

Part I

# Second-generation *p*-values: Introduction and Applications

---

Jeffrey D. Blume, and Megan H. Murray,

PhD

PhD

School of Data Science  
University of Virginia

Statistician at  
Eli Lilly and Company

## About Us

- Jeffrey
  - Associate Dean and Professor in Data Science (UVA)
  - Areas of research include statistical inference, likelihood methods, second-generation *p*-values, prediction modeling, ROC curves, mediation modeling, missing data in prediction problems, and false discovery rates.
  - Website: [www.statisticalevidence.com](http://www.statisticalevidence.com)
  - UVA Profile: [www.datascience.virginia.edu/people/jeffrey-blume](http://www.datascience.virginia.edu/people/jeffrey-blume)
- Megan
  - Project Statistician at Eli Lilly since September 2022
  - PhD in Biostatistics at Vanderbilt: August 2022
  - Dissertation: ‘On second-generation *p*-values for equivalence testing and study planning, and flexible false discovery rate computation for classical *p*-values’

## Course Layout

- Slides Part I:
  - Motivation and definition
  - Examples
  - Live Coding Demonstration
  - Code for all examples available in GitHub Repo
- Slides Part II:
  - Statistical Properties
  - False discovery rates
  - Study Planning
  - Connection with Equivalence tests
  - SGPV Variable Selection (if we have time)
  - Code for all examples available in GitHub Repo

ASA SGPV Short Course

Blume and Murray, 2022

## Resources

- GitHub with Slides and Code
  - [https://github.com/murraymegan/SGPV\\_CSP\\_2023](https://github.com/murraymegan/SGPV_CSP_2023)
- RStudio Desktop
  - [www.rstudio.com/products/rstudio/download](http://www.rstudio.com/products/rstudio/download)
- Interrupt if you have any questions!

ASA SGPV Short Course

Blume and Murray, 2022

## Synopsis

- Classical  $p$ -values are
  - Ubiquitous, Sacrosanct, Imperfect, Misused
  - Misunderstood (Significance vs. Hypothesis testing roles)
  - Openly debated in practice and theory
- Trend toward estimation in reporting of results
  - Report an estimation interval (e.g., confidence interval)
  - Does interval contain only clinically significant values?
- Second-generation  $p$ -values (SGPVs)
  - Embody and formalize this trend
  - Maintain and improve error rate control
  - Define clinically significant before looking at the data

ASA SGPV Short Course

Blume and Murray, 2022

## Outline

- Evidential Metrics
- Second-generation  $p$ -value
- Introductory examples
- High-dimensional examples
- Live Coding using R



ASA SGPV Short Course

Blume and Murray, 2022

## Evidential metrics

Example:  
Diagnostic Test

1. Measure of the strength evidence
  - Axiomatic and intuitive justification
  - Summary statistic, yardstick
2. Propensity to collect data that will yield a misleading #1
  - Error rates
  - Properties of the study design (!)
3. Probability that an observed #1 is misleading
  - False Discovery rate, False Confirmation rate
  - Chance that an observed result is mistaken
  - Properties of the observed data (!)

ASA SGPV Short Course

Blume and Murray, 2022

## Testing

Evidential Metric	What it measures	Hypothesis Testing	Significance Testing
1	strength of the evidence	Absent	Tail-area probability ( <i>p</i> -value)
2	propensity for study to yield misleading evidence	Tail-area probability (error rates)	Absent
3	propensity for observed results to be misleading	misinterpret #2	misinterpret #1

- The tail-area probability is used to measure three distinct metrics

ASA SGPV Short Course

Blume and Murray, 2022

## Second-generation *p*-value

Evidential Metric	What it measures	SGPV
1	Summary measure	$p_\delta$
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	$P(H_0   p_\delta = 0)$ $P(H_1   p_\delta = 1)$

- Statistical properties important for study planning
- False discovery rates <> Reliability of results

ASA SGPV Short Course

Blume and Murray, 2022

## The *p*-value (what it is)

- Number between 0 and 1
- Smaller  $\Rightarrow$  support for an alternative hypothesis
- Larger  $\Rightarrow$  data are inconclusive
- Clinical significance is ignored
- Sample size confounds comparisons
- Interpretation
  - awkward
  - assumes null hypothesis true
  - rooted in inductive reasoning
- Not clear if/when ‘adjustments’ are necessary

ASA SGPV Short Course

Blume and Murray, 2022

*2<sup>nd</sup>-generation*

## The $\hat{p}$ -value (what we want it is)

*Version 2.0*

- ✓ Number between 0 and 1 →
  - near 0 supports alt
  - near 1 supports null
  - near  $\frac{1}{2}$  inconclusive
- ✓ Smaller ⇒ support for an alternative hypothesis
  - Larger ⇒ data are ~~inconclusive~~ support null
  - Clinical significance is ~~ignored~~ incorporated
- ✗ ~~Sample size confounds comparisons~~
  - Interpretation → Fraction of data-supported hypotheses that are null
    - ~~awkward straightforward~~
    - ~~assumes null hypothesis true conditions on observed data~~
    - ~~rooted in inductive reasoning~~ descriptive, summarizes
    - ~~Not clear if/when ‘adjustments’ are necessary~~

*Ideally, never*

ASA SGPV Short Course      Blume and Murray, 2022

## Illustration

The diagram illustrates the relationship between different types of null hypotheses and the resulting confidence intervals.

**Point Null Hypothesis:** A single vertical line labeled  $H_0$  on a horizontal axis. A light blue bracket above it represents a confidence interval centered around a point  $\hat{H}$ . The distance from  $H_0$  to  $\hat{H}$  is labeled "confidence interval".

**Interval Null Hypothesis:** A double-headed arrow labeled  $H_0^-$  and  $H_0^+$  on the horizontal axis. A light blue bracket below it represents a confidence interval labeled  $[CI^-, CI^+]$ . The width of this interval is labeled  $\delta$ . The region between  $H_0^-$  and  $H_0^+$  is labeled "interval null". The overlap between the "interval null" and the "confidence interval" is labeled "overlap".

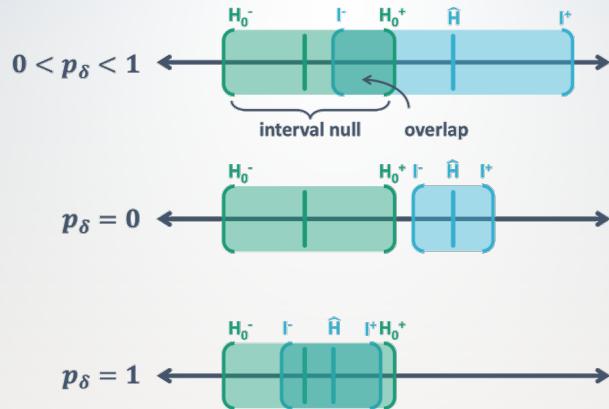
**Labels:**

- Point Null Hypothesis:  $H_0$
- Interval Null Hypothesis:  $H_0^-$ ,  $H_0^+$ ,  $\delta$
- Confidence interval:  $CI^-$ ,  $CI^+$
- Data-supported hypothesis:  $\hat{H}$

Point null hypothesis  $H_0$  and interval null hypothesis  $[H_0^-, H_0^+]$   
 Data-supported hypothesis  $\hat{H}$  and confidence interval  $[CI^-, CI^+]$

ASA SGPV Short Course      Blume and Murray, 2022

## Illustration



Works with confidence, credible, and support intervals

ASA SGPV Short Course

Blume and Murray, 2022

## Second-generation $p$ -value

- SGPV is in  $[0,1]$  and denoted by  $p_\delta$
- $\delta$  for scientific significance
  - $p_\delta = 0 \Rightarrow$  null **incompatible** with data
  - $p_\delta = 1 \Rightarrow$  null **compatible** with data
  - $0 < p_\delta < 1 \Rightarrow$  data are **inconclusive**



- Fraction of data-supported hypotheses that are null
- Retains strict error control, all rates  $\rightarrow 0$

ASA SGPV Short Course

Blume and Murray, 2022

Funny?

P-VALUE	INTERPRETATION
0.001	HIGHLY SIGNIFICANT
0.01	HIGHLY SIGNIFICANT
0.02	HIGHLY SIGNIFICANT
0.03	HIGHLY SIGNIFICANT
0.04	SIGNIFICANT
0.049	SIGNIFICANT
0.050	OH CRAP. REDO CALCULATIONS.
0.051	ON THE EDGE OF SIGNIFICANCE
0.06	ON THE EDGE OF SIGNIFICANCE
0.07	HIGHLY SUGGESTIVE,
0.08	SIGNIFICANT AT THE P<0.10 LEVEL
0.09	SIGNIFICANT AT THE P<0.10 LEVEL
0.099	HEY, LOOK AT THIS INTERESTING SUBGROUP ANALYSIS
≥0.1	THIS INTERESTING SUBGROUP ANALYSIS

ASA SGPV Short Course      Blume and Murray, 2022

## Definition

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

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

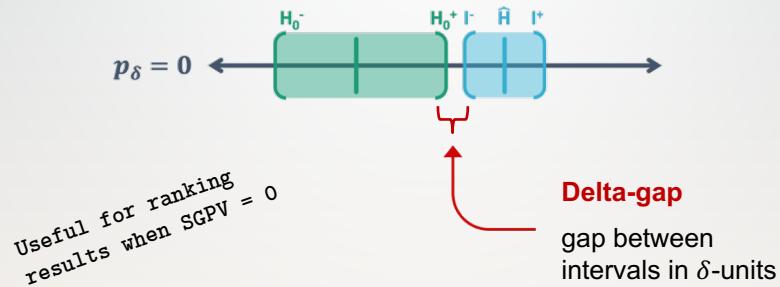
**Proportion** of data-supported hypotheses that are also null hypotheses

**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

## The Delta-gap

When  $\text{SGPV}=0$ , there is a gap between the intervals. The length of that gap, in  $\delta$ -units is the **delta-gap**.



ASA SGPV Short Course

Blume and Murray, 2022

## Steps

- SGPV  $\sim$  the fraction of data-supported hypotheses that are null or practically null
1. Specify an interval null hypothesis or a point null with indifference zone
  2. Find confidence, support or credible interval
  3. Measure the fraction of interval (#2) that is in the null interval
  4. Apply small-sample correction factor, as necessary

ASA SGPV Short Course

Blume and Murray, 2022

**COVID Clinical Trial**

**medRxiv** THE PREPRINT SERVER FOR HEALTH SCIENCES    CSH Cold Spring Harbor Laboratory    BMJ Yale

**Preprint**

**Ivermectin for Treatment of Mild-to-Moderate COVID-19 in the Outpatient Setting: A Decentralized, Placebo-controlled, Randomized, Platform Clinical Trial**

Accelerating COVID-19 Therapeutic Interventions and Vaccines (ACTIV)-6 Study Group, Susanna Naglie  
doi: <https://doi.org/10.1101/2022.06.10.22276252>

This article is a preprint and has not been certified by peer review [what does this mean?]. It reports new medical research that has yet to be evaluated and so should not be used to guide clinical practice.

4 0 0 0 24 0 2103

Abstract Full Text Info/History Metrics Preview PDF

Previous Next

Posted June 12, 2022.

Download PDF    Email  
 Print/Save Options    Share  
 Author Declarations    Citation Tools  
 Supplementary Material  
 Data/Code

[Tweet](#) [Like 222](#)

**COVID-19 SARS-CoV-2 preprints from medRxiv and bioRxiv**

ASA SGPV Short Course      Blume and Murray, 2022

**COVID Clinical Trial**

- 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
- Patients reported each day their symptoms and severity, health care visits, and medications.

**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

**Uncertainty Data Interval: (0.15, 0.82) days**

Difference in mean time unwell between ivermectin treatment and placebo.

Hypothesis	Indifference or Null Zone	SGPV ( $p_\delta$ )	Inference Outcome
3 hours difference	[−0.125, 0.125] days	$p_\delta = 0$	Consistent with alternative zone effects
12 hours difference	[−0.5, 0.5] days	$p_\delta = 0.522$	Inconclusive
18 hours difference	[−0.75, 0.75] days	$p_\delta = 0.896$	Inconclusive
1 day difference	[−1, 1] days	$p_\delta = 1$	Consistent with null zone effects
2 days difference	[−2, 2] days	$p_\delta = 1$	Consistent with null zone effects

ASA SGPV Short Course

Blume and Murray, 2022

# COVID Clinical Trial

**Uncertainty Data Interval: (0.15, 0.82) days**

Difference in mean time unwell between ivermectin treatment and placebo.

Hypothesis	Indifference or Null Zone	SGPV ( $p_\delta$ )	Inference Outcome
3 hours difference	[−0.125, 0.125] days	$p_\delta = 0$	Consistent with alternative zone effects
12 hours difference	[−0.5, 0.5] days	$p_\delta = 0.522$	Inconclusive
18 hours difference	[−0.75, 0.75] days	$p_\delta = 0.896$	Inconclusive
1 day difference	[−1, 1] days	$p_\delta = 1$	Consistent with null zone effects
2 days difference	[−2, 2] days	$p_\delta = 1$	Consistent with null zone effects

ASA SGPV Short Course

Blume and Murray, 2022

## Systolic Blood Pressure

- SBP is reported to the nearest 2 mmHg
- Null Hypothesis: mean SPB is 145 mmHg
- Interval Null hypothesis: mean is 143 to 147 mmHg
- Results from 8 mock studies

ASA SGPV Short Course

Blume and Murray, 2022

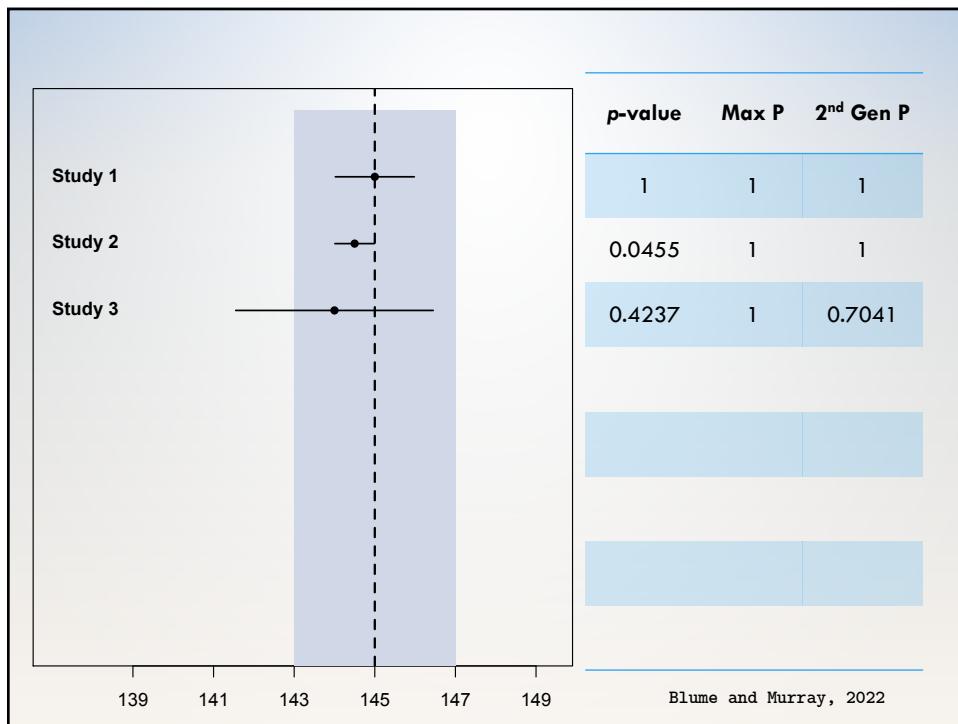
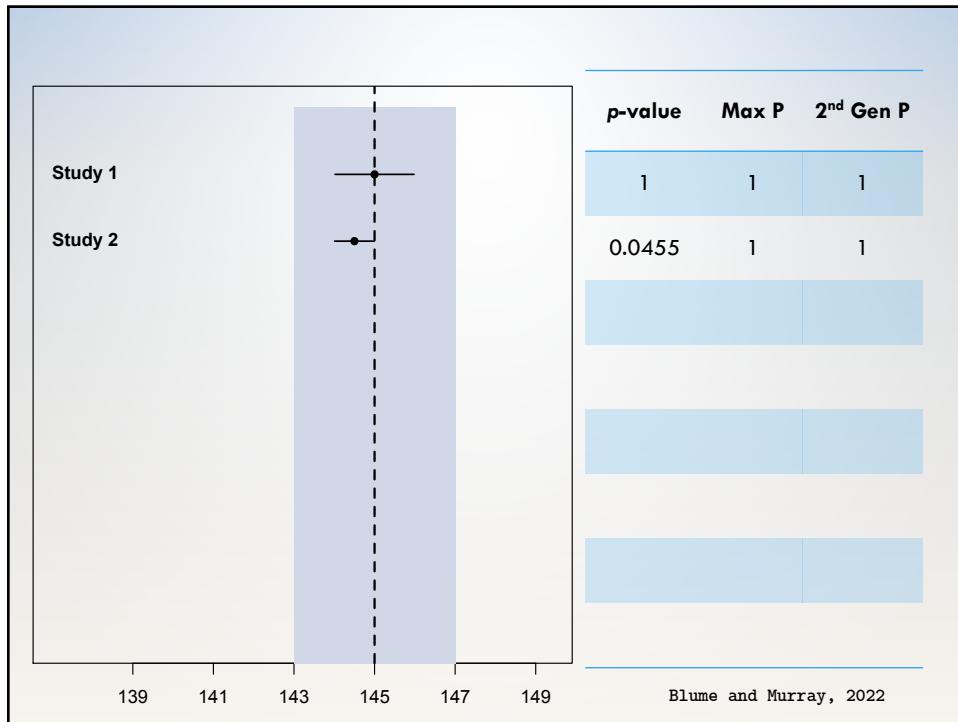
Study 1

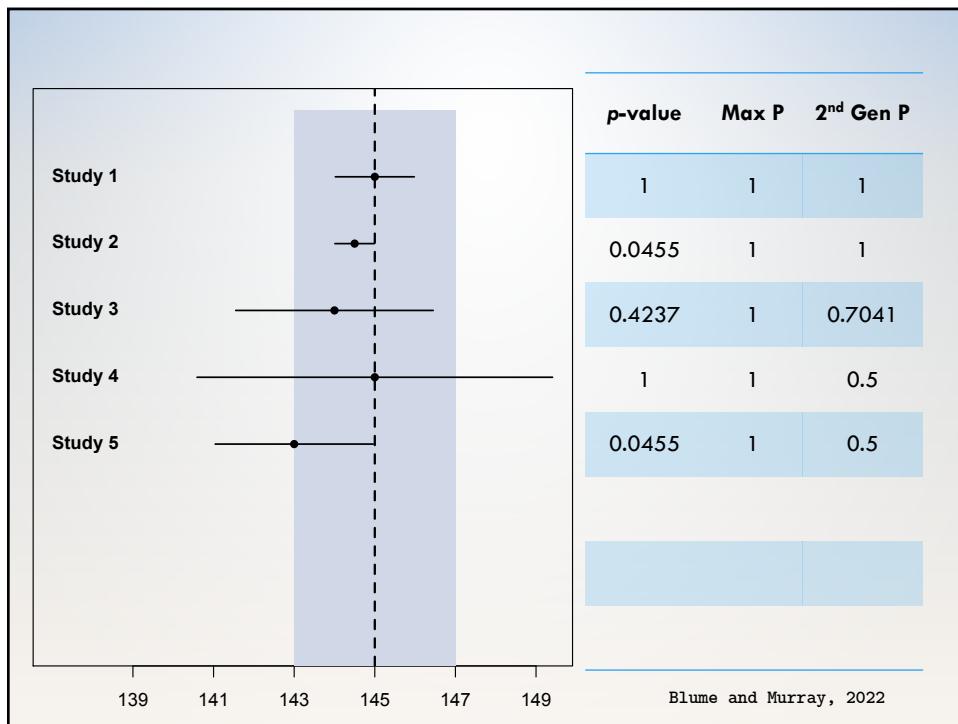
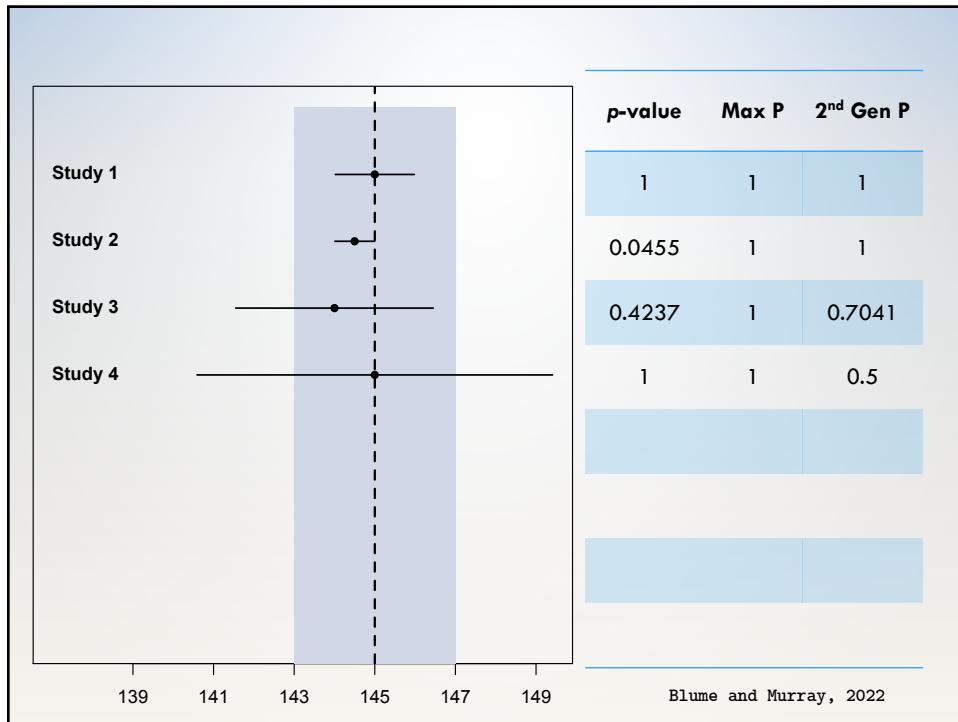
139 141 143 145 147 149

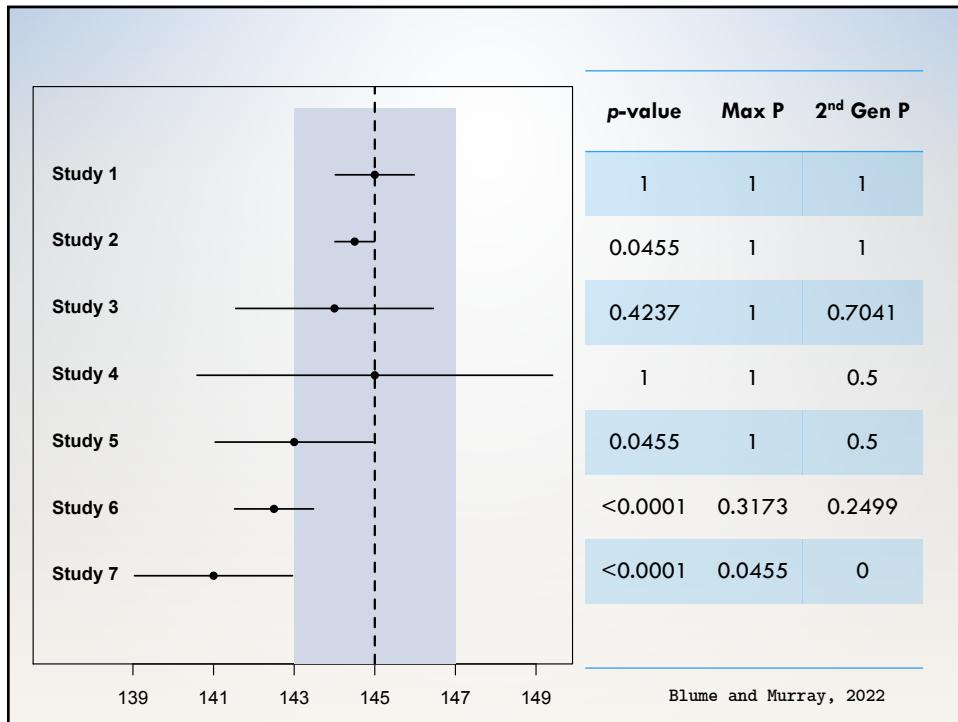
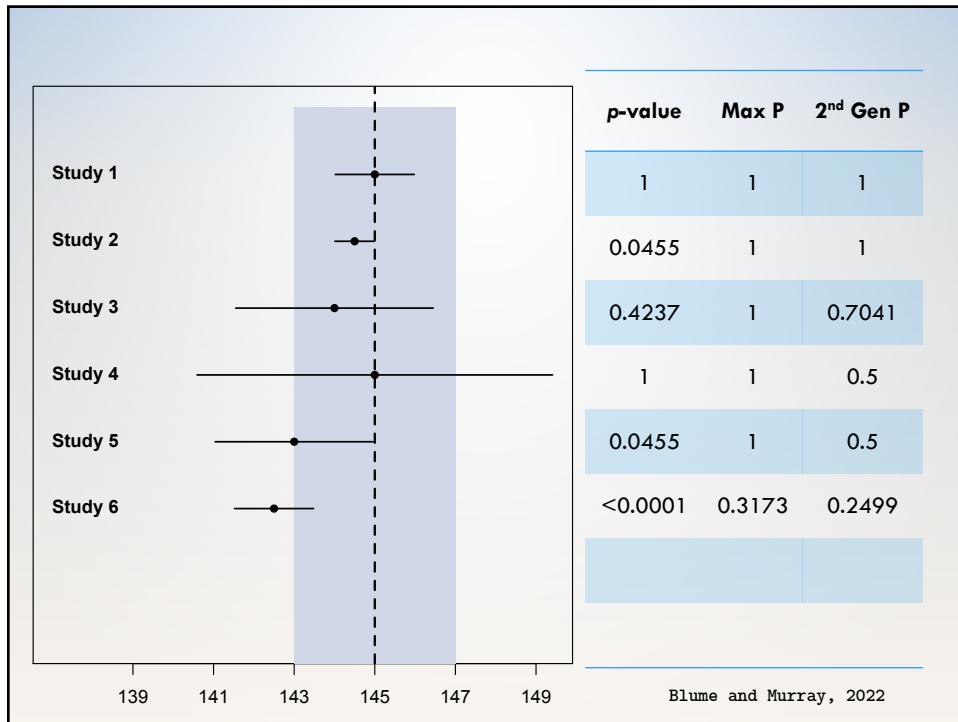
p-value Max P 2<sup>nd</sup> Gen P

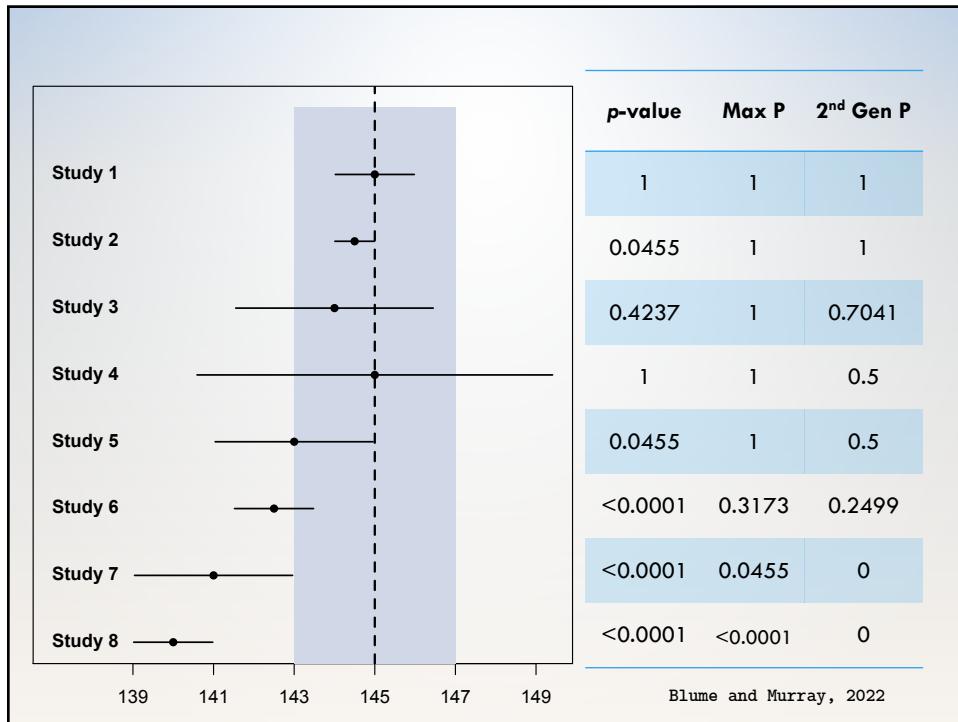
1 1 1

Blume and Murray, 2022









<u>2x2 Tables &amp; Odds Ratios</u>		
Exposure	Outcome	
	No	Yes
Exposed	35	65
Unexposed	50	50

OR = 1.86  
95% CI: (1.05, 3.29)

Null: (0.9, 1.11)

$p_\delta = \frac{(1.11 - 1.05)}{(3.29 - 1.05)}(1) = 0.024$

---

$\log(\text{or}) = 0.62$   
95% CI: (0.05, 1.19)

Null: (-0.1, 0.1)

$p_\delta = \frac{(0.1 - 0.05)}{(1.19 - 0.05)}(1) = 0.044$

ASA SGPV Short Course      Blume and Murray, 2022

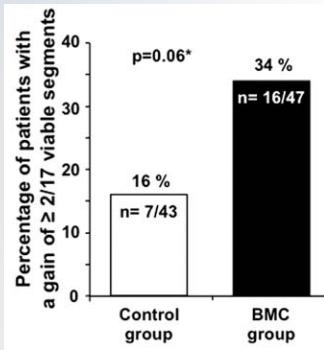
## Bone Marrow in Acute Myocardial Infarction (BOMAMI)

- European Heart Journal (2011)
- Randomized multicenter study
- Intracoronary administration of autologous bone marrow cells (BMCs) can lead to a modest improvement in cardiac function
- Aim: Evaluate the effect of BMC therapy on myocardial viability in patients with decreased left ventricular ejection fraction (LVEF) after acute myocardial infarction (AMI)

ASA SGPV Short Course

Blume and Murray, 2022

## BOMAMI Trial



Odds ratios	95% confidence interval	P-value
2.654	0.967 - 7.286	0.06

Null Interval: (0.9, 1.11)

$$p_\delta = \frac{(1.11 - 0.967)}{(7.286 - 0.967)} (15) = 0.34$$

ASA SGPV Short Course

Blume and Murray, 2022

# Effect Measures for BOMAMI

	BMC	Control	Total
Gain	16	7	23
No Gain	31	36	67
Total	47	43	90
Risk	0.34	0.16	

## Null Hypotheses

OR/RR: (0.9, 1.11)

RD: (-0.05, 0.05)

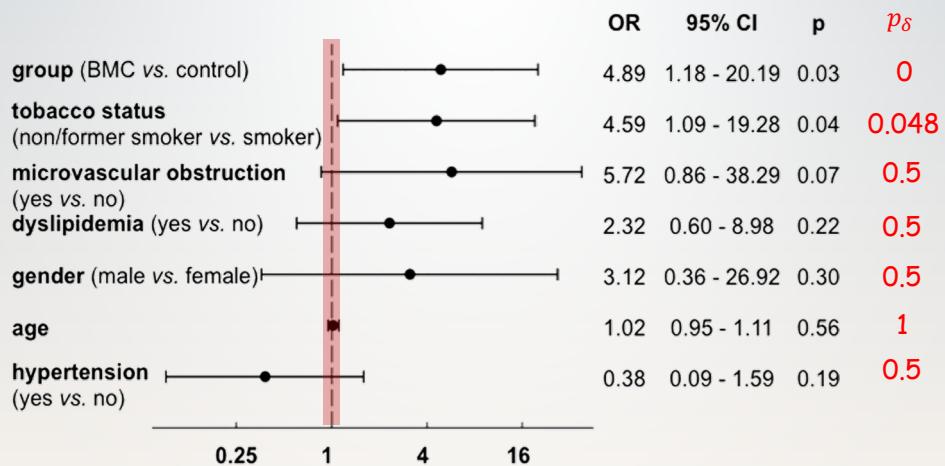
	Estimate	CI Lower	CI Upper	SGPV
Odds Ratio	2.65	0.967	7.286	0.34
Risk Ratio	2.09	0.953	4.589	0.37
Risk Difference	0.18	0.003	0.352	0.24

ASA SGPV Short Course

Blume and Murray, 2022

# BOMAMI Trial

## Logistic Regression with Null Zone: (0.9, 1.11)



ASA SGPV Short Course

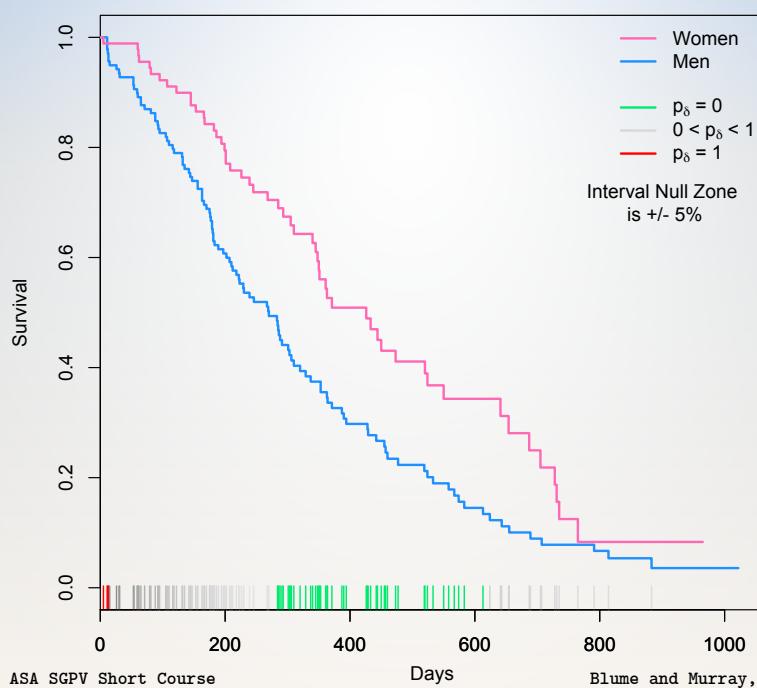
Blume and Murray, 2022

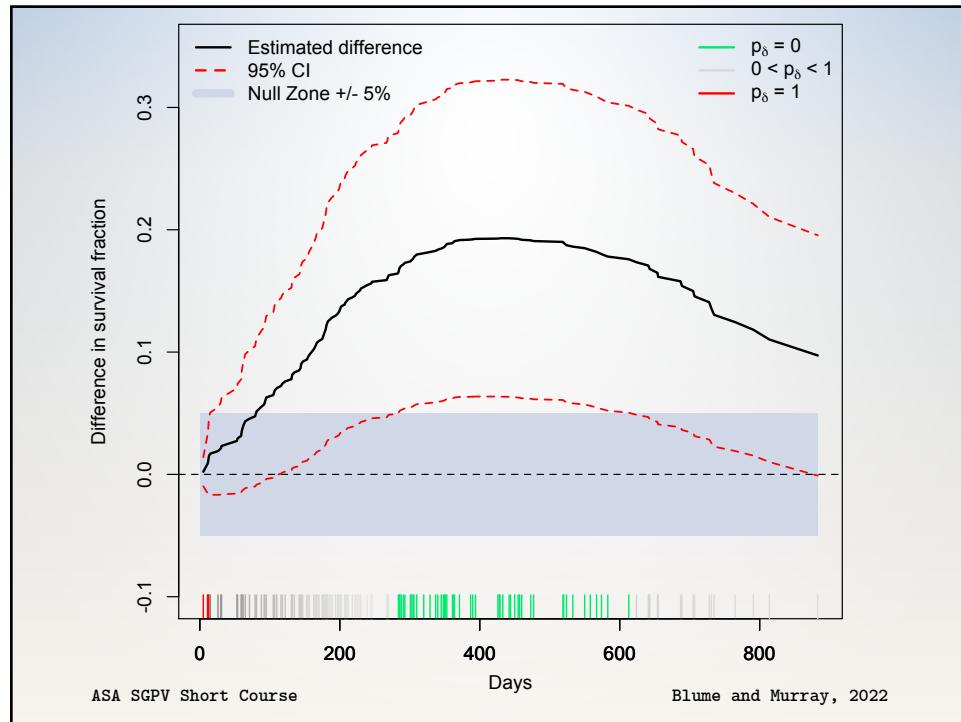
## Lung Cancer Survival

- Survival time in patients with advanced lung cancer (days)
- Potential for gender dissimilarities
- Trial by North Central Cancer Treatment Group (1994)

ASA SGPV Short Course

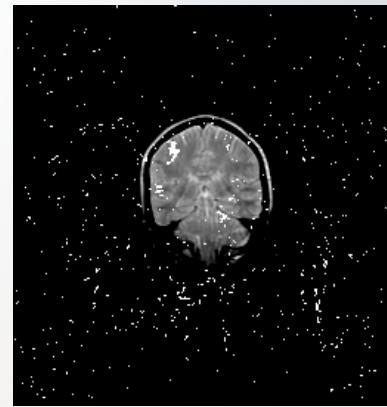
Blume and Murray, 2022

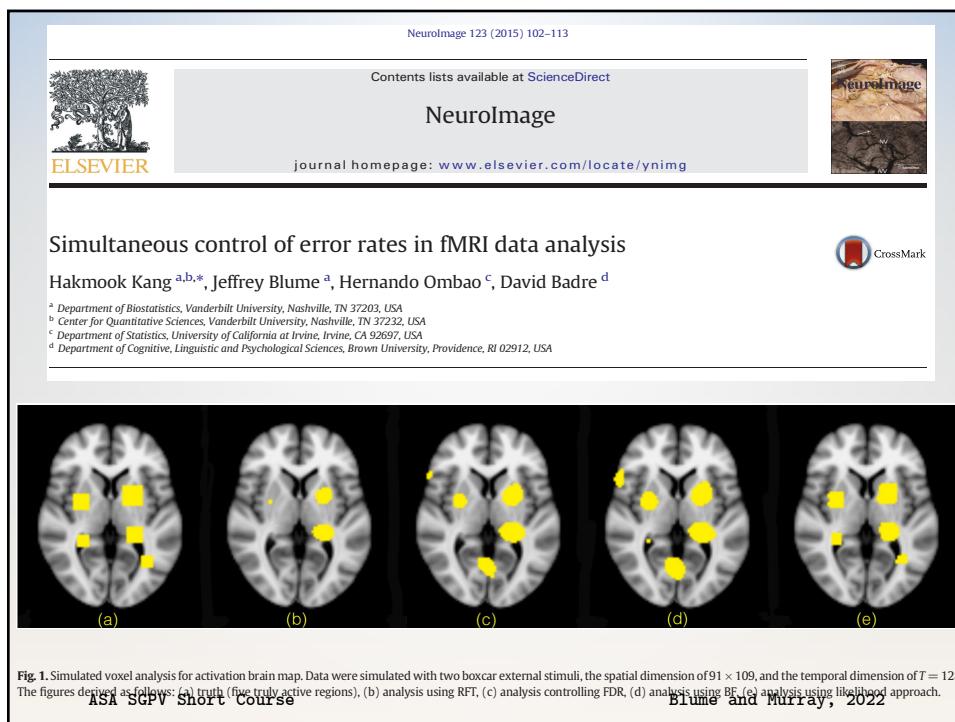
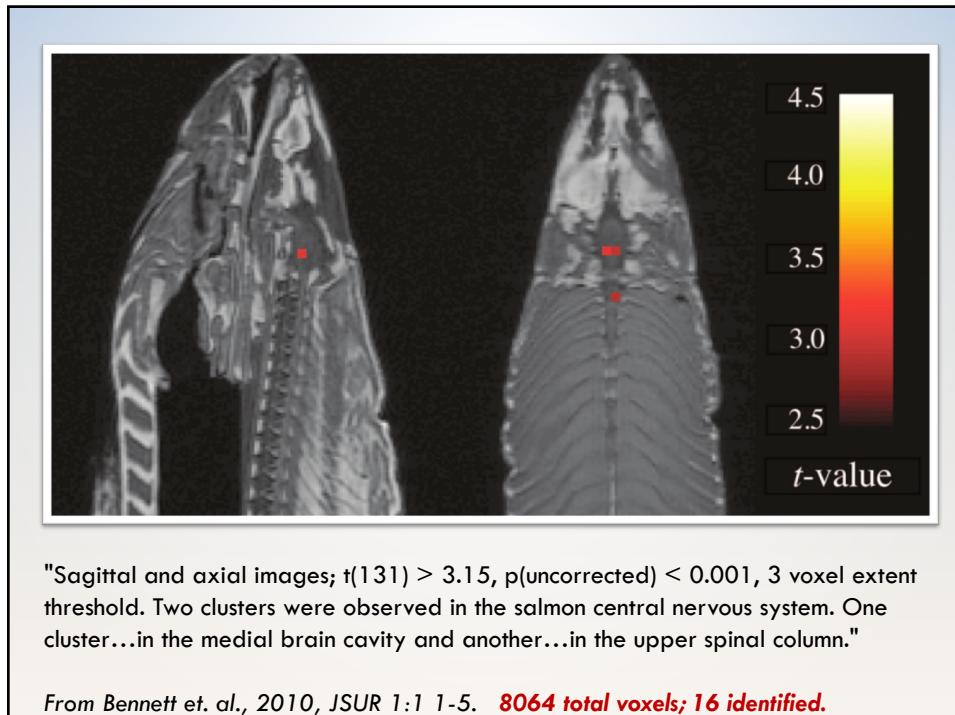




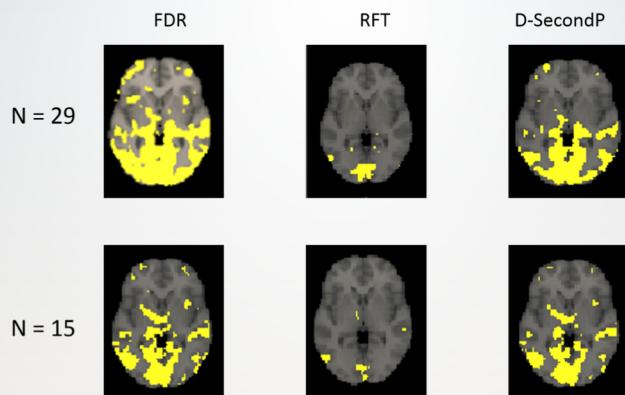
## Setting interval null

- Before analyzing data (!)
- Measurement error
- Subject matter knowledge
- Impact of findings
- Community standard
- Get creative (fMR example)
- Width not critical, buffer
- *The Atlantic salmon imaging saga*





## SGPVs for Functional MRI Data



Lisa Lin Dissertation 2021: Data decimation results of activation maps for the 37th axial slice. The yellow blobs indicate activated areas resulted from FDR (first column), RFT (second column) and D-SecondP (third column). The first row corresponds to the activation maps with full sample size at 29. The second row corresponds to the activation maps with a randomly selected sample of size 15.

Blume and Murray, 2022

## Reference of real world SGPV Example

- “Comparing pathologic outcomes for robotic versus laparoscopic Surgery in rectal cancer resection: a propensity adjusted analysis of 7616 patients”
  - By Hopkins et. al
  - Surgical Endoscopy 2020
  - <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8117669/>

## SGPV in Action!

 HHS Public Access  
 Author manuscript  
 Peer-reviewed and accepted for publication  
[About author manuscripts](#)      [Submit a manuscript](#)

Surg Endosc. Author manuscript; available in PMC 2021 Jun 1.  
*Published in final edited form as:*  
Surg Endosc. 2020 Jun; 34(6): 2613–2622.  
*Published online* 2019 Jul 25. doi: [10.1007/s00464-019-07032-1](https://doi.org/10.1007/s00464-019-07032-1)

PMCID: PMC8117669  
 NIHMSID: NIHMS1700788  
 PMID: [31346754](https://pubmed.ncbi.nlm.nih.gov/31346754/)

Comparing pathologic outcomes for robotic versus laparoscopic Surgery in rectal cancer resection: a propensity adjusted analysis of 7616 patients

M. Benjamin Hopkins,<sup>1</sup> Timothy M. Geiger,<sup>1</sup> Alva J. Bethurum,<sup>1</sup> Molly M. Ford,<sup>1</sup> Roberta L. Muldoon,<sup>1</sup> David E. Beck,<sup>1</sup> Thomas G. Stewart,<sup>2</sup> and Alexander T. Hawkins<sup>1</sup>

► [Author information](#) ► [Copyright and License information](#)    [Disclaimer](#)

The publisher's final edited version of this article is available at [Surg Endosc](#)

ASA SGPV Short Course      Blume and Murray, 2022

## Statistical Analysis Paragraph 1

- “...The primary hypothesis was tested using a second generation p value ( $p\delta$ ), in which the confidence interval for the log-odds of the outcome (and marginalized difference in probability) for robotic versus laparoscopic approach was compared to the region of practical equivalence specified prior to the analysis [11, 12] ...”

## Statistical Analysis Paragraph 2

- "... We implemented the second generation p value in this analysis because it avoids a common limitation of many traditional statistical tests related to the interpretation of null results. Take the t test as an example..."

ASA SGPV Short Course

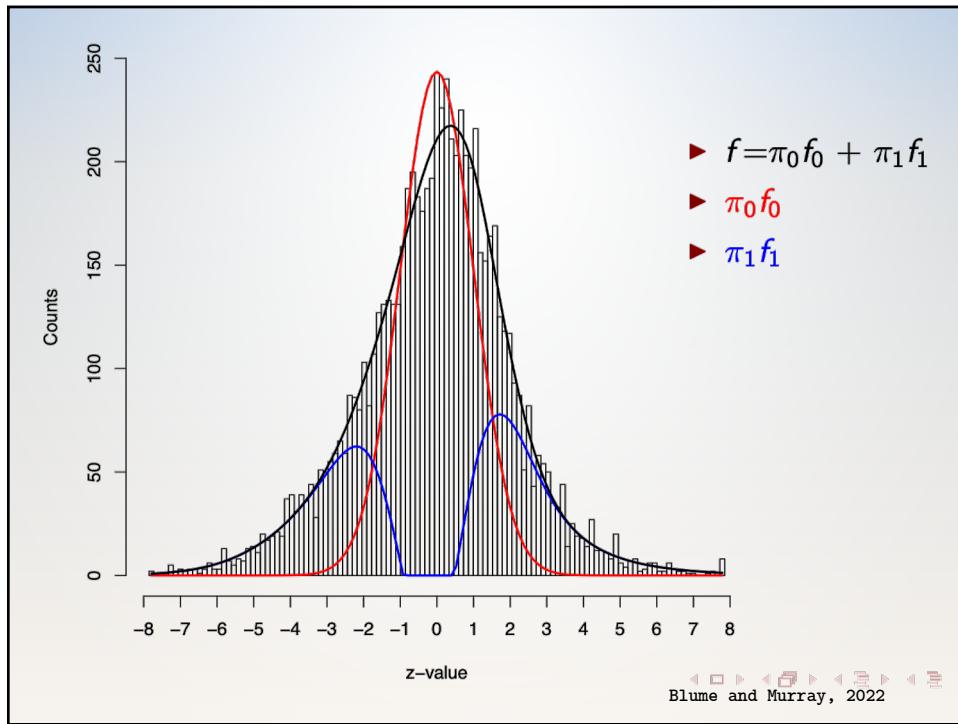
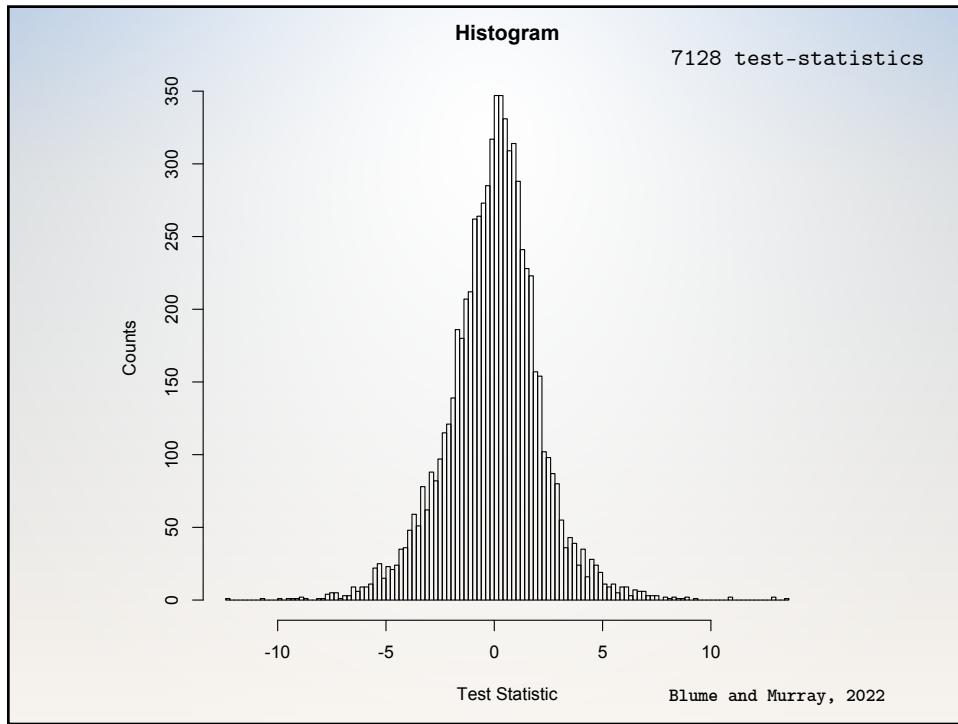
Blume and Murray, 2022

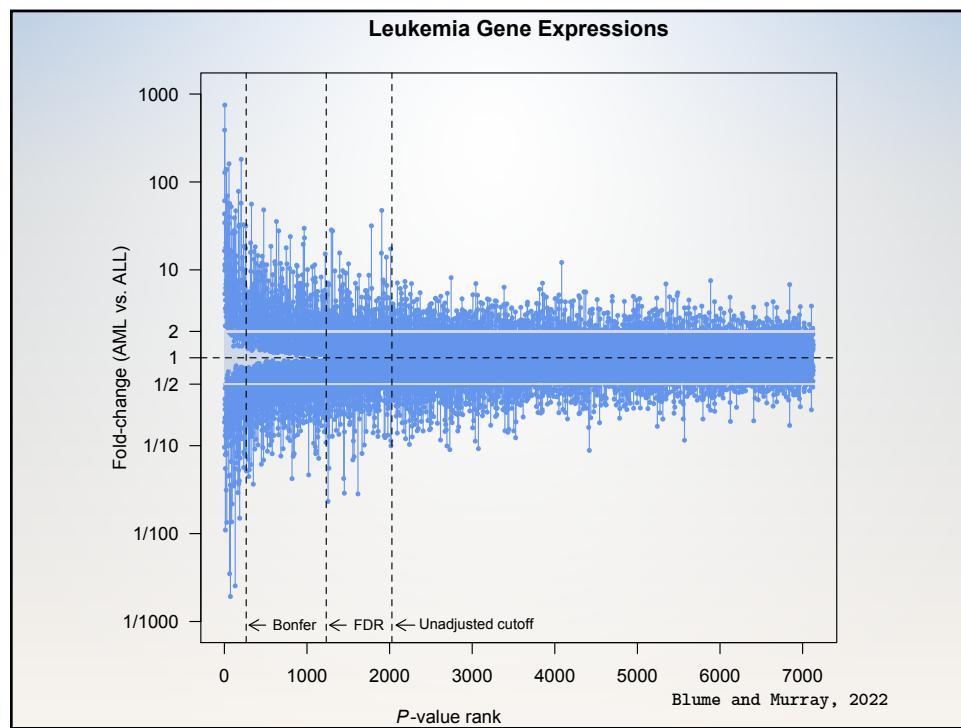
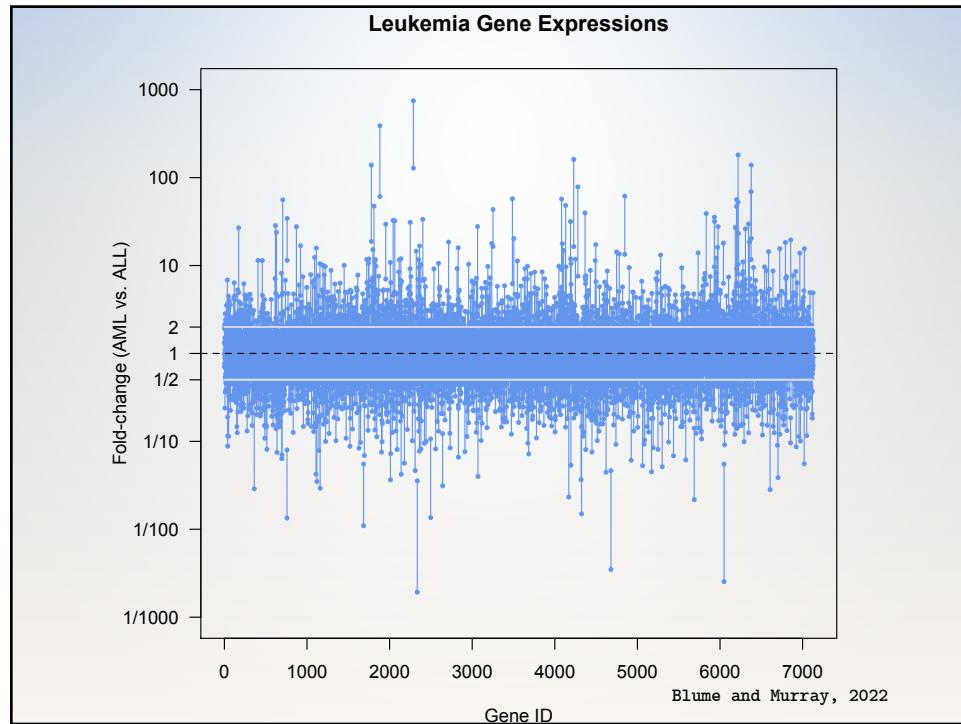
## High-Dimensional Data: Leukemia gene expression

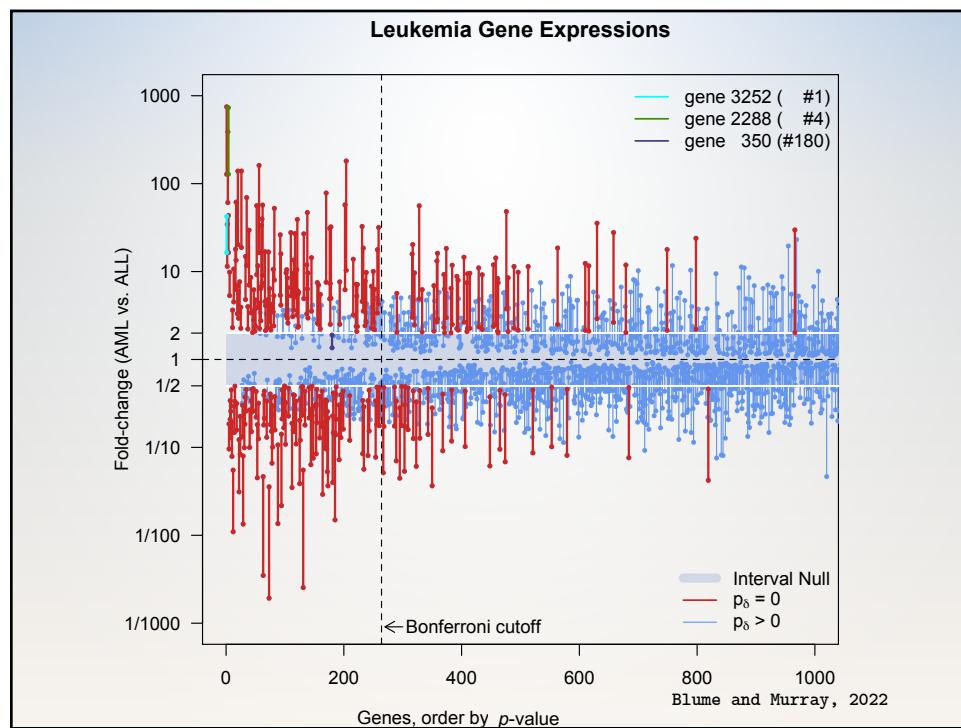
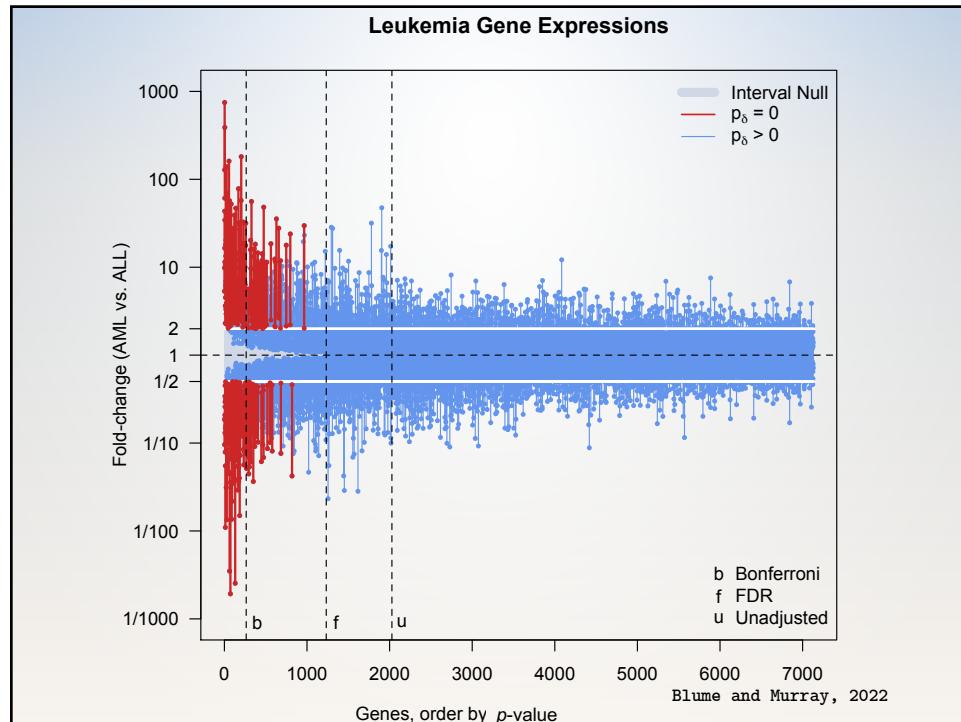
- Classifying acute leukemia by precursors (Golub 1999, *Science*)
- 7128 genes ; 72 patients (47 ALL and 25 AML)
- Affymetrix chip collected expression levels
- Goal: Identify 'interesting' genes whose expression levels differ between All and AML subjects.
- Looking for fold changes of 2 or more

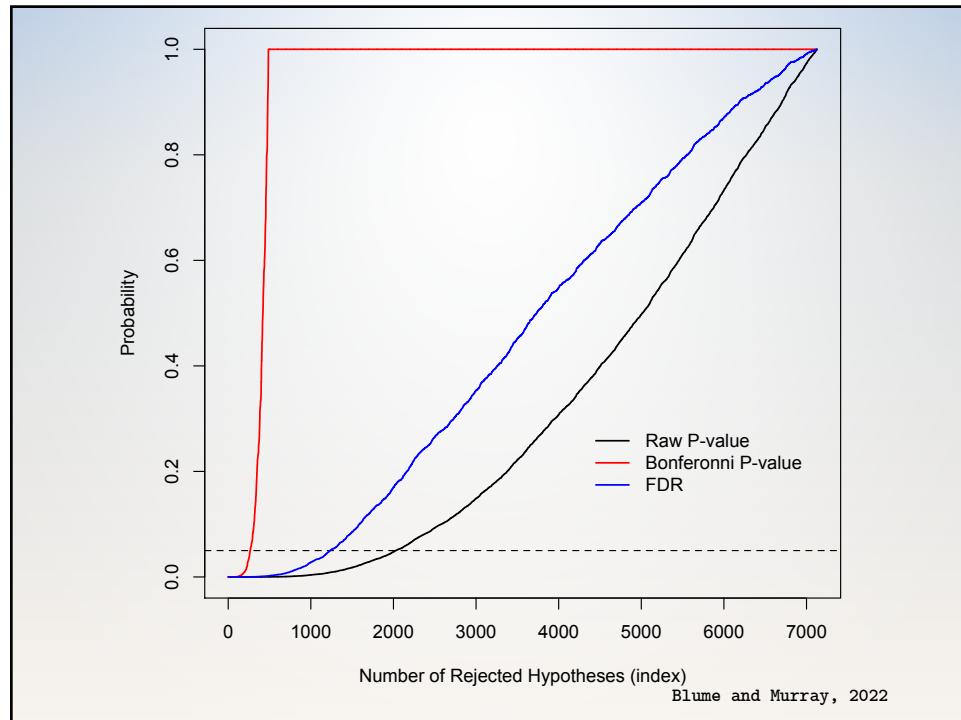
ASA SGPV Short Course

Blume and Murray, 2022









### Cross-Tabulation of Leukemia Results

- Bonferroni vs Second Generation  $p$ -values

		<b><math>1/2 &lt; \text{Fold Change} &lt; 2</math> (<math>\delta = 0.3</math>)</b>			
		$p_\delta = 0$	$p_\delta > 0$	<b>Total</b>	
$p_{bon} < 0.05$	164	100	264		
$p_{bon} > 0.05$	65	6799	6864		
<b>Total</b>	229	6899	7128		

## Leukemia study findings

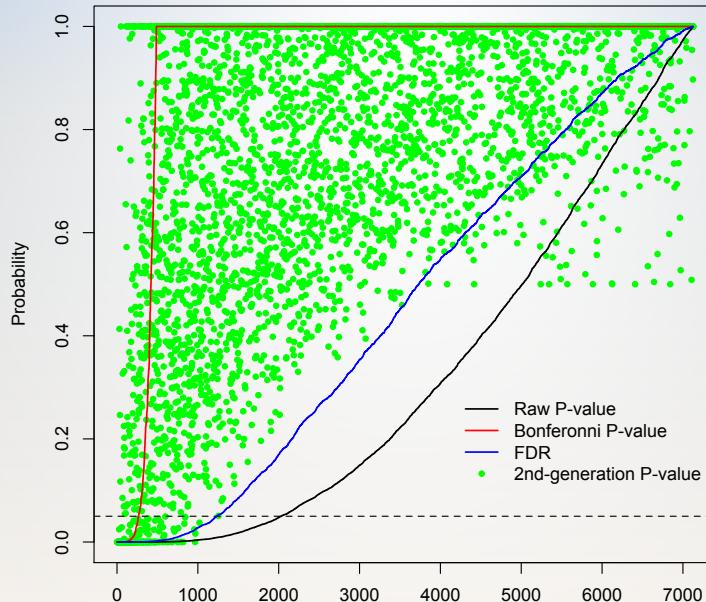
- Findings: Bonferroni 264, SGPV 229
  - Agree on 164 findings
  - Bonferroni +100, SGPV +65
- Effective Type I error rate: 0.037 vs. 0.032
- FDR of 2.45% captures all  $p_\delta = 0$ , 737 others
- Moving cutoff trades Type I for Type II errors
- SGPV changes the *ranking* of findings
  - Three categories now: null, alt, inconclusive
  - Null findings not illustrated here

Some SGPV findings  
have a priori  
published validation

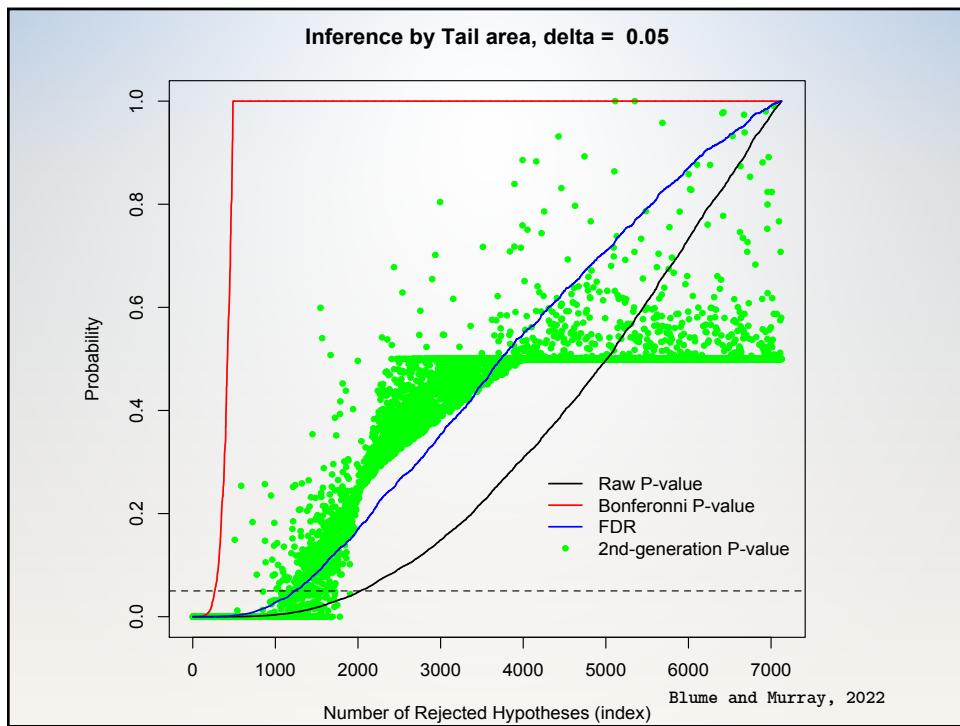
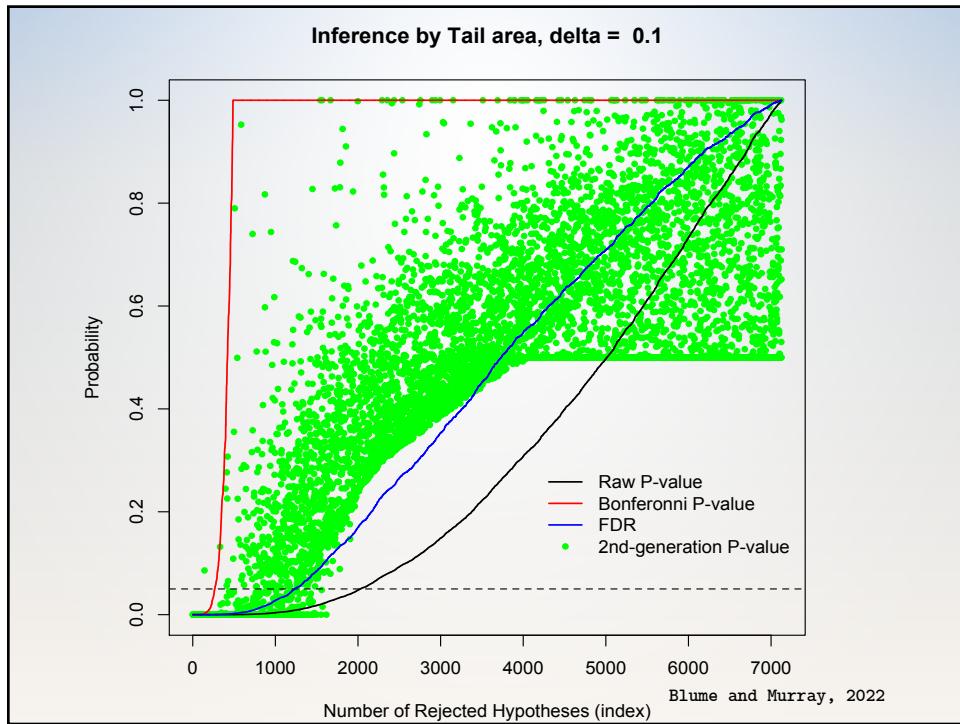
ASA SGPV Short Course

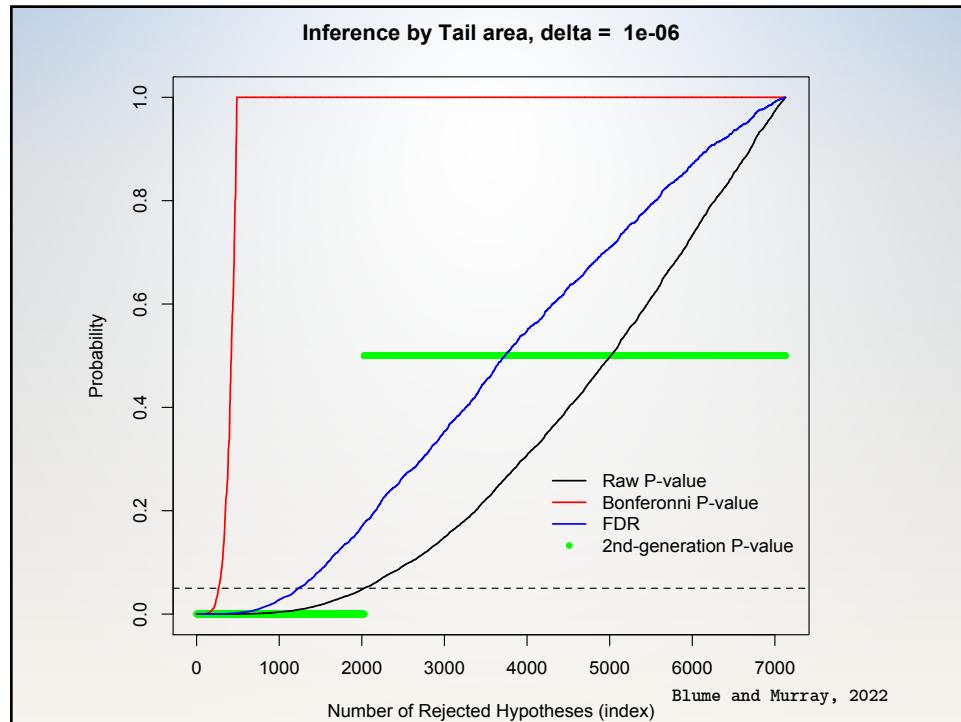
Blume and Murray, 2022

Inference by Tail area, delta = 0.3



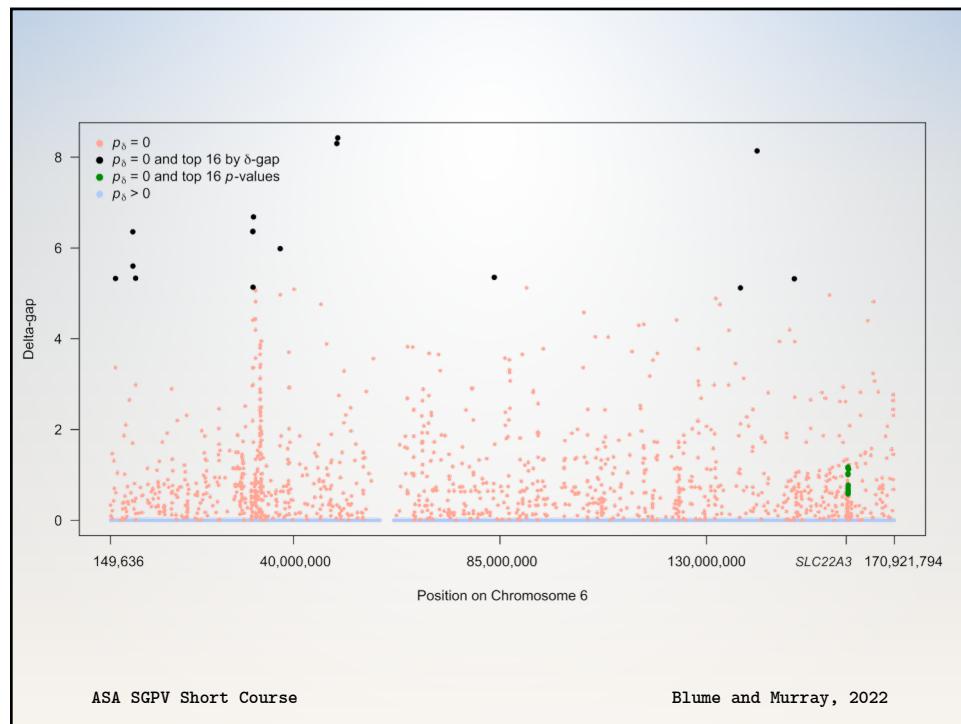
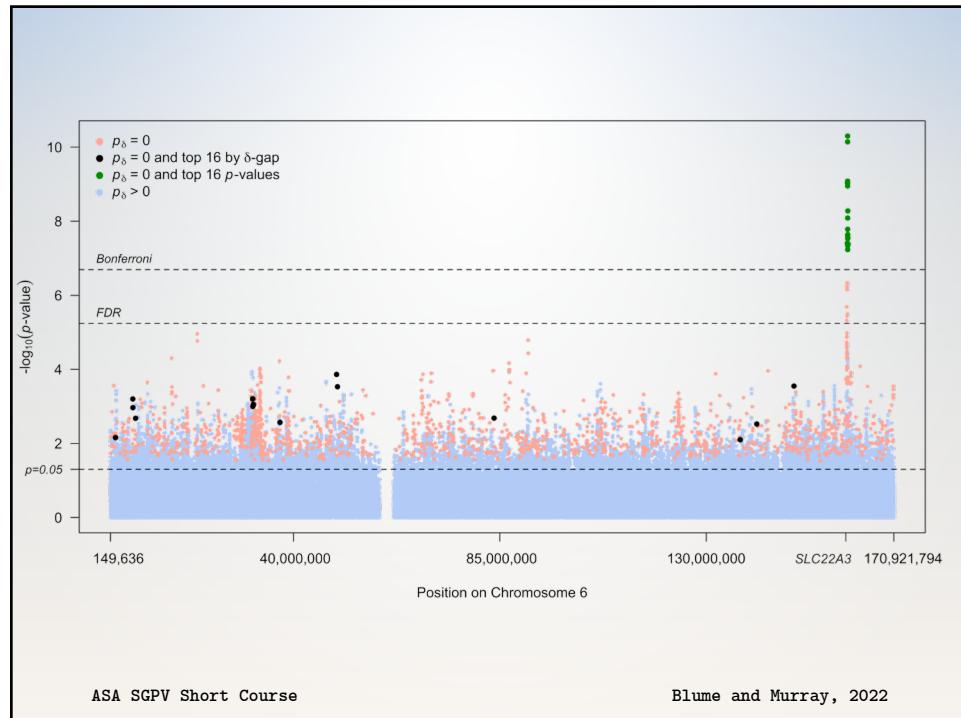
Blume and Murray, 2022

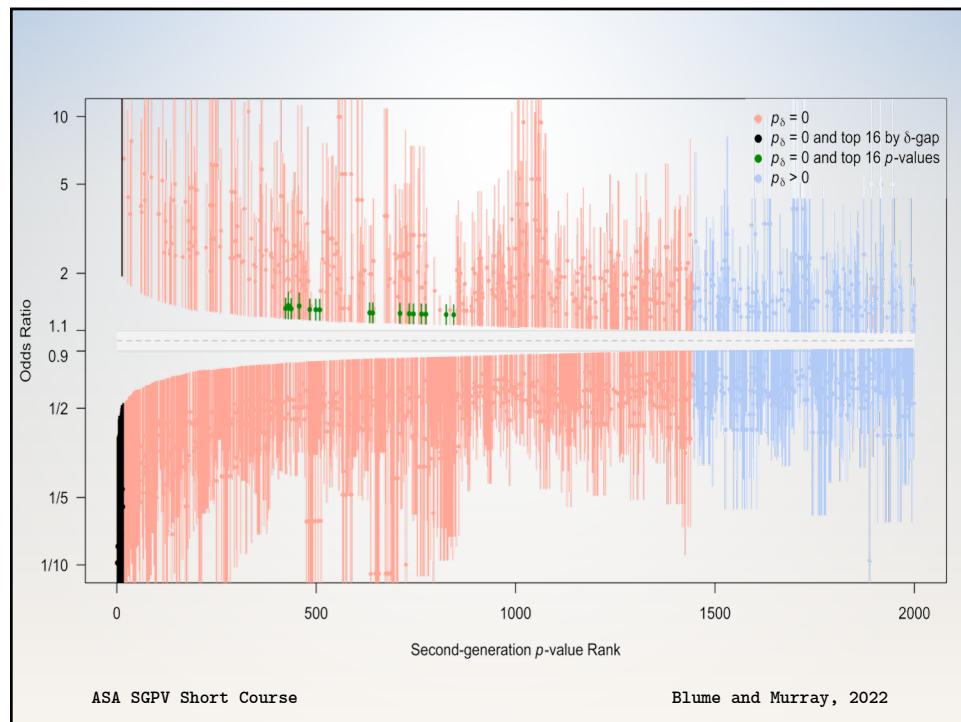
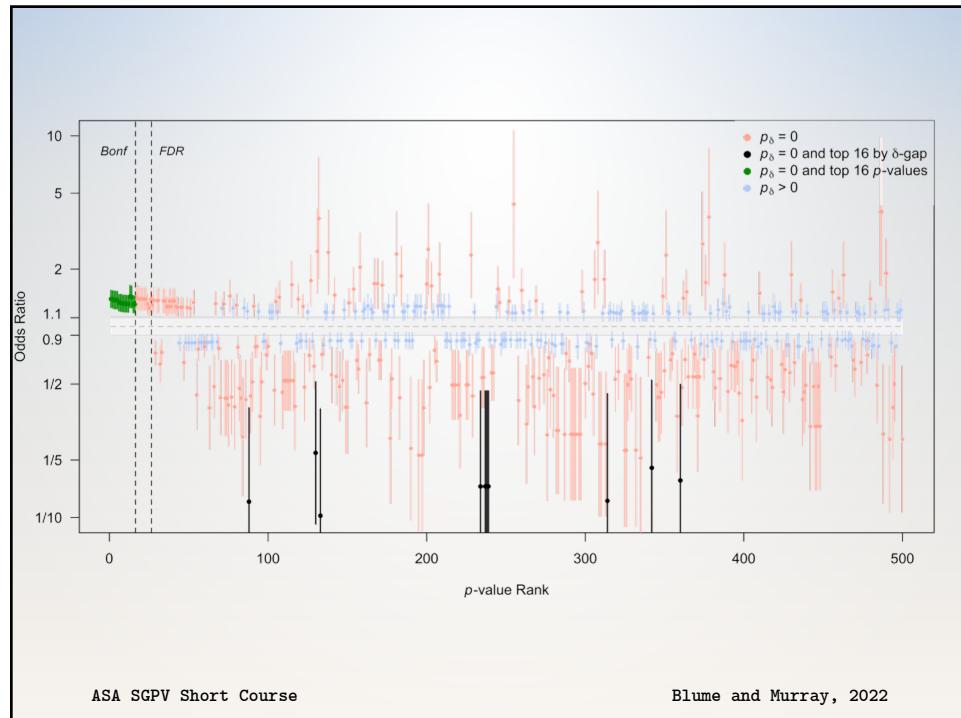


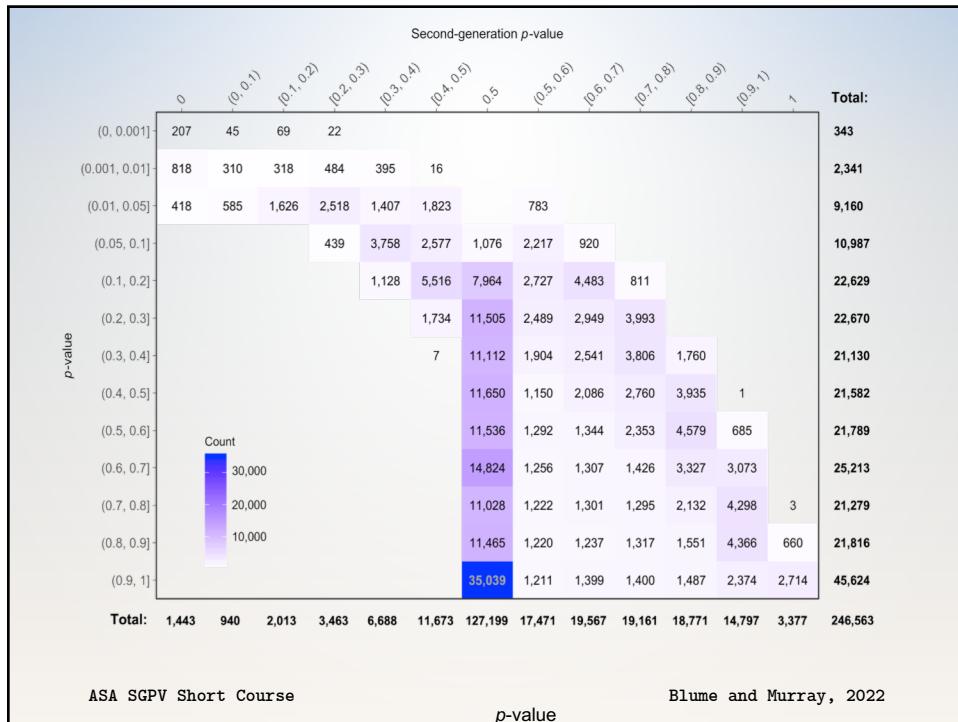


## Prostate Cancer SNPs

- International Consortium for Prostate Cancer Genetics (Schaid and Chang 2005; ICPCG 2018)
- 3,894 subjects: 2,511 cases & 1,383 controls
- 247,000 single-nucleotide polymorphisms (SNPs) from Chromosome 6
- Goal: Identify 'interesting' SNPs potentially associated with prostate cancer
- Looking for odds ratios of  $<0.9$  or  $>1.11$







# Outrageous Claim

SGPVs achieve the inferential properties that many scientists hope, or believe, are attributes of the classic *p*-value.

## Discussion?

**Time for live coding  
example in R!**

ASA SGPV Short Course

Blume and Murray, 2022

**End of Part 1!**

ASA SGPV Short Course

Blume and Murray, 2022