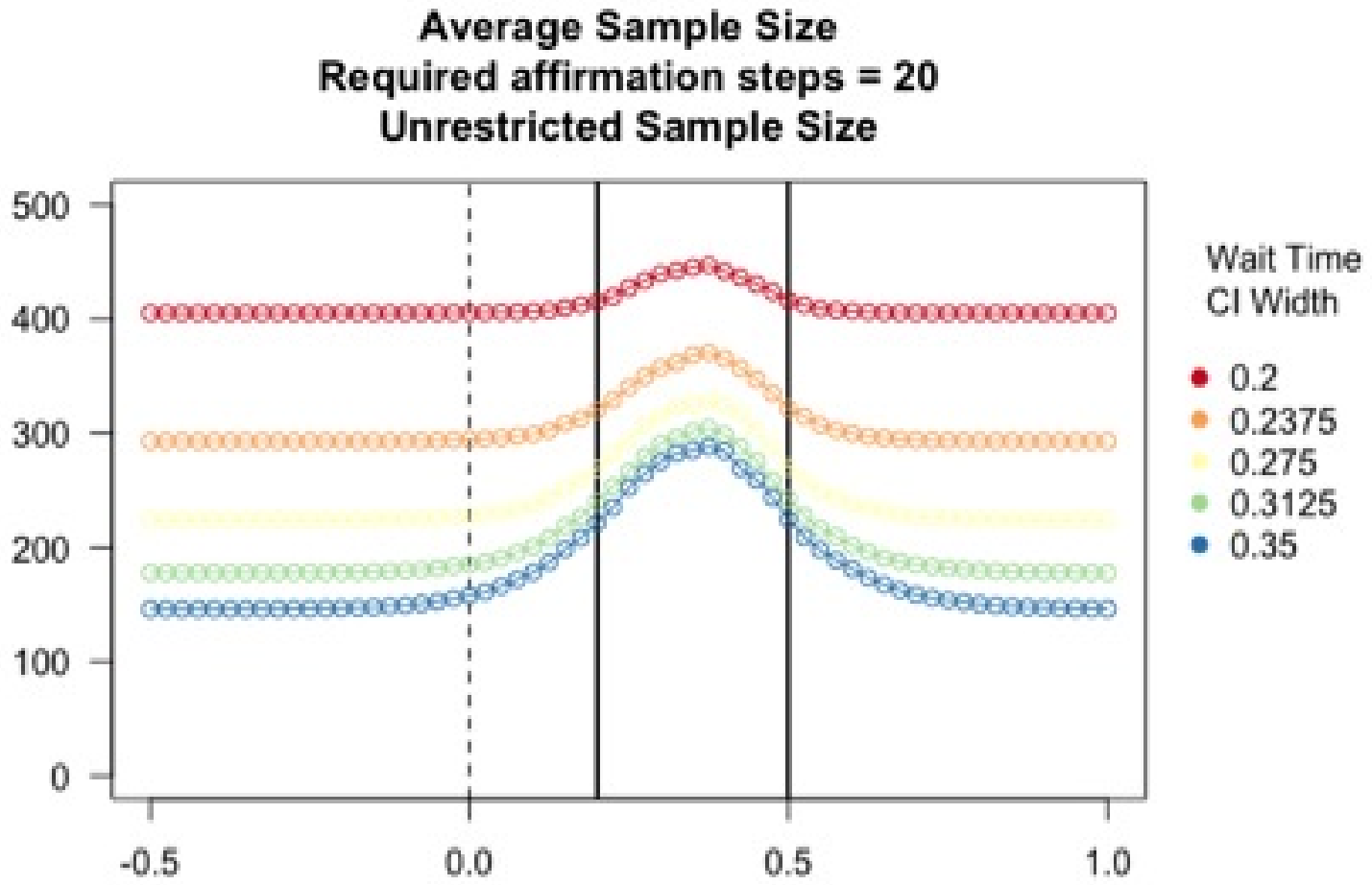
sgpvAM package
1. Uses Valerie's spgv package on github
2. Vignette with examples

– Example 1-sided study with at most trivial effects = 0.2 and minimal highly actionable effect of 0.5, lag of 50 observations, will monitor no more frequently than 10 patients.

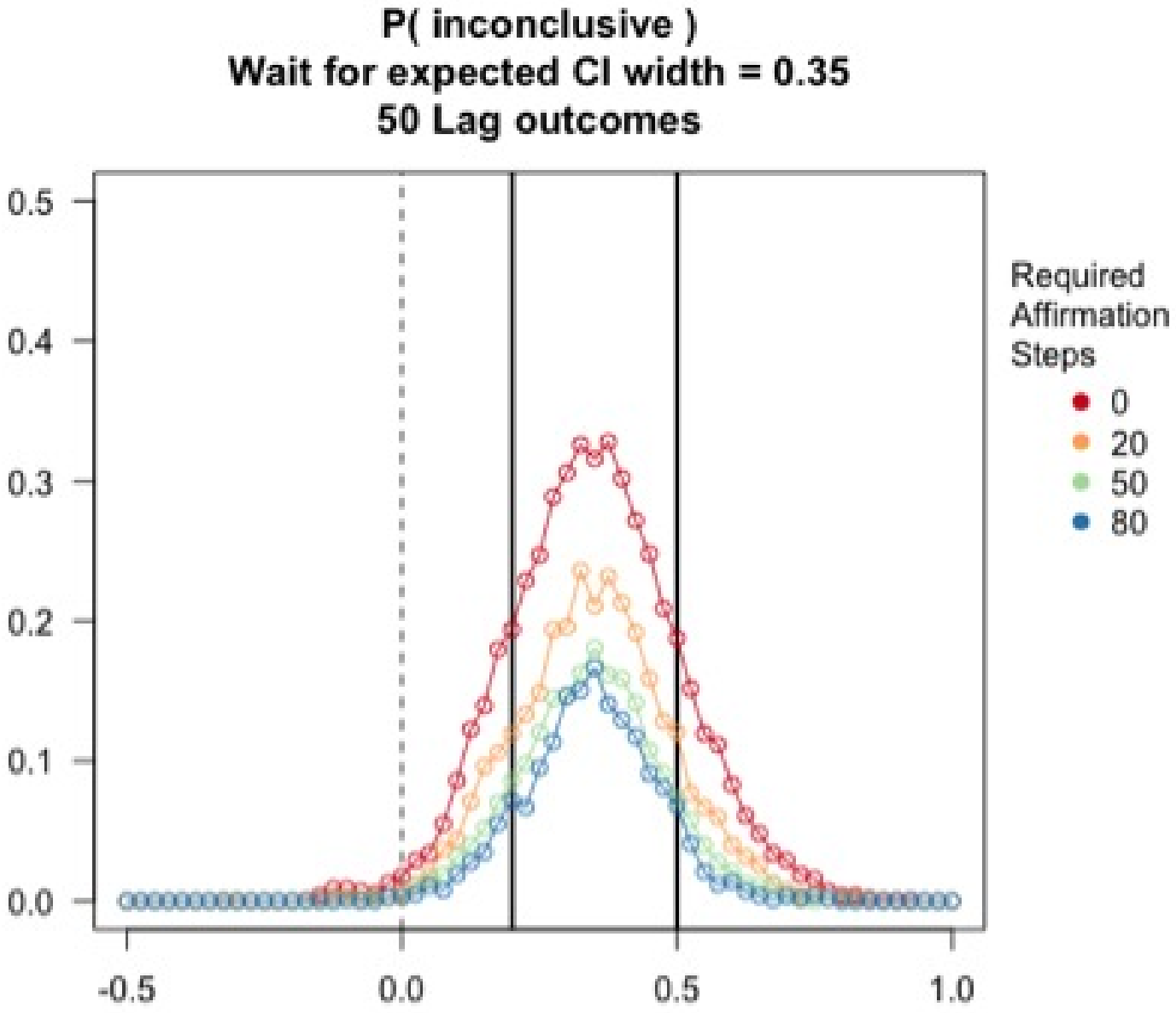
sgpvAM package



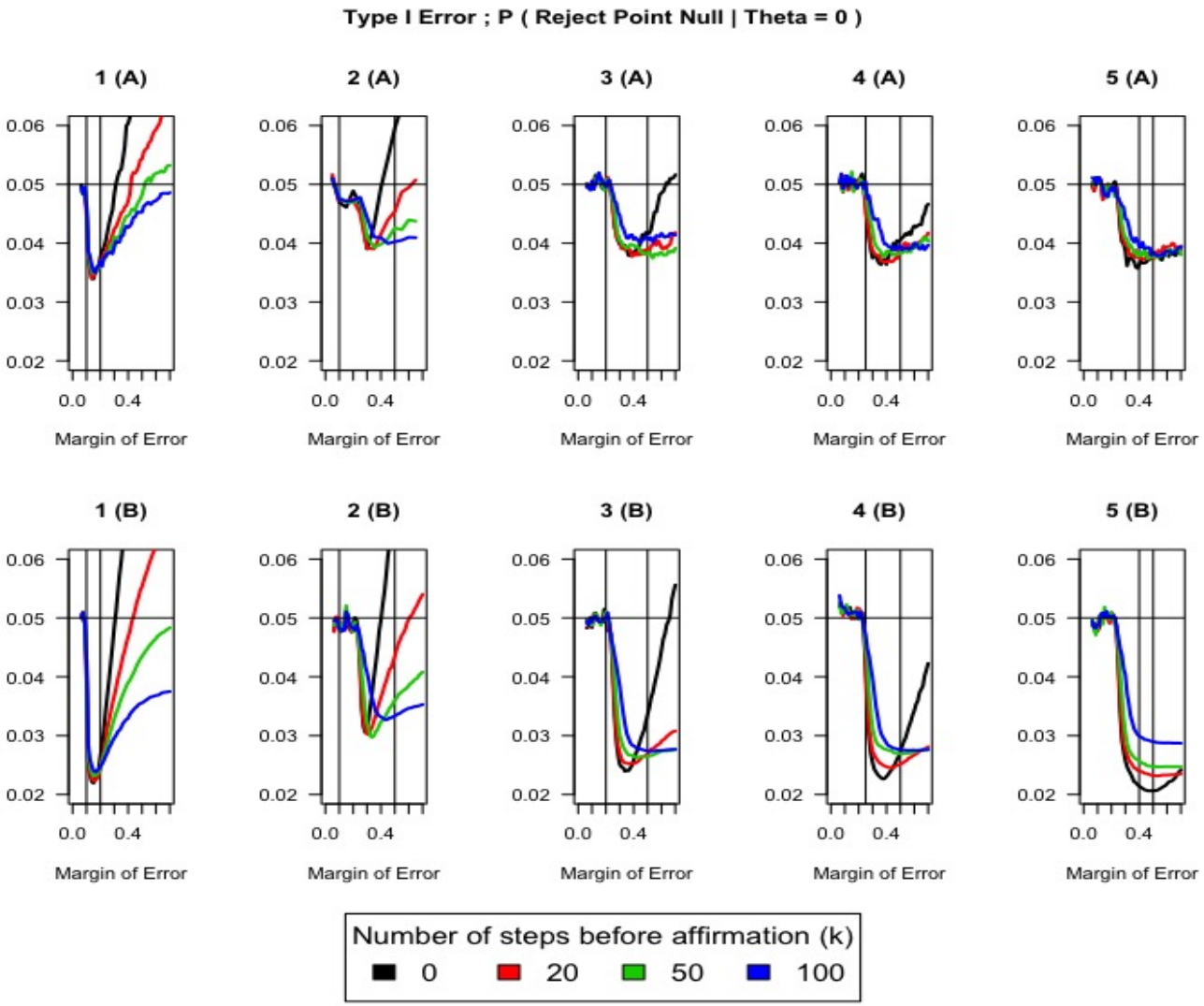
sgpvAM package



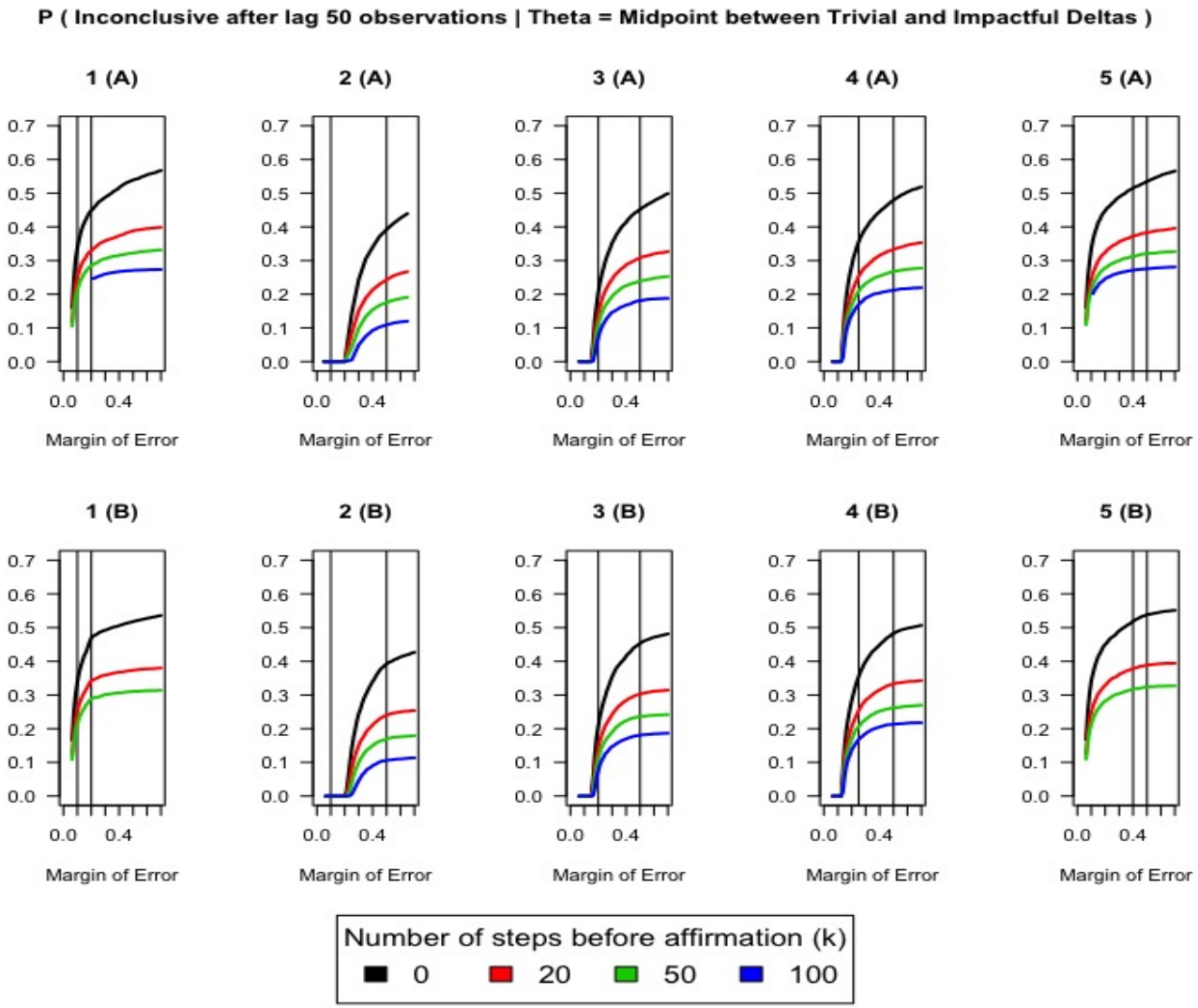
sgpvAM package



sgpvAM package



sgpvAM package



References