

An introduction to second-generation *p*-values

Jeffrey D. Blume, PhD

School of Data Science
University of Virginia

blume@virginia.edu

www.statisticalevidence.com

Evidential metrics

Example:
Diagnostic Test

1. Measure of the strength evidence

- Axiomatic and intuitive justification
- Summary statistic, yardstick

Positive Test
Negative Test

2. Propensity to collect data that will yield a misleading #1

- Error rates
- Properties of the study design (!)

Sensitivity
Specificity

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 (!)

PPV
NPV

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

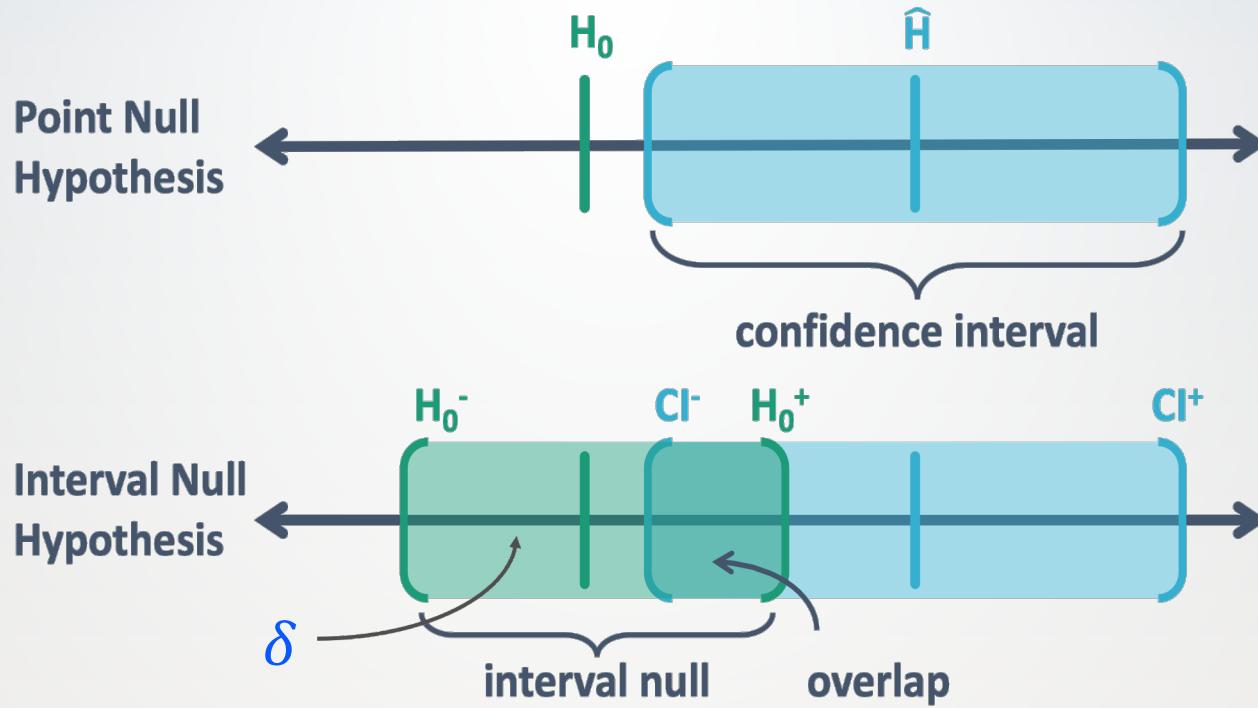
The ^{2nd-generation} ^p-value (what it is)

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

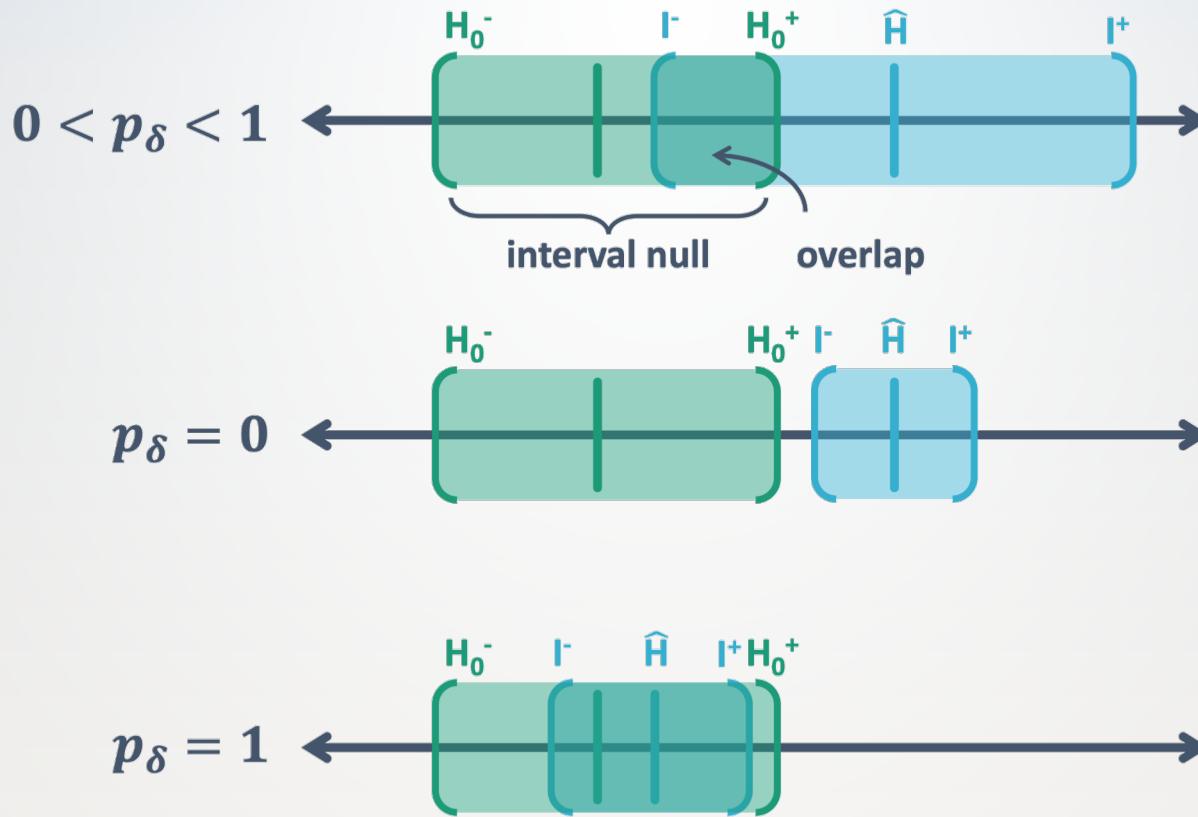
Illustration



Point null hypothesis H_0 and **interval null hypothesis** $[H_0^-, H_0^+]$

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

Illustration



Works with confidence, credible, and support intervals

Steps

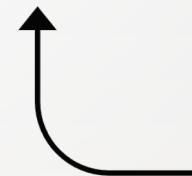
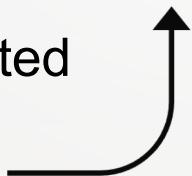
- SGPV ~ the fraction of data-supported hypotheses that are null or practically null
1. Specify an interval null hypothesis or a point null with indifference zone
 2. Find confidence, support or credible interval
 3. Measure the fraction of interval (#2) that is in the null interval
 4. Apply small-sample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

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

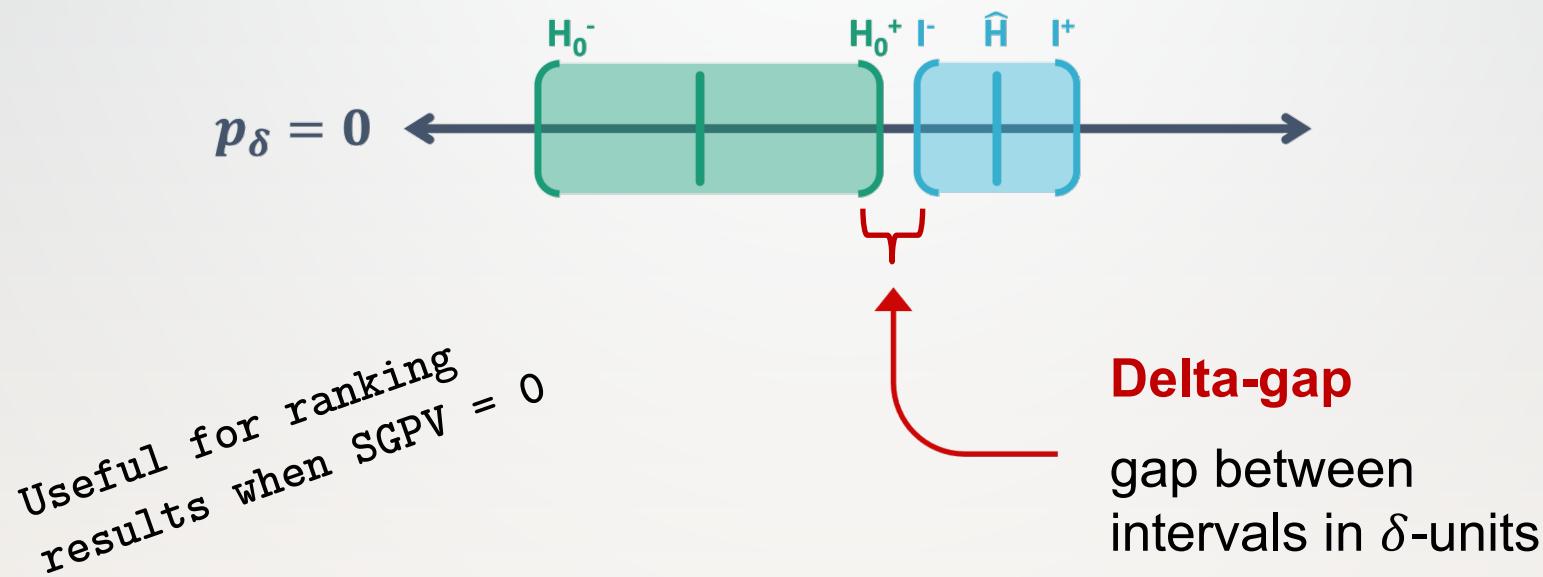


**Small-sample
correction factor**
shrinks proportion
to $\frac{1}{2}$ when $|I|$ wide

when $|I| > 2|H_0|$

The Delta-gap

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



COVID Clinical Trial



BMJ Yale

HOME | SUBMIT | FAQ | BLOG | ALERTS / RSS | ABOUT

Search



Advanced Search

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.



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

Abstract

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).”

COVID Clinical Trial

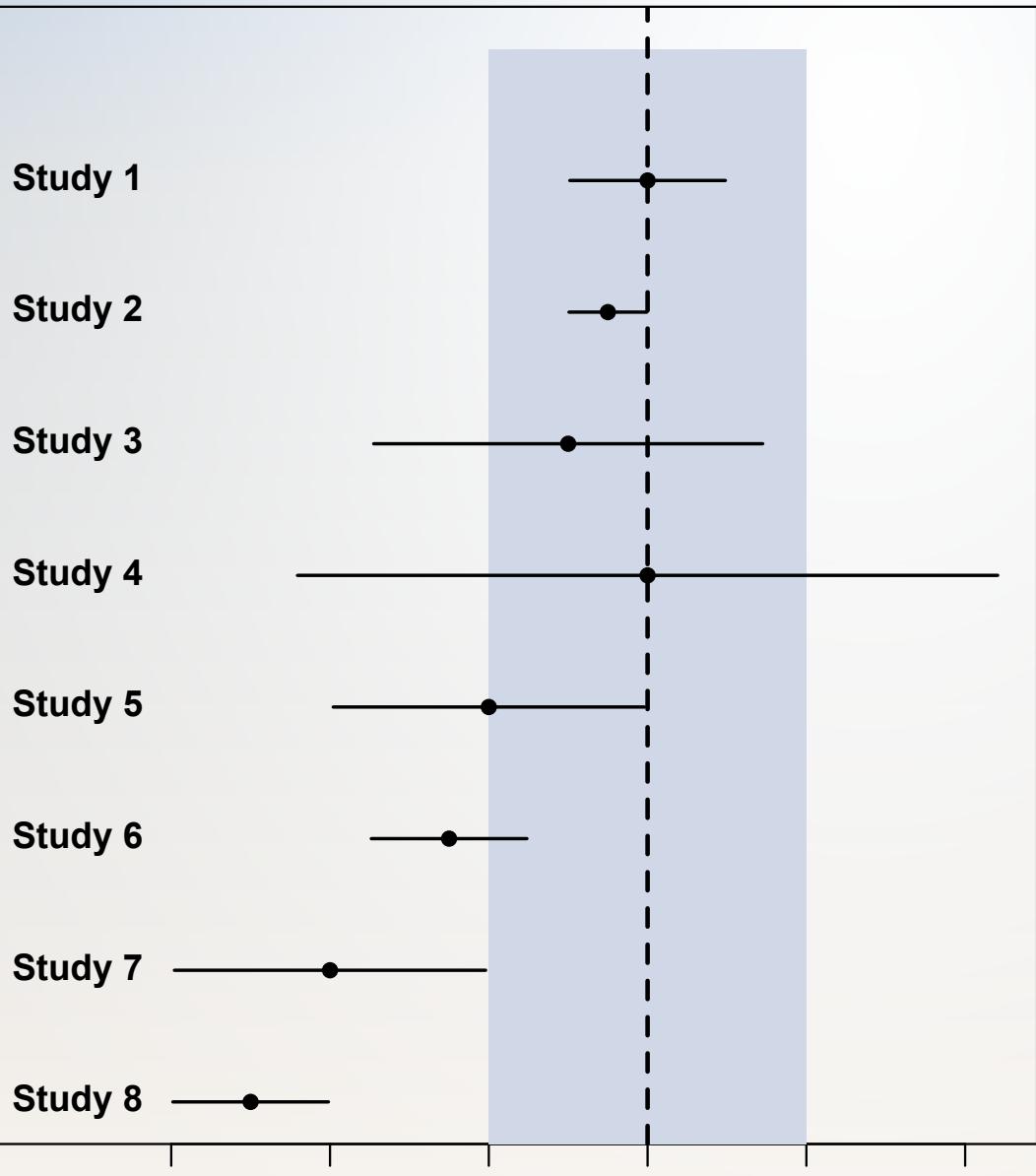
Uncertainty Data Interval: (0.15, 0.82) days

Difference in mean time unwell between ivermectin treatment and placebo.

Hypothesis	Indifference or Null Zone	SGPV (p_δ)	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

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

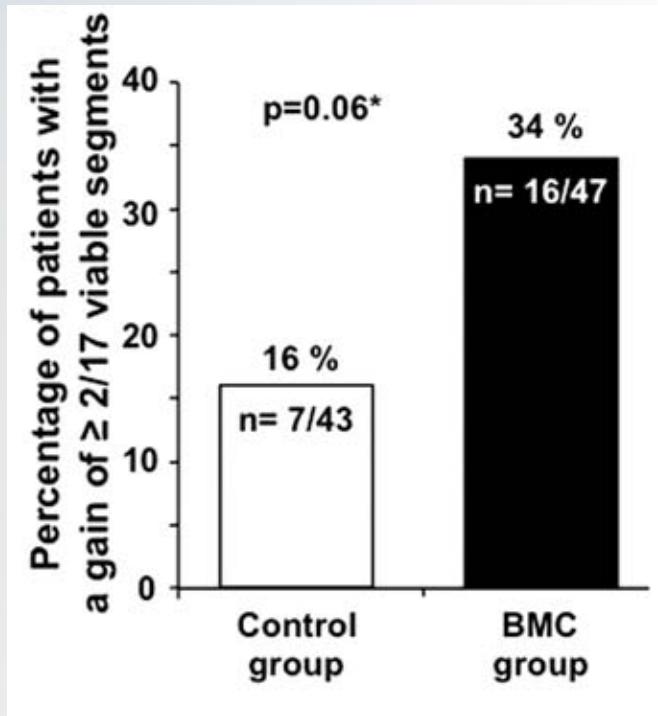


p-value	Max P	2nd Gen P
1	1	1
0.0455	1	1
0.4237	1	0.7041
1	1	0.5
0.0455	1	0.5
<0.0001	0.3173	0.2499
<0.0001	0.0455	0
<0.0001	<0.0001	0

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)

BOMAMI Trial



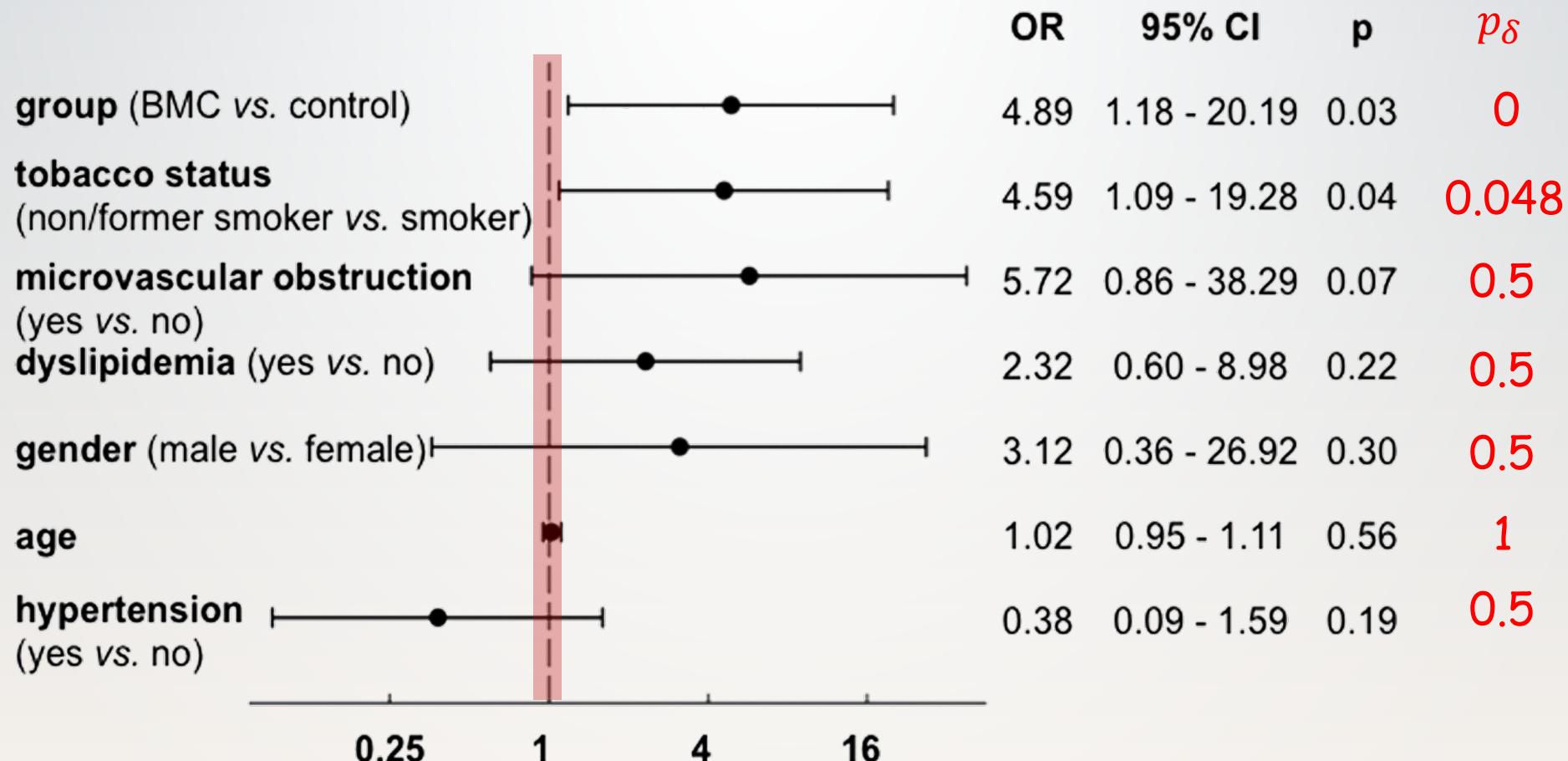
Odds ratios	95% confidence interval	P-value
2.654	0.967 - 7.286	0.06

Null Interval: (0.9, 1.11)

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

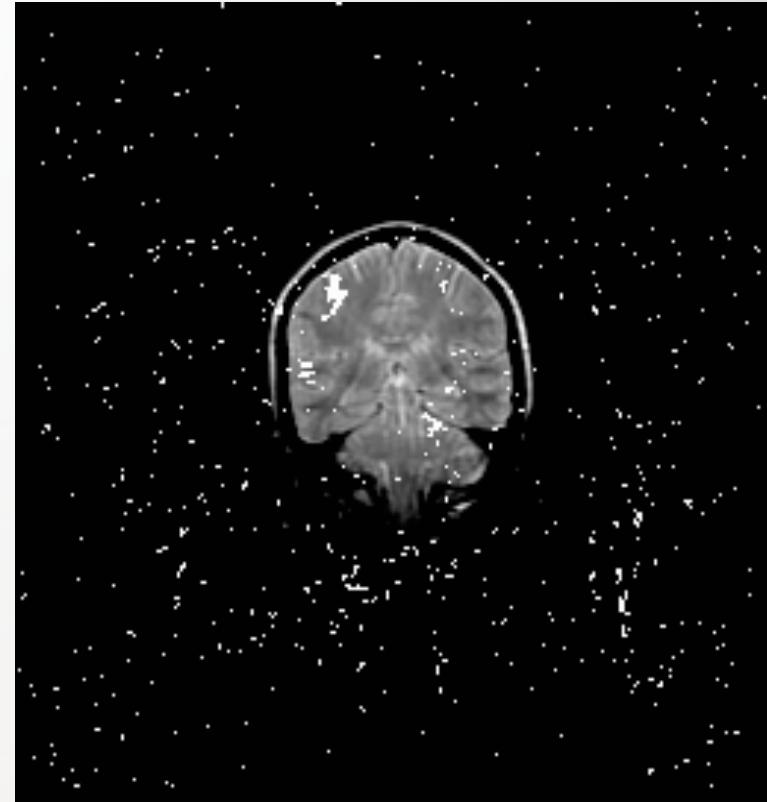
BOMAMI Trial

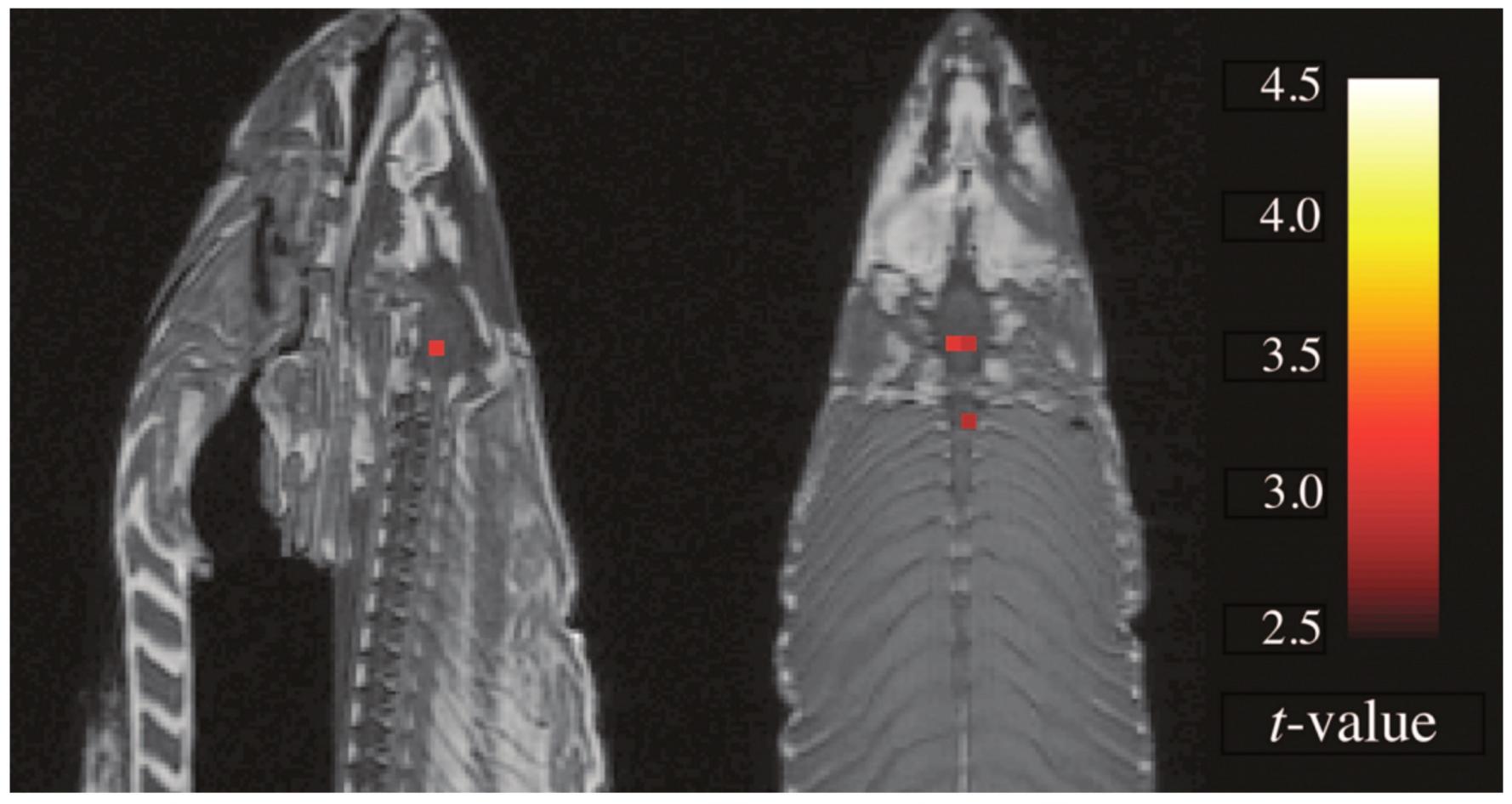
Logistic Regression with Null Zone: (0.9, 1.11)



Setting interval null

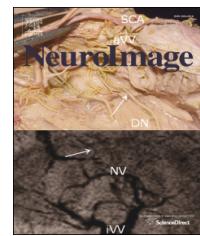
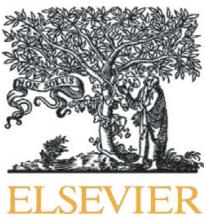
- Before analyzing data (!)
- Measurement error
- Subject matter knowledge
- Impact of findings
- Community standard
- Get creative (fMR example)
- Width not critical, buffer
- *The Atlantic salmon imaging saga*





"Sagittal and axial images; $t(131) > 3.15$, $p(\text{uncorrected}) < 0.001$, 3 voxel extent threshold. Two clusters were observed in the salmon central nervous system. One cluster...in the medial brain cavity and another...in the upper spinal column."

From Bennett et. al., 2010, JSUR 1:1 1-5. **8064 total voxels; 16 identified.**



Simultaneous control of error rates in fMRI data analysis



Hakmook Kang ^{a,b,*}, Jeffrey Blume ^a, Hernando Ombao ^c, David Badre ^d

^a Department of Biostatistics, Vanderbilt University, Nashville, TN 37203, USA

^b Center for Quantitative Sciences, Vanderbilt University, Nashville, TN 37232, USA

^c Department of Statistics, University of California at Irvine, Irvine, CA 92697, USA

^d Department of Cognitive, Linguistic and Psychological Sciences, Brown University, Providence, RI 02912, USA

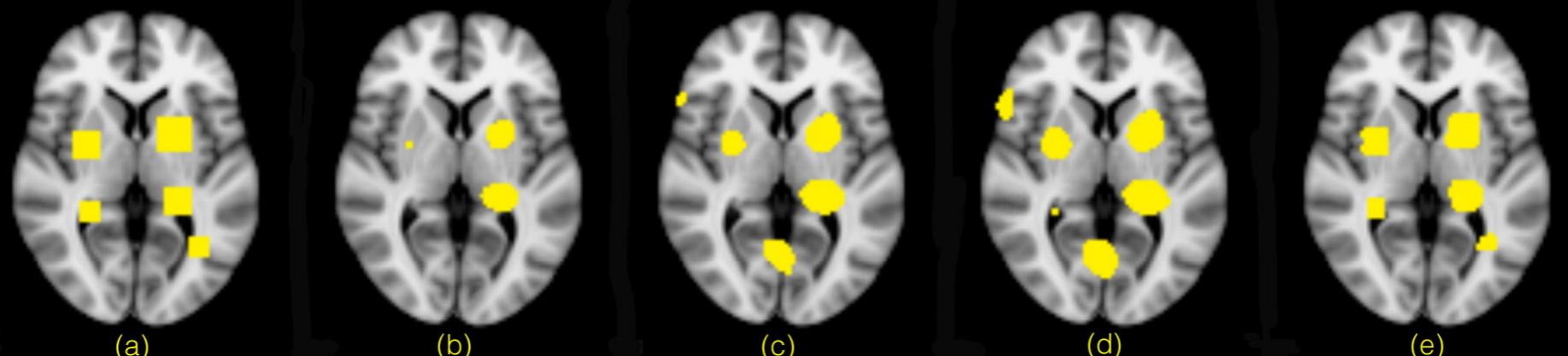
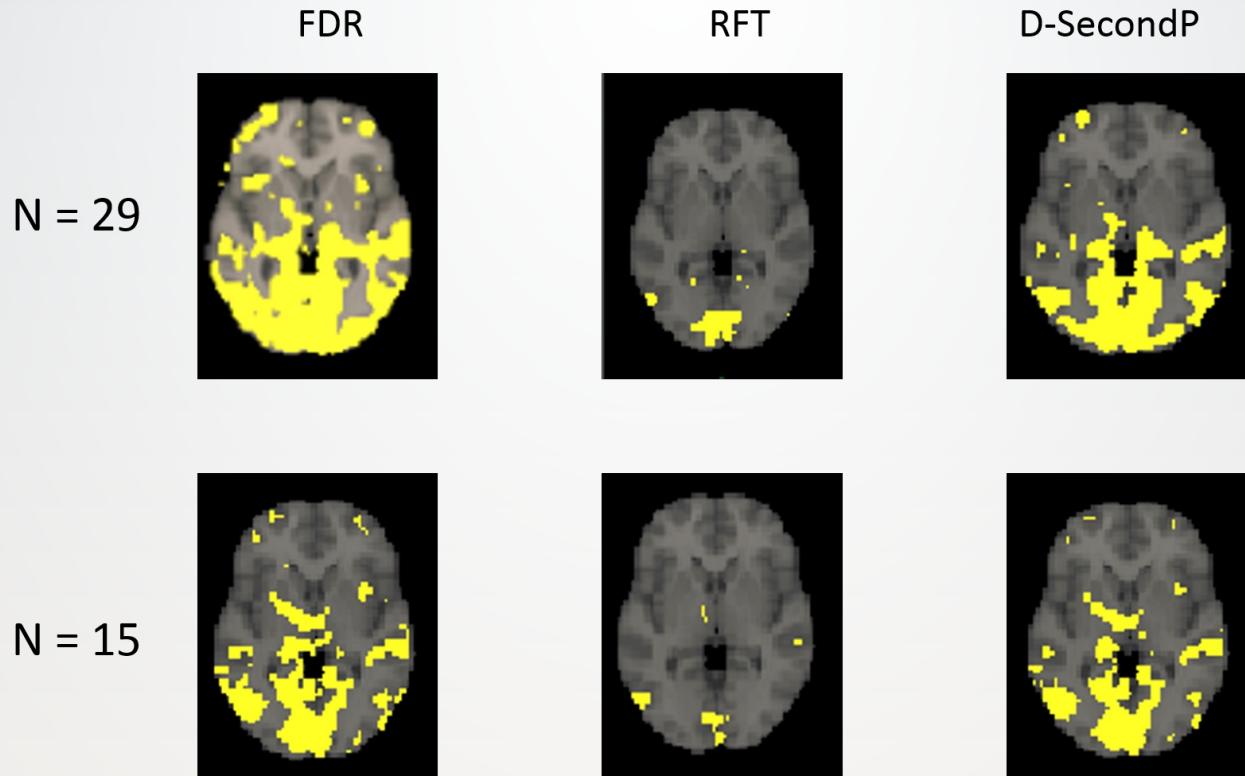


Fig. 1. Simulated voxel analysis for activation brain map. Data were simulated with two boxcar external stimuli, the spatial dimension of 91×109 , and the temporal dimension of $T = 128$. The figures derived as follows: (a) truth (five truly active regions), (b) analysis using RFT, (c) analysis controlling FDR, (d) analysis using BF, (e) analysis using likelihood approach.

SGPVs for Functional MRI Data



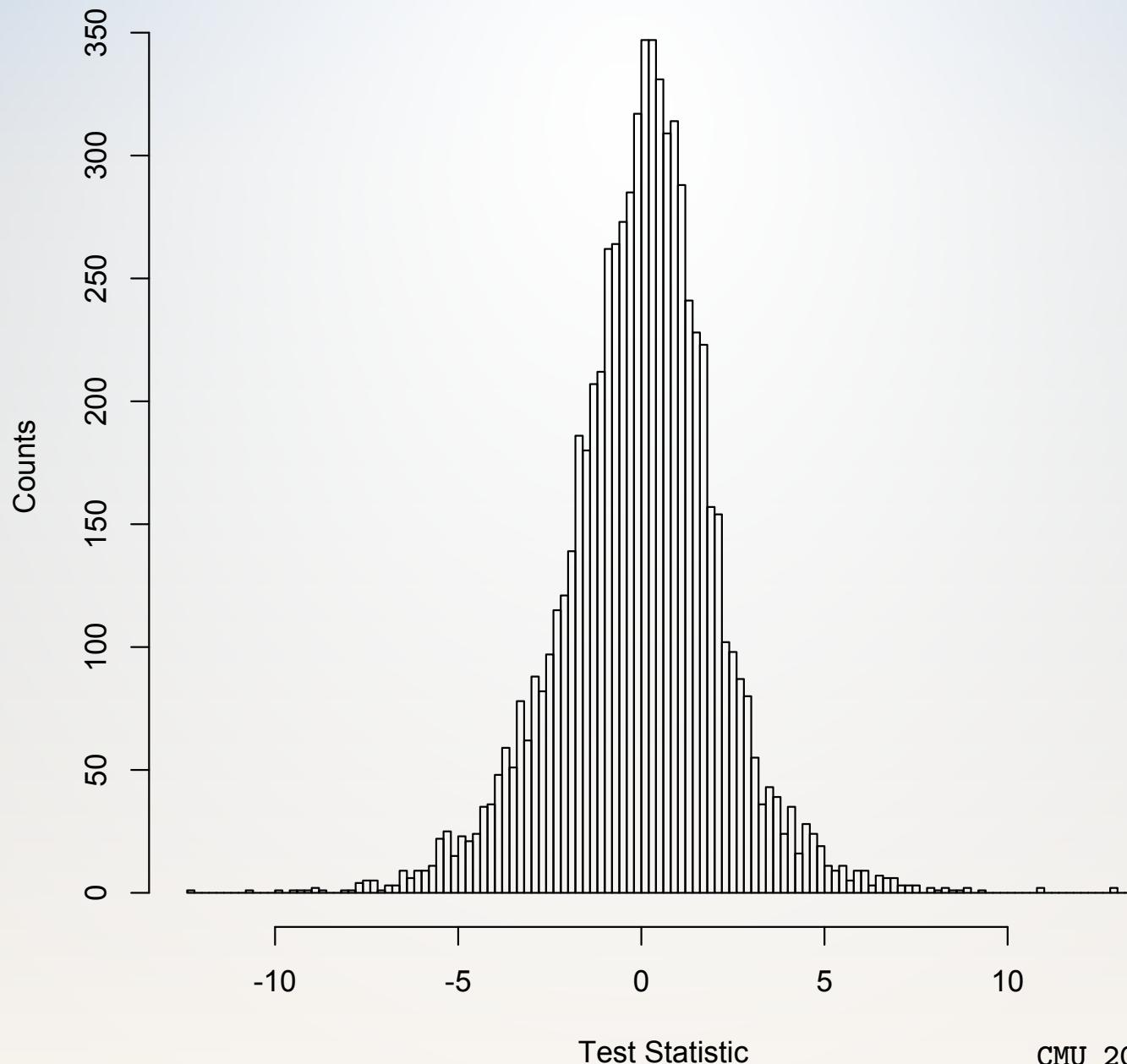
Lisa Lin Dissertation 2021: Data decimation results of activation maps for the 37th axial slice. The yellow blobs indicate activated areas resulted from FDR (first column), RFT (second column) and D-SecondP (third column). The first row corresponds to the activation maps with full sample size at 29.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

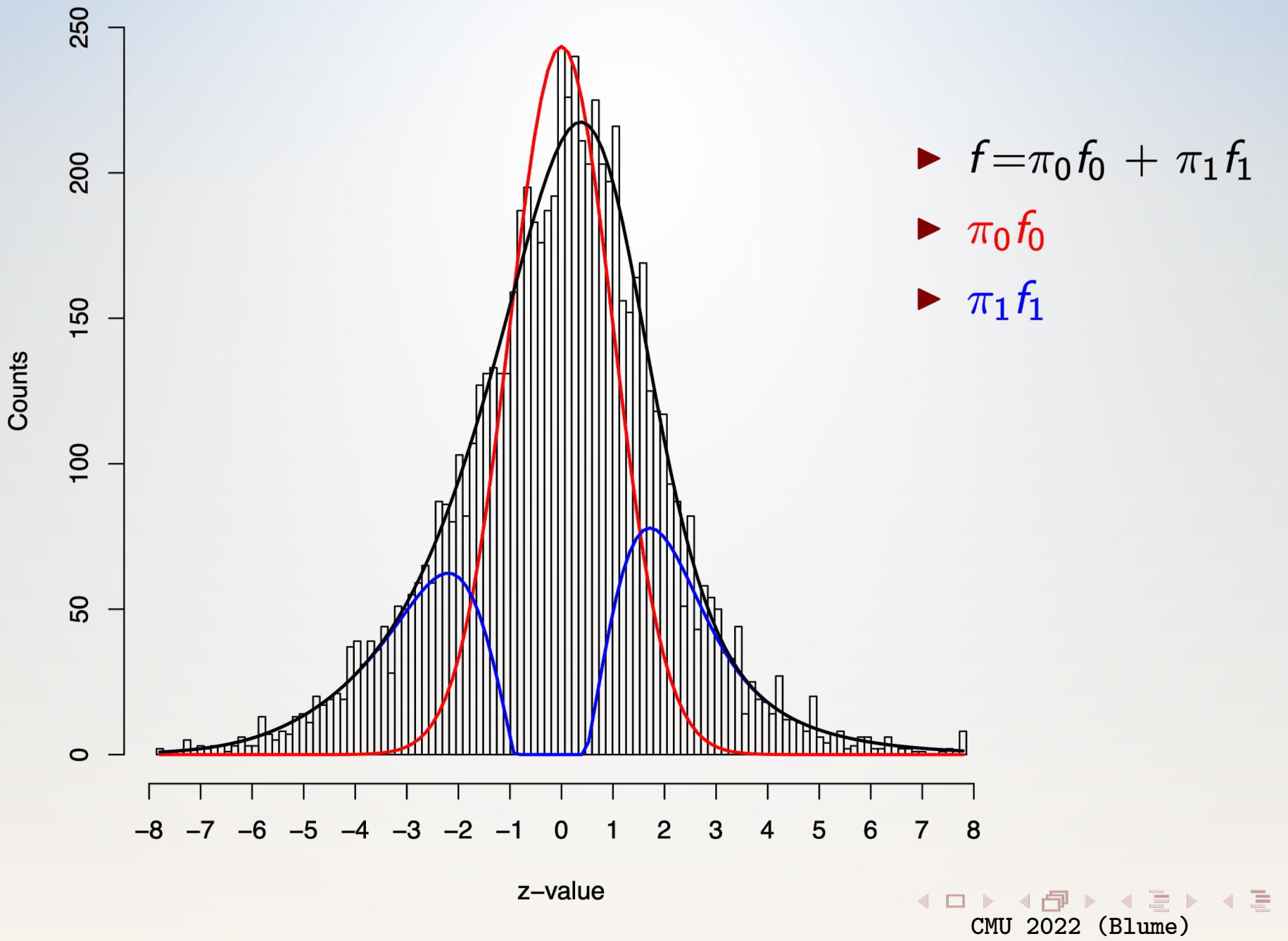
High-Dimensional Data: Leukemia gene expression

- 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

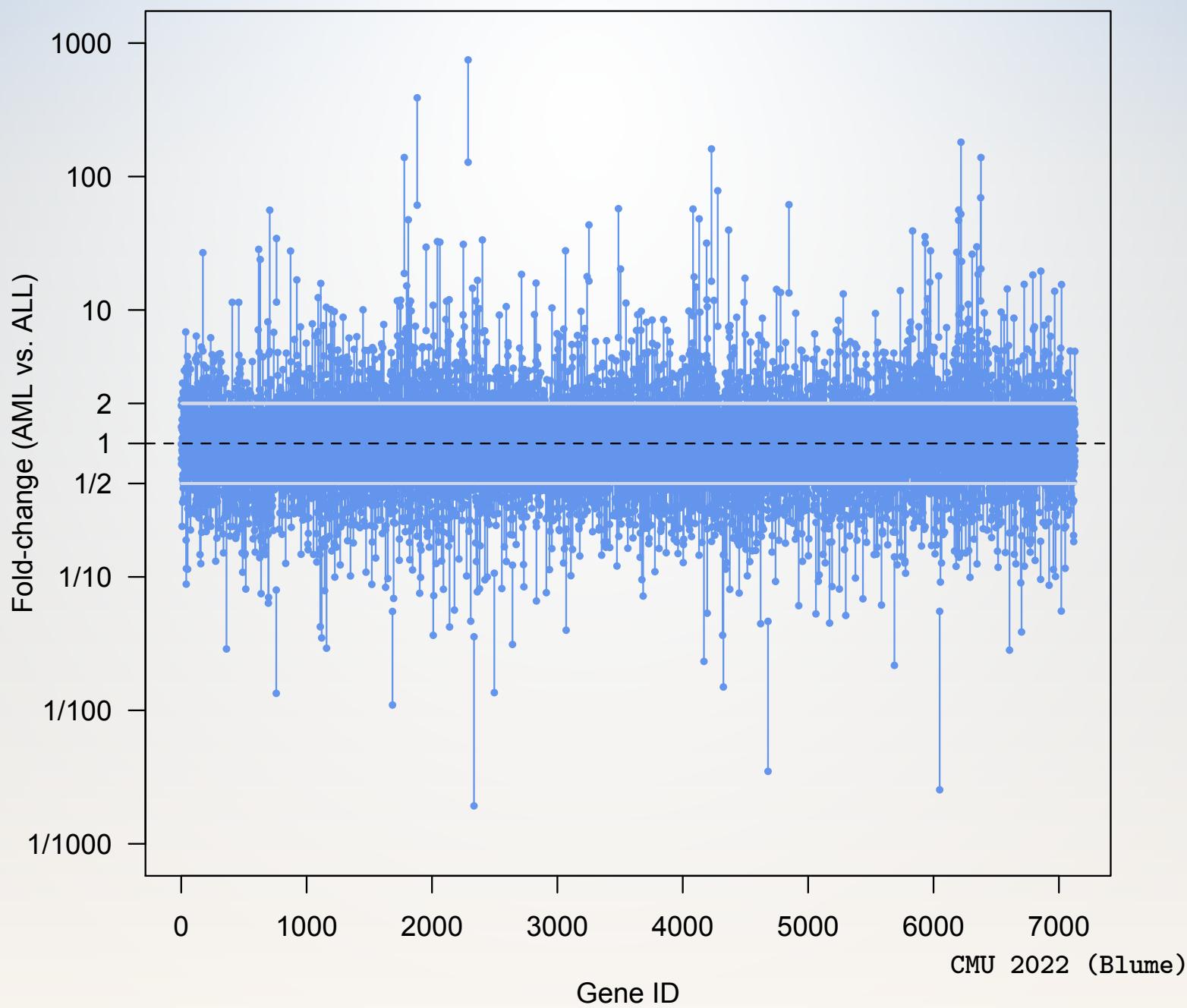
Histogram

7128 test-statistics

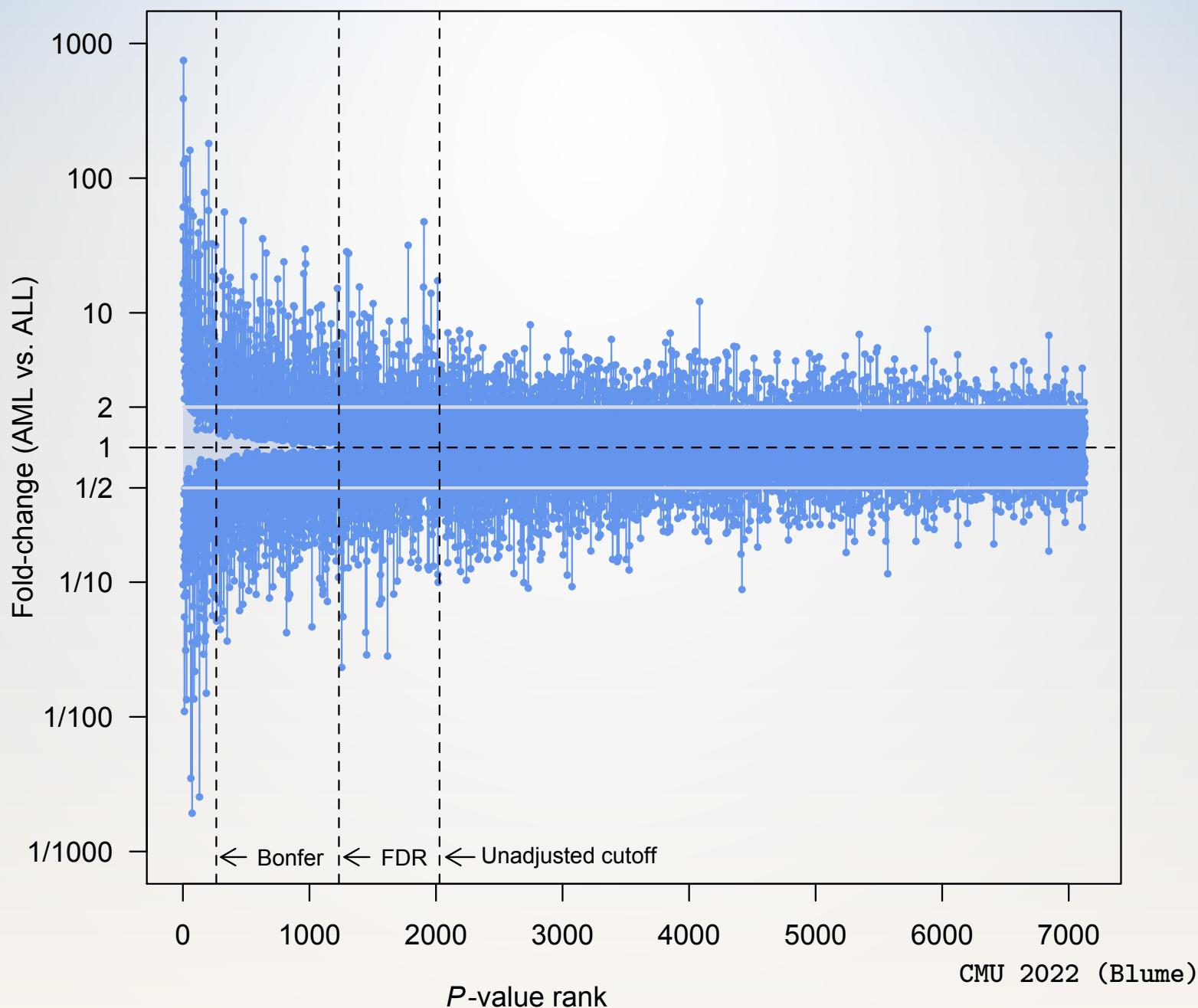




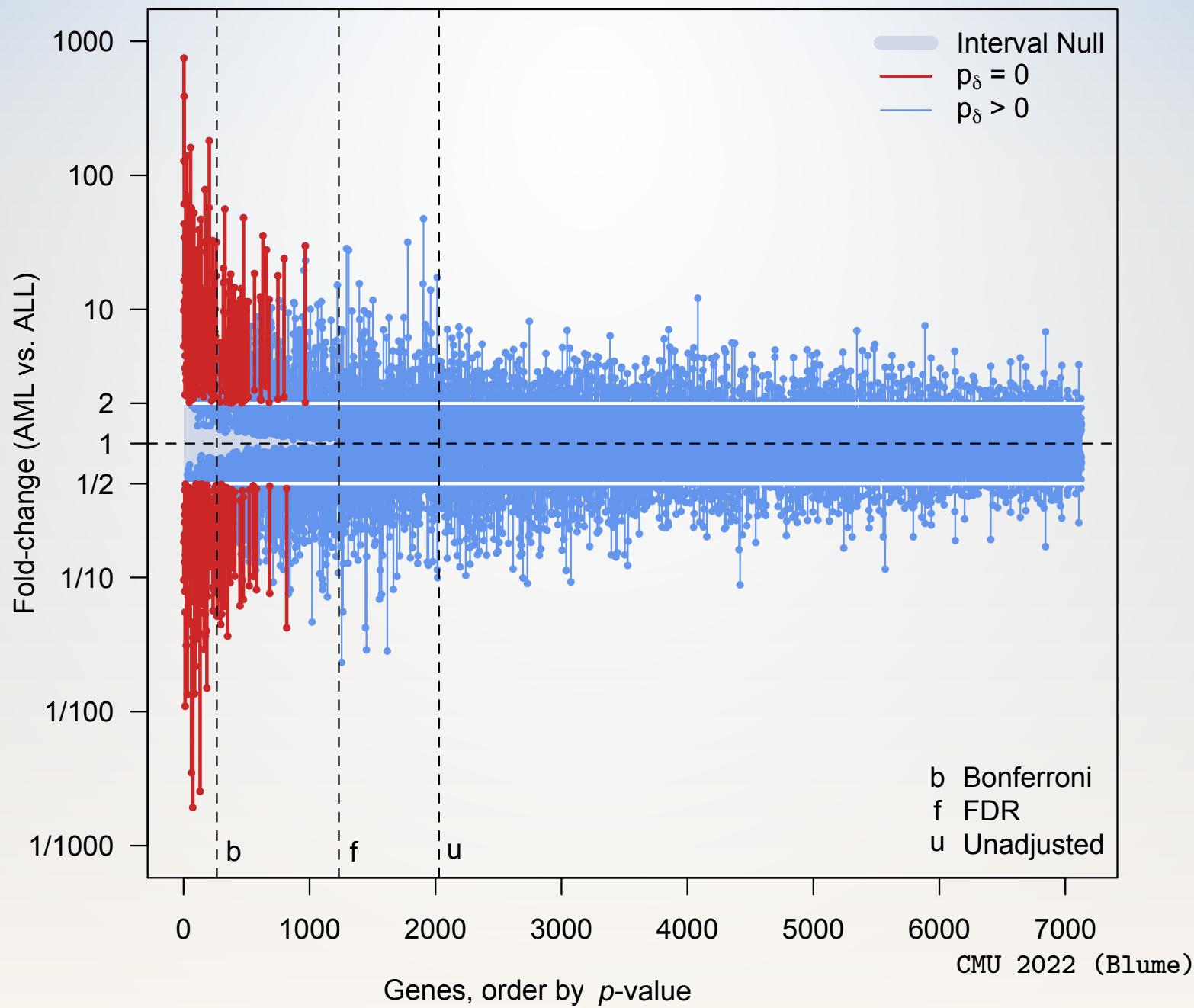
Leukemia Gene Expressions



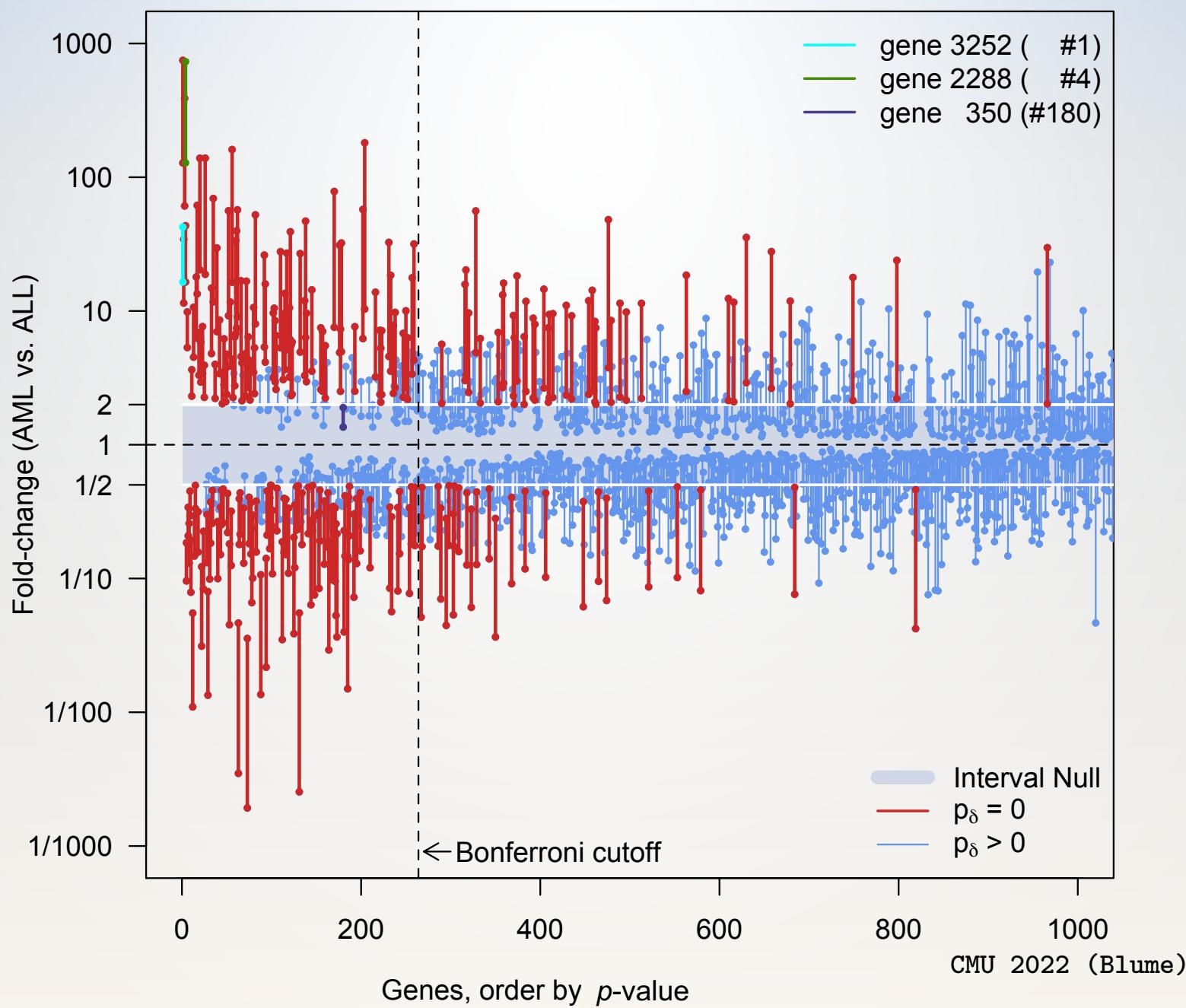
Leukemia Gene Expressions



Leukemia Gene Expressions



Leukemia Gene Expressions



Cross-Tabulation of Leukemia Results

- Bonferroni vs Second Generation p -values

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

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

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**)

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

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

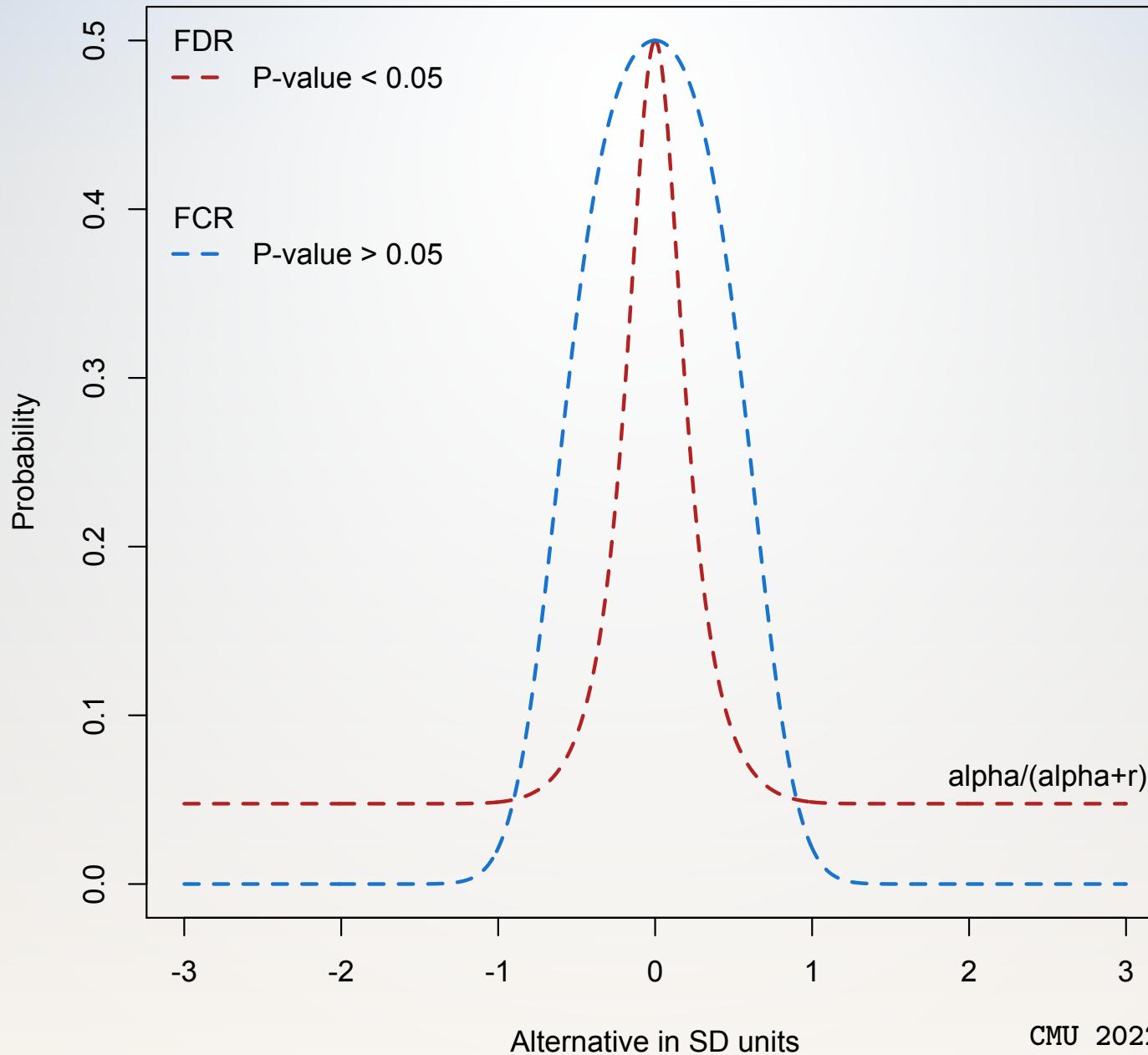
Error rates



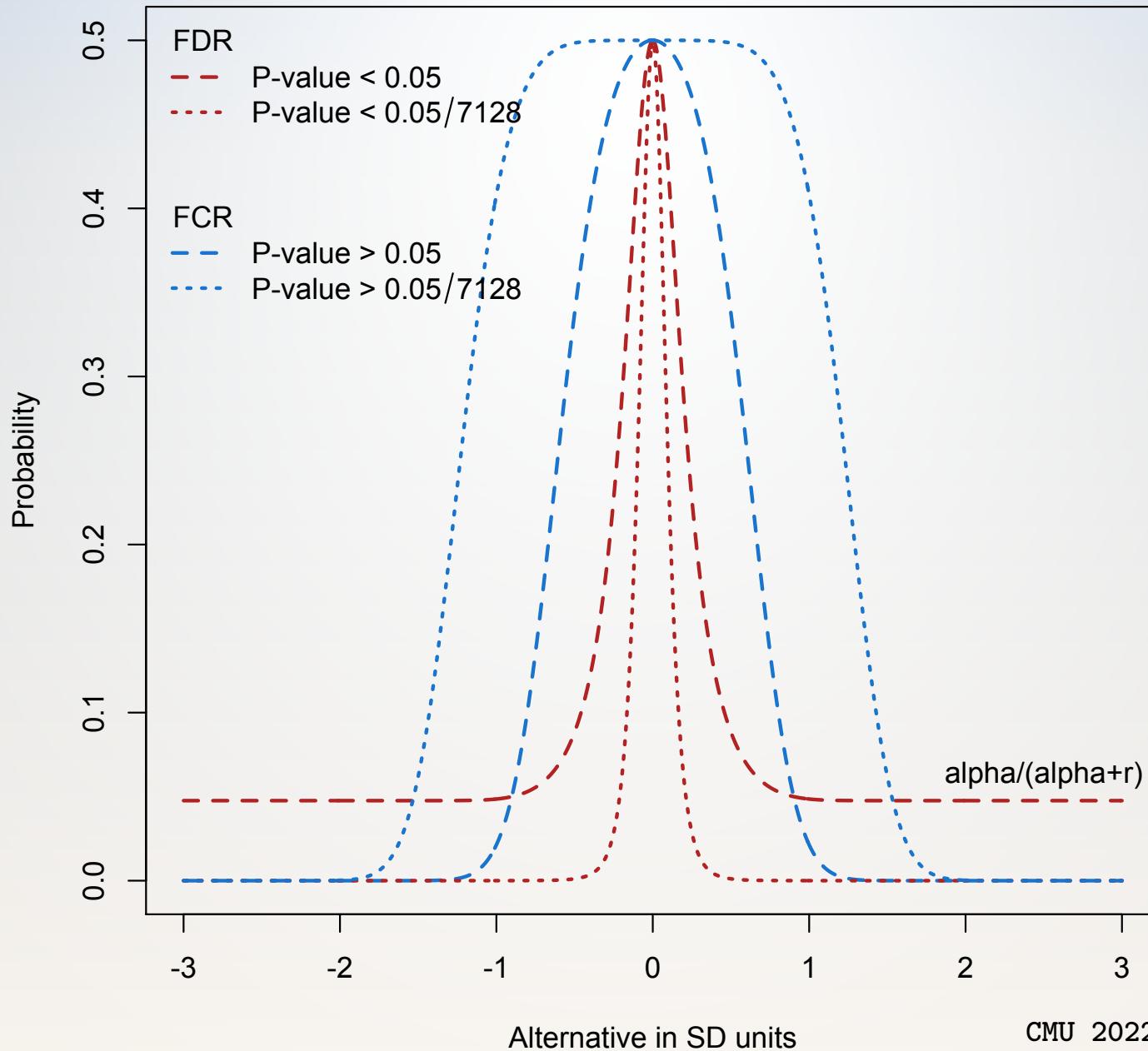
Error rates



False discovery and confirmation rates



False discovery and confirmation rates



False discovery rates

- Second-generation p -values
- False Discovery Rate (**FDR**)



Error Rates

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

- False Confirmation Rate (**FCR**)

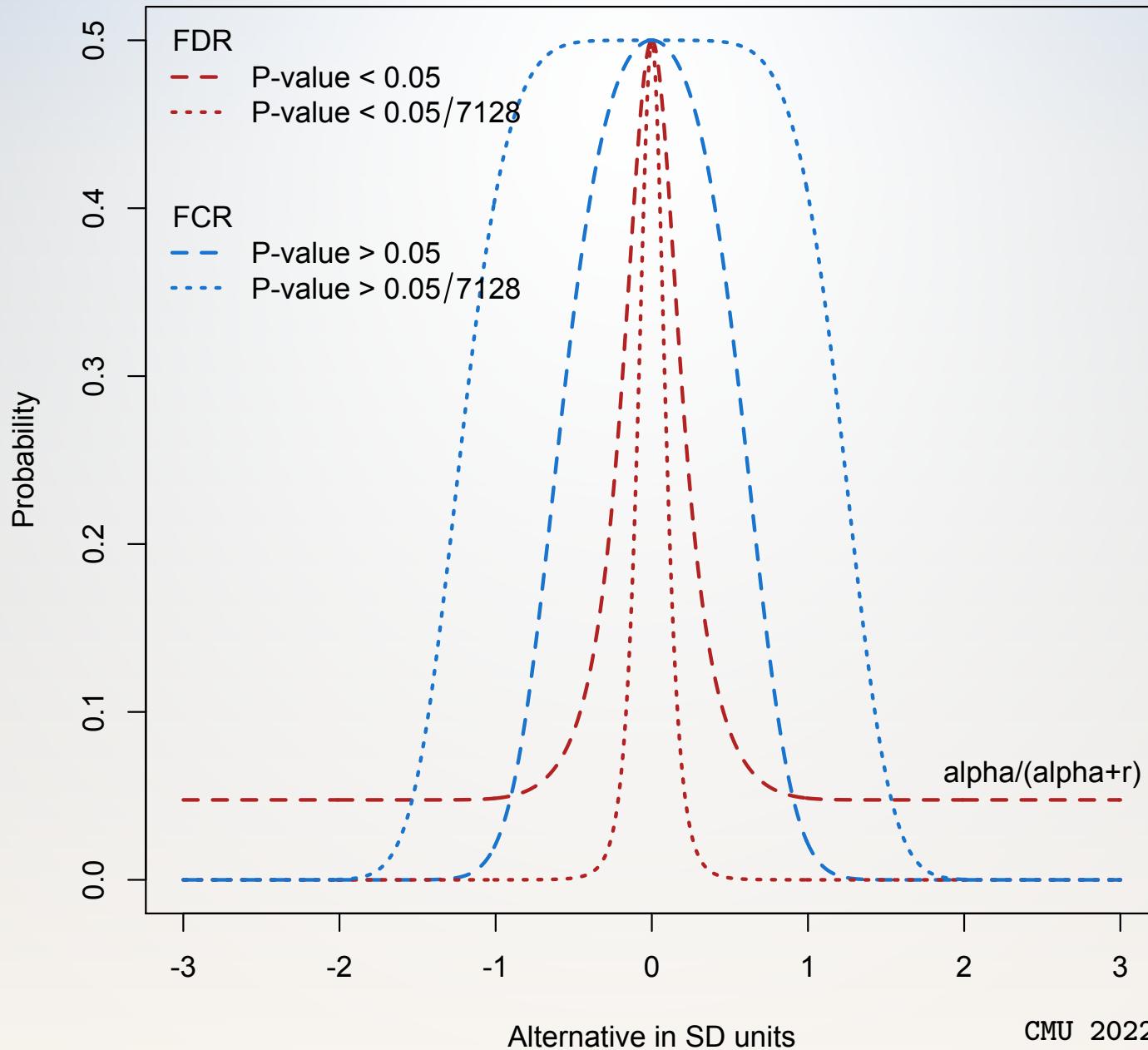


Error Rates

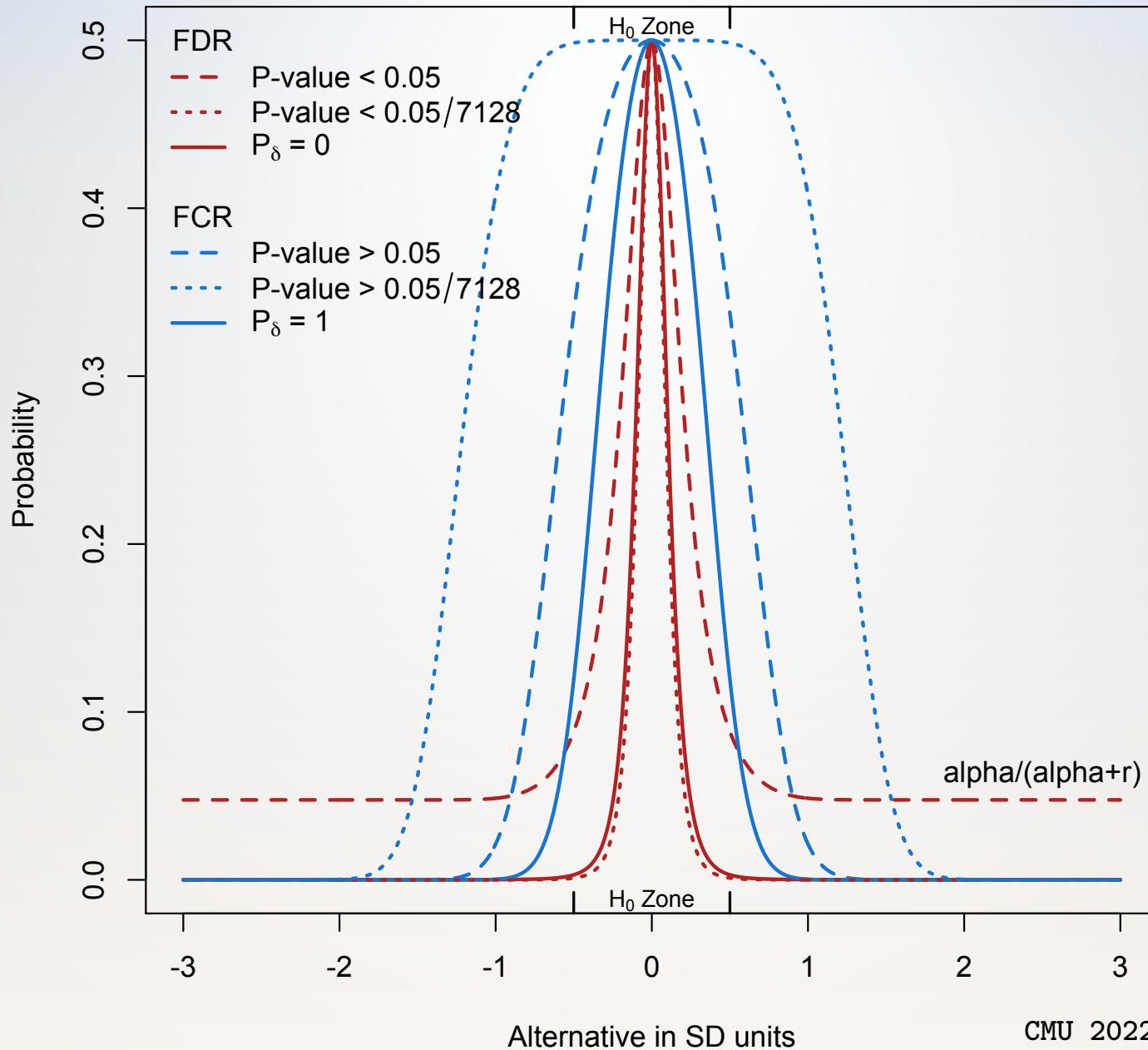
$$r = P(H_1)/P(H_0)$$

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

False discovery and confirmation rates



False discovery and confirmation rates



Beyond Simple Examples

- Connection to shrinkage and bayes methods
- Connection with equivalence testing
- Study planning and ‘Optimal’ null interval shrinkage
- False discovery rates for SGPVs
- Shortcourse Github with examples and code
 - <http://www.github.com/murraymegan/SGPV-ASA-Short-Course-2022>
- Variable Selection with SGPVs
 - Papers (Zuo and Blume)
 - <https://www.tandfonline.com/doi/full/10.1080/00031305.2021.1946150>
 - f1000research.com/articles/11-58
 - ProSGPV GitHub
 - <https://github.com/zuoyi93/ProSGPV>
 - <https://cran.r-project.org/web/packages/ProSGPV/vignettes/linear-vignette.html>
 - <https://cran.r-project.org/web/packages/ProSGPV/vignettes/glm-cox-vignette.html>