

Part I

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

Jeffrey D. Blume, and Megan H. Murray,

PhD almost PhD (July 6th defense)

School of Data Science

Department of Biostatistics

University of Virginia

Vanderbilt University

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
 - UVA Profile: www.datascience.virginia.edu/people/jeffrey-blume
- Megan
 - PhD Candidate in Biostatistics at Vanderbilt: July 6th Defense
 - Dissertation: ‘On second-generation *p*-values for equivalence testing and study planning, and flexible false discovery rate computation for classical *p*-values’
 - After graduation: Research Scientist at Eli Lilly Pharmaceuticals

Course Layout

- Slides Part I
 - Introduction and applications
- Coding Part I
- Slides Part II
 - Equivalence tests, statistical properties, and false discovery rates
- Coding Part II
- Questions and Discussion

ASA SGPV Short Course

Blume and Murray, 2022

Resources

- GitHub with Slides and Code
 - www.github.com/murraymegan/SGPV-ASA-Short-Course
- RStudio Desktop
 - www.rstudio.com/products/rstudio/download
- Interrupt or use Zoom chat for questions!
- For technical difficulties email Megan
 - megan.c.hollister@vanderbilt.edu

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 a 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 (4)
- Code Part 1a
- High-dimensional example, 7128 Genes
 - $\alpha=0.05$ vs $\alpha=0.05/7128$ vs SG *p*-value
- High-dimensional example
 - Prostate Cancer SNP data (~247,000)
- Outrageous claim
- Code Part 1b



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

- StatisticalEvidence.com
- Examine statistical properties in module 2
- Retains strict error control

Evidential Metric	What it measures	SGPV
1	Summary measure	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

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

2nd-generation

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

Version 2.0

- ✓ Number between 0 and 1 → $\begin{cases} \text{near 0 supports alt} \\ \text{near 1 supports null} \\ \text{near } \frac{1}{2} \text{ inconclusive} \end{cases}$
- ✓ 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. Above the axis, the text "Point Null Hypothesis" is written. To the right of the point, a light blue horizontal interval labeled \hat{H} is shown, with a bracket below it labeled "confidence interval".

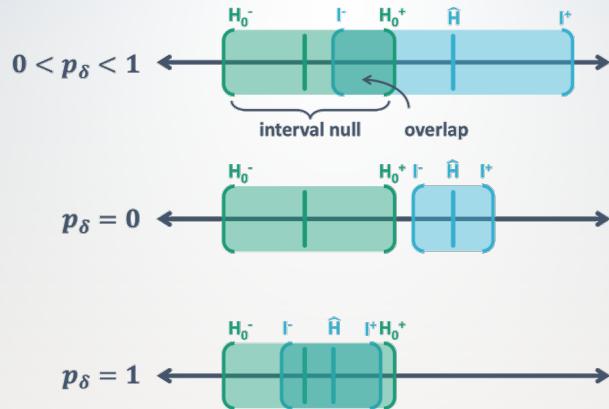
Interval Null Hypothesis: A double-headed arrow labeled H_0^- at the left end and H_0^+ at the right end. Below the axis, the text "Interval Null Hypothesis" is written. To the left of the interval, a green shaded region is labeled δ . Below the axis, the text "interval null" is written. To the right of the interval, a light blue horizontal interval labeled CI^- and CI^+ is shown, with a bracket below it labeled "overlap".

Text Labels:

- 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_δ
- δ 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

Definition

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

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

Proportion of data-supported hypotheses that are also null hypotheses

Small-sample correction factor
shrinks proportion to $\frac{1}{2}$ when $|I|$ wide

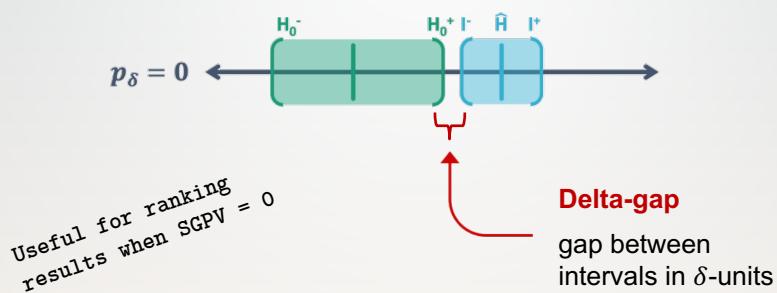
when $|I| > 2|H_0|$

ASA SGPV Short Course

Blume and Murray, 2022

The Delta-gap

When SGPV=0, there is a gap between the intervals. The length of that gap, in δ -units is the **delta-gap**.



ASA SGPV Short Course

Blume and Murray, 2022

Steps

- SGPV ~ the fraction of data-supported hypotheses that are null or practically null
1. Specify an the 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

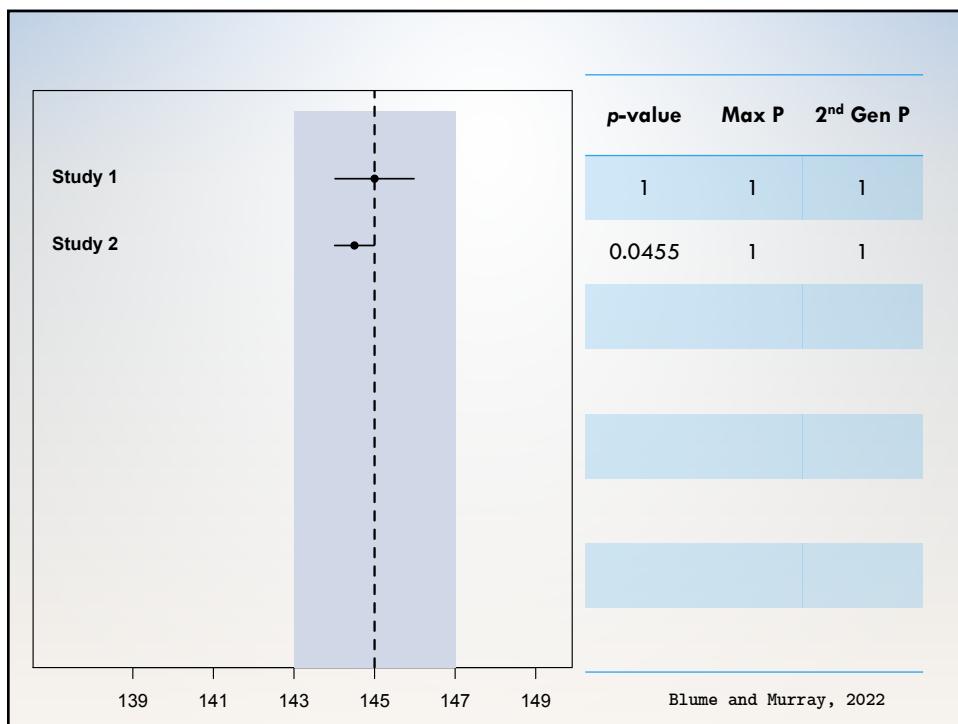
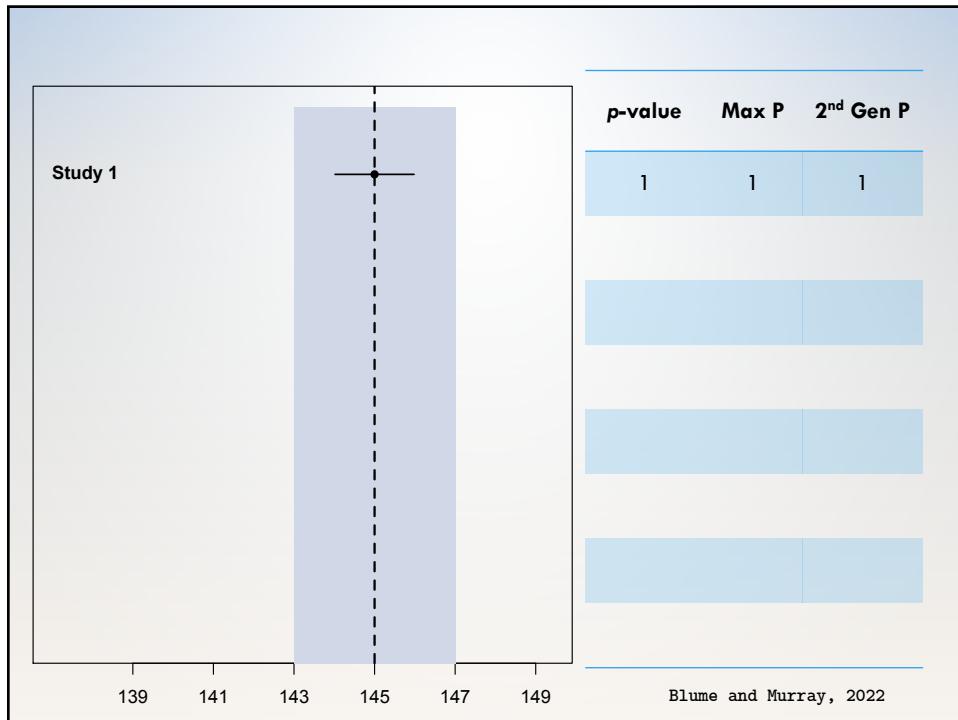
Blume and Murray, 2022

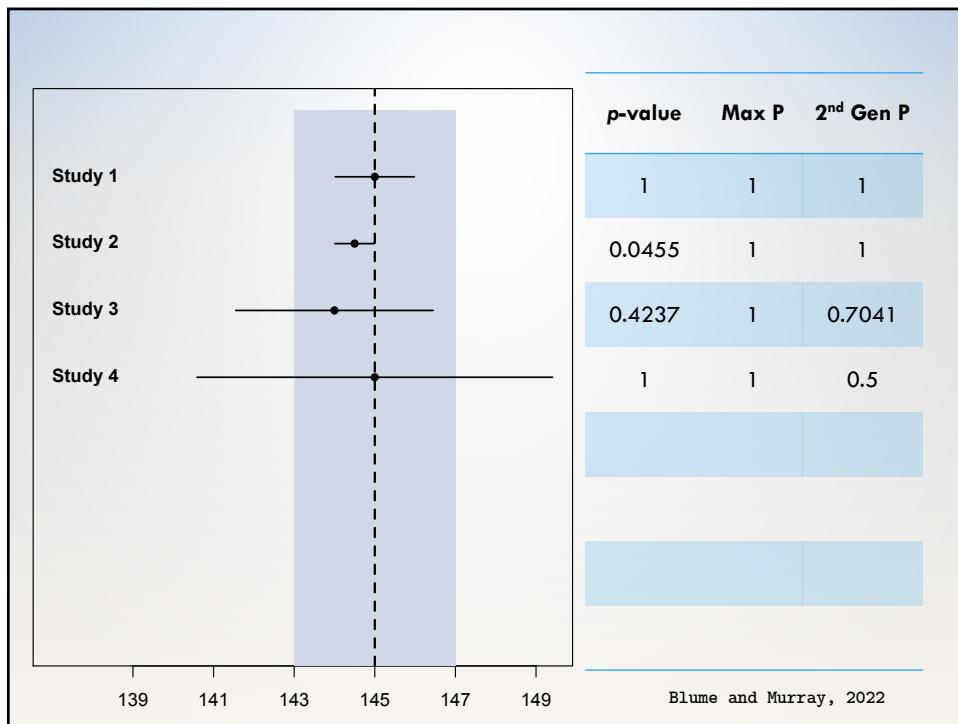
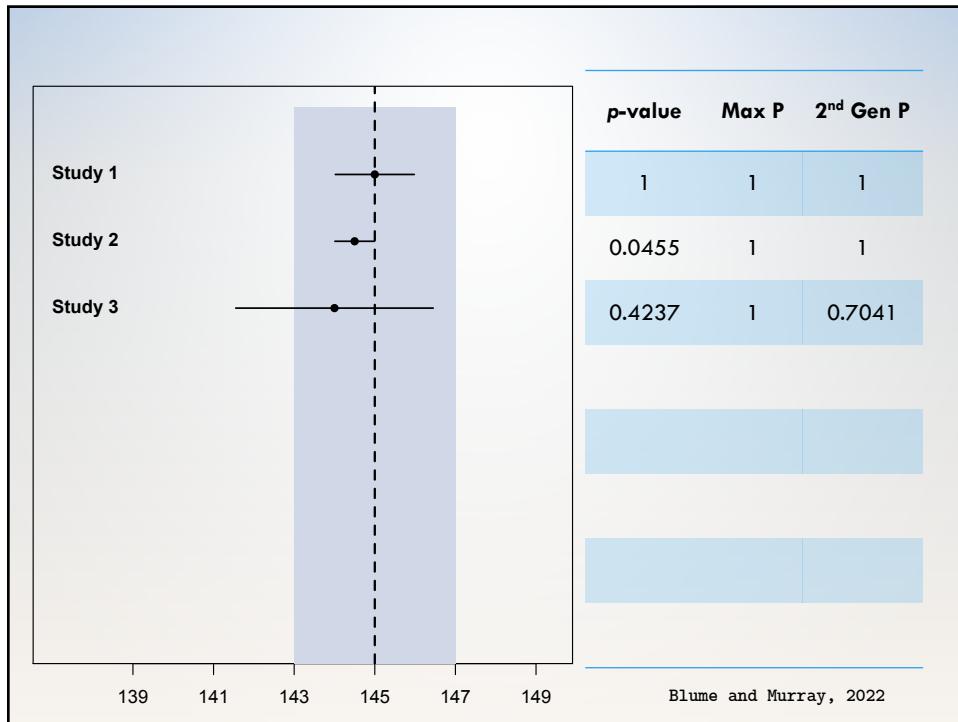
Systolic Blood Pressure

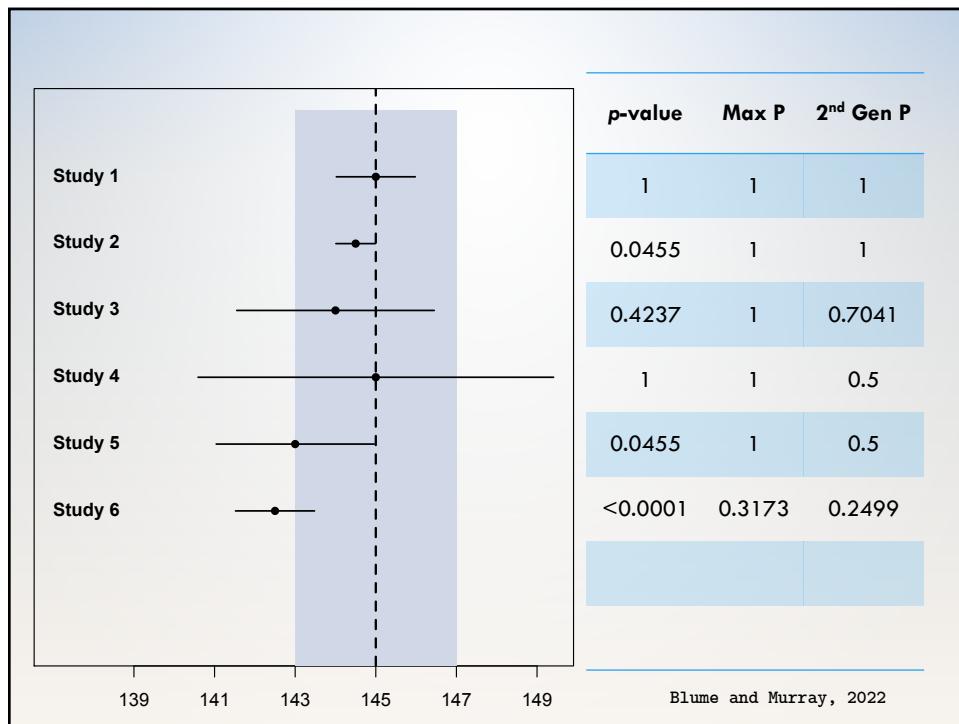
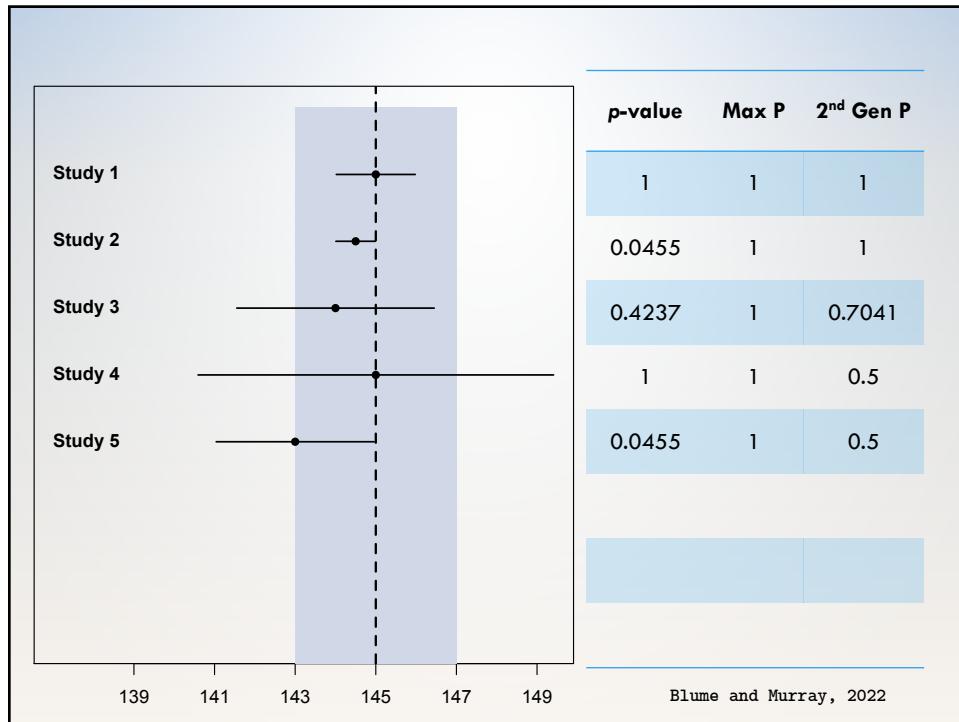
- SBP is reported to the nearest 2 mmHg
- Null Hypothesis: mean SPB is 145 mmHg
- Interval Null hypothesis: mean is 143 to 147 mmHg
- Results from 8 mock studies

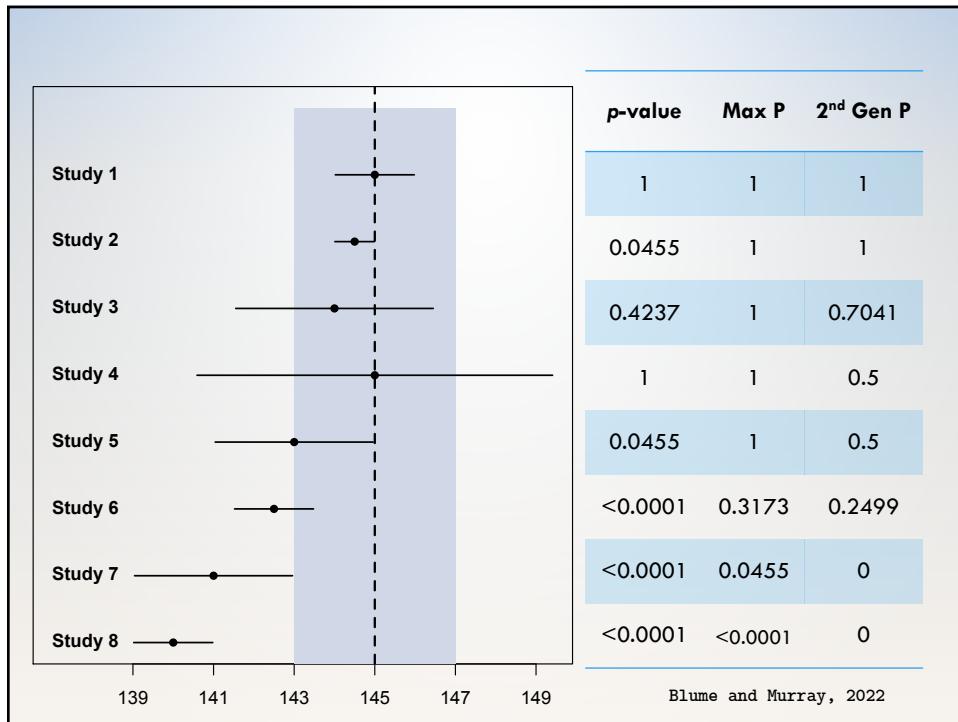
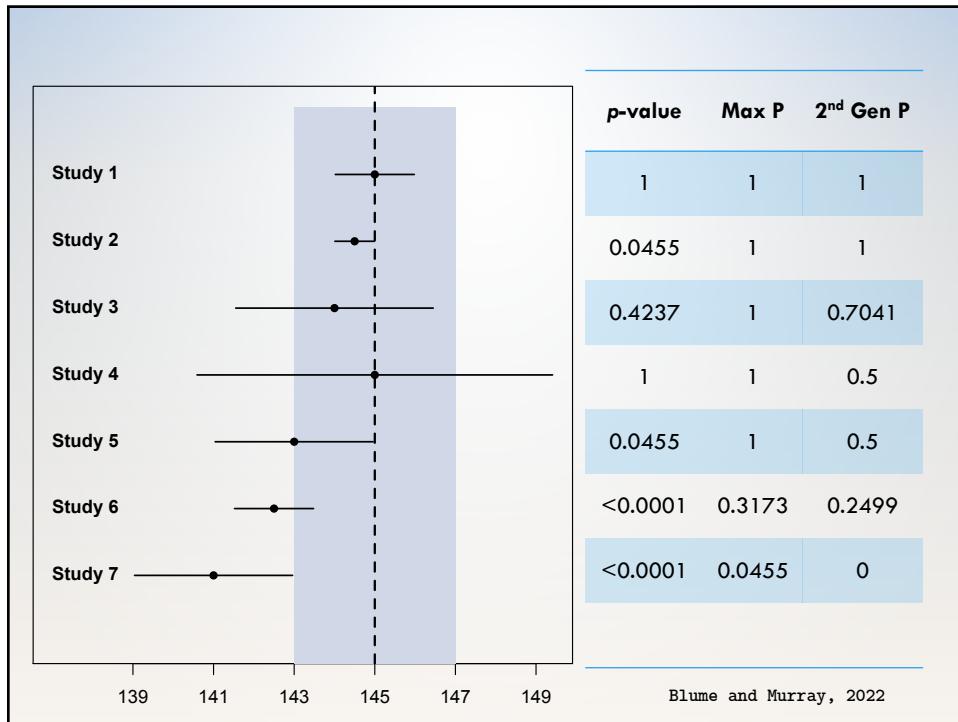
ASA SGPV Short Course

Blume and Murray, 2022









2x2 Tables & Odds Ratios

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(or) = 0.62
95% CI: (0.05, 1.19)

Null: (-0.1, 0.1)

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

- Scale can be important for inconclusive results
- A p_δ of 0 or 1 will always remain the same
- An inconclusive p_δ will vary slightly

ASA SGPV Short Course

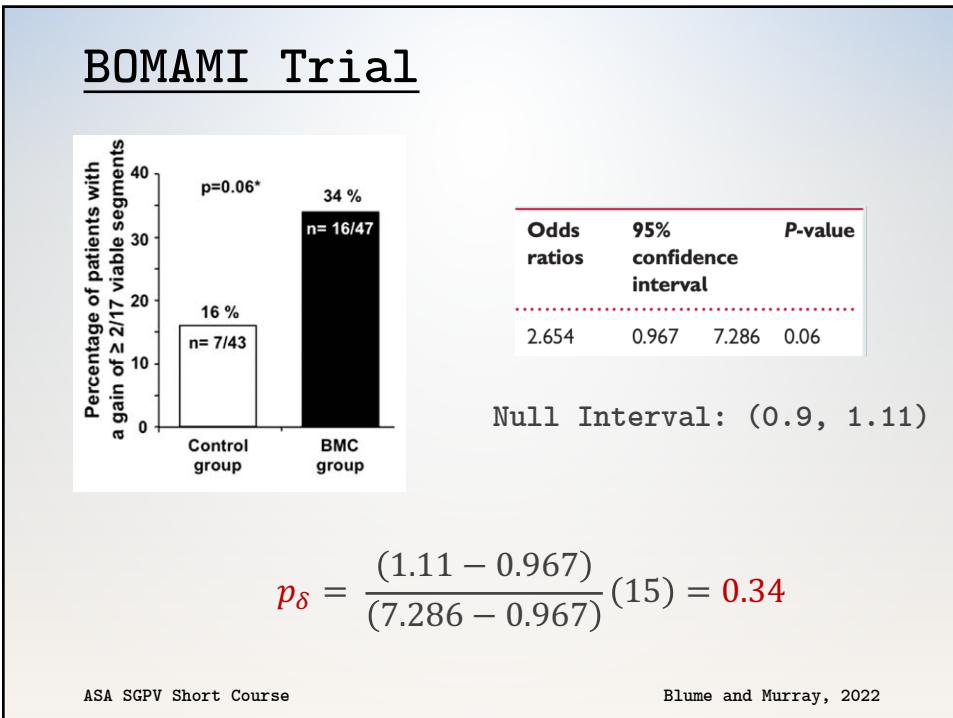
Blume and Murray, 2022

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

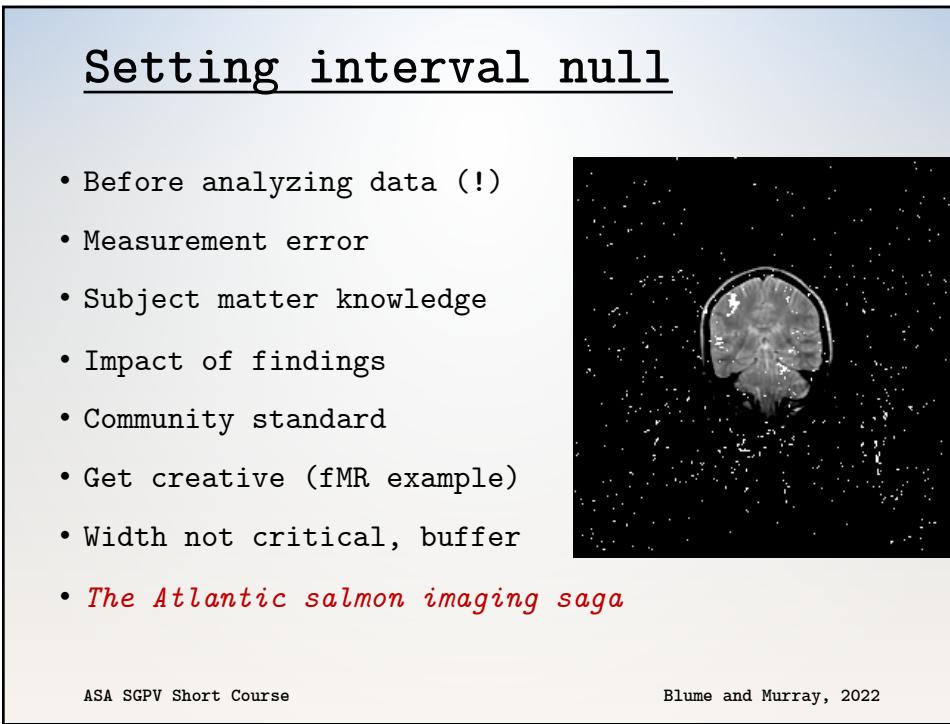
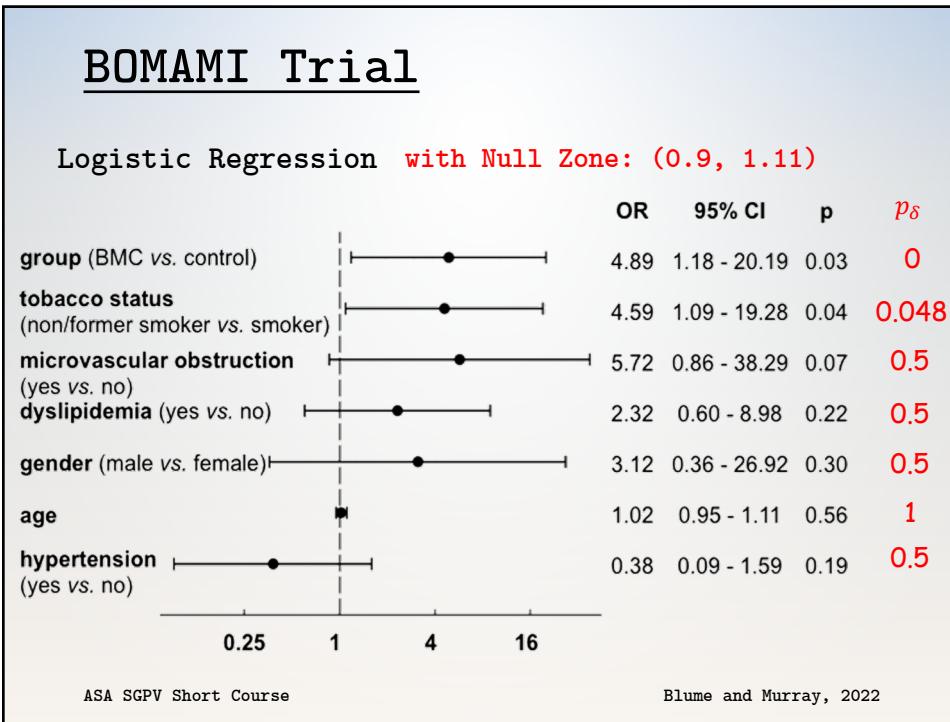


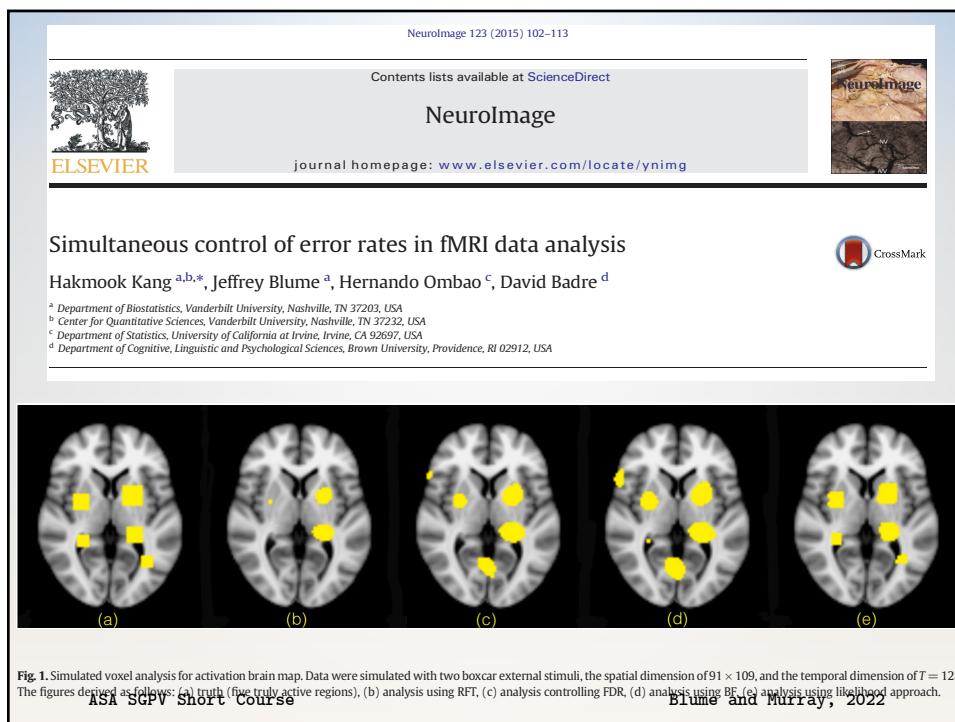
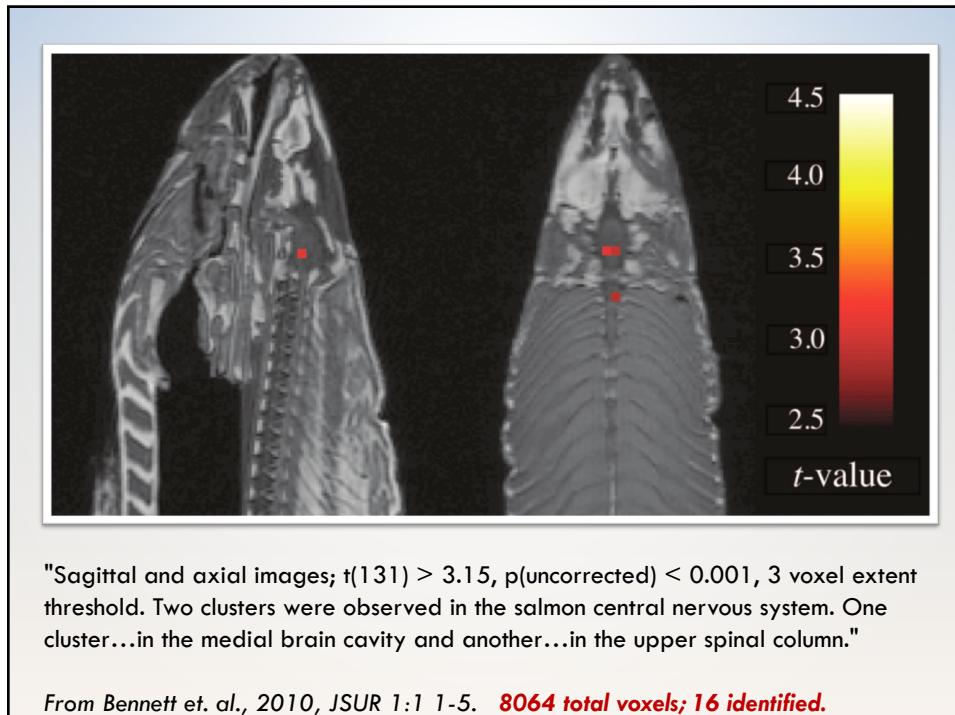
Effect Measures for BOMAMI

	BMC	Control	Total	Null Hypotheses
Gain	16	7	23	OR/RR: (0.9, 1.11)
No Gain	31	36	67	RD: (-0.05, 0.05)
Total	47	43	90	
Risk	0.34	0.16		

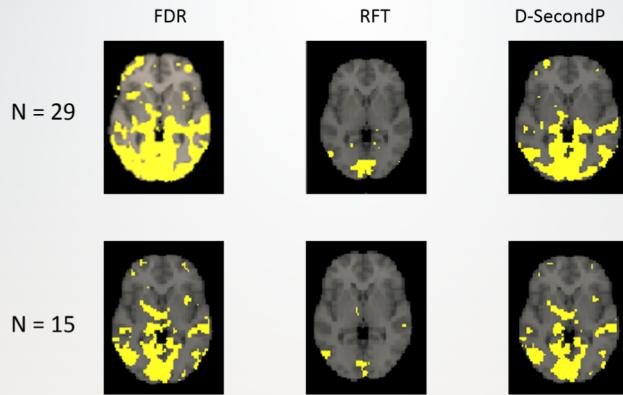
	Estimate	CI Lower	CI Upper	SGPV
Odds Ratio	2.65	0.967	7.286	0.34
Risk Ratio	2.09	0.953	4.589	0.37
Risk Difference	0.18	0.003	0.352	0.24

ASA SGPV Short Course Blume and Murray, 2022





SGPVs For Functional Magnetic Resonance Imaging Data

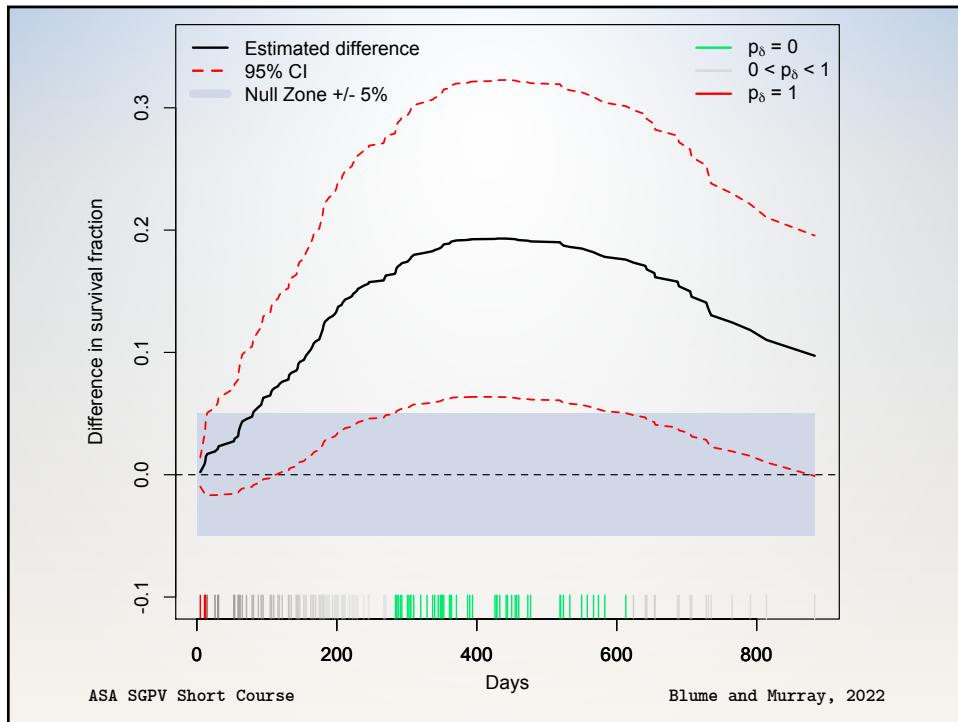
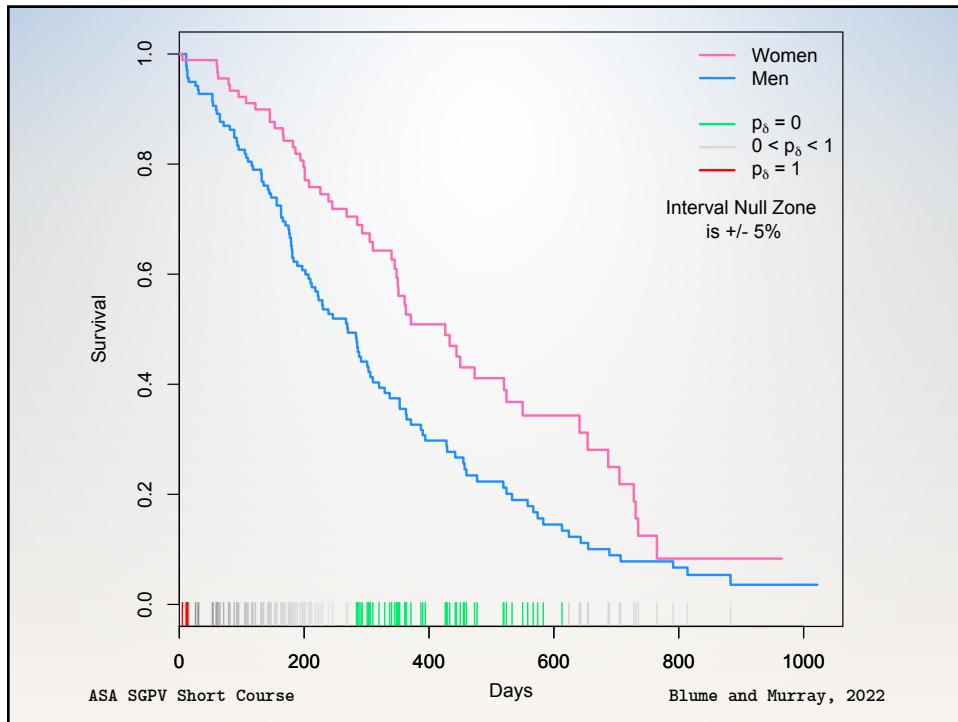


Lisa Lin Dissertation 2021: Data decimation results of activation maps for the 37th axial slice. The yellow blobs indicate activated areas resulted from FDR (first column), RFT (second column) and D-SecondP (third column). The first row corresponds to the activation maps with full sample size at 29.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

Blume and Murray, 2022

Lung Cancer Survival

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



Time for Code Part 1a!

ASA SGPV Short Course

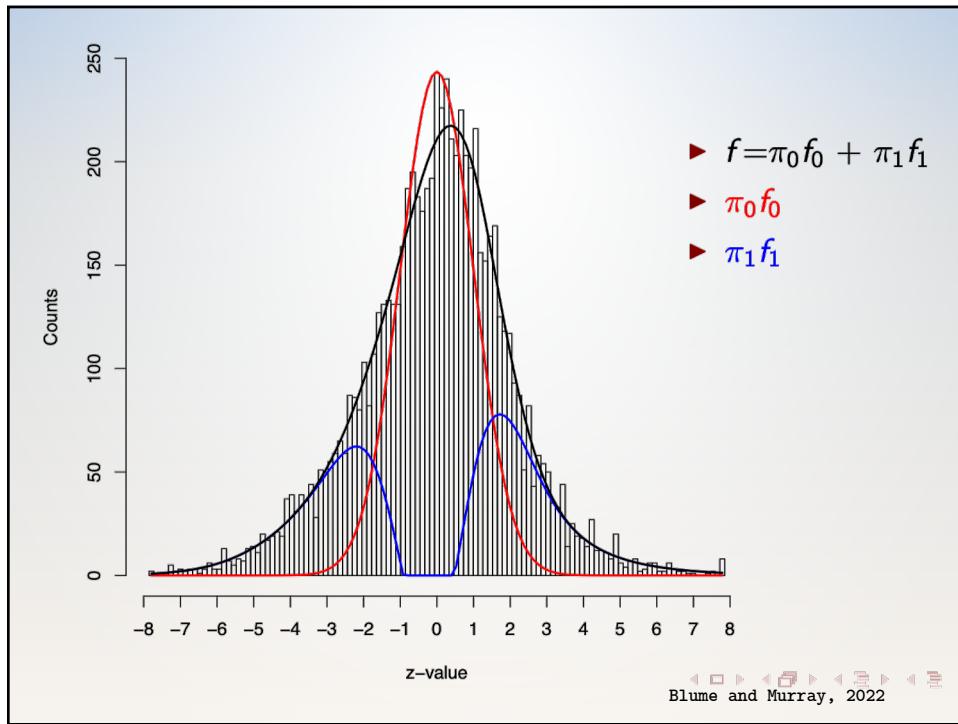
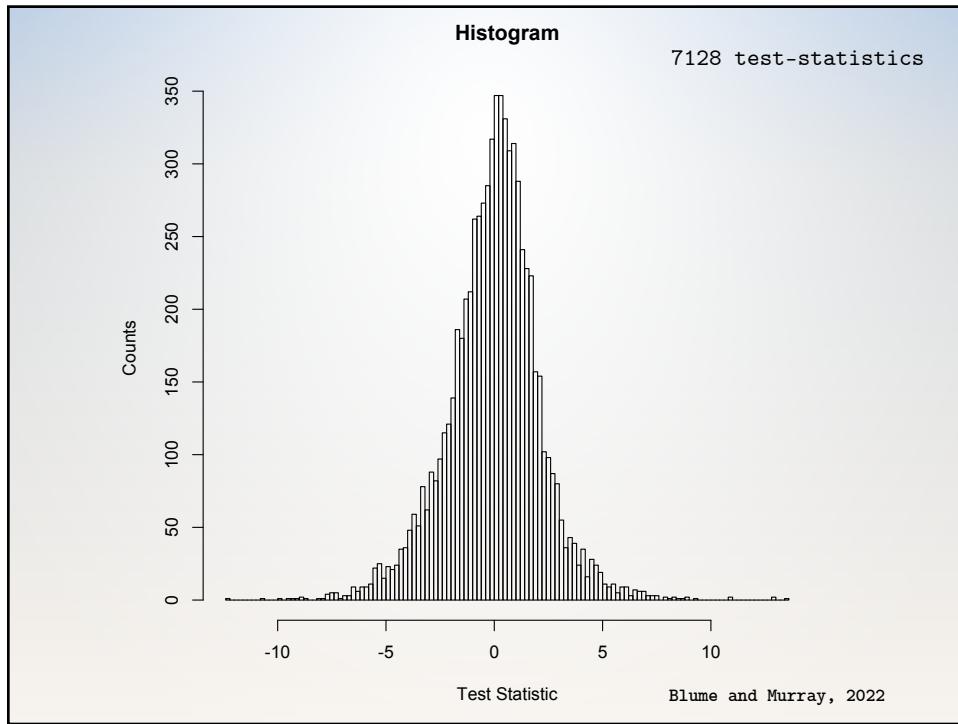
Blume and Murray, 2022

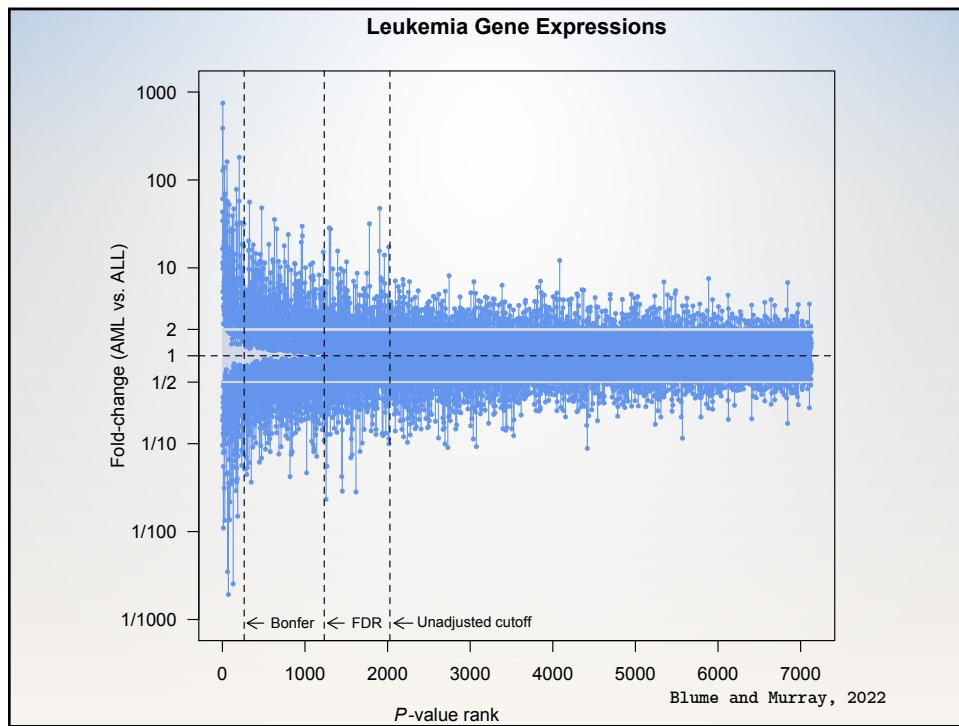
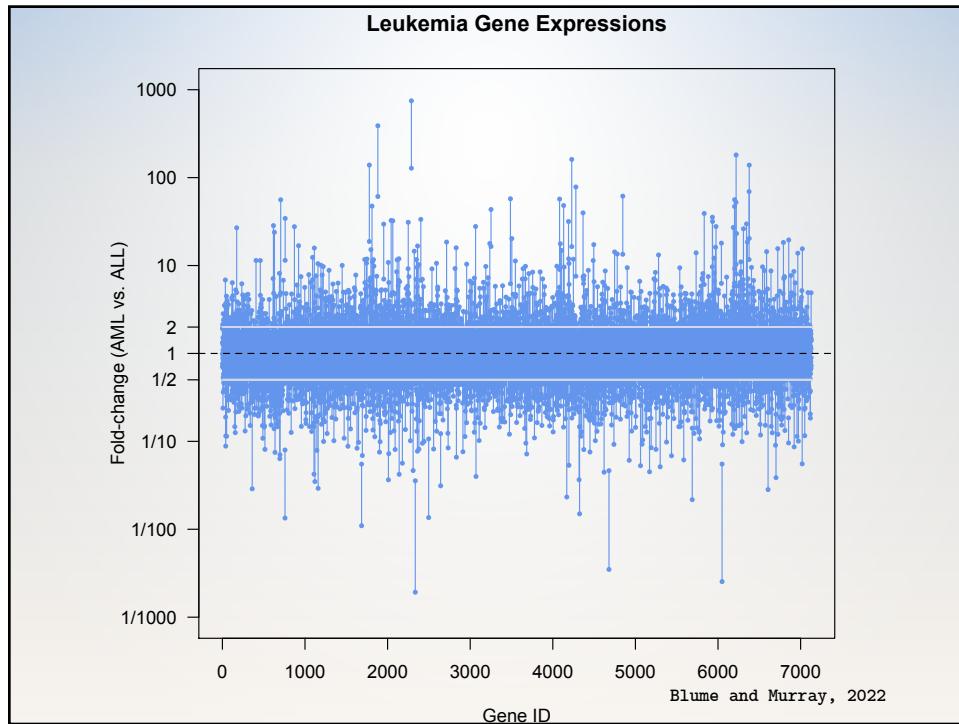
High-Dimensional Data: Leukemia gene expression

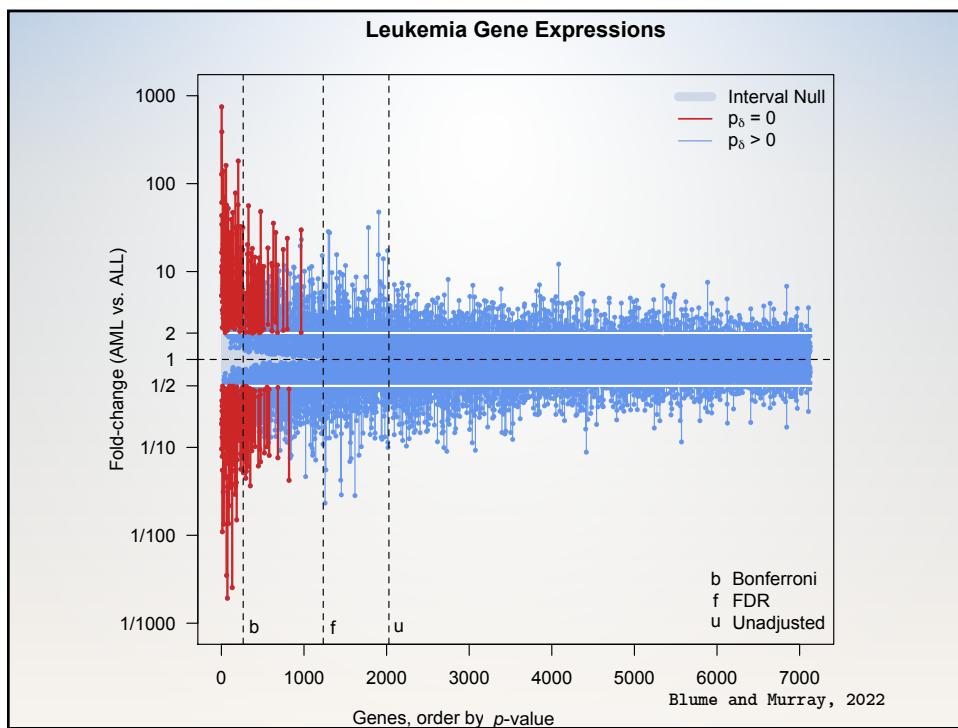
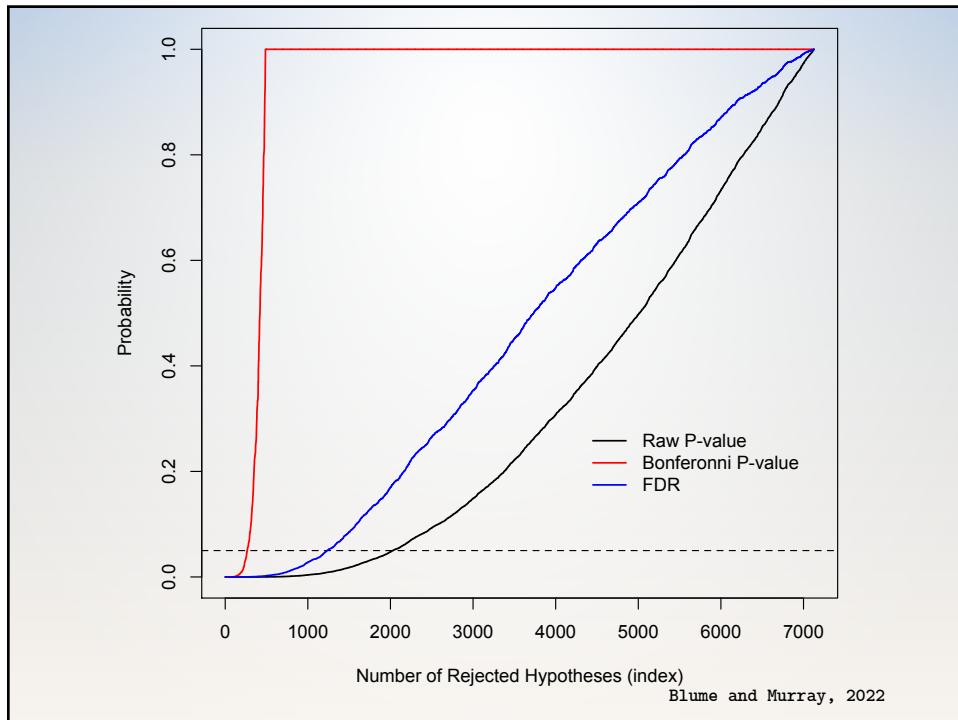
- Classifying acute leukemia by precursors
(Golub 1999, *Science*)
- 7128 genes ; 72 patients (47 ALL and 25 AML)
- Affymetrix chip collected expression levels
- Goal: Identify 'interesting' genes whose expression levels differ between All and AML subjects.
- Looking for fold changes of 2 or more

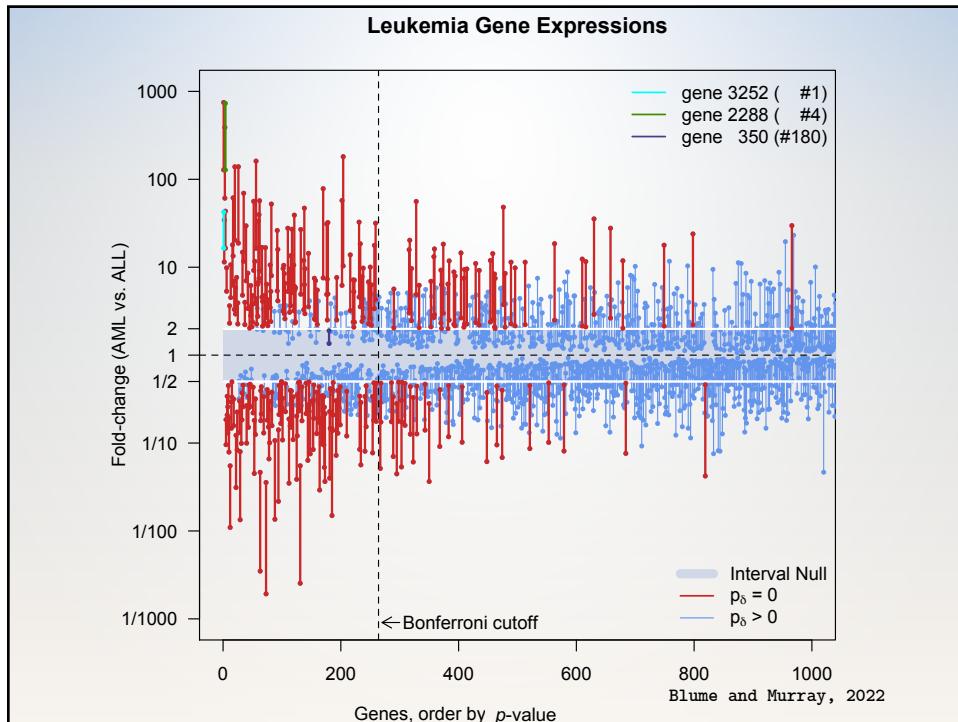
ASA SGPV Short Course

Blume and Murray, 2022









Cross-Tabulation of Leukemia Results

- Bonferroni vs ASA SGpv Short Course

	$1/2 < \text{Fold Change} < 2$ ($\delta = 0.3$)		$1/1.915 < \text{Fold Change} < 1.915$ ($\delta = 0.282$)	
	$p_\delta = 0$	$p_\delta > 0$	$p_\delta = 0$	$p_\delta > 0$
$p_{bon} < 0.05$	164	100	182	82
$p_{bon} > 0.05$	65	6799	82	6782
Total	229	6899	264	6864

Leukemia study findings

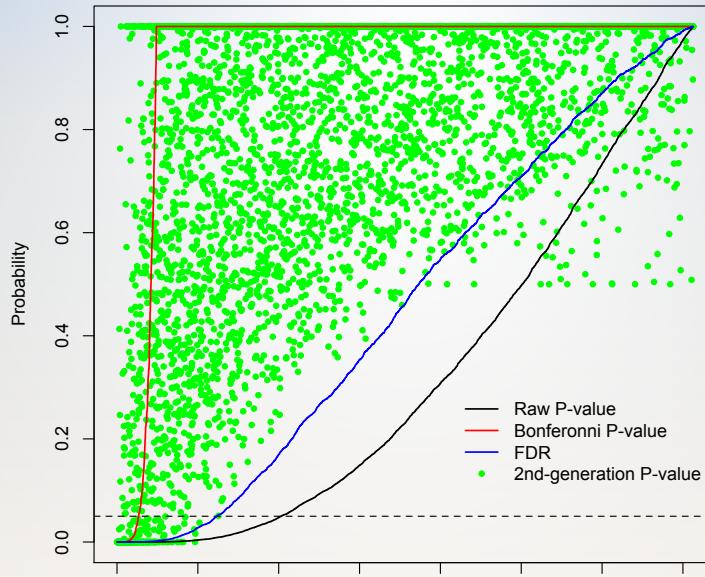
- Findings: Bonferroni 264, SGPV 229
 - Agree on 164 findings
 - Bonferroni +100, SGPV +65
- Effective Type I error rate: 0.037 vs. 0.032
- FDR of 2.45% captures all $p_\delta = 0$, 737 others
- Moving cutoff trades Type I for Type II errors
- SGPV changes the *ranking* of findings
 - Three categories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

Some SGPV findings
have a priori
published validation

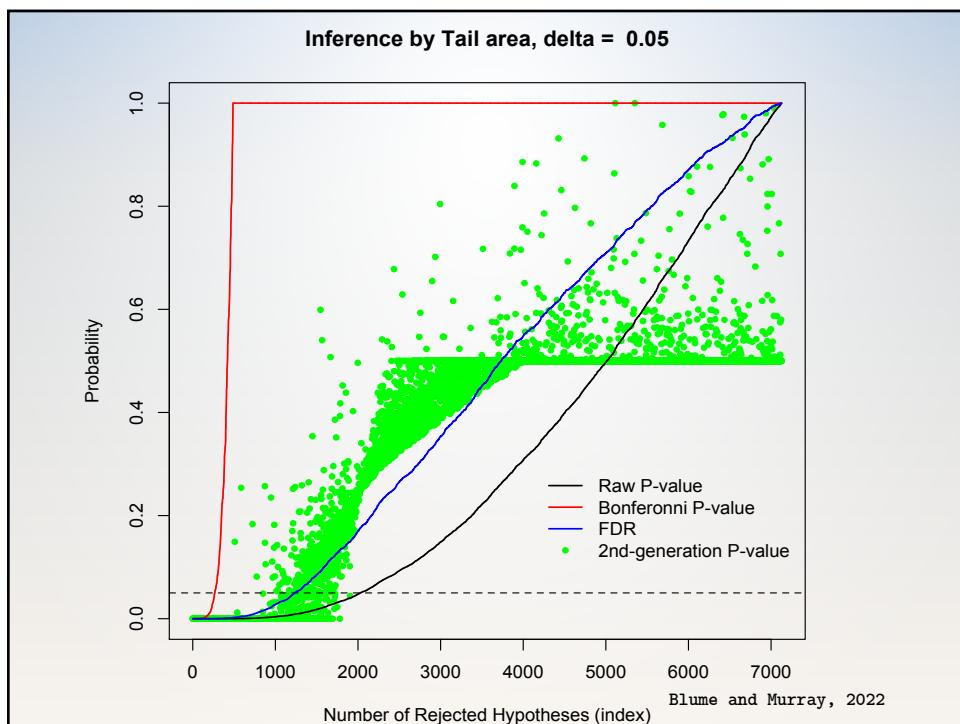
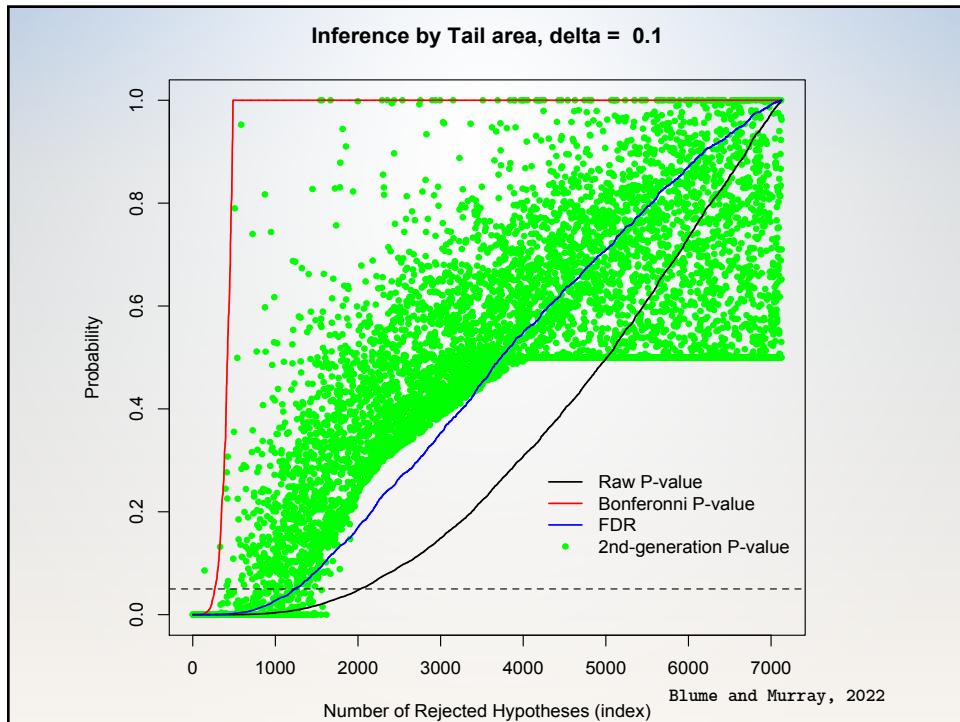
ASA SGPV Short Course

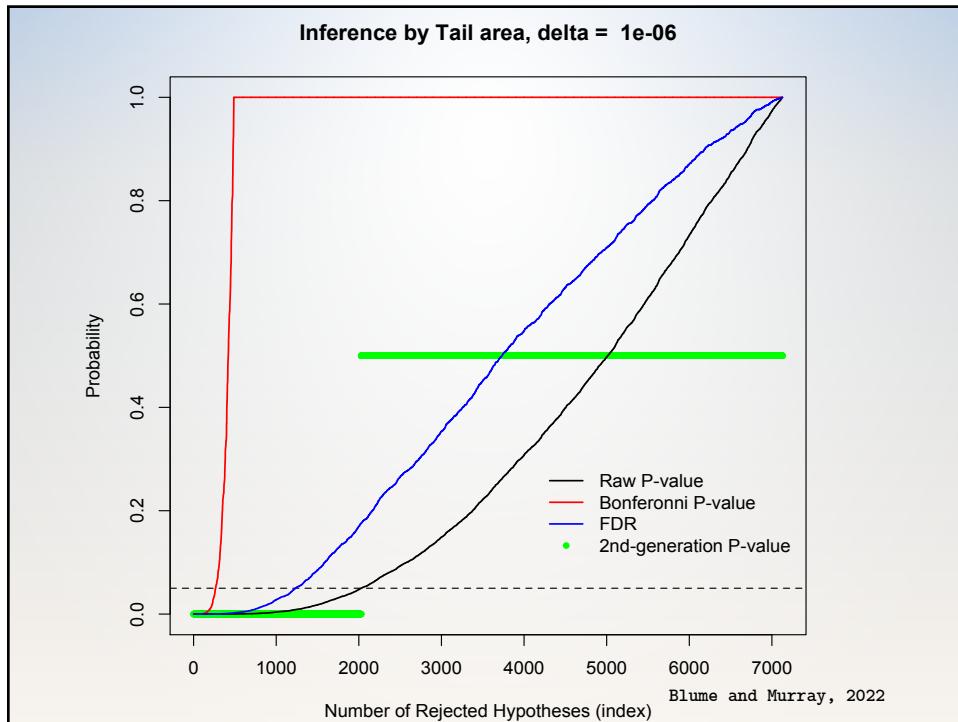
Blume and Murray, 2022

Inference by Tail area, delta = 0.3



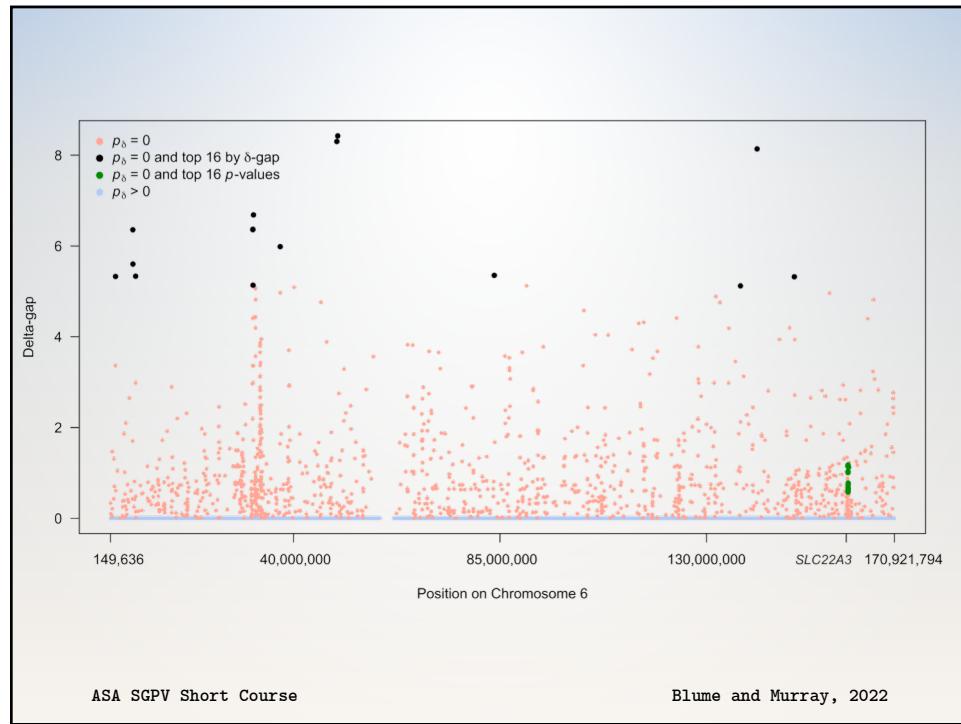
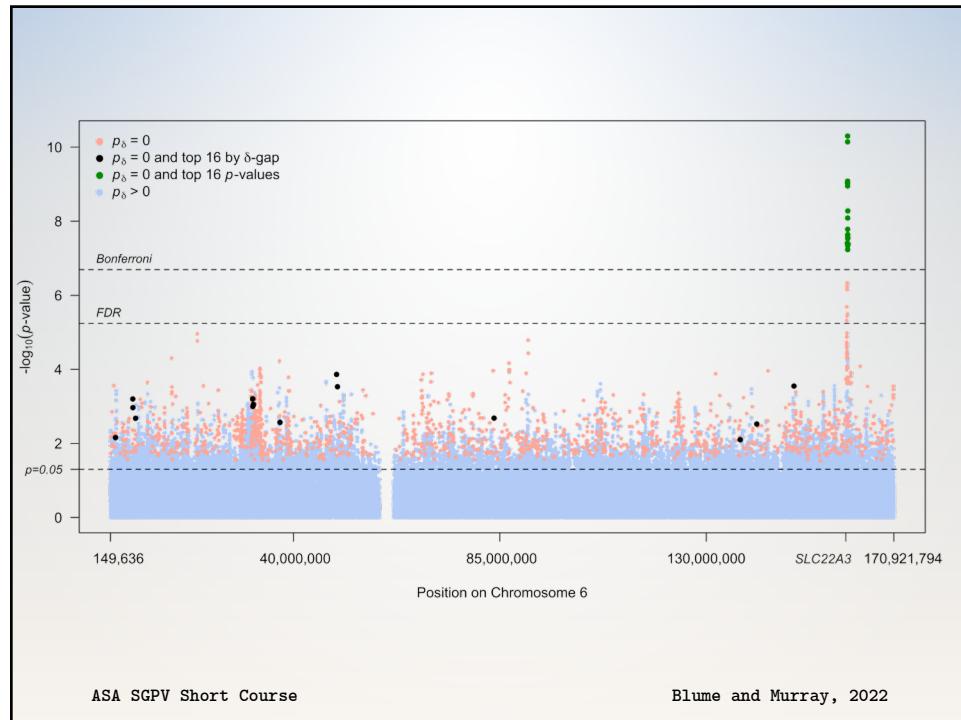
Blume and Murray, 2022

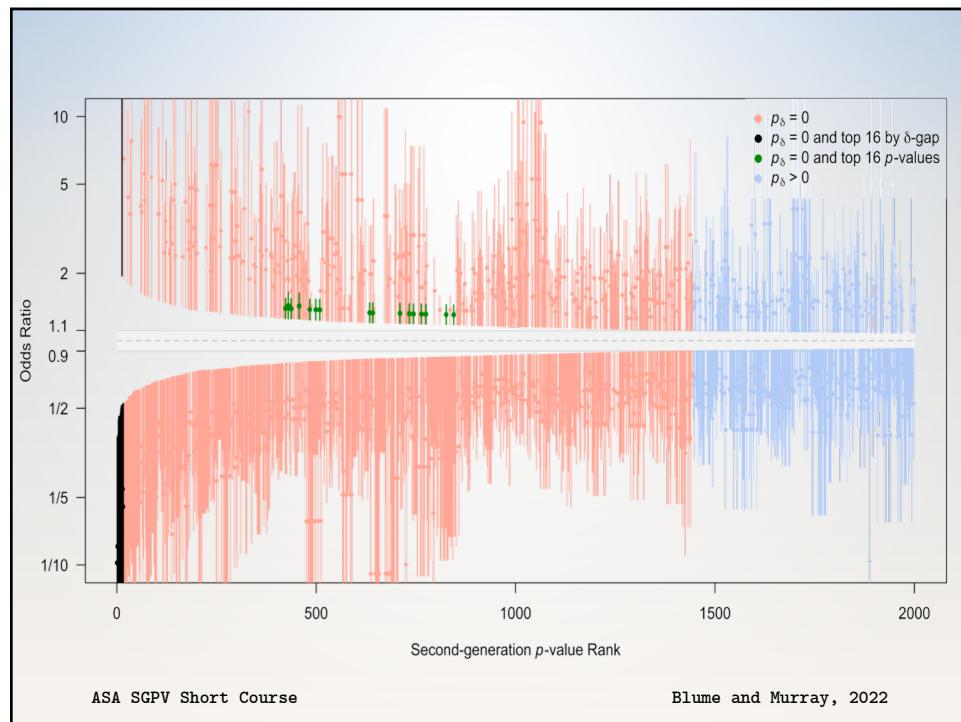
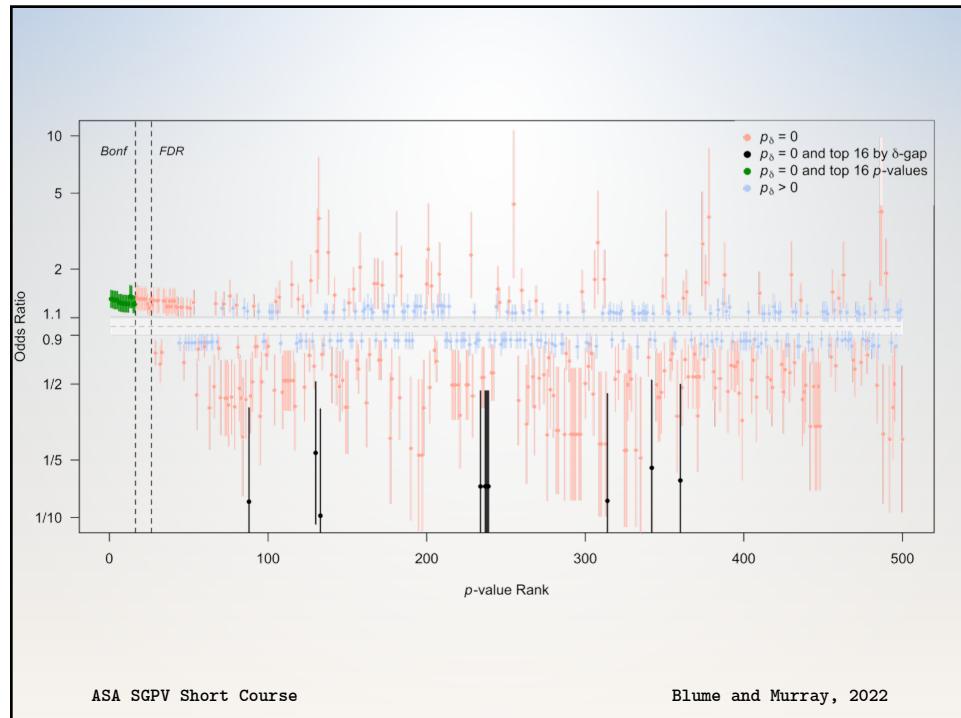


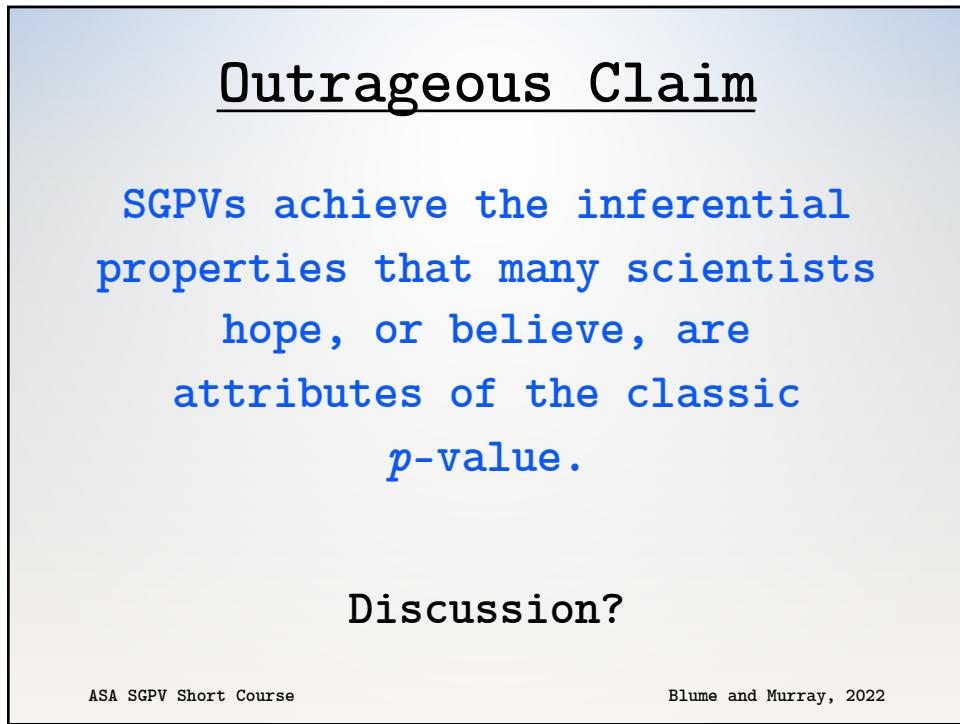
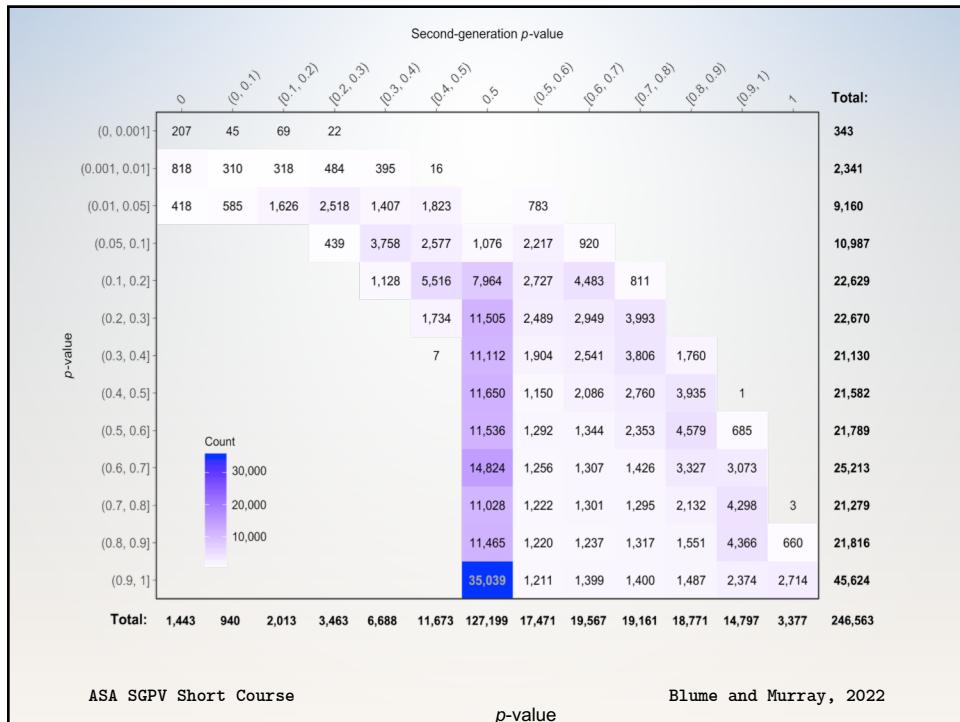


Prostate Cancer SNPs

- International Consortium for Prostate Cancer Genetics (Schaid and Chang 2055; 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







Time for Code Part 1b!