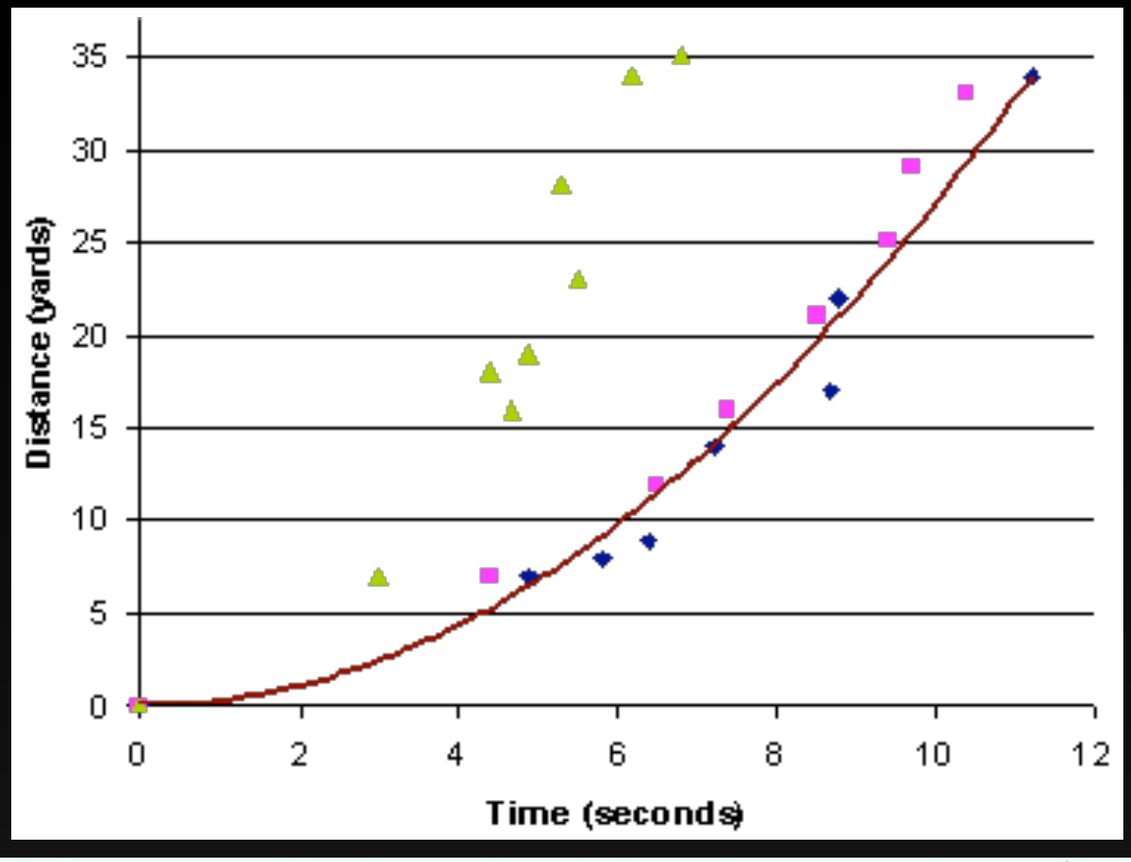
Statistical Reasoning

Robert Brown, PhD
Montreal Neurological Institute

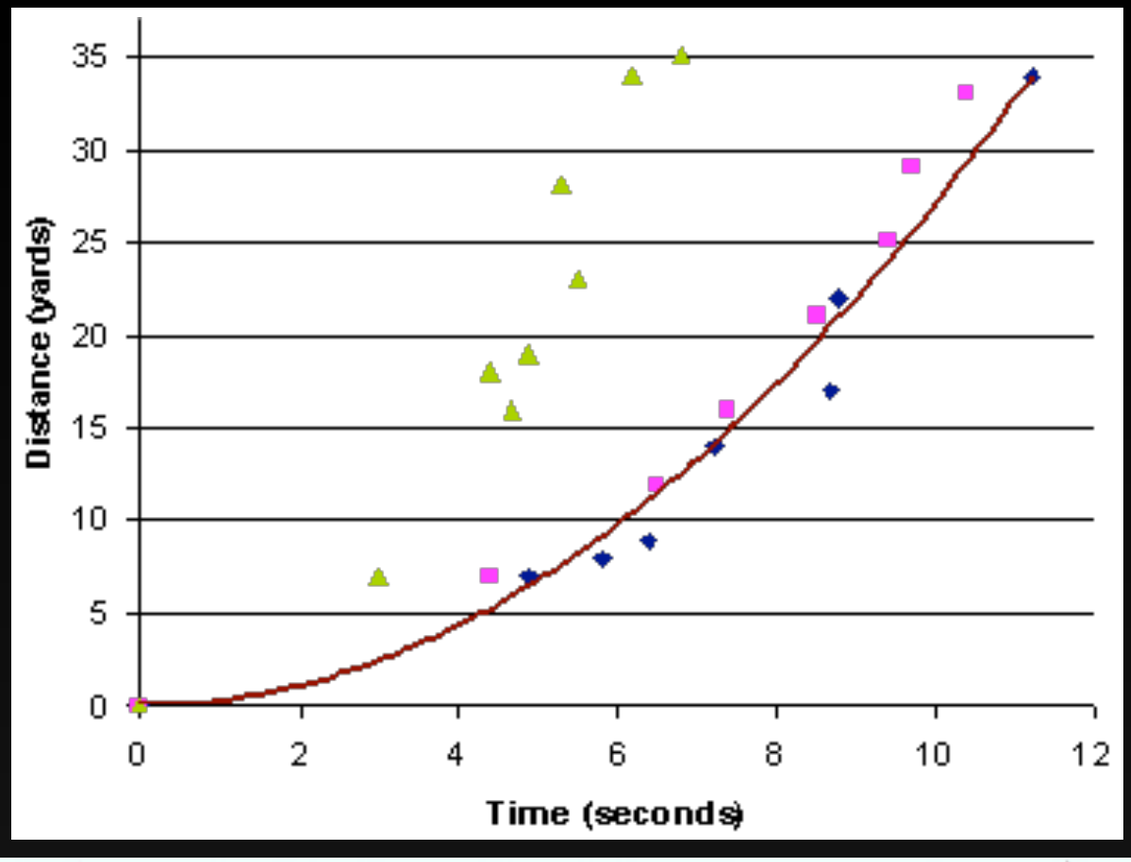
“Philosophy is written in this grand book, the universe which stands continually open to our gaze. But the book cannot be understood unless one first learns to comprehend the language and read the letters in which it is composed. It is written in the language of mathematics, and its characters are triangles, circles and other geometric figures without which it is humanly impossible to understand a single word of it; without these, one wanders about in a dark labyrinth.”

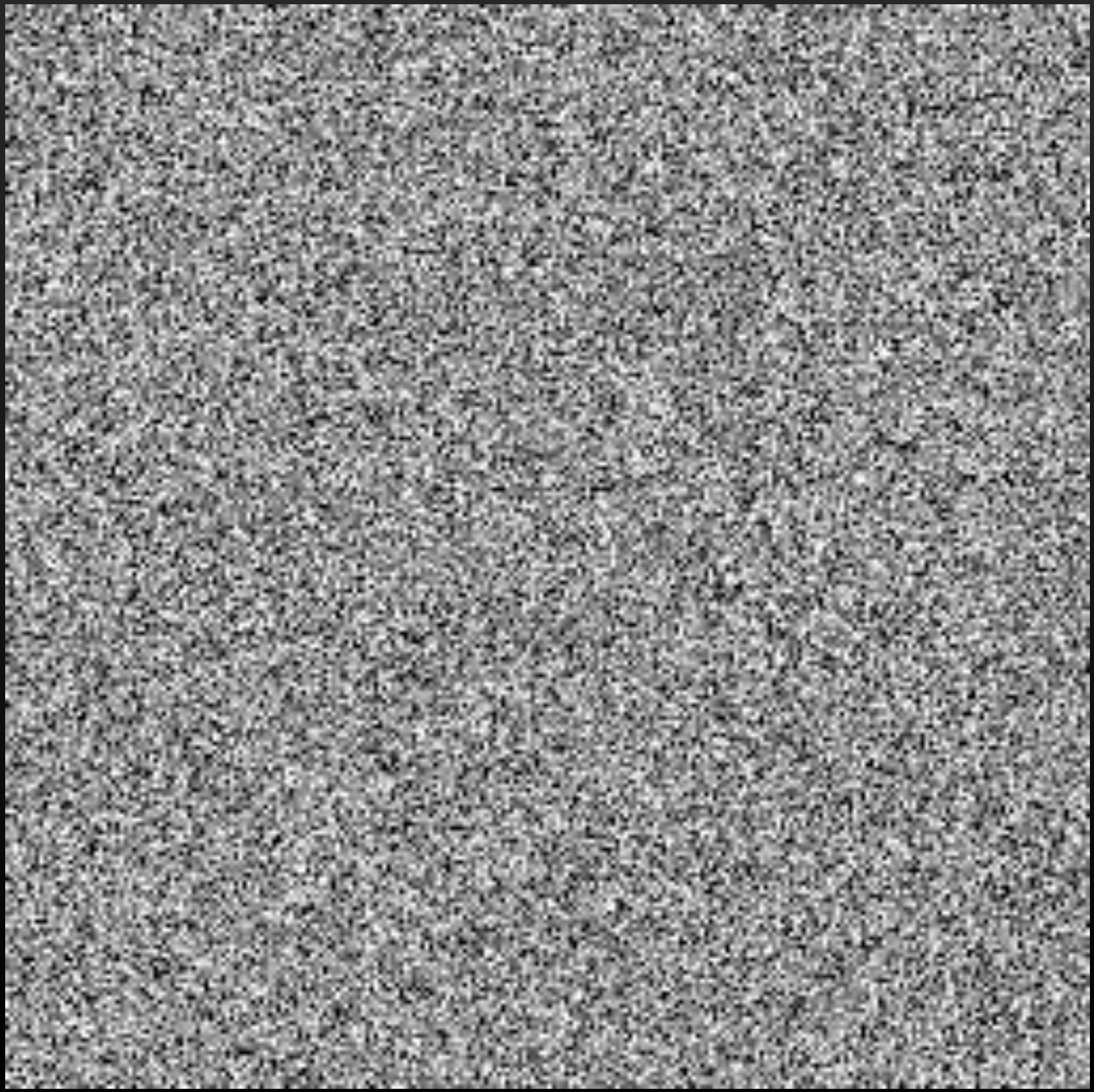
-- Galileo Galilee in Assayer

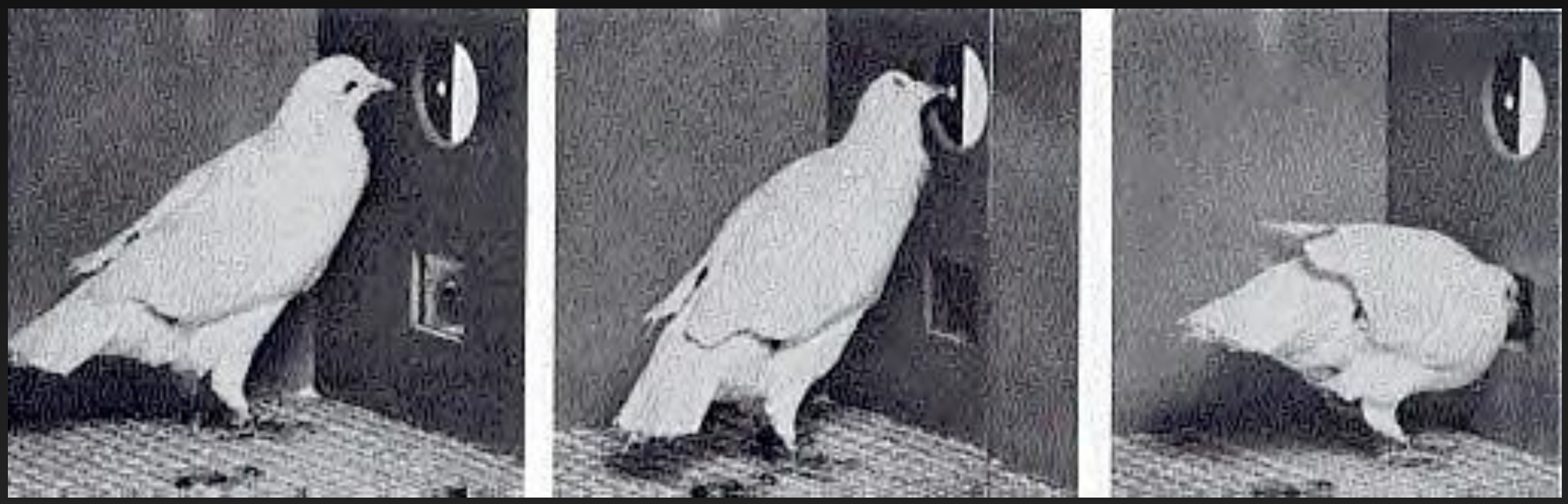

Distance (yards)
0
5
10
15
20
25
30
35
0
2
4
6
8
10
12
Time (seconds)

All measurements are wrong.

$$\rho(a,c) - \rho(b,a) - \rho(b,c) \leq 1$$


Distance (yards)
0
5
10
15
20
25
30
35
0
2
4
6
8
10
12
Time (seconds)



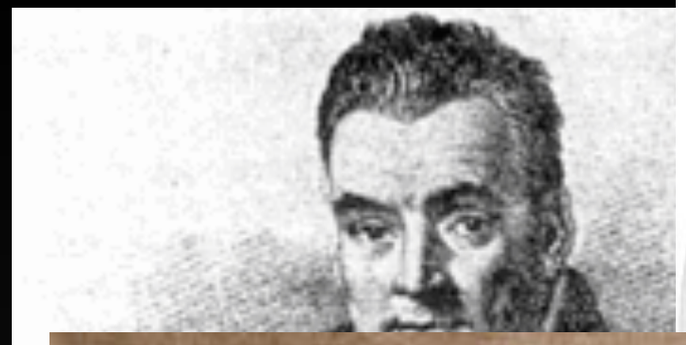


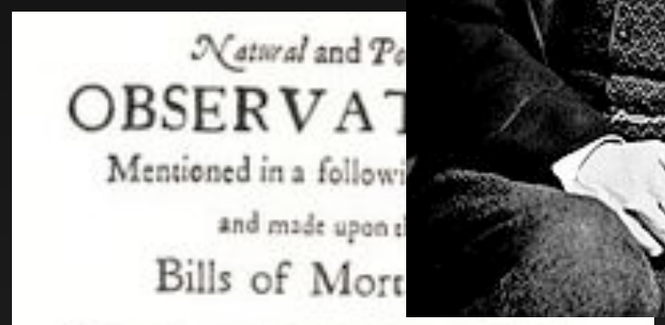
'Superstition' in the pigeon. Skinner, B. F.
Journal of Experimental Psychology, Vol 38(2), Apr 1948, 168-172

All measurements are wrong.

We’re bad at interpreting data

What do we do?





Natural and Pe

OBSERVAT

Mentioned in a followi

and made upon t

Bills of Mort

By JOHN GRAUNT,

Citizen of

LONDON.

With reference to the Government, Religion, Trade, Growth, Ayre, Diseases, and the several Changes of the said CITY.

—— Nec, ne se miretur Turba, laboro,
Contentus paucis Lectoribus ——

LONDON,

Printed by Tho: Roycroft, for John Martin, James Allestry, and Tho: Dicas, at the Sign of the Bell in St. Paul's Church-yard, MDCLXII.

Mathematics is the language of the Universe.

Statistics is the dialect it speaks.

So how well do we speak it?

Why Most Published Research Findings Are False
Ioannidis, JPA,
PLoS Medicine Vol 2, Aug 2005

Modelled scientific studies with bias and errors

- Non-randomized studies: 80% wrong
- Randomized studies: 25% wrong
- Large randomized trials: 10% wrong

These rates are close to what is observed in practice

Contradicted and Initially Stronger Effects in Highly Cited Clinical Research,
Ioannidis, JPA,
JAMA Vol 294, 2005

Followup: 49 of the most highly regarded findings in medicine over the last 13 years

45 claimed to find effective interventions

34 have been retested

14 (41%) have been shown to be wrong or exaggerated

Bayer: fails to replicate about 2/3 of studies identifying drug targets

Believe it or not: how much can we rely on published data on potential drug targets? Nature Reviews Drug Discovery, Vol 10

Amgen: could not replicate 47/53 promising results

Drug development: Raise standards for preclinical cancer research, Nature Vol 483

A simple (but common) statistical error

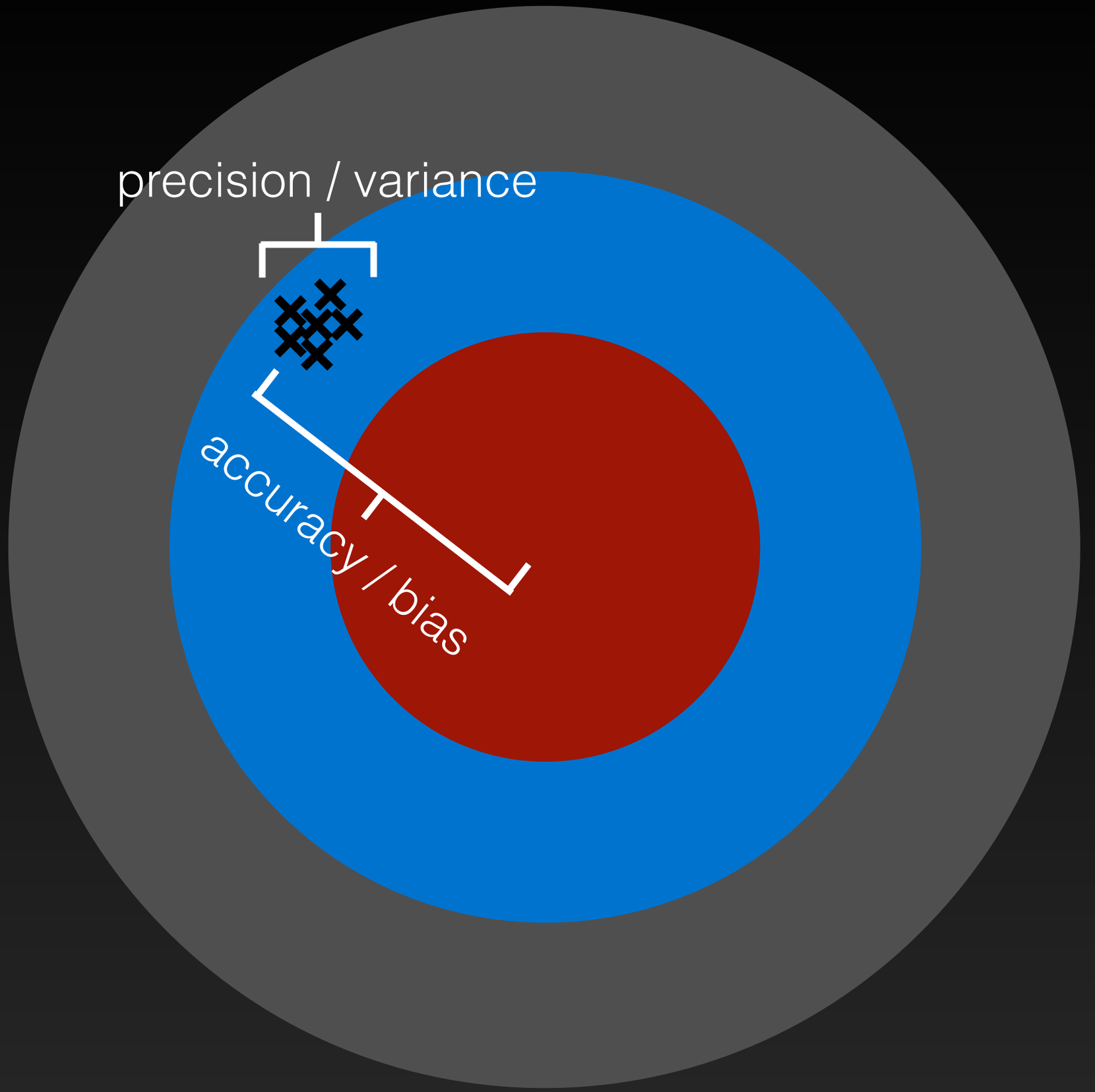
	nature	Science AAAS	nature neuroscience	Neuron	*The Journal of Neuroscience*
Correct	3	9	17	13	36
Incorrect	7	11	16	15	30
% correct	30 %	45 %	52 %	46 %	55 %

Erroneous analyses of interactions in neuroscience: a problem of significance
Nieuwenhuis, S, *et. al.*, *Nature Neuroscience*, Aug 2011


17 59
GUINNESS
DRAUGHT

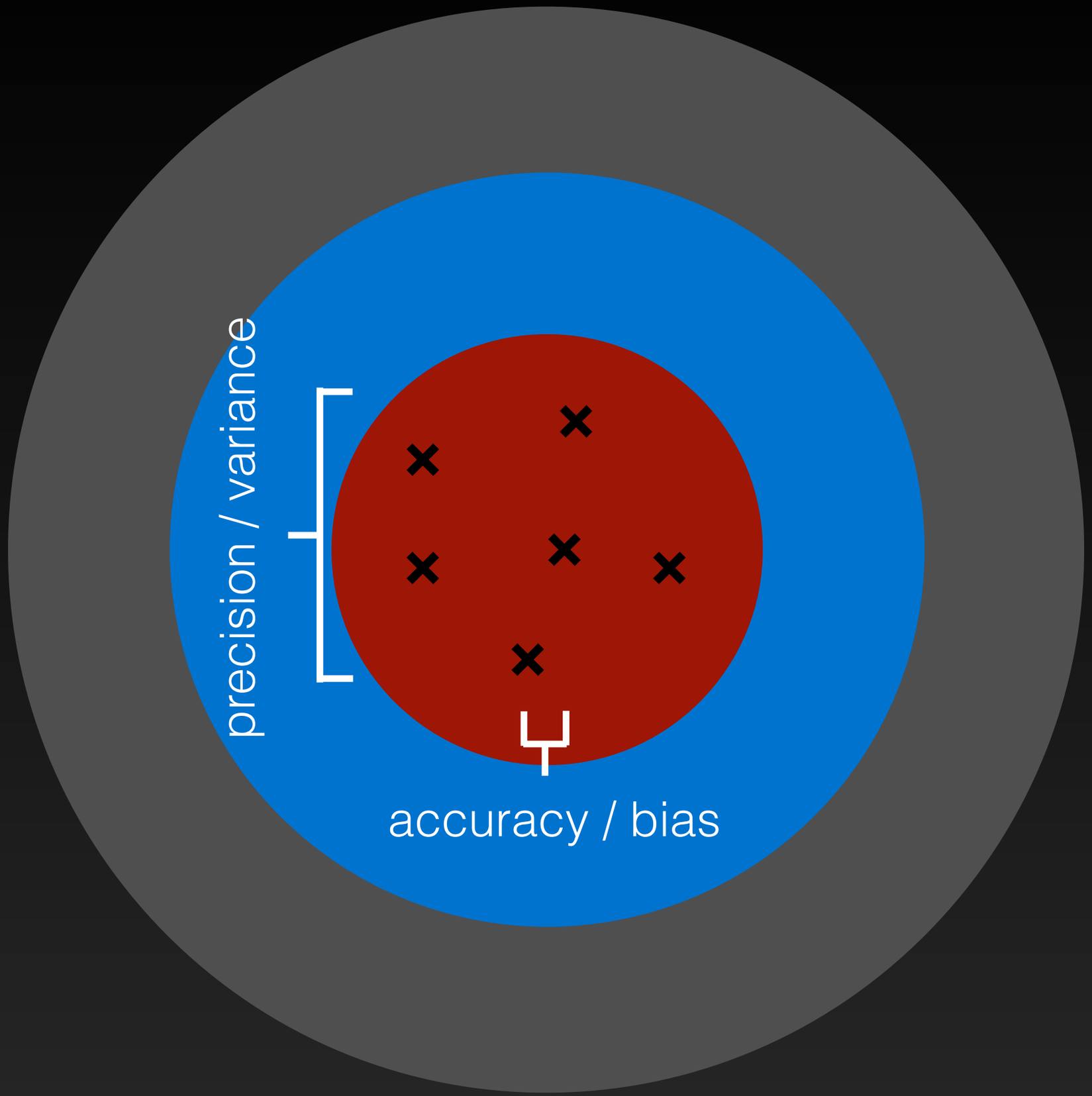
High precision,
low accuracy

Low variance,
high bias

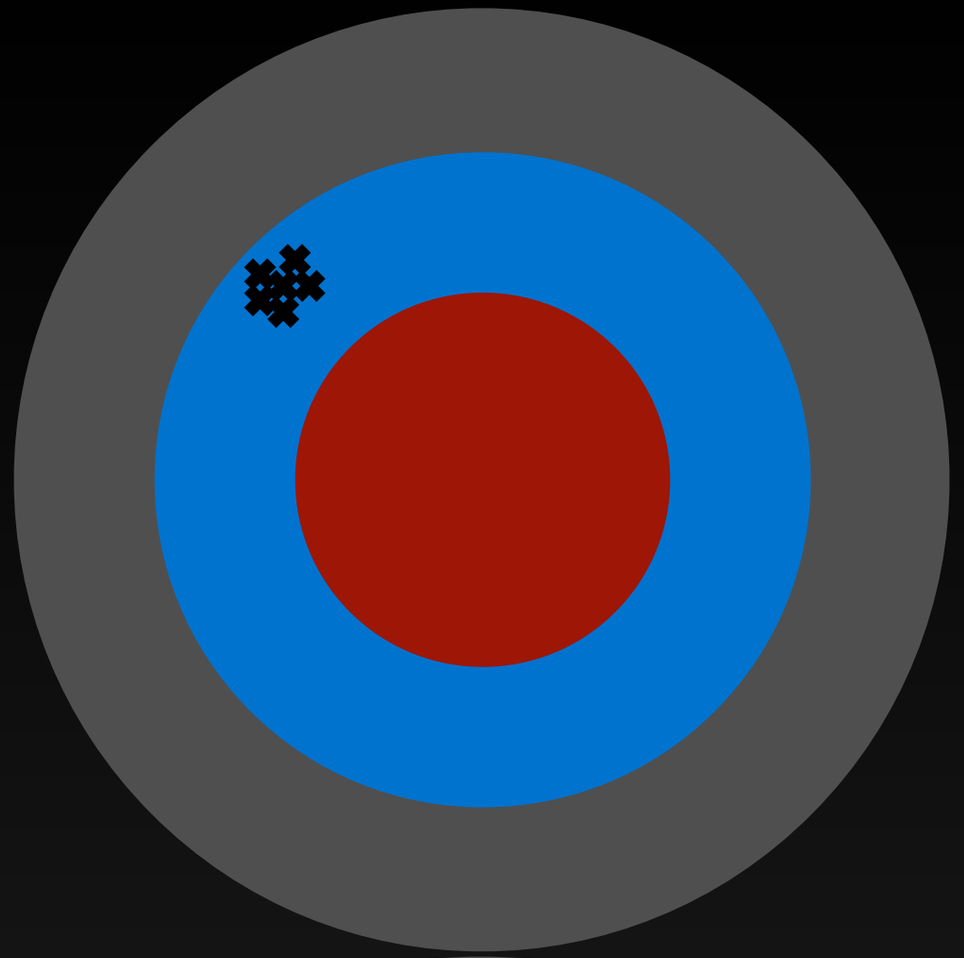


Low precision,
high accuracy

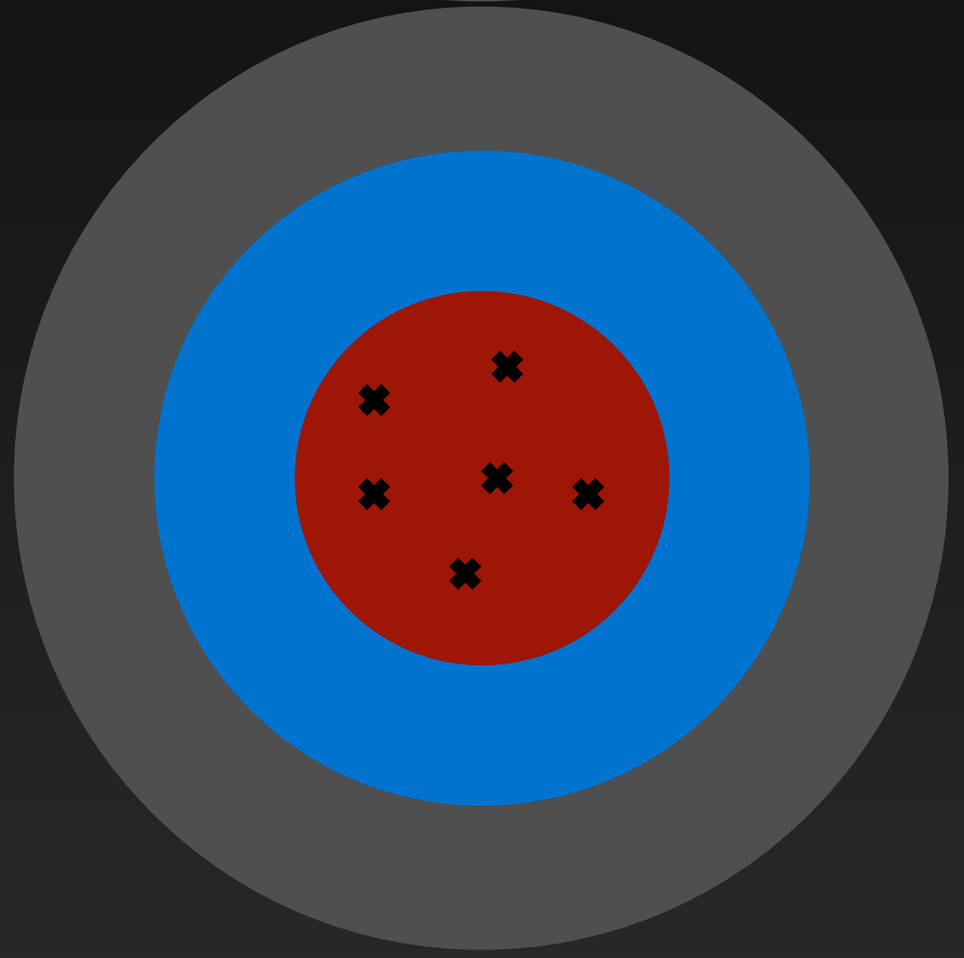
High variance,
low bias



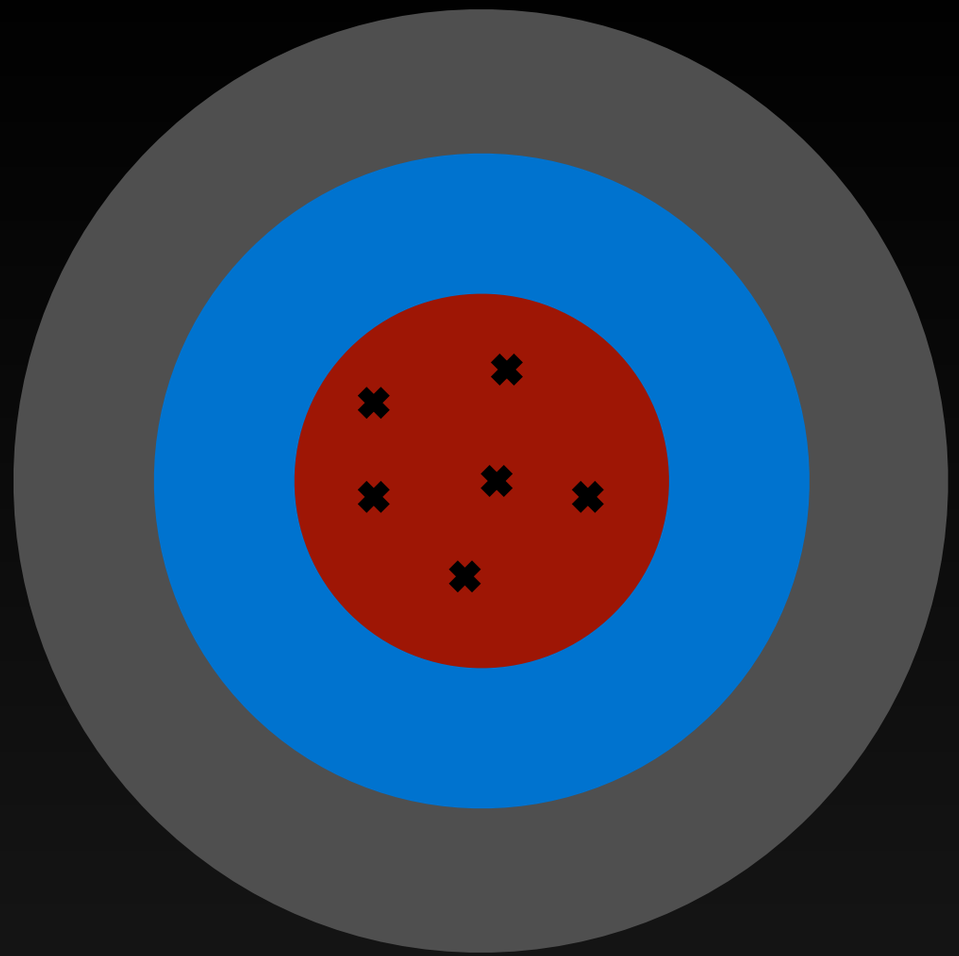
You’re confidently wrong



You’re tentatively correct

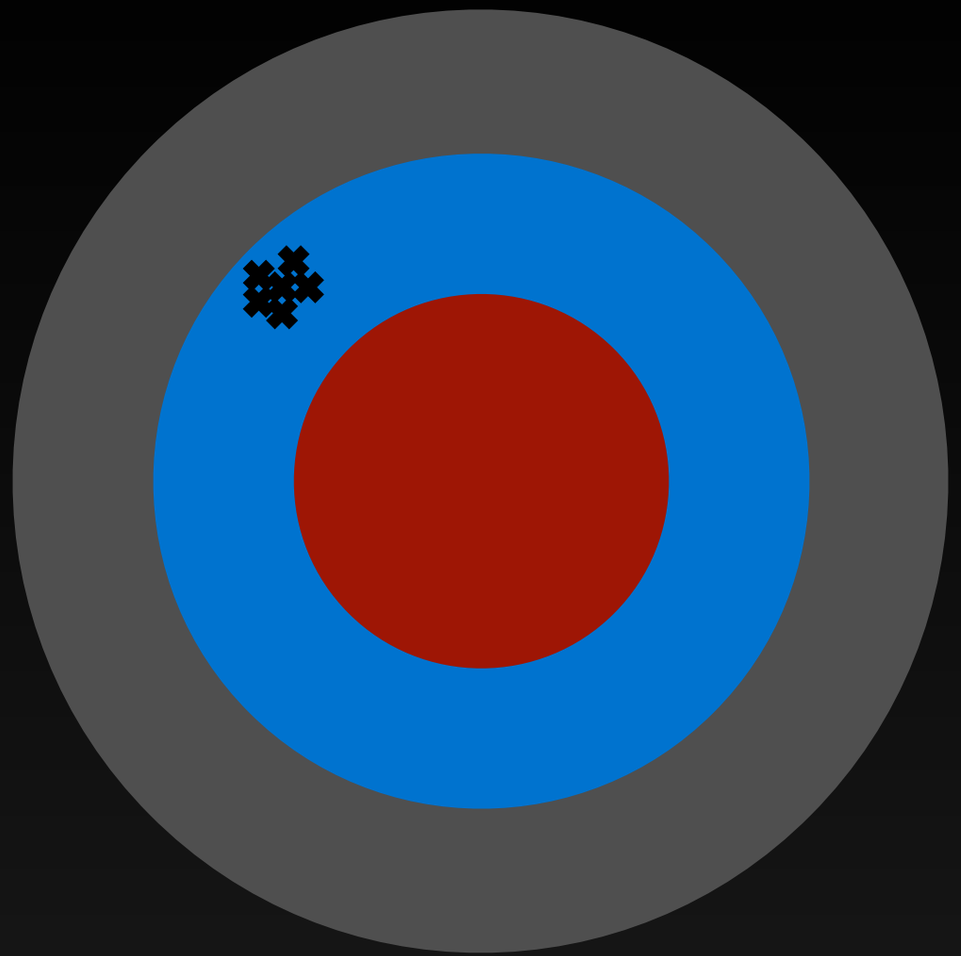


The mean converges to the correct value. Variance is handled by statistics



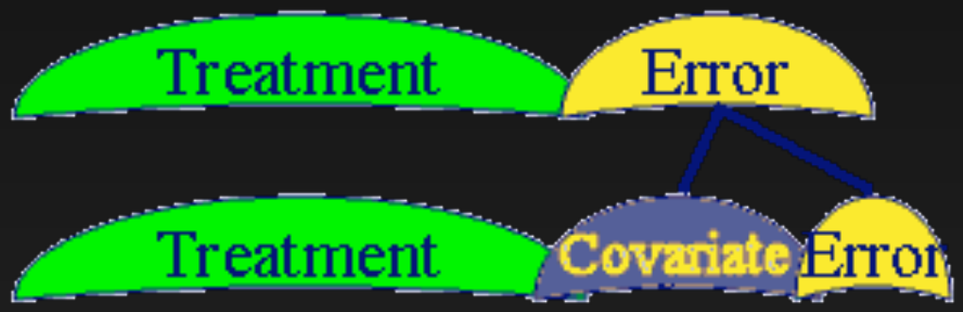
Improve your measurement to reduce variance, or just collect more data

The mean converges to the wrong value.
Bias is not handled by statistics



Handling Bias

Match subjects



Use covariates

Do both!


KEEP
CALM
AND
RANDOMIZE
IT

Handling Bias

Have a hypothesis
(and an experimental design)

Blind everybody

An Experiment

Introduction

Hypothesis

Methods

Results

Conclusions

Case Study

Journal of Cerebral Blood Flow & Metabolism (2009) 29, 1867–1878

www.jcbfm.com

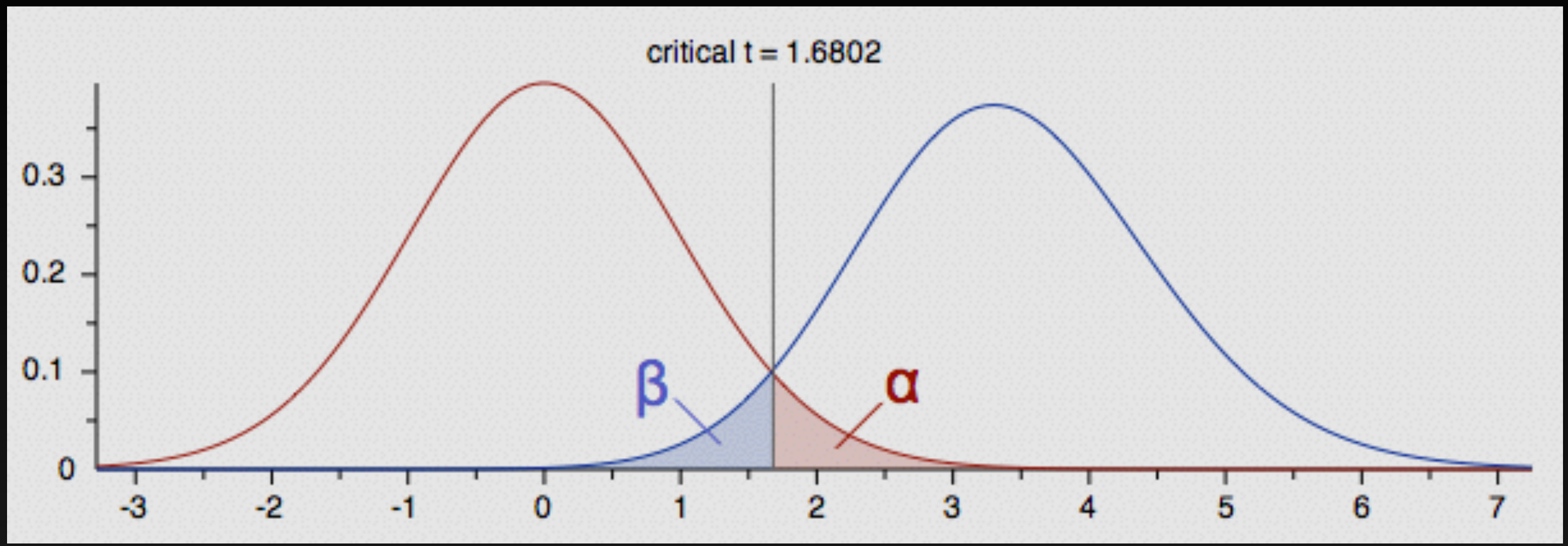
Review Article

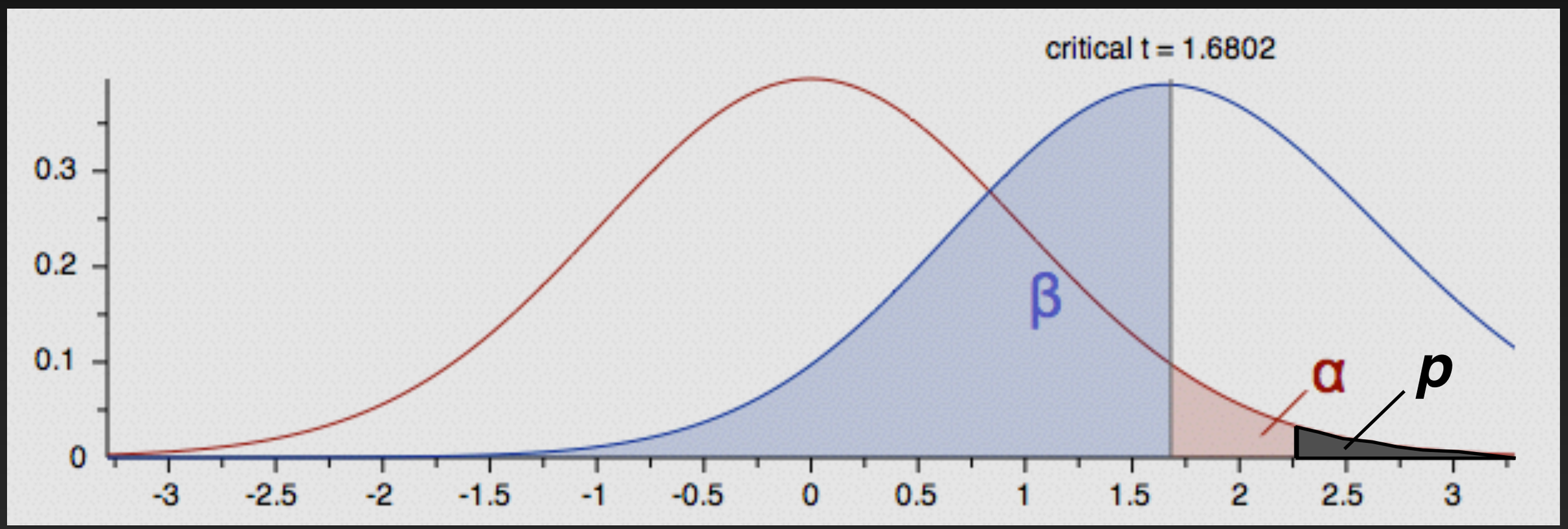
Anomalous venous blood flow and iron deposition in multiple sclerosis

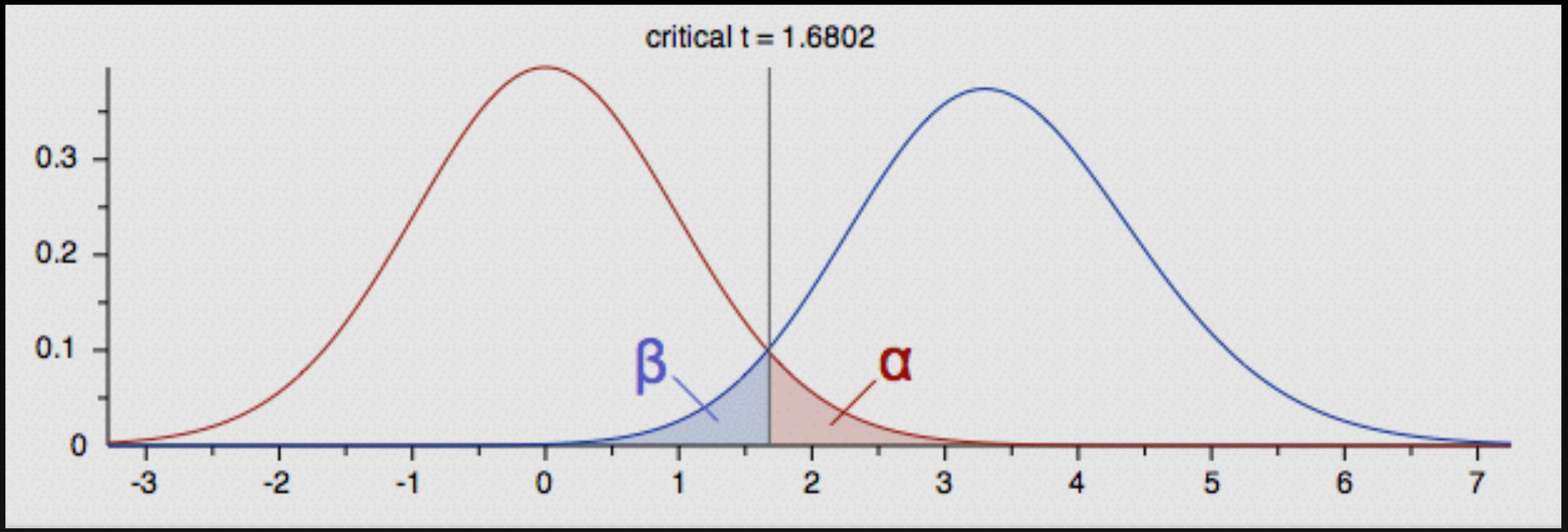
Ajay Vikram Singh[1] and Paolo Zamboni[2]

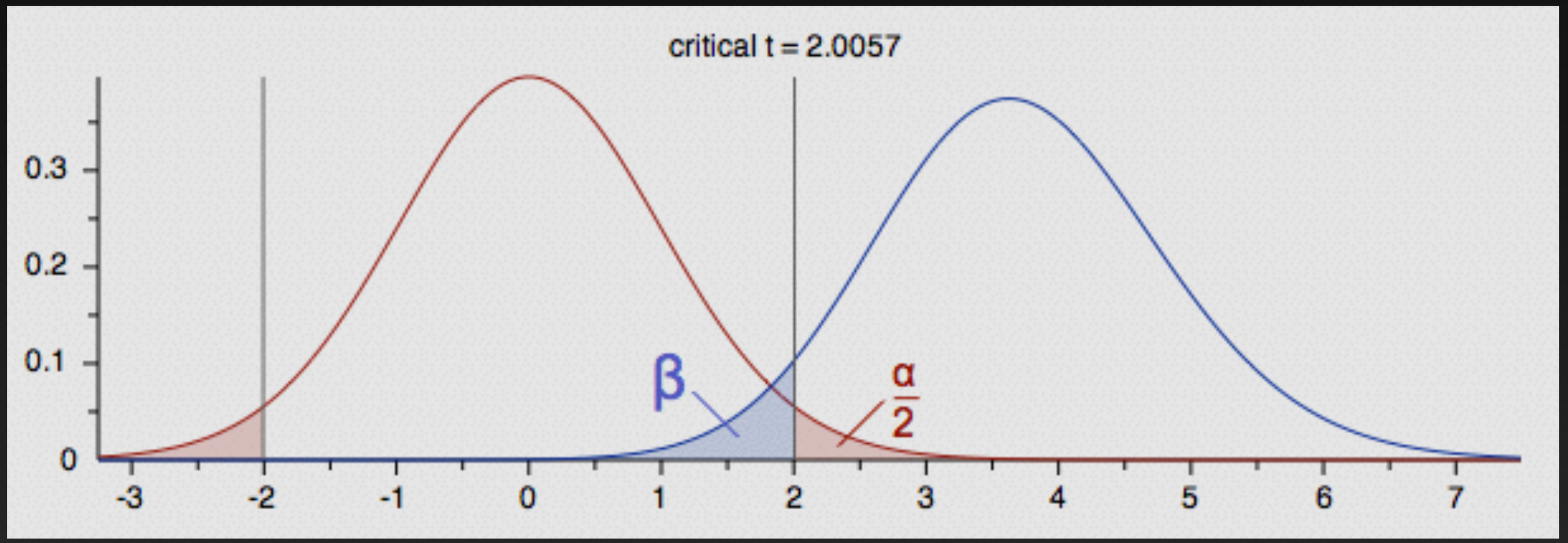
[1]*Department of Physics, European School of Molecular Medicine (SEMM), IFOM-IEO Campus, Centro Interdisciplinare Materiali e Interfacce Nanostrutturati (CIMAINA), University of Milan, Milan, Italy;*
[2]*Vascular Diseases Center, University of Ferrara, Ferrara, Italy*

Power and P-values

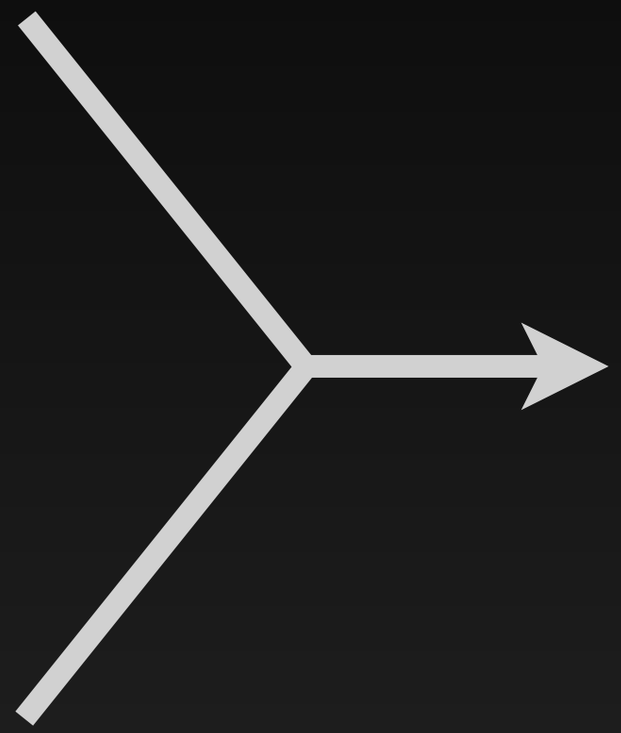





critical t = 1.6802
0.3
0.2
0.1
0
β
α
-3
-2
-1
0
1
2
3
4
5
6
7


critical t = 2.0057
0.3
0.2
0.1
0
β
α
2
-3
-2
-1
0
1
2
3
4
5
6
7

Example Power Calculation



t-test

$N = ?$

Test family

t tests

Statistical test

Means: Difference between two independent means (two groups)

Type of power analysis

A priori: Compute required sample size – given α, power, and effect size

Input parameters

Determine

Tail(s) Two

Effect size d 0.5

α err prob 0.05

Power (1−β err prob) 0.9

Allocation ratio N2/N1 1

Output parameters

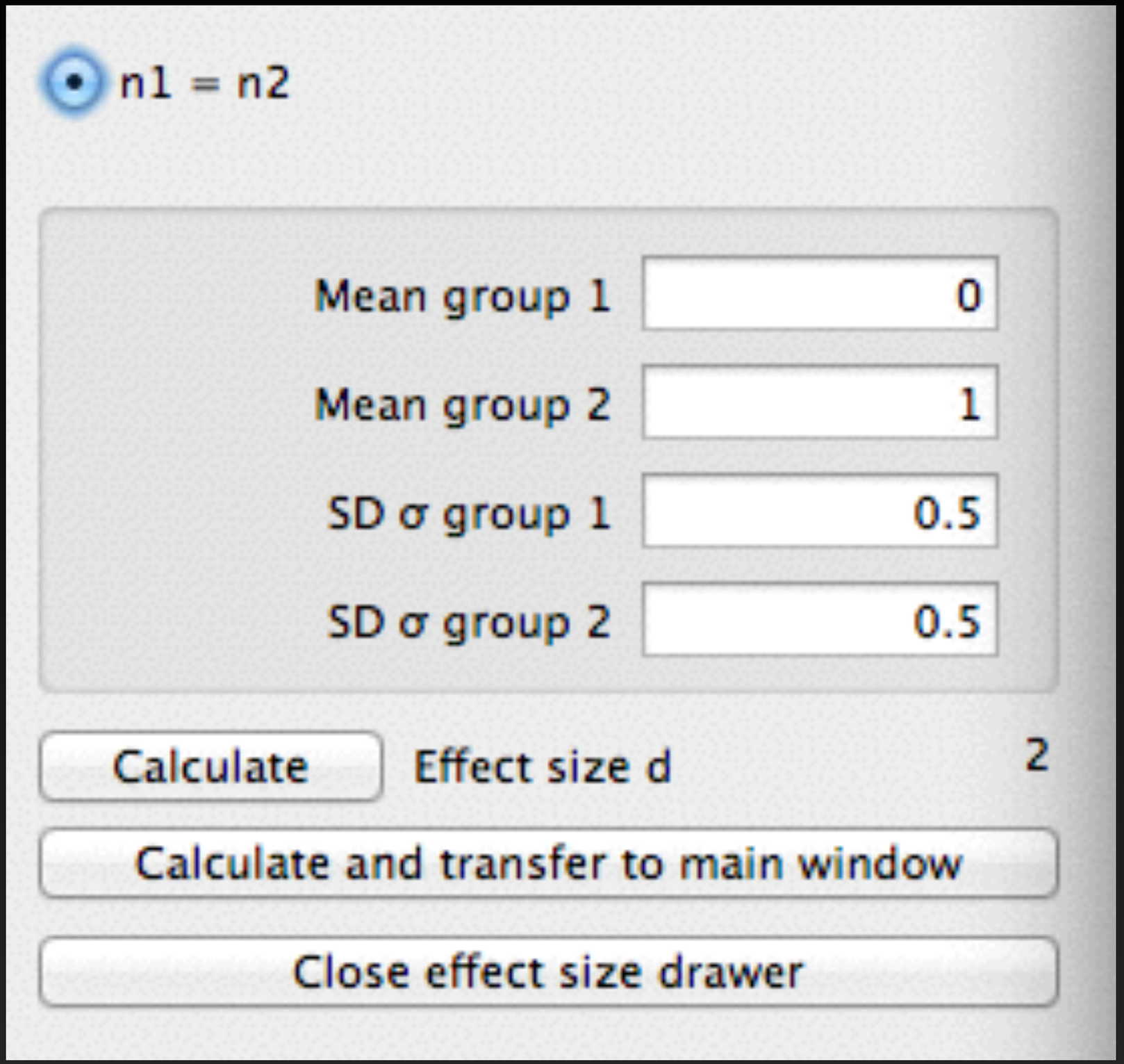
Noncentrality parameter δ	3.2787193
Critical t	1.9740167
Df	170
Sample size group 1	86
Sample size group 2	86
Total sample size	172
Actual power	0.9032300

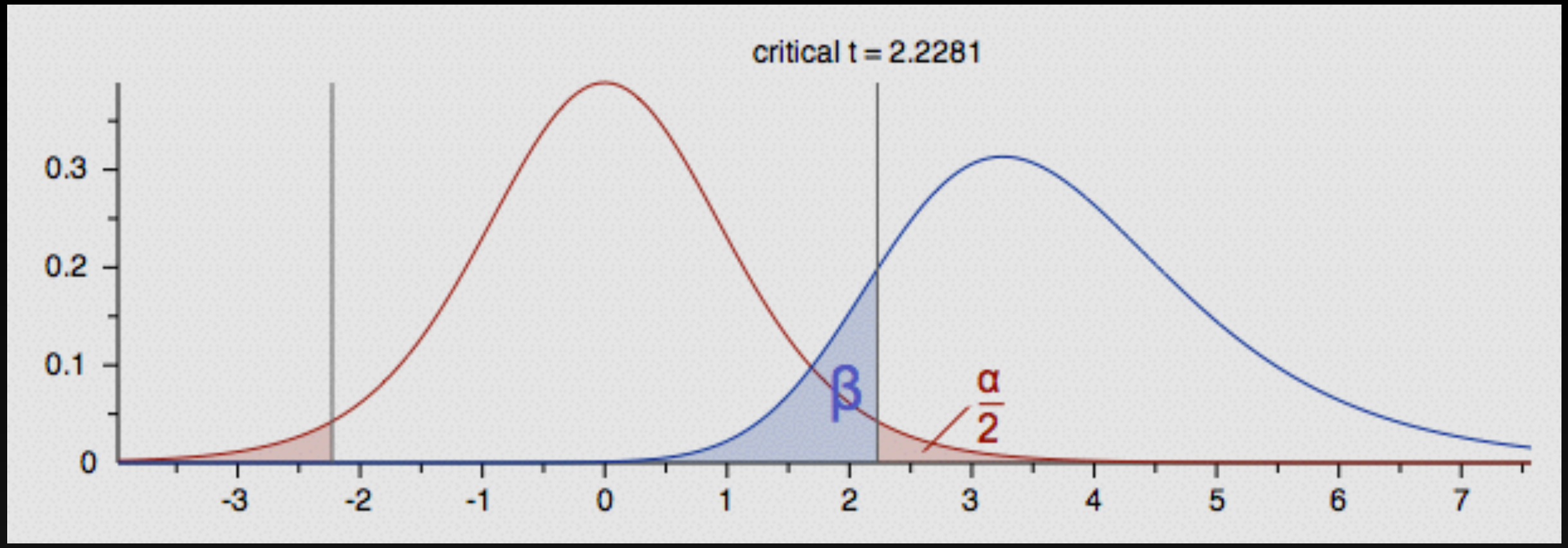
Effect size

α

β (power)

Ratio


n1 = n2
Mean group 1
0
Mean group 2
1
SD σ group 1
0.5
SD σ group 2
0.5
Calculate
Effect size d
2
Calculate and transfer to main window
Close effect size drawer



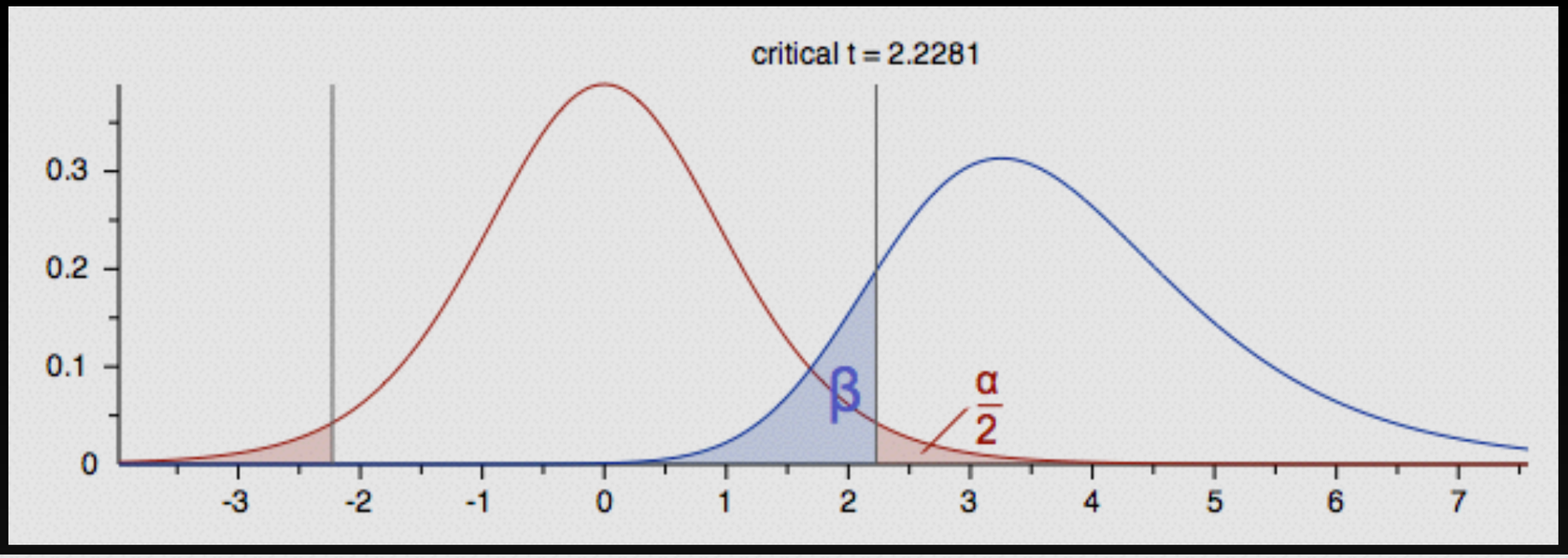
Input parameters

Tail(s)	Two
Effect size d	2
α err prob	0.05
Power (1−β err prob)	0.8
Allocation ratio N2/N1	1

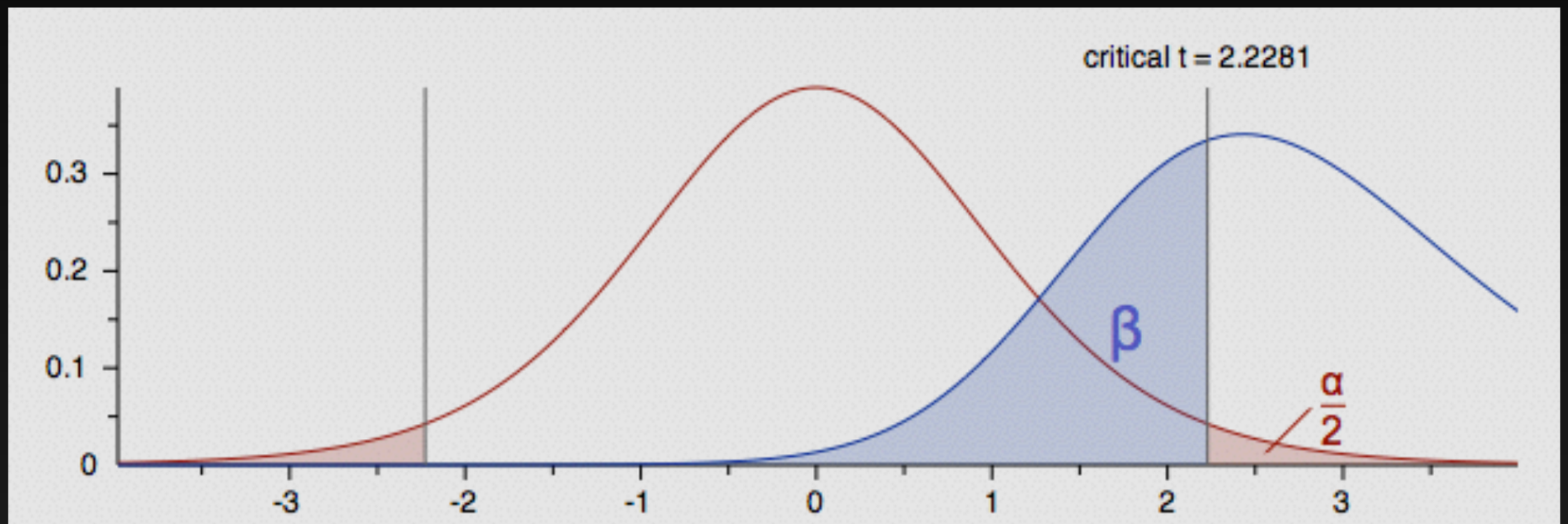
Determine

Output parameters

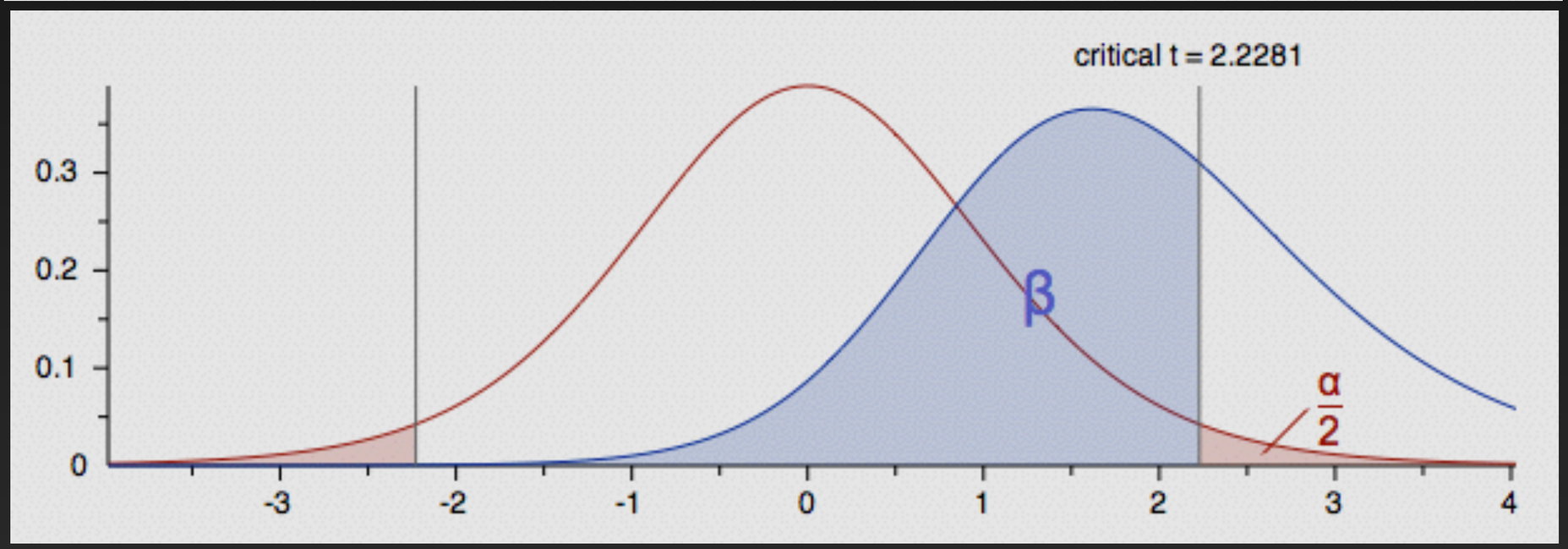
Noncentrality parameter δ	3.4641016
Critical t	2.2281389
Df	10
Sample size group 1	6
Sample size group 2	6
Total sample size	12
Actual power	0.8764178



d = 2
N = 12
β = 0.12



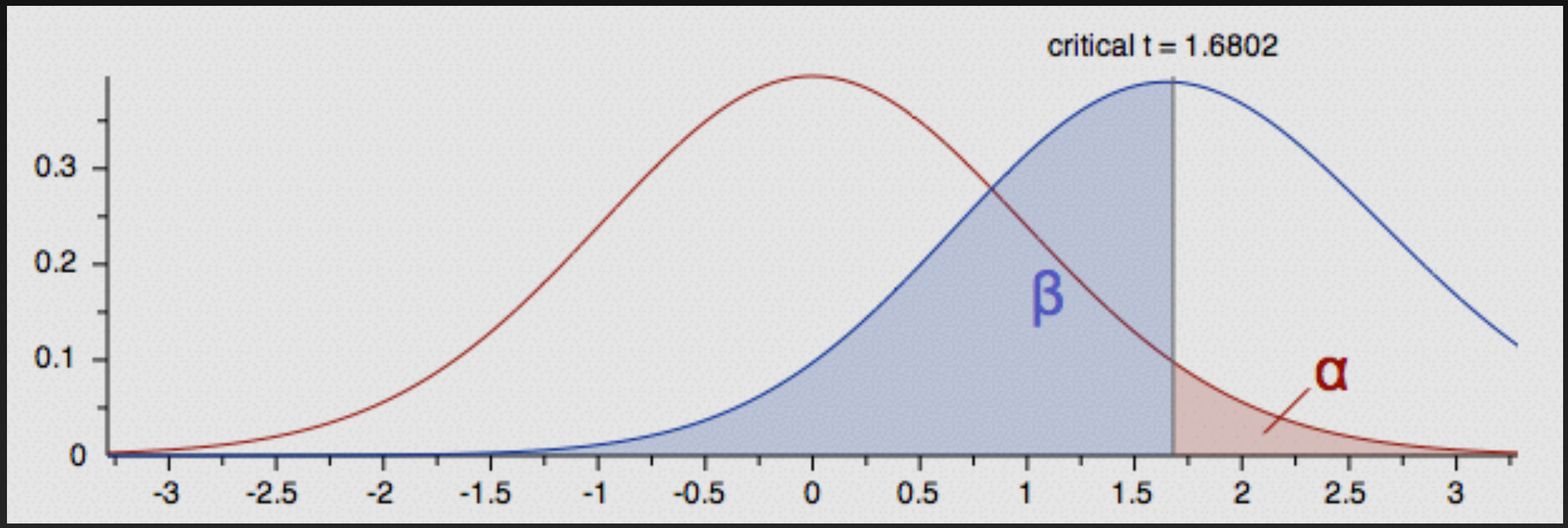
d = 1.5
N = 12
β = 0.35



d = 1
N = 12
β = 0.65

The most important thing to remember:

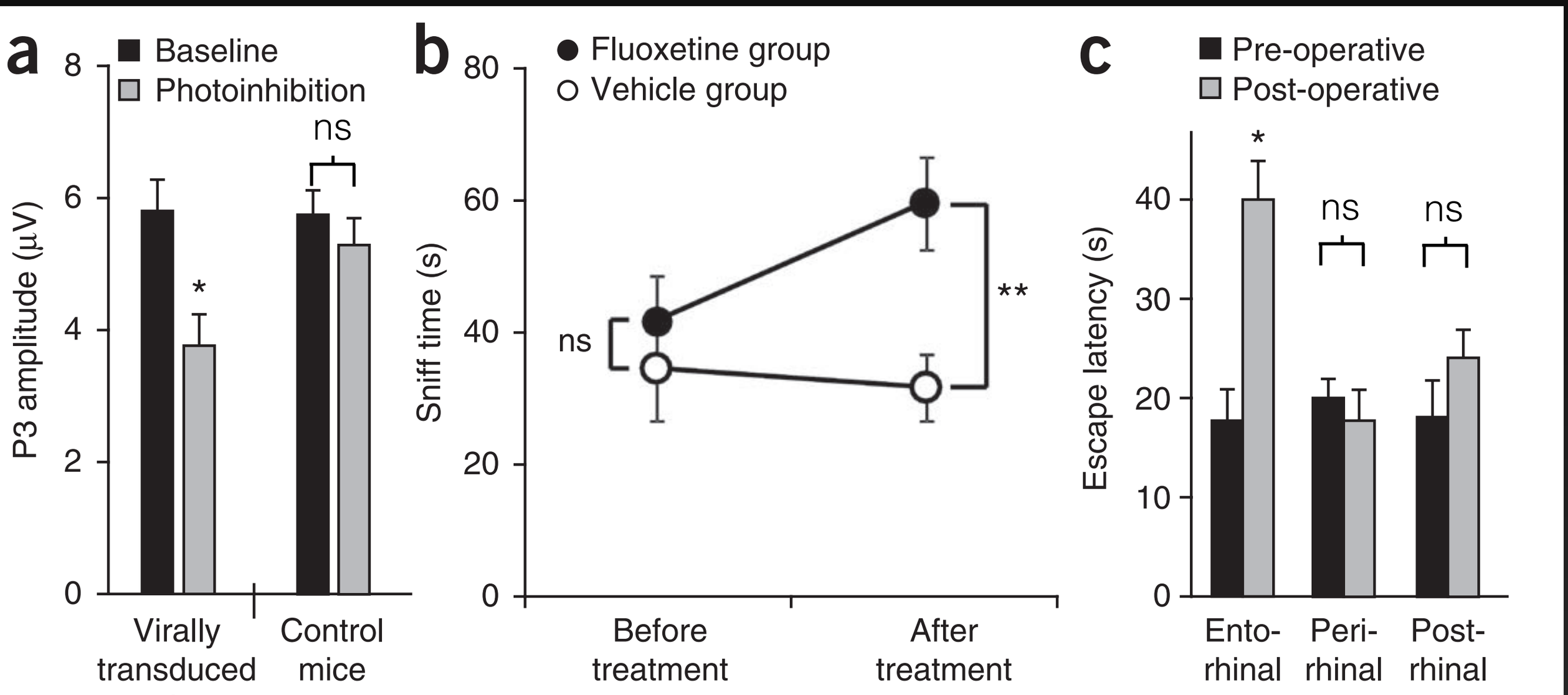
A non-significant result does NOT mean there is no effect.



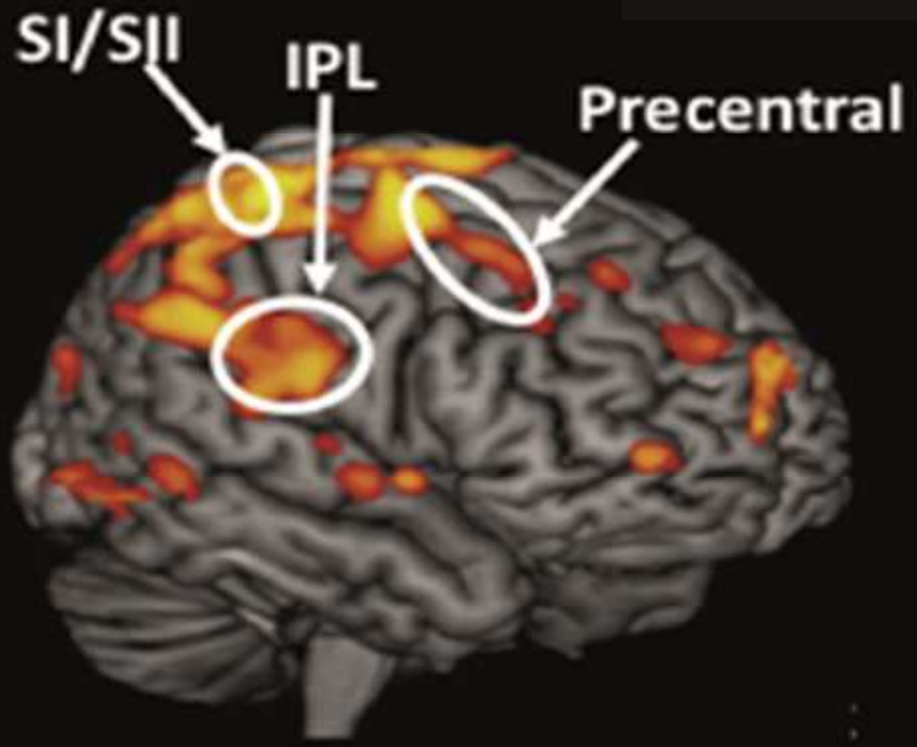
Multiple Comparisons

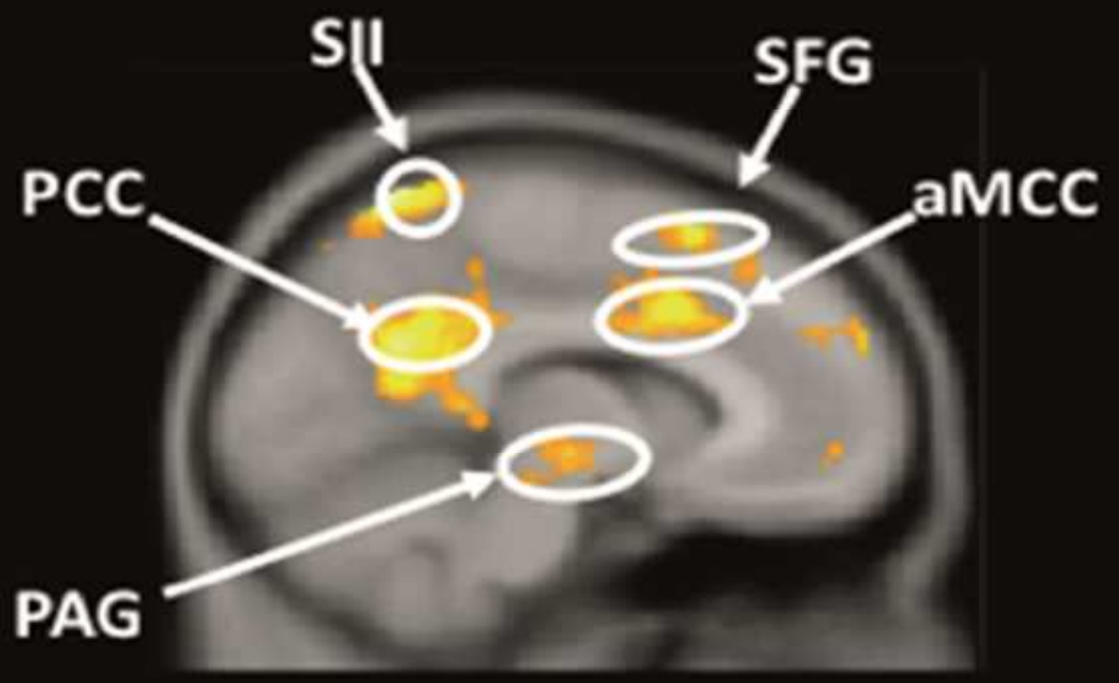
Test

Question 1:


a
Baseline
Photoinhibition
P3 amplitude (μV)
0
2
4
6
*
ns
Virally transduced mice
Control mice
b
Fluoxetine group
Vehicle group
Sniff time (s)
0
20
40
60
ns
**
Before treatment
After treatment
c
Pre-operative
Post-operative
Escape latency (s)
0
10
20
30
40
*
ns
ns
Ento-rhinal
Peri-rhinal
Post-rhinal

Normal Control





Autism Spectrum Disorder