

Part II

Second-generation *p*-values: equivalence tests, statistical properties, and false discovery rates

Jeffrey D. Blume, and Megan H. Murray,

PhD

PhD

School of Data Science

Department of Biostatistics

University of Virginia

Vanderbilt University

Course Layout

- Slides Part I: Introduction and applications
 - Coding Part I
- Lunch (12:00-1:00pm)
- Slides Part II: Statistical Properties, Equivalence tests, false discovery rates, and study planning
 - Coding Part II
- Slides Part III: SGPV Variable Selection
 - Coding Part III
- Questions and Discussion

Outline

- Statistical Properties cont.
- Equivalence Tests
 - Two One-Sided Tests (TOST)
 - Comparison to SGPVs
- False Discovery Rates
 - R Packages
 - `sgpv::fdrisk()`
 - `FDRestimation::p.fdr()`
- SGPV Study Planning

Blume and Murray, 2022

Second-generation p -value

- Statistical properties in TAS & PLOS One
- Retains strict error control

Evidential Metric	What it measures	SGPV
1	Summary measure	SGPV (p_δ)
2	Operating characteristics	$P(p_\delta = 0 H_0)$ $P(p_\delta = 1 H_1)$ $P(0 < p_\delta < 1 H)$
3	False discovery rates	$P(H_0 p_\delta = 0)$ $P(H_1 p_\delta = 1)$

ASA SGPV Short Course

Blume and Murray, 2022

Statistical Properties

Suppose interval I has coverage probability $1-\alpha$, then

Three ‘Error’ Rates

1. $P(p_\delta = 0|H_0) \leq \alpha$ and $\rightarrow 0$ as $n \rightarrow \infty$
2. $P(p_\delta = 1|H_1) \leq \alpha$ and $\rightarrow 0$ as $n \rightarrow \infty$
3. $P(0 < p_\delta < 1|H)$ controlled through sample size

Will examine
these first

Two False Discovery Rates

1. $P(H_0 | p_\delta = 0)$
2. $P(H_1 | p_\delta = 1)$

Will graph to
illustrate

Statistical Properties

- Three Inferential Categories

1. $p_\delta = 0 \Rightarrow$ data **incompatible** with null
2. $p_\delta = 1 \Rightarrow$ data **compatible** with null
3. $0 < p_\delta < 1 \Rightarrow$ data are **inconclusive**

- Three ‘error’ rates

1. $P(p_\delta = 0|H_0)$ when H is null
2. $P(p_\delta = 1|H_1)$ when H is not null
3. $P(0 < p_\delta < 1|H)$ when H is either

- Assume H makes statements about a parameter θ

- Large sample setting

Statistical Properties

- How often are the data incompatible with null?
- Examine $P(p_\delta = 0|\theta)$ as θ varies
 - Power function
- This probability
 - converges to one for alternatives not near the edge of interval null
 - converges to zero for null hypotheses not near the edge of the null set
 - converges to alpha for hypotheses approaching or on the edge of the null set

ASA SGPV Short Course

Blume and Murray, 2022

‘Power’ Function

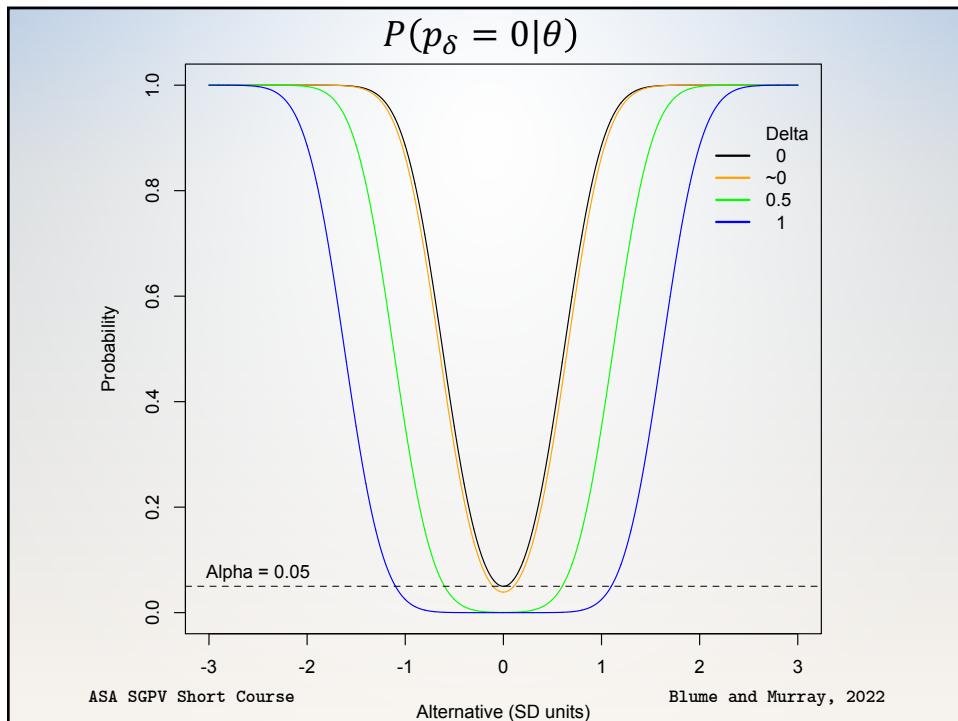
- θ_0 : point null, σ : standard deviation
- δ : half-width of indifference zone

$$P(p_\delta = 0|\theta) = \Phi \left[\frac{\sqrt{n}(\theta_0 - \theta)}{\sigma} - \frac{\sqrt{n}\delta}{\sigma} - Z_{\alpha/2} \right] + \Phi \left[-\frac{\sqrt{n}(\theta_0 - \theta)}{\sigma} - \frac{\sqrt{n}\delta}{\sigma} - Z_{\alpha/2} \right]$$

$$P_{\theta_0}(p_\delta = 0|\theta_0) = 2\Phi \left[-\frac{\sqrt{n}\delta}{\sigma} - Z_{\alpha/2} \right]$$

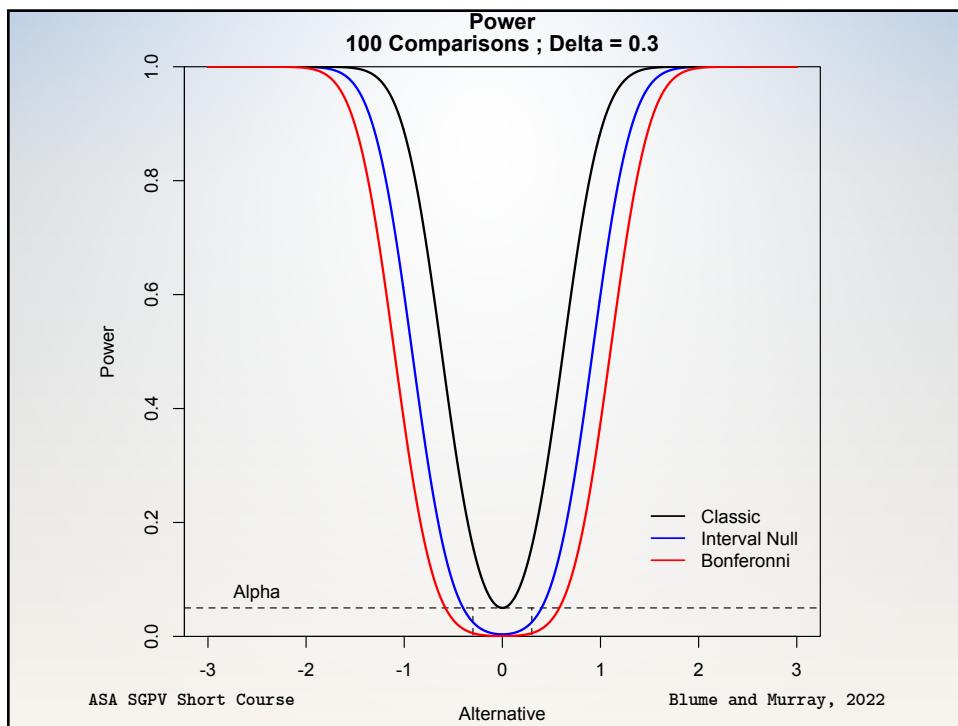
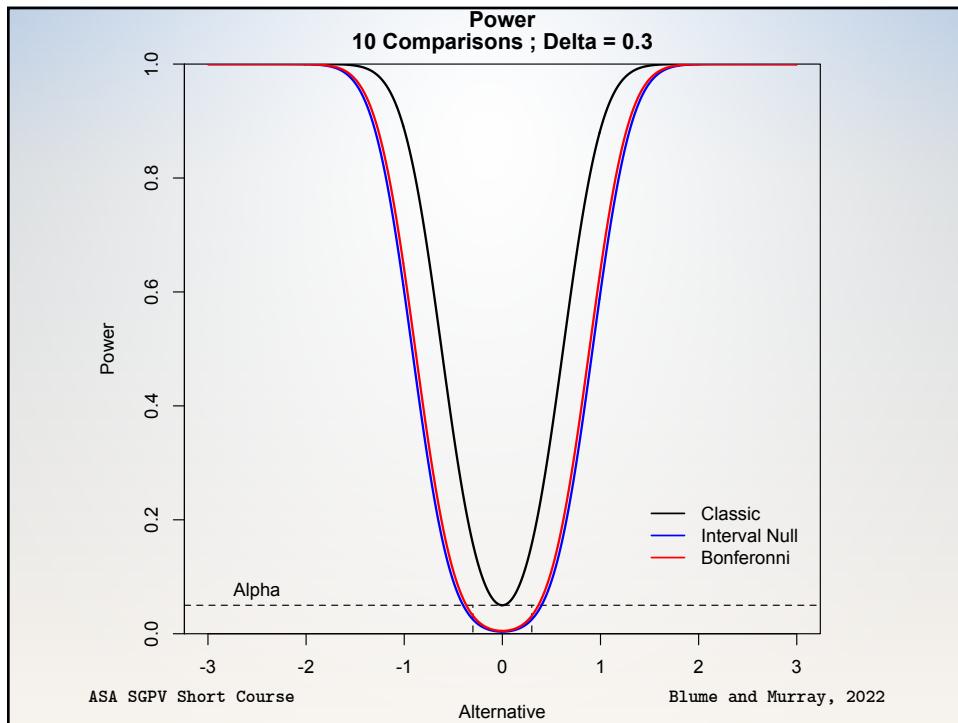
ASA SGPV Short Course

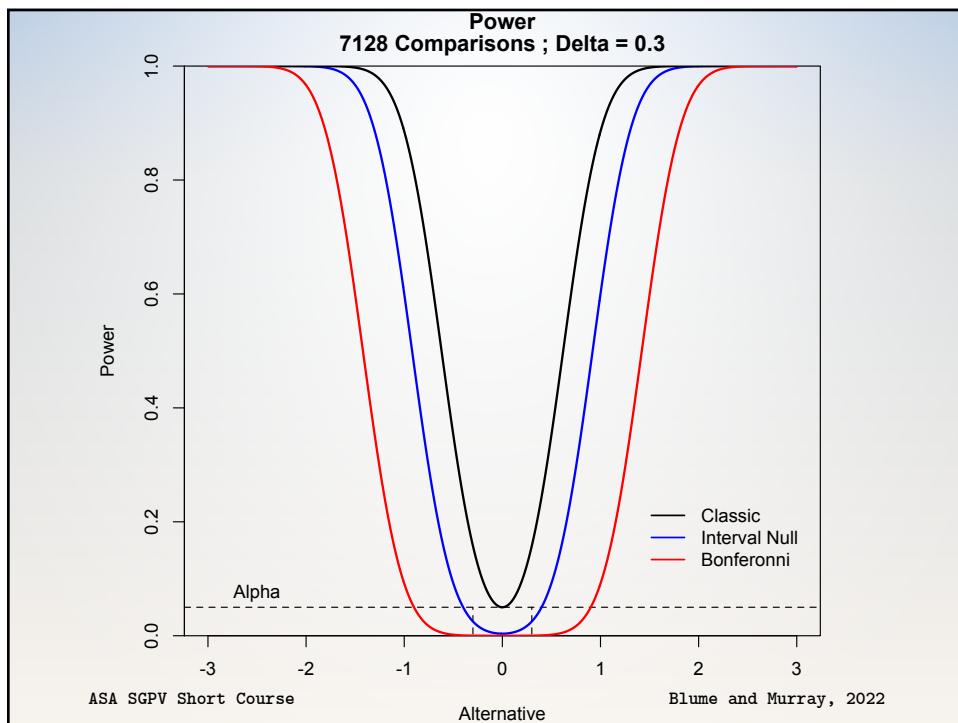
Blume and Murray, 2022



Compare with standard methods

- Second-generation p -value vs. Bonferroni correction
 - Adjusted for $\{10, 100, 7128\}$ comparisons
 - Leukemia data example
- Remember SGPV are not adjusted for comparisons
- Discuss?





Compatible with Null

- How often are the data compatible with null?
- Examine $P(p_\delta = 1|\theta)$ when θ is null or practically null
 - Essentially opposite of power function
- Sample size must be large enough to allow the null interval to contain the interval estimate
- This probability
 - converges to zero or one quickly for alternatives not near the edge of interval null

‘Null Power’ Function

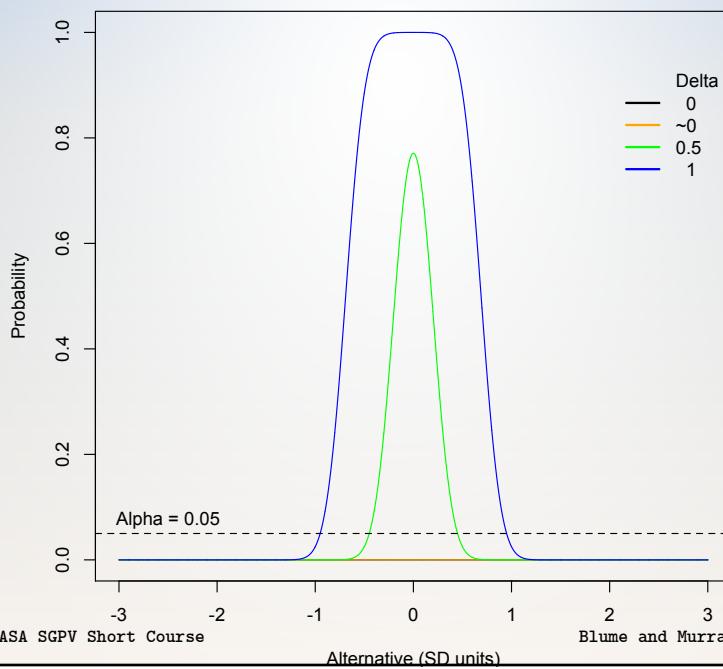
- How often are the data compatible with null?
- Sample size must be large enough to allow null interval to contain the interval estimate, so $(\delta > Z_{\alpha/2}/\sqrt{n})$ or $(\sqrt{n} > Z_{\alpha/2}/\delta)$
- This probability converges to 0 or 1 quickly

$$P(p_\delta = 1|\theta) = \Phi \left[\frac{\sqrt{n}(\theta_0 + \delta)}{\sigma} - \frac{\sqrt{n}\theta}{\sigma} - Z_{\alpha/2} \right] - \Phi \left[\frac{\sqrt{n}(\theta_0 - \delta)}{\sigma} - \frac{\sqrt{n}\theta}{\sigma} + Z_{\alpha/2} \right]$$

when $\delta > Z_{\alpha/2}/\sqrt{n}$

ASA SGPV Short Course

Blume and Murray, 2022

 $P(p_\delta = 1|\theta)$ 

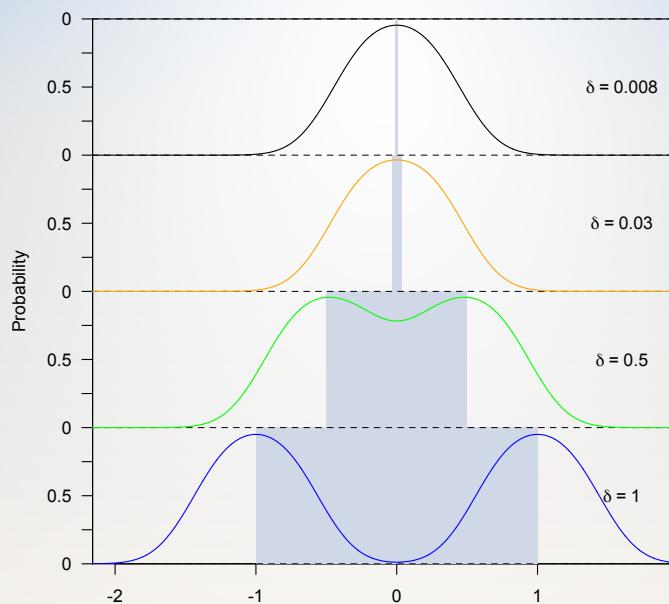
Probability of Inconclusive Data

- How often are the data inconclusive?
- Examine $P(0 < p_\delta < 1|\theta)$ for various θ
- This probability
 - drives sample size projections
 - is maximized when H is near the interval null edge
 - decreases quickly as H moves away from edge of null
- $P(0 < p_\delta < 1|\theta) = 1 - P(p_\delta = 0|\theta) - P(p_\delta = 1|\theta)$

ASA SGPV Short Course

Blume and Murray, 2022

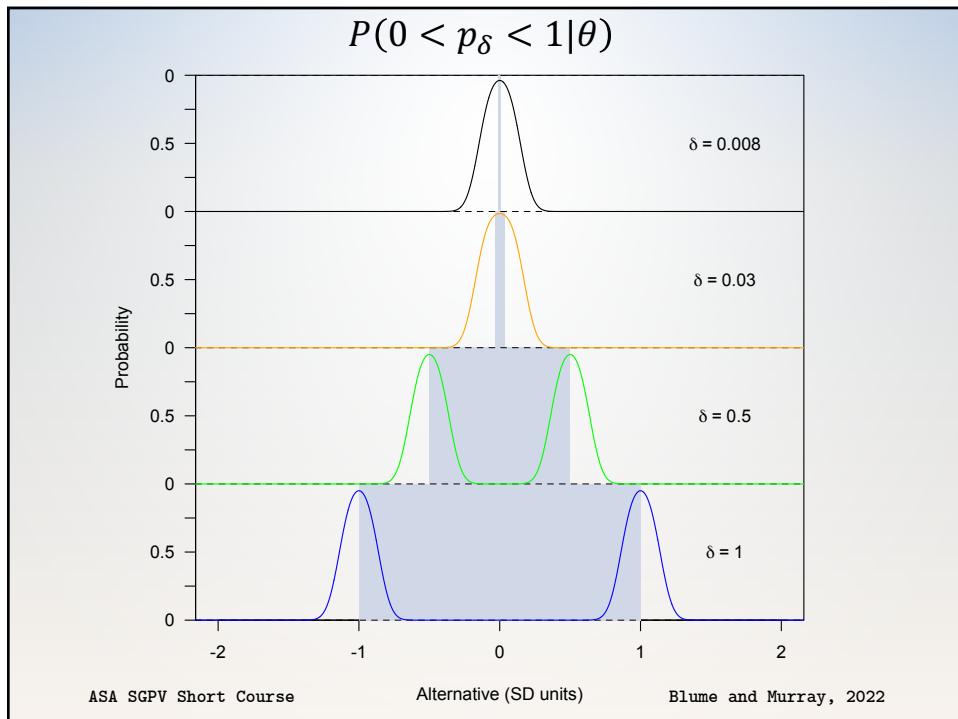
$$P(0 < p_\delta < 1|\theta)$$



ASA SGPV Short Course

Alternative (SD units)

Blume and Murray, 2022



Statistical Properties

Suppose interval I has coverage probability $1-\alpha$, then

Three ‘Error’ Rates

1. $P(p_\delta = 0 | H_0) \leq \alpha$ and $\rightarrow 0$ as $n \rightarrow \infty$
2. $P(p_\delta = 1 | H_1) \leq \alpha$ and $\rightarrow 0$ as $n \rightarrow \infty$
3. $P(0 < p_\delta < 1 | H)$ controlled through sample size

Two False Discovery Rates (next section!)

1. $P(H_0 | p_\delta = 0)$
2. $P(H_1 | p_\delta = 1)$

Remarks

- Second-generation p -values...
 - Has three ‘Error’ rates
 - Allows Type I and II rate to converge to zero
 - Control changes of inconclusive results
 - Controls error rate using *science*
 - Reduces the false discovery rate (next section)
- Anchoring the scale of the effect size...
 - Eliminates most Type I Errors
 - Improves scientific translation of statistical model

ASA SGPV Short Course

Blume and Murray, 2022

Time for Code Part 2a!

ASA SGPV Short Course

Blume and Murray, 2022

10 Minute Break!

ASA SGPV Short Course

Blume and Murray, 2022

Equivalence Tests

- Establish bioequivalence between data and an established equivalence range or interval null
- Example: A pharmaceutical company tests for drug approval by comparing new drug's performance to an approved drug's performance
- Uses an interval null or equivalence range
 - $H_0 = [\theta^-, \theta^+]$

ASA SGPV Short Course

Blume and Murray, 2022

TOST Test

- Most popular frequentist test is the Two One-Sided t-tests (TOST) (Schuirmann 1987)
 - Flips the null and alternative (be careful)
 - Uses the $(1-2\alpha)\%$ confidence interval (be careful)
- Tests are ordinary, one-sided, α -level t-tests
- If *both* one-sided tests reject then conclude the evidence is contained in the equivalence range

ASA SGPV Short Course

Blume and Murray, 2022

TOST Test

- **Fail to reject the (original) null:** The confidence interval is outside of the indifference zone

$$H_0: (\theta < \theta^- \text{ or } \theta > \theta^+)$$
- **Reject the (original) null:** The confidence interval is contained within the indifference zone

$$H_1: (\theta \geq \theta^- \text{ and } \theta \leq \theta^+)$$
- Reported p -value is the p -value of largest magnitude from the two one-sided tests

$$p_T = \max\{p_{T_1}, p_{T_2}\}$$

ASA SGPV Short Course

Blume and Murray, 2022

SGPV Definition

**Second-generation
p-value (SGPV)**

$$\rightarrow p_\delta = \frac{|I \cap H_0|}{|I|} \times \max \left\{ \frac{|I|}{2|H_0|}, 1 \right\}$$

Proportion of data-supported hypotheses that are also null hypotheses

**Small-sample
correction factor**

shrinks proportion to $\frac{1}{2}$ when $|I|$ wide

when $|I| > 2|H_0|$

ASA SGPV Short Course

Blume and Murray, 2022

TOST vs. SGPV comparison

		SGPV Outcomes		
		Consistent with the alternative (SGPV near 0)	Inconclusive (SGPV near $\frac{1}{2}$)	Consistent with the null (SGPV near 1)
Equivalence Tests Outcomes	Consistent with the alternative (p-value is unable to indicate this)	Not applicable A	Not applicable B	Not applicable C
	Inconclusive (p-value is non-significant)	Can occur D	Can occur E	Never occurs F
	Consistent with the null (p-value is significant)	Never occurs H	Can occur in small samples I	Can occur J

ASA SGPV Short Course

Blume and Murray, 2022

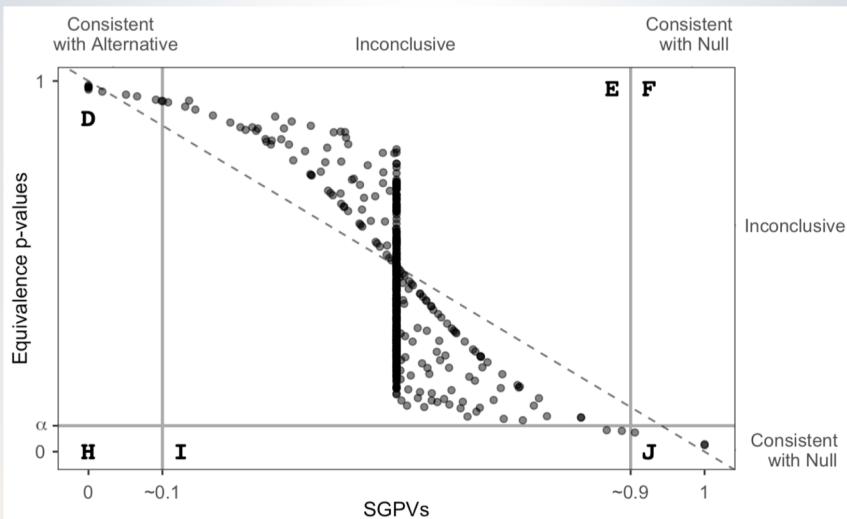
TOST v SGPV Simulation

- Simulate TOST and SGPV reported p -values
 - Data generated under the null, $N(0,1)$
 - Sample size of $n=6$
 - Yields 70% power for $\Delta = 1$ with 5% type 1 error
 - Indifference zone is $[\theta^-, \theta^+] = [-0.375, 0.375]$
 - Uncertainty interval is 95% confidence interval
 - 500 iterations (for illustration)
 - TOSTER R package ([Link](#))

ASA SGPV Short Course

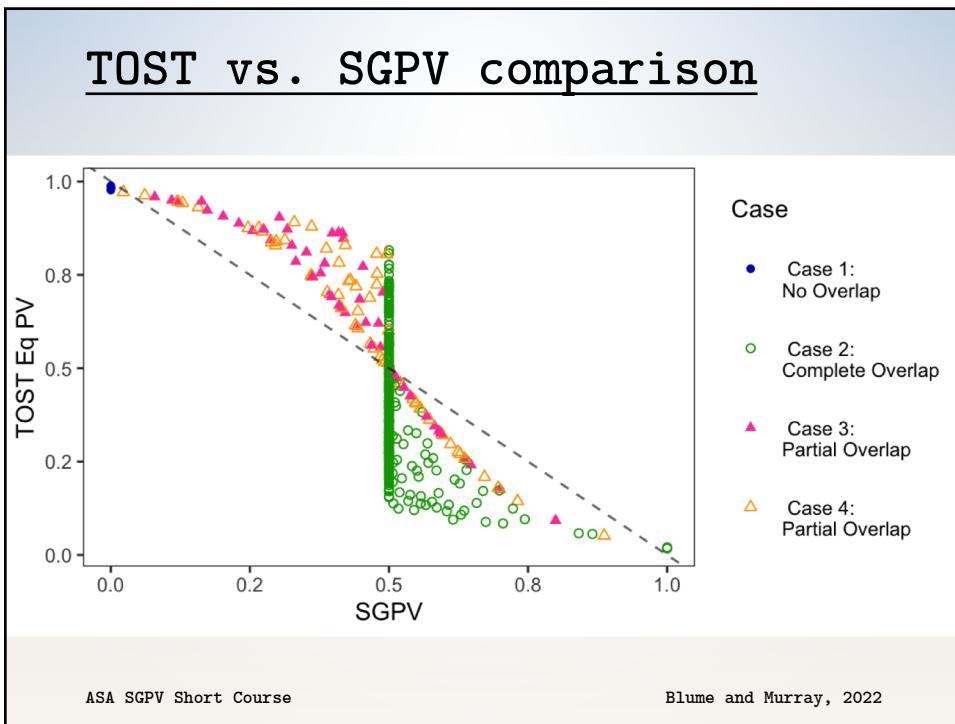
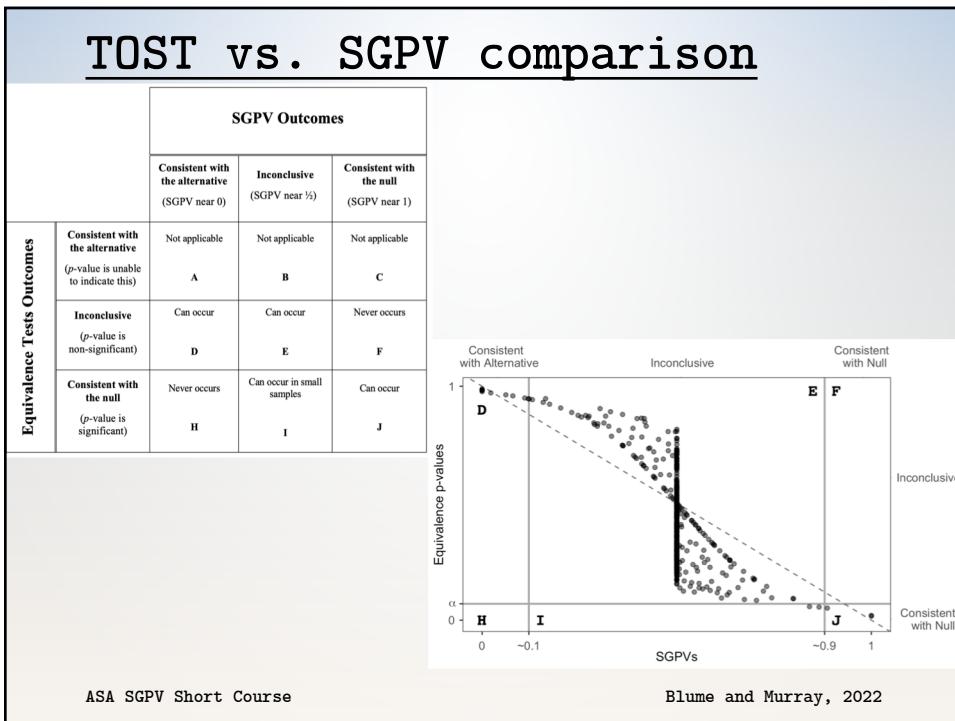
Blume and Murray, 2022

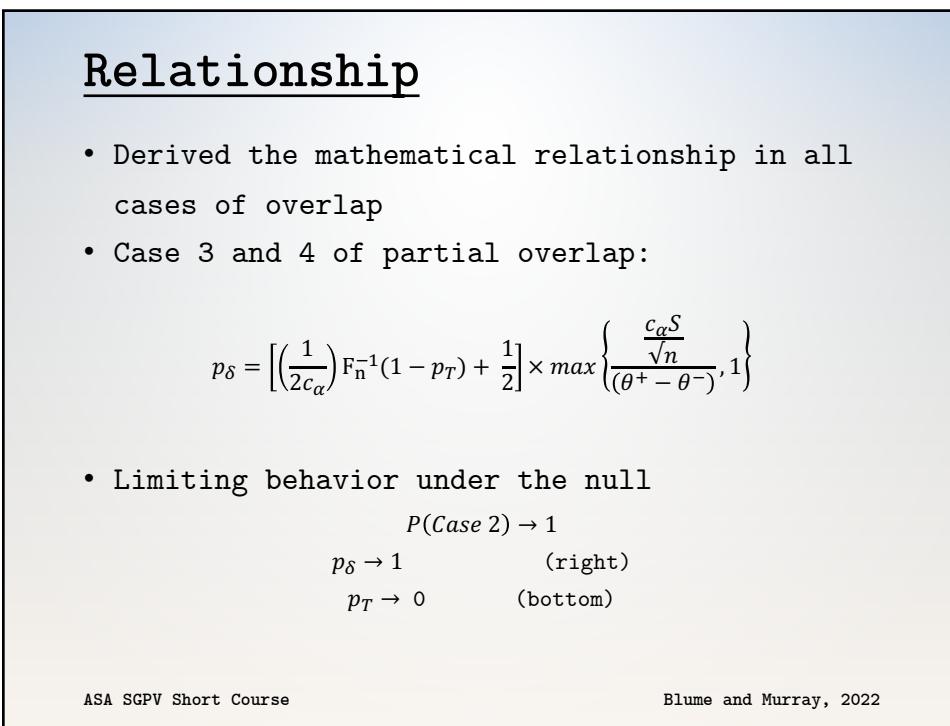
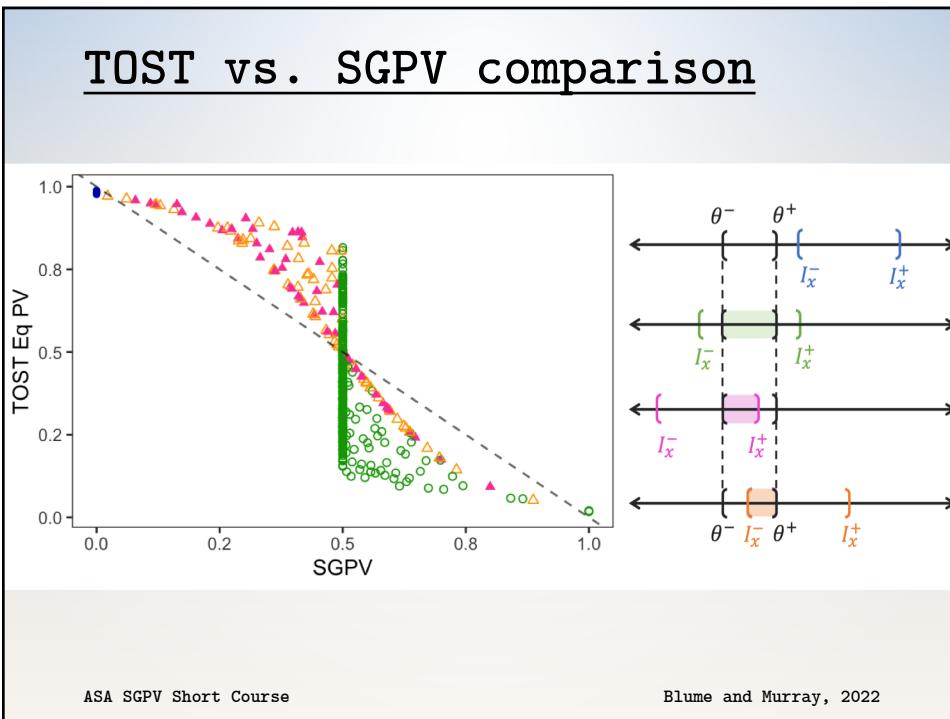
TOST vs. SGPV comparison



ASA SGPV Short Course

Blume and Murray, 2022





TOST vs. SGPV comparison

TOST	SGPV
• 2 inference outcomes	• 3 inference outcomes
• Conclusions only about $(1 - 2\alpha)\%$ confidence interval	• Any uncertainty data interval can be used
• Type I Error is ultra-conservative (distribution of p_T is non-uniform)	• Type I error is accurately assessed (limited by width of data interval)
• Not uniformly most powerful	• Indicates when data agree with null or alternative without additional testing
• No measure of overlap included in computation	• Includes overlap in reported p -value

ASA SGPV Short Course

Blume and Murray, 2022

Comparison

- TOST and SGPV are not one-to-one unless the variance is known
- TOST has significant limitations
- *SGPV is more flexible and easier to interpret*

ASA SGPV Short Course

Blume and Murray, 2022

Time for Code Part 2b!

ASA SGPV Short Course

Blume and Murray, 2022

10 Minute Break!

ASA SGPV Short Course

Blume and Murray, 2022

Statistical Properties

Suppose interval I has coverage probability $1-\alpha$, then

Three ‘Error’ Rates

1. $P(p_\delta = 0 | H_0) \leq \alpha$ and $\rightarrow 0$ as $n \rightarrow \infty$
2. $P(p_\delta = 1 | H_1) \leq \alpha$ and $\rightarrow 0$ as $n \rightarrow \infty$
3. $P(0 < p_\delta < 1 | H)$ controlled through sample size

Two False Discovery Rates

1. $P(H_0 | p_\delta = 0)$
2. $P(H_1 | p_\delta = 1)$

ASA SGPV Short Course

Blume and Murray, 2022

False Discovery Rates

- FDR for 5 SGPV=0 findings; computed under various null and alternative configurations (w/ flat prior).

SNP ID	SGPV rank	p-value rank	OR	1/8 SI lower limit	1/8 SI upper limit	FDR ₁	FDR ₂	FDR ₃
kgp4568244_C	1	133	0.10	0.03	0.37	2.9%	17.1%	3.3%
kgp8051290_G	13	2002	15.58	1.95	124.68	4.3%	30.3%	4.9%
kgp4497498_A	28	255	4.37	1.80	10.64	2.5%	8.6%	3.1%
rs3123636_G	423	1	1.39	1.26	1.55	0.01%	0.1%	0.4%
kgp7460928_G	1443	3310	1.78	1.11	2.87	2.4%	2.0%	3.0%

ASA SGPV Short Course

Blume and Murray, 2022

False discovery rates

- Impact of $\alpha=0.05$ vs $\alpha=0.05/7128$ (7128 comparisons)

- False Discovery Rate (**FDR**)

$$P(H_0|p < \alpha) = \left[1 + \frac{(1 - \beta)}{\alpha} r \right]^{-1}$$

- False Confirmation Rate (**FCR**)

$$P(H_1|p > \alpha) = \left[1 + \frac{(1 - \alpha)}{\beta} \frac{1}{r} \right]^{-1}$$

$$r = P(H_1)/P(H_0)$$

Error rates

Error rates

ASA SGPV Short Course

Blume and Murray, 2022

False discovery rates

- Second-generation p -values

- False Discovery Rate (**FDR**)

$$P(H_0|p_\delta = 0) = \left[1 + \frac{P(p_\delta = 0|H_1)}{P(p_\delta = 0|H_0)} r \right]^{-1}$$

- False Confirmation Rate (**FCR**)

$$P(H_1|p_\delta = 1) = \left[1 + \frac{P(p_\delta = 1|H_0)}{P(p_\delta = 1|H_1)} \frac{1}{r} \right]^{-1}$$

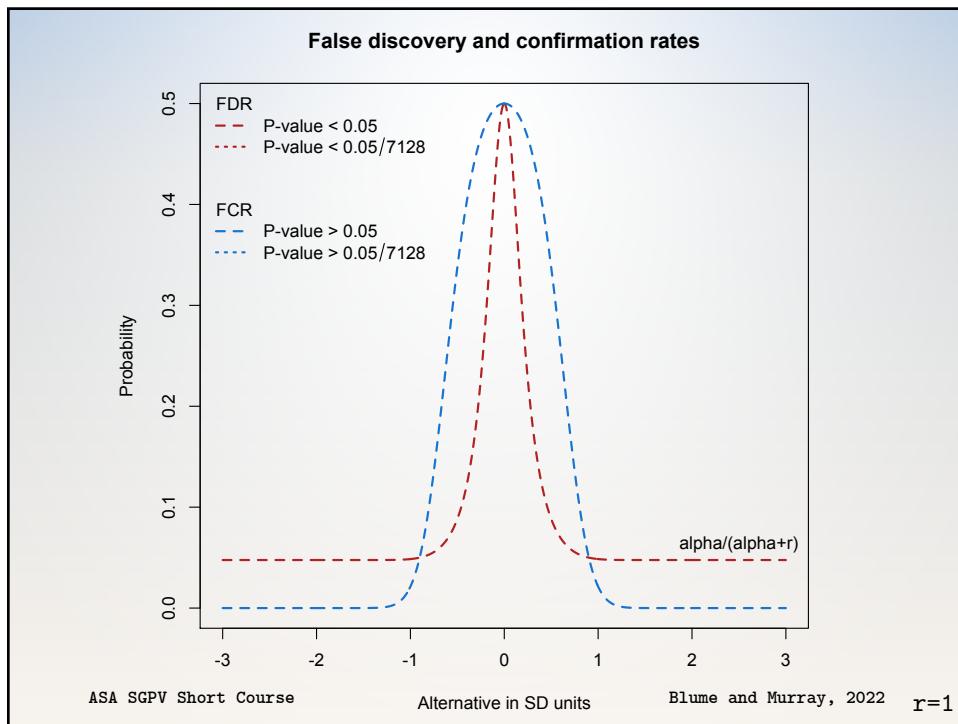
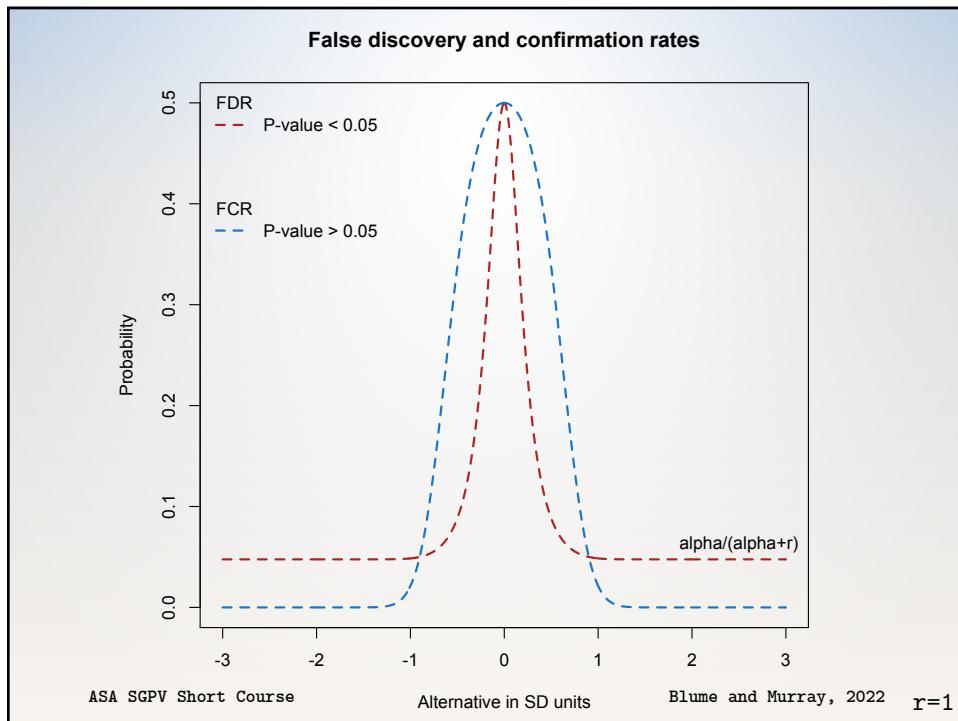
$$r = P(H_1)/P(H_0)$$

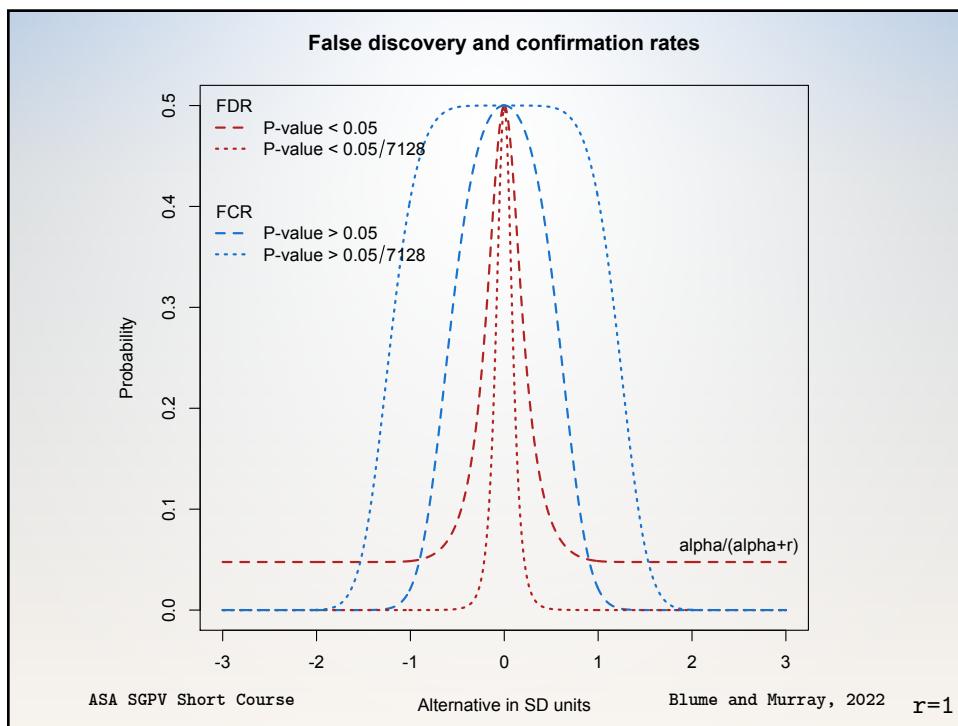
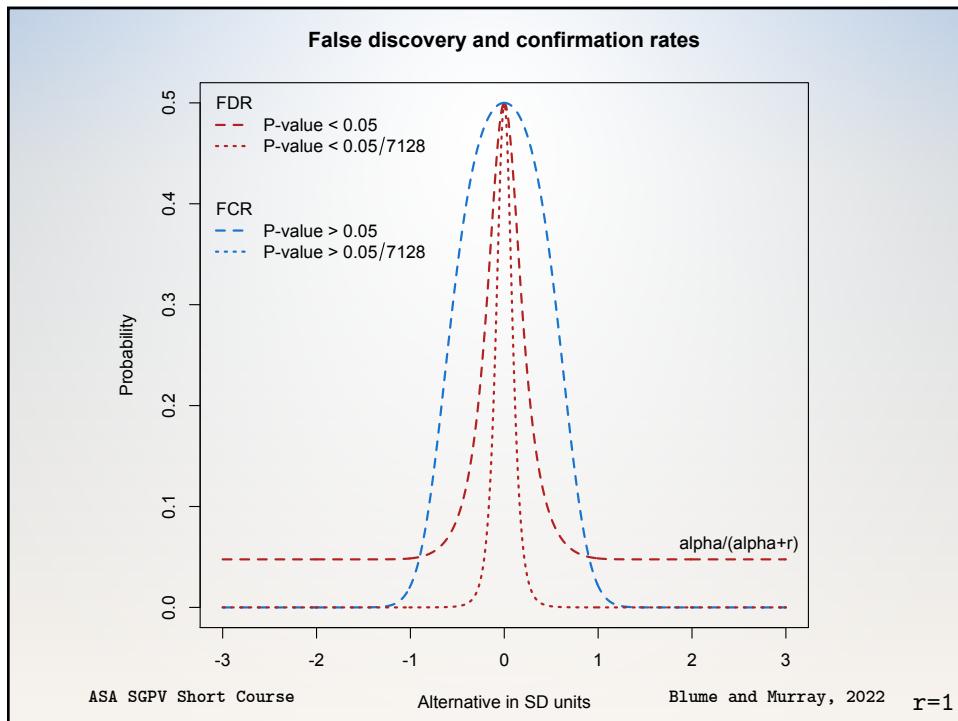
Error Rates

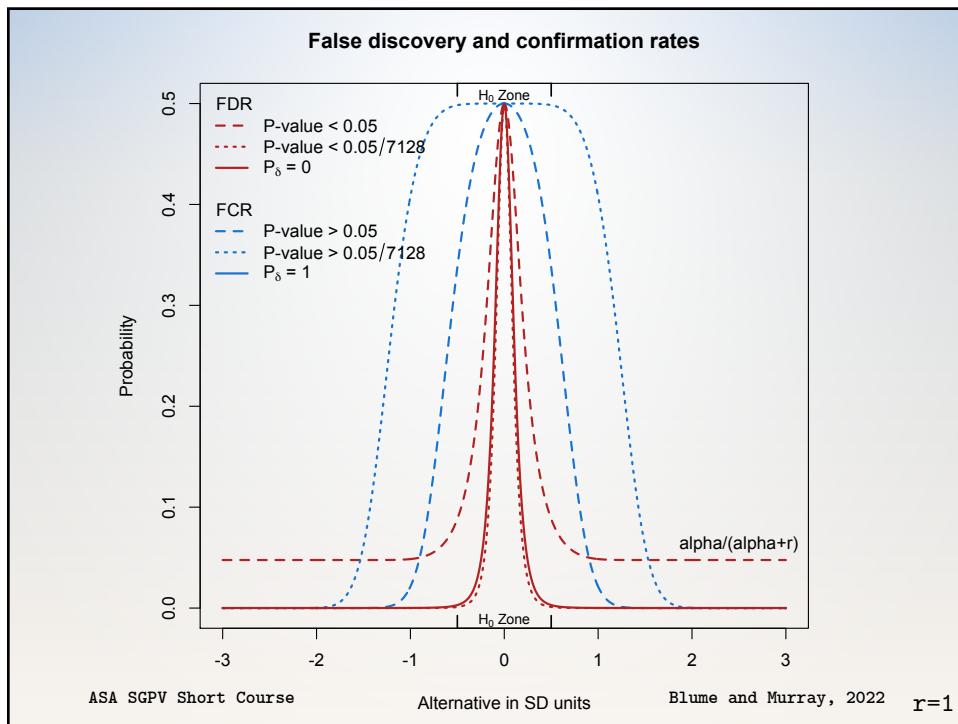
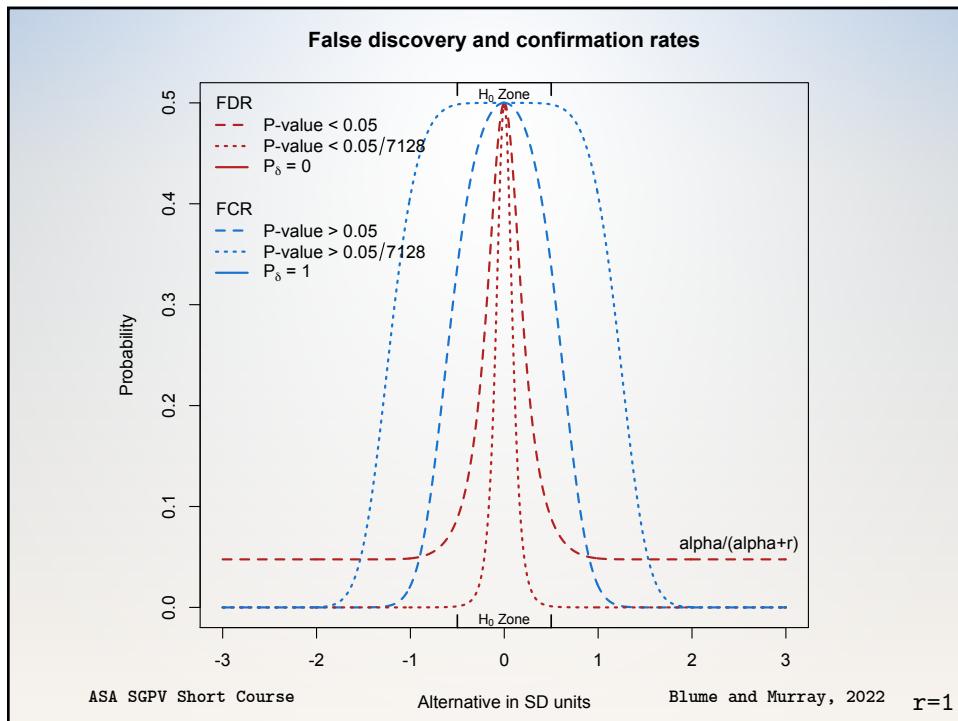
Error Rates

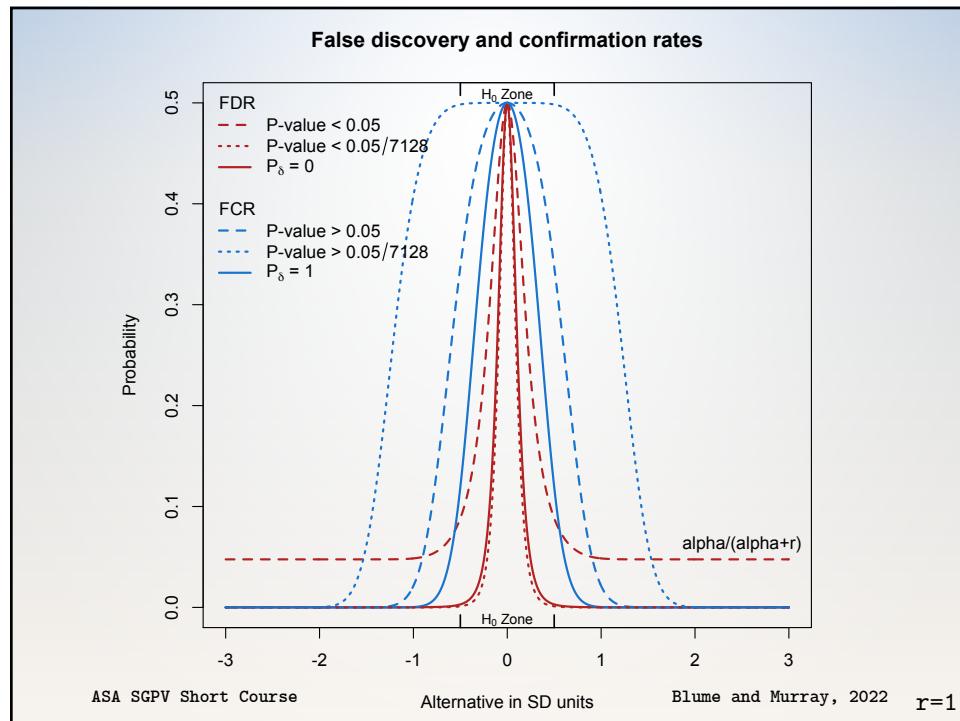
ASA SGPV Short Course

Blume and Murray, 2022









FDR R Packages

- SGPVs
 - Valerie Welty
 - `sgpv::fdrisk()`
 - This function computes the false discovery risk (sometimes called the "empirical bayes FDR") for a second-generation p -value of 0, or the false confirmation risk for a second-generation p -value of 1.
- Raw p-values
 - `FDRestimation::p.fdr()`
 - This function computes FDRs and Method Adjusted p-values.
 - Methods include: Benjamini-Hochberg, Benjamini-Yekutieli, Bonferroni, Holm, Hochberg, and Sidak.

Time for Code Part 2c!

ASA SGPV Short Course

Blume and Murray, 2022

10 Minute Break!

ASA SGPV Short Course

Blume and Murray, 2022

Study Planning for SGPs

- Will edit paper and submit by end of 2022
- Purpose: Evaluate how different techniques of setting the indifference zone effect the SGpv study properties.
- What options does a collaborator have when they are uncertain of the indifference zone?

ASA SGpv Short Course

Blume and Murray, 2022

Traditional p-value vs. SGpv

		Null is true		Alternative is true	
		Traditional p-value		Traditional p-value	
Inference	Data are consistent with the alternative	Type I error $P_{\theta_0}(p \leq 0.05)$		Power or (1 - Type II error) $P_{\theta_1}(p \leq 0.05)$	
	Data are inconclusive	1 - Type I error $P_{\theta_0}(p > 0.05)$		Type II error $P_{\theta_1}(p > 0.05)$	
	Data are consistent with the null	NA		NA	

ASA SGpv Short Course

Blume and Murray, 2022

Traditional p-value vs. SGPV

		Null is true		Alternative is true	
		SGPV		SGPV	
Inference	Data are consistent with the alternative		β_0 $P_{\theta_0}(p_\delta = 0)$		β_1 $P_{\theta_1}(p_\delta = 0)$
	Data are inconclusive		γ_0 $P_{\theta_0}(0 < p_\delta < 1)$		γ_1 $P_{\theta_1}(0 < p_\delta < 1)$
	Data are consistent with the null		ω_0 $P_{\theta_0}(p_\delta = 1)$		ω_1 $P_{\theta_1}(p_\delta = 1)$

ASA SGPV Short Course

Blume and Murray, 2022

Traditional p-value vs. SGPV

		Null is true		Alternative is true	
		Traditional p-value	SGPV	Traditional p-value	SGPV
Inference	Data are consistent with the alternative	Type I error $P_{\theta_0}(p \leq 0.05)$	β_0 $P_{\theta_0}(p_\delta = 0)$	Power or (1 - Type II error) $P_{\theta_1}(p \leq 0.05)$	β_1 $P_{\theta_1}(p_\delta = 0)$
	Data are inconclusive	1 - Type I error $P_{\theta_0}(p > 0.05)$	γ_0 $P_{\theta_0}(0 < p_\delta < 1)$	Type II error $P_{\theta_1}(p > 0.05)$	γ_1 $P_{\theta_1}(0 < p_\delta < 1)$
	Data are consistent with the null	NA	ω_0 $P_{\theta_0}(p_\delta = 1)$	NA	ω_1 $P_{\theta_1}(p_\delta = 1)$

ASA SGPV Short Course

Blume and Murray, 2022

Traditional p-value vs. SGPV

		Null is true		Alternative is true	
		Traditional p-value	SGPV	Traditional p-value	SGPV
Inference	Data are consistent with the alternative	Type I error $P_{\theta_0}(p \leq 0.05)$	β_0 $P_{\theta_0}(p_\delta = 0)$	Power or (1 - Type II error) $P_{\theta_1}(p \leq 0.05)$	β_1 $P_{\theta_1}(p_\delta = 0)$
	Data are inconclusive	1 - Type I error $P_{\theta_0}(p > 0.05)$	γ_0 $P_{\theta_0}(0 < p_\delta < 1)$	Type II error $P_{\theta_1}(p > 0.05)$	γ_1 $P_{\theta_1}(0 < p_\delta < 1)$
	Data are consistent with the null	NA	ω_0 $P_{\theta_0}(p_\delta = 1)$	NA	ω_1 $P_{\theta_1}(p_\delta = 1)$

ASA SGPV Short Course

Blume and Murray, 2022

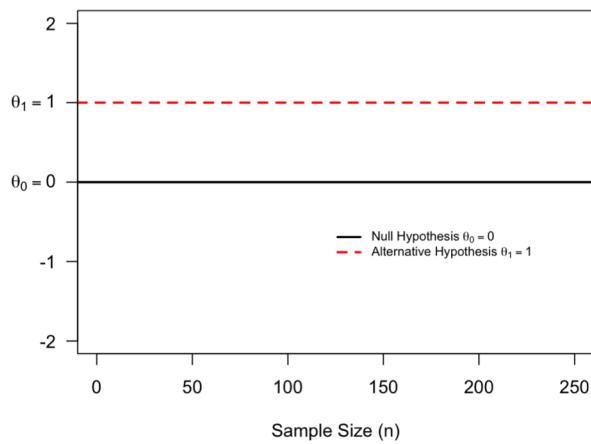
Traditional p-value vs. SGPV

		Null is true		Alternative is true	
		Traditional p-value	SGPV	Traditional p-value	SGPV
Inference	Data are consistent with the alternative	Type I error $P_{\theta_0}(p \leq 0.05)$	β_0 $P_{\theta_0}(p_\delta = 0)$	Power or (1 - Type II error) $P_{\theta_1}(p \leq 0.05)$	β_1 $P_{\theta_1}(p_\delta = 0)$
	Data are inconclusive	1 - Type I error $P_{\theta_0}(p > 0.05)$	γ_0 $P_{\theta_0}(0 < p_\delta < 1)$	Type II error $P_{\theta_1}(p > 0.05)$	γ_1 $P_{\theta_1}(0 < p_\delta < 1)$
	Data are consistent with the null	NA	ω_0 $P_{\theta_0}(p_\delta = 1)$	NA	ω_1 $P_{\theta_1}(p_\delta = 1)$

ASA SGPV Short Course

Blume and Murray, 2022

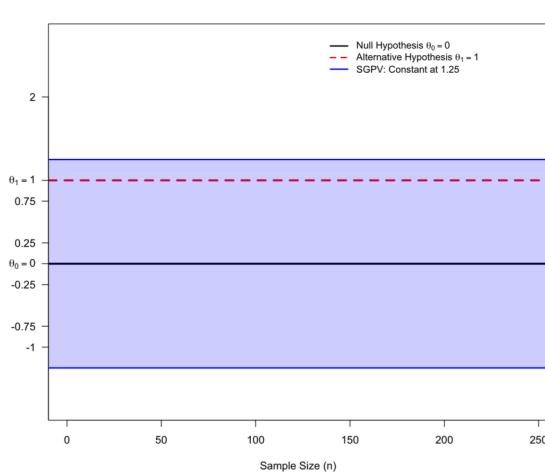
Standard Assumptions



ASA SGPV Short Course

Blume and Murray, 2022

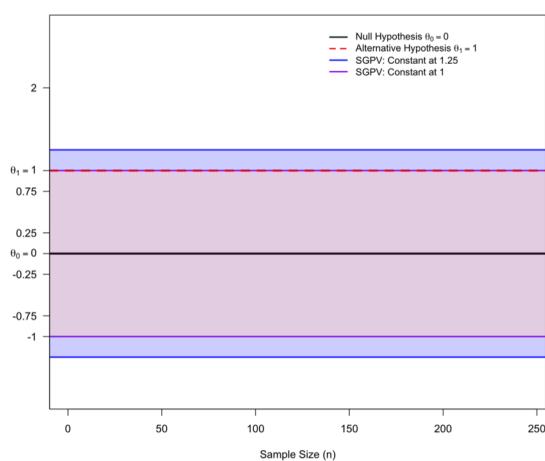
Fixed Intervals



ASA SGPV Short Course

Blume and Murray, 2022

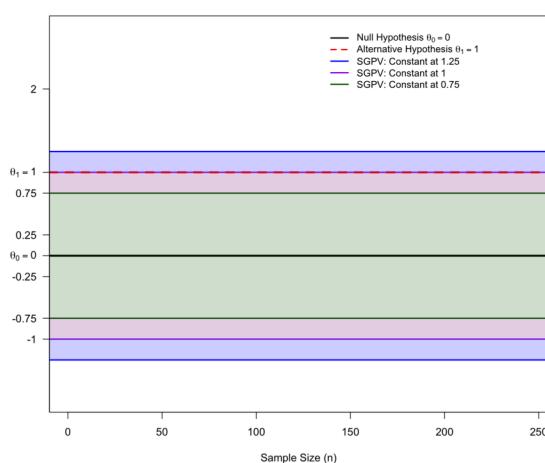
Fixed Intervals



ASA SGPV Short Course

Blume and Murray, 2022

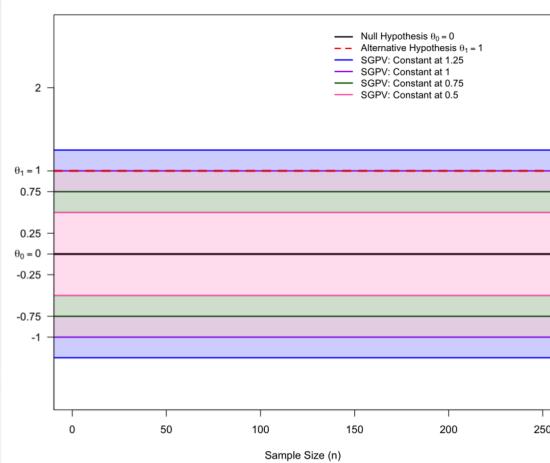
Fixed Intervals



ASA SGPV Short Course

Blume and Murray, 2022

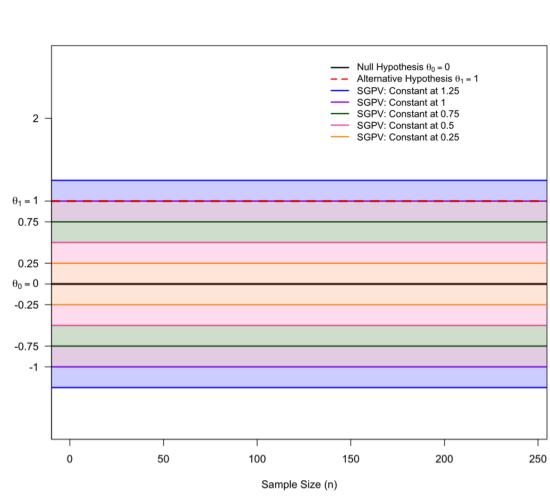
Fixed Intervals



ASA SGPV Short Course

Blume and Murray, 2022

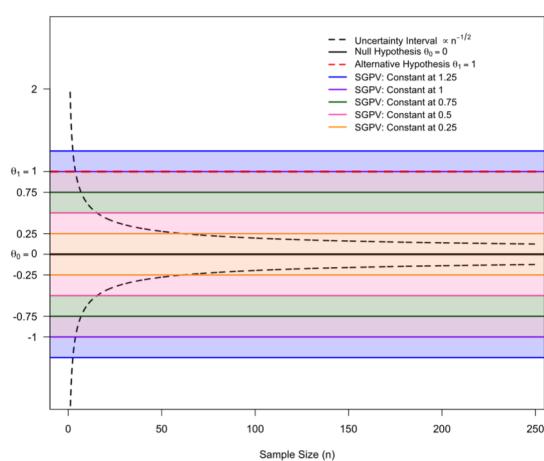
Fixed Intervals



ASA SGPV Short Course

Blume and Murray, 2022

Fixed Intervals



ASA SGPV Short Course

Blume and Murray, 2022

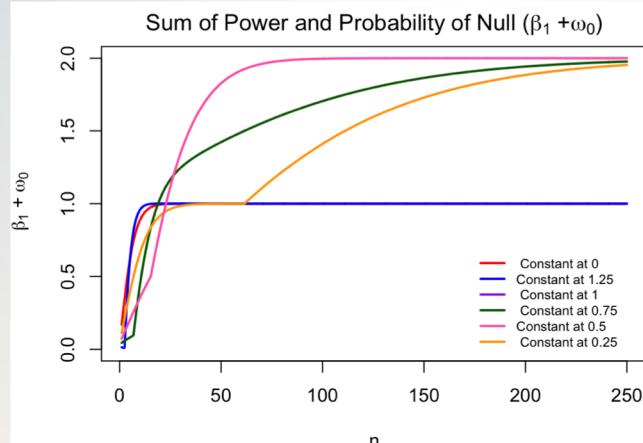
Traditional p-value vs. SGPV

		Null is true		Alternative is true	
		Traditional <i>p</i> -value	SGPV	Traditional <i>p</i> -value	SGPV
Inference	Data are consistent with the alternative	Type I error $P_{\theta_0}(p \leq 0.05)$	β_0 $P_{\theta_0}(p_\delta = 0)$	Power or (1 - Type II error) $P_{\theta_1}(p \leq 0.05)$	β_1 $P_{\theta_1}(p_\delta = 0)$
	Data are inconclusive	1 - Type I error $P_{\theta_0}(p > 0.05)$	γ_0 $P_{\theta_0}(0 < p_\delta < 1)$	Type II error $P_{\theta_1}(p > 0.05)$	γ_1 $P_{\theta_1}(0 < p_\delta < 1)$
	Data are consistent with the null	NA	ω_0 $P_{\theta_0}(p_\delta = 1)$	NA	ω_1 $P_{\theta_1}(p_\delta = 1)$

ASA SGPV Short Course

Blume and Murray, 2022

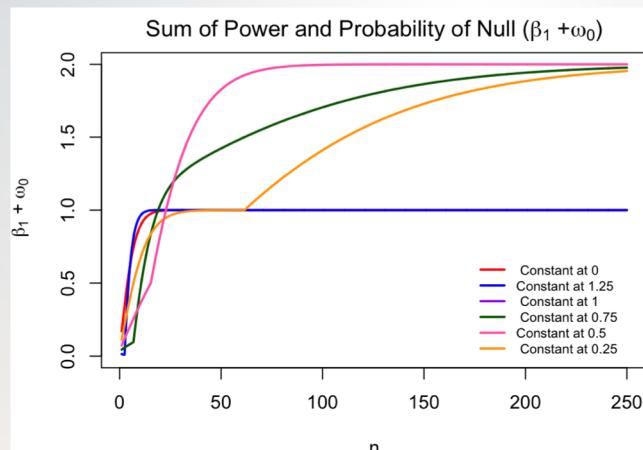
Fixed Intervals



ASA SGPV Short Course

Blume and Murray, 2022

Fixed Intervals



ASA SGPV Short Course

Blume and Murray, 2022

Shrinking Intervals

- Collaborator is uncertain but comfortable with a wider interval to begin

For example: $[-0.9, 0.9]$

- Limit shrinking to half distance between $(\theta^-$ or $\theta^+)$ and θ_0 $[-0.45, 0.45]$
- We can let the indifference zone shrink over sample size at different rates:

$$\delta_1(n) \propto \frac{\sqrt{V}}{n^{1/4}}$$

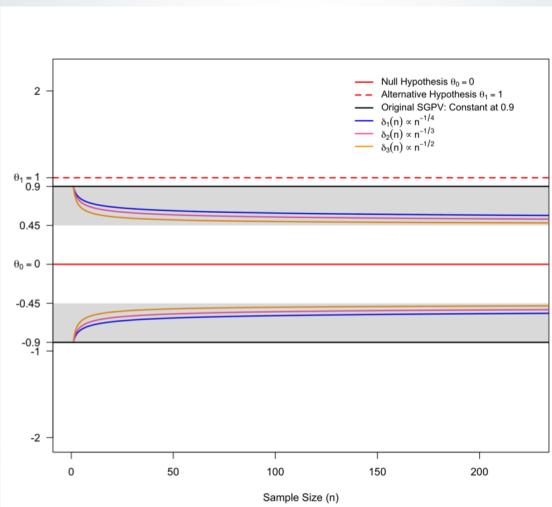
$$\delta_2(n) \propto \frac{\sqrt{V}}{n^{1/3}}$$

$$\delta_3(n) \propto \frac{\sqrt{V}}{n^{1/2}}$$

ASA SGPV Short Course

Blume and Murray, 2022

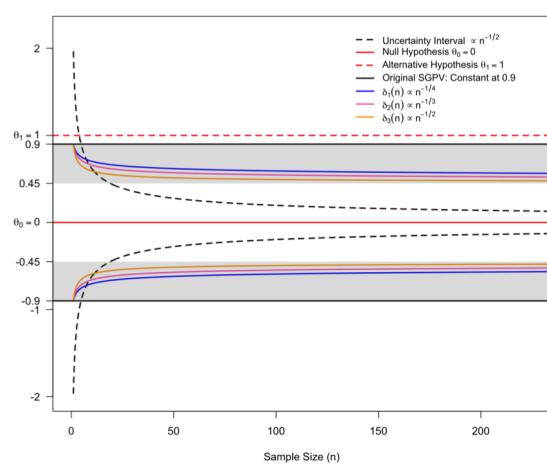
Shrinking Intervals



ASA SGPV Short Course

Blume and Murray, 2022

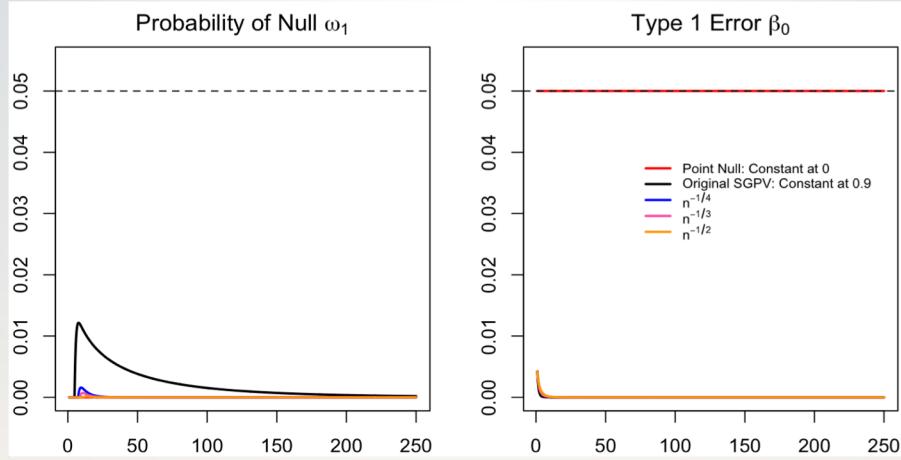
Shrinking Intervals



ASA SGPV Short Course

Blume and Murray, 2022

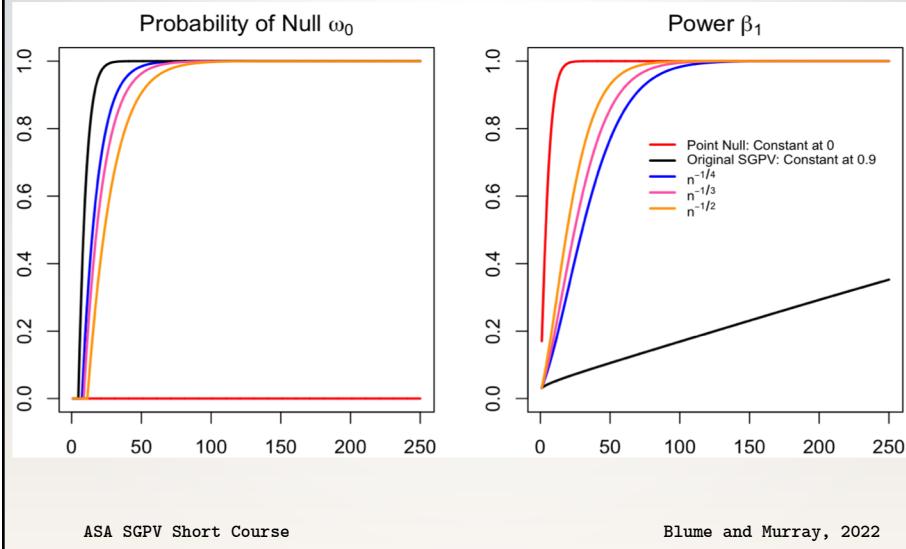
Incorrect probabilities (Error)



ASA SGPV Short Course

Blume and Murray, 2022

Correct probabilities (Power)



ASA SGPV Short Course

Blume and Murray, 2022

Recommendations for Collaborator Uncertainty

Collaborator Hypothesis	Suggested SGPV Analysis	Outcome
Confident in a null zone.	Use it!	Ideal
Confident in the alternative point closest to null.	Use the halfway point as the null zone.	Great
Uncertain of null zone.	Use the null zone that shrinks.	Good
Uncertain of alternative point.	Use the null zone that shrinks.	Ok
Cannot identify.	Do NOT use SGPV! (but can't ever measure evidence for null)	Poor

ASA SGPV Short Course

Blume and Murray, 2022

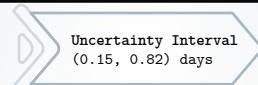
COVID Clinical Trial Example

- Randomized 1,591 patients to ivermectin treatment or placebo
- Mean time spent unwell was estimated using a longitudinal ordinal regression model; range was 0 to 14 days
- Results: “The difference in the amount of time spent feeling unwell with COVID was estimated to be 0.49 days in favor of ivermectin with a 95% credible interval of (0.15, 0.82).”

ASA SGPV Short Course

Blume and Murray, 2022

COVID Clinical Trial Example



Collaborator Hypothesis	SGPV results	Outcome
Confident in a null zone, $[\theta^-, \theta^+] = [-1, 1]$.	$p_\delta = 1$	Data are consistent with the null.
Confident in an alternative point, $\theta_1 = 1$, use half standard deviation $[-0.5, 0.5]$.	$p_\delta = 0.522$	Inconclusive
Uncertain of wide small sample null zone, $[\theta^-, \theta^+] = [-1.5, 1.5]$, shrunk to $[-0.753, 0.753]$.	$p_\delta = 0.991$	Data are almost fully consistent with the null.
Uncertain of small sample alternative point, $\theta_1 = 0.75$, shrunk to $[-0.377, 0.377]$.	$p_\delta = 0.384$	Inconclusive
Cannot identify.	NA	NA

SGPV Study Planning Conclusions

- Goal is to balance statistical properties
- Results and inference will vary depending on the chosen indifference zone
- Promotes conversation between collaborator and statistician
 - Point null
 - Indifference zone or closest alternative point of interest
 - Sample size
 - If shrinking is necessary

ASA SGPV Short Course

Blume and Murray, 2022

Time for Code Part 2d!

ASA SGPV Short Course

Blume and Murray, 2022

10 Minute Break!

ASA SGPV Short Course

Blume and Murray, 2022