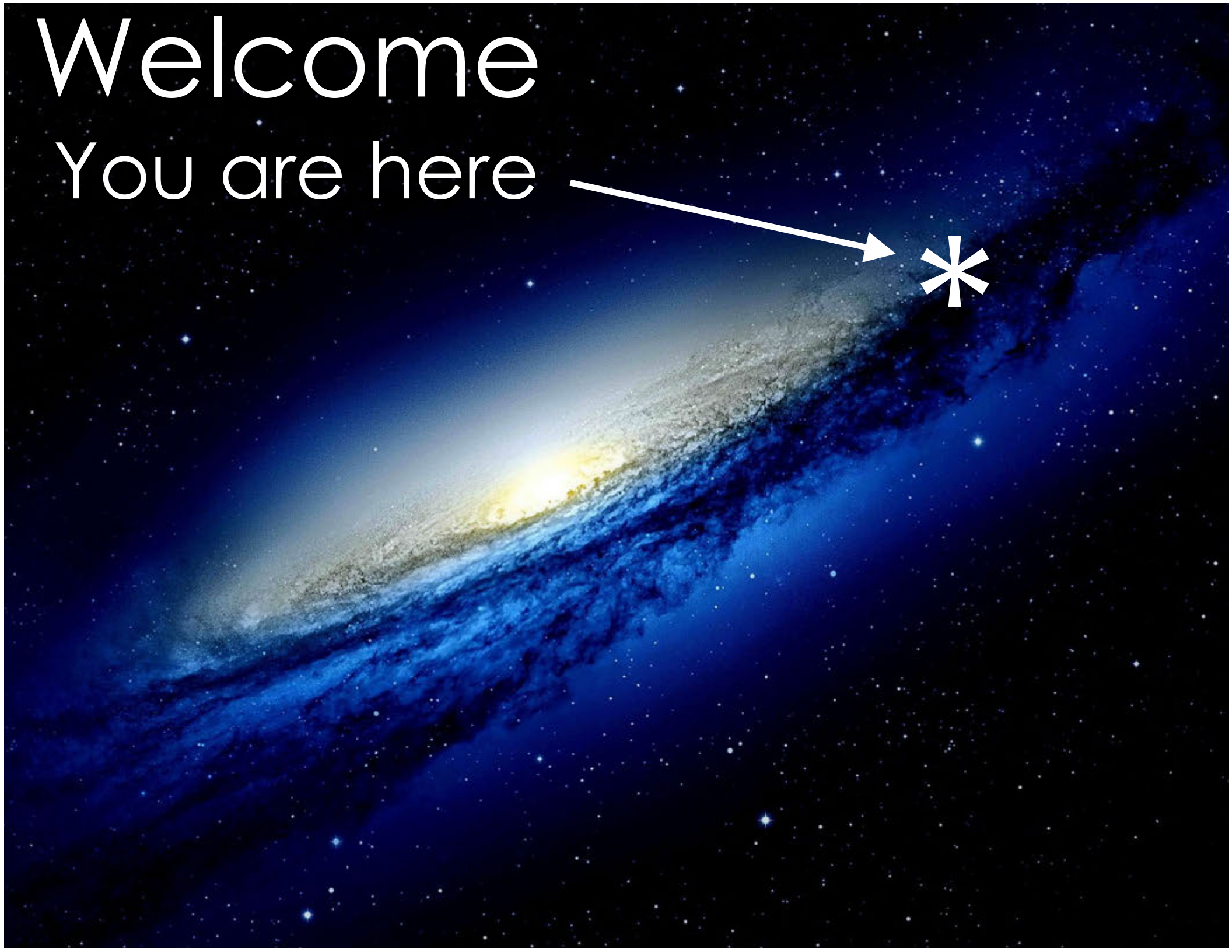
Hello, Statistics*

*REFORM

Rex B Kline

Concordia University, Montréal

10 November 2014


Welcome
You are here
*

http://tinyurl.com/statreform

Hello, Statistics Reform

Rex B Kline, Psychology, Concordia University
rex.kline@concordia.ca; http://tinyurl.com/rexkline
November 10, 2014
University of Nebraska–Lincoln

Reform content

- Flaws of significance testing
- Effect sizes with replication (do, not just talk)
- Focus on substantive significance

Broken significance tests

- *p* values wrong in most studies
- Assumptions are implausible and unverified
- Trained incapacity (abilities as blind spots)
- Design-analysis mismatch
- Random sampling is assumed
- So is no other type of error
- E.g., measurement, specification, implementation
- Unverified statistical requirements
- Most researchers do not bother to check
- Incorrect methods (statistical tests of assumptions)
- Significance tests usually incorrectly conducted
- Most researchers do not estimate power
- Selection of arbitrary level of α is inappropriate
- Legitimizes trivial topics ("significant" results)
- Researcher *df* (Simmons et al., 2011)
- False-positive psychology (junkyard of false +s)
- Widely misinterpreted
- Professors no better than students
- Multigenerational cycle of misinformation

Electronic version at http://tinyurl.com/statreform

Big 5 cognitive errors and consequences

1. Odds against chance
2. Inverse probability
3. Local Type I error

} p

4. Replicability
5. Validity

} $1 - p$

- Overall effect: Overinterpret results
- False confidence in findings
- Little sense of need to replicate

Fallacy	ψ profs[a]	Ugrads
Odds against chance	—	72%
Inverse probability	17–36%	35
Local Type I error	67–73	45
Replicability	37–60	42
Validity	33–66	15

[a]Haller and Krauss (2002), Oakes (1986).

Significance test requirements

- Study probability samples
- Control all errors except sampling error
- Estimate costs of Type I vs. Type II error
- Estimate power
- Set α intelligently, not arbitrarily
- Verify all assumptions
- Nil hypothesis is plausible
- Not misinterpreted
- Study is replicated

Lambdin's (2012) criticisms

- Statistical buffoonery, sorcery, shamanism
- Intellectually dishonest
- Unscientific research literature

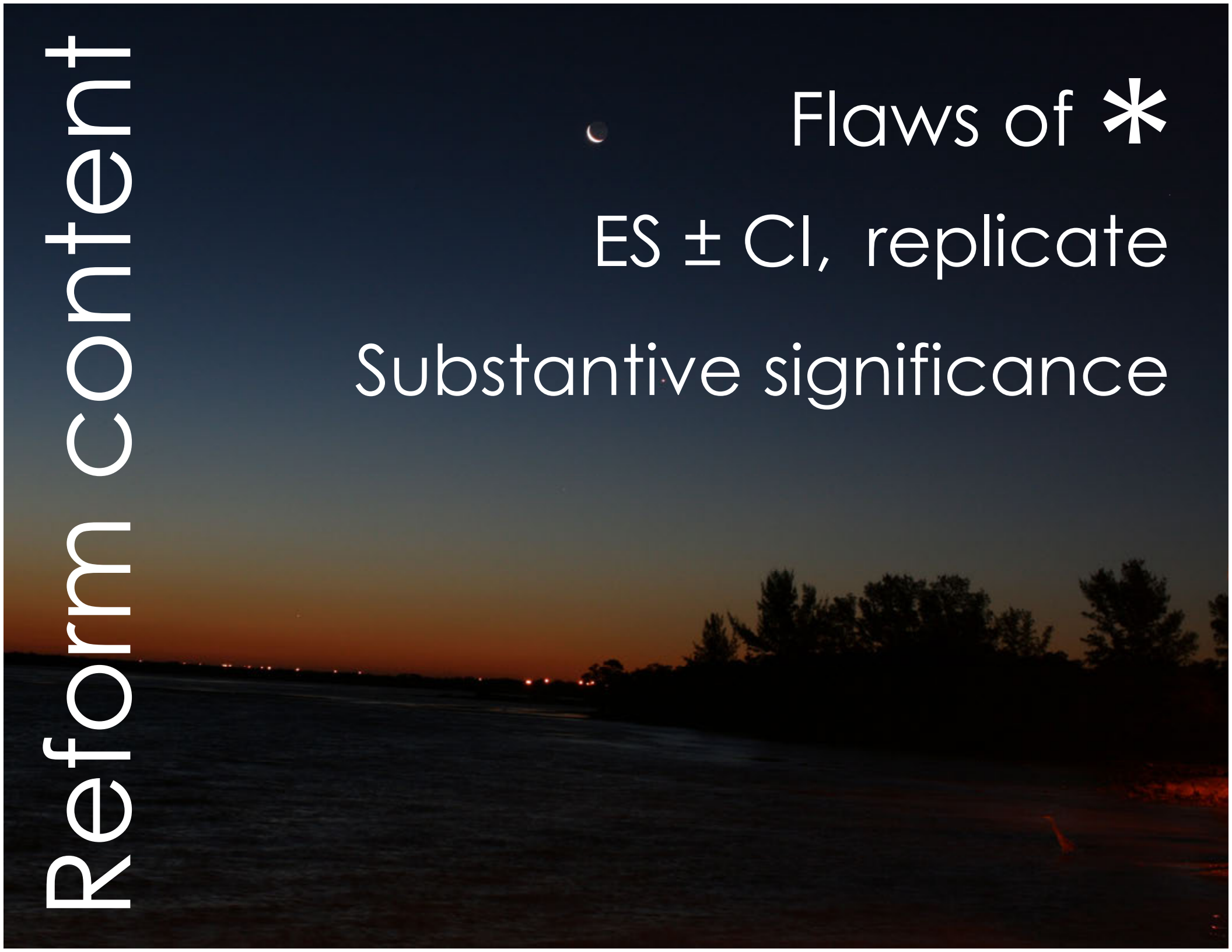
sta·tis·tics re·form [st*uh*-**tis**-tiks ree-**fawrm**]

noun

1. effort to improve quantitative literacy among researchers not formally trained in statistics

2. aims to improve results comprehension and quality of published studies

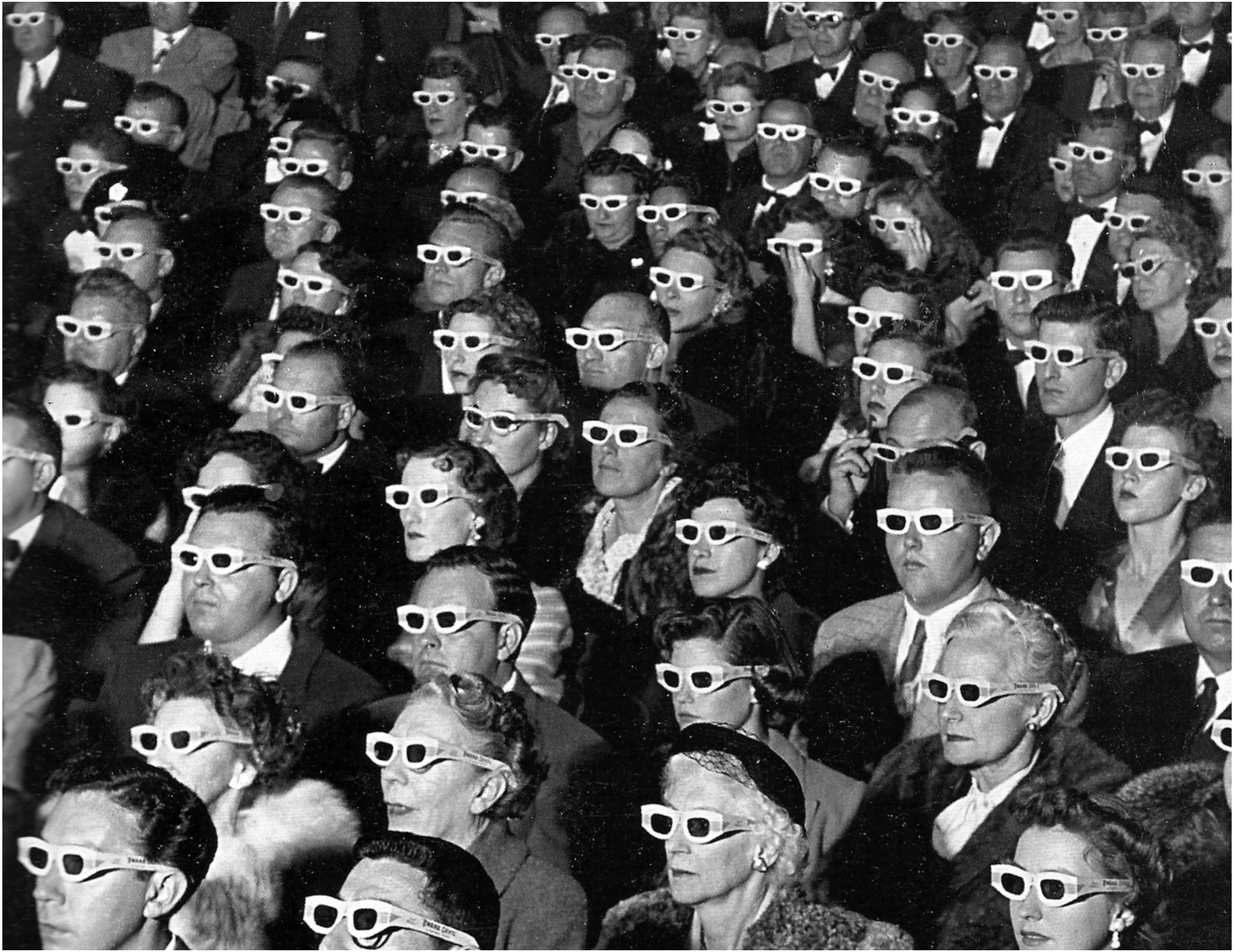
Kline (2013)


Reform content
Flaws of *
ES ± CI, replicate
Substantive significance


It's *significant!*


CAUTION


broken
wrong
fantasy
unscientific
junkyard
deluded
uncritical




DO NOT
DISTURB

*$p < .05$

Which one is like all the rest?


KEEP
CALM
AND
THROW A
CHAIR

A reformer is a guy who rides through the sewer in a glass bottom boat.

Jimmy Walker


All aboard!

Broken * tests

p values wrong

Implausible assumptions

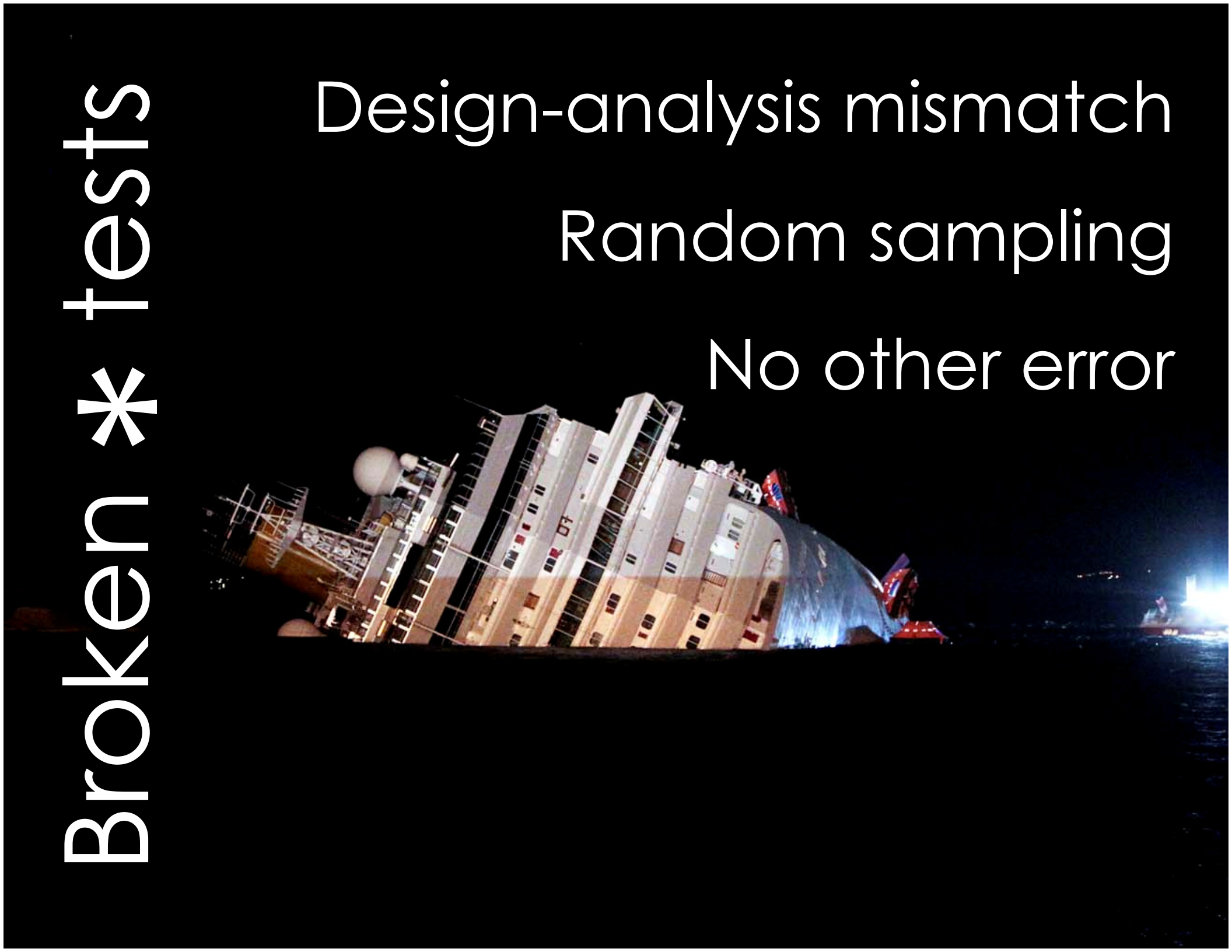
Trained incapacity

Broken * tests

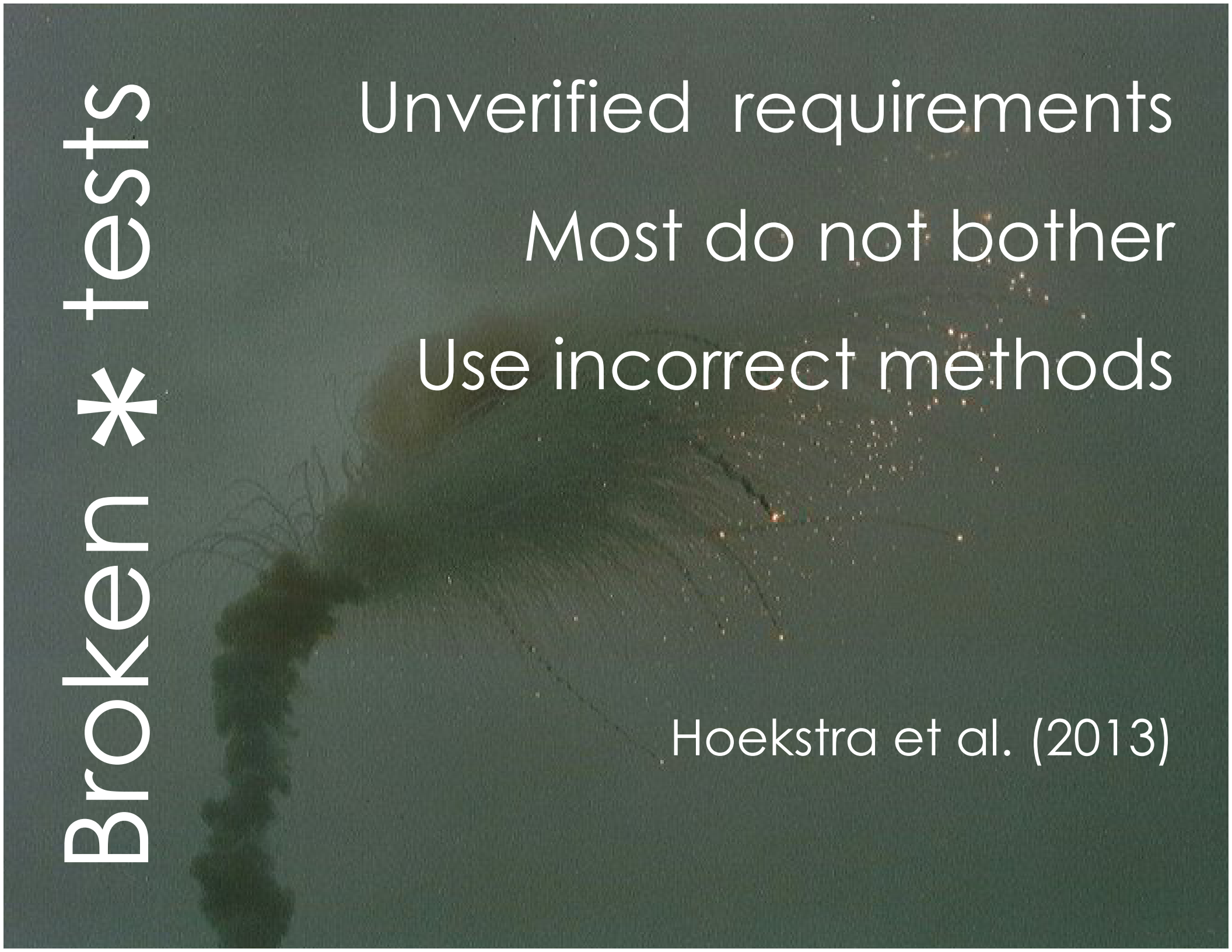
Design-analysis mismatch

Random sampling

No other error




Broken * tests
No
measurement
specification
implementation
error


Broken * tests
Unverified requirements
Most do not bother
Use incorrect methods
Hoekstra et al. (2013)

Researchers should not rely on statistical tests to check assumptions because of the frequency with which they produce inaccurate results.

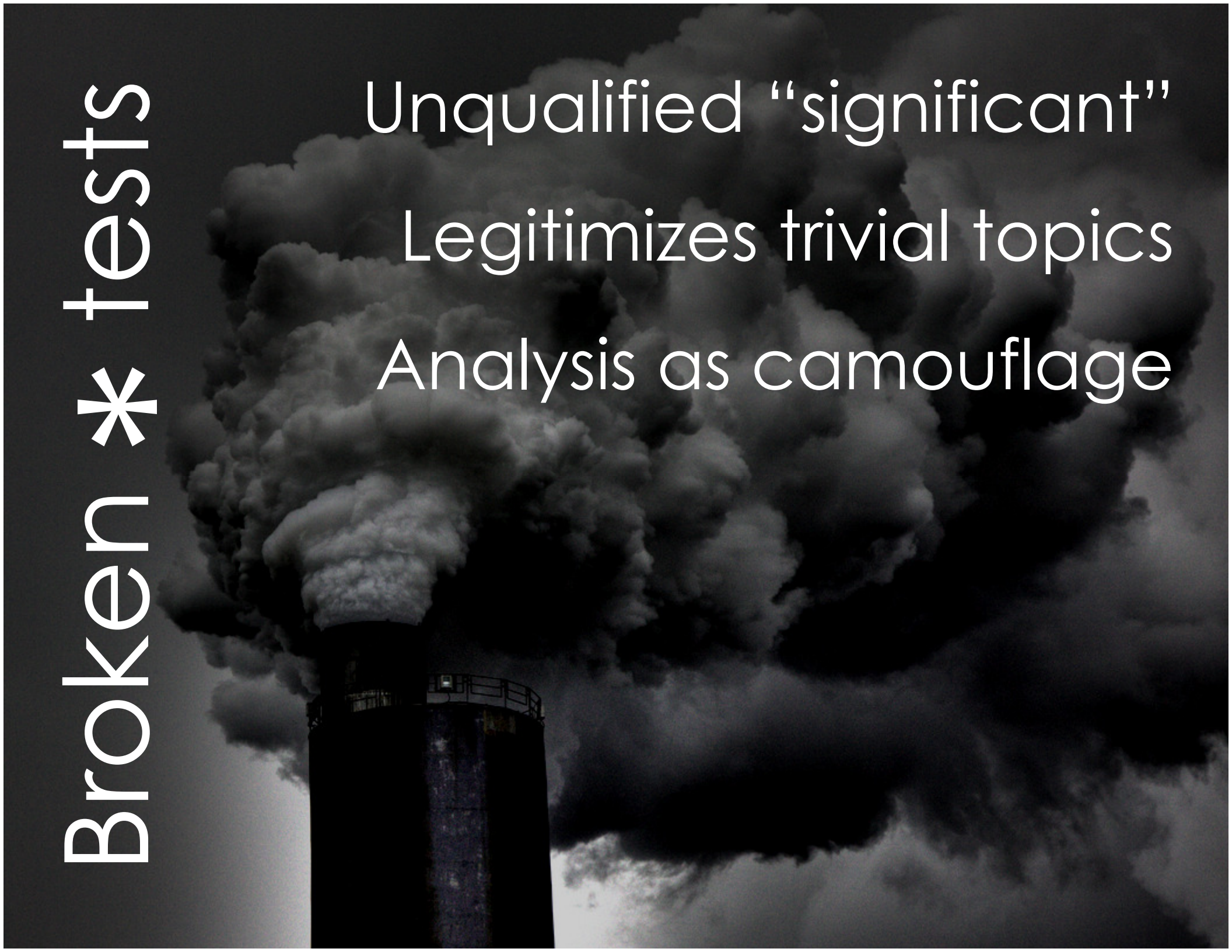
Erceg-Hurn & Mirosevich (2008)

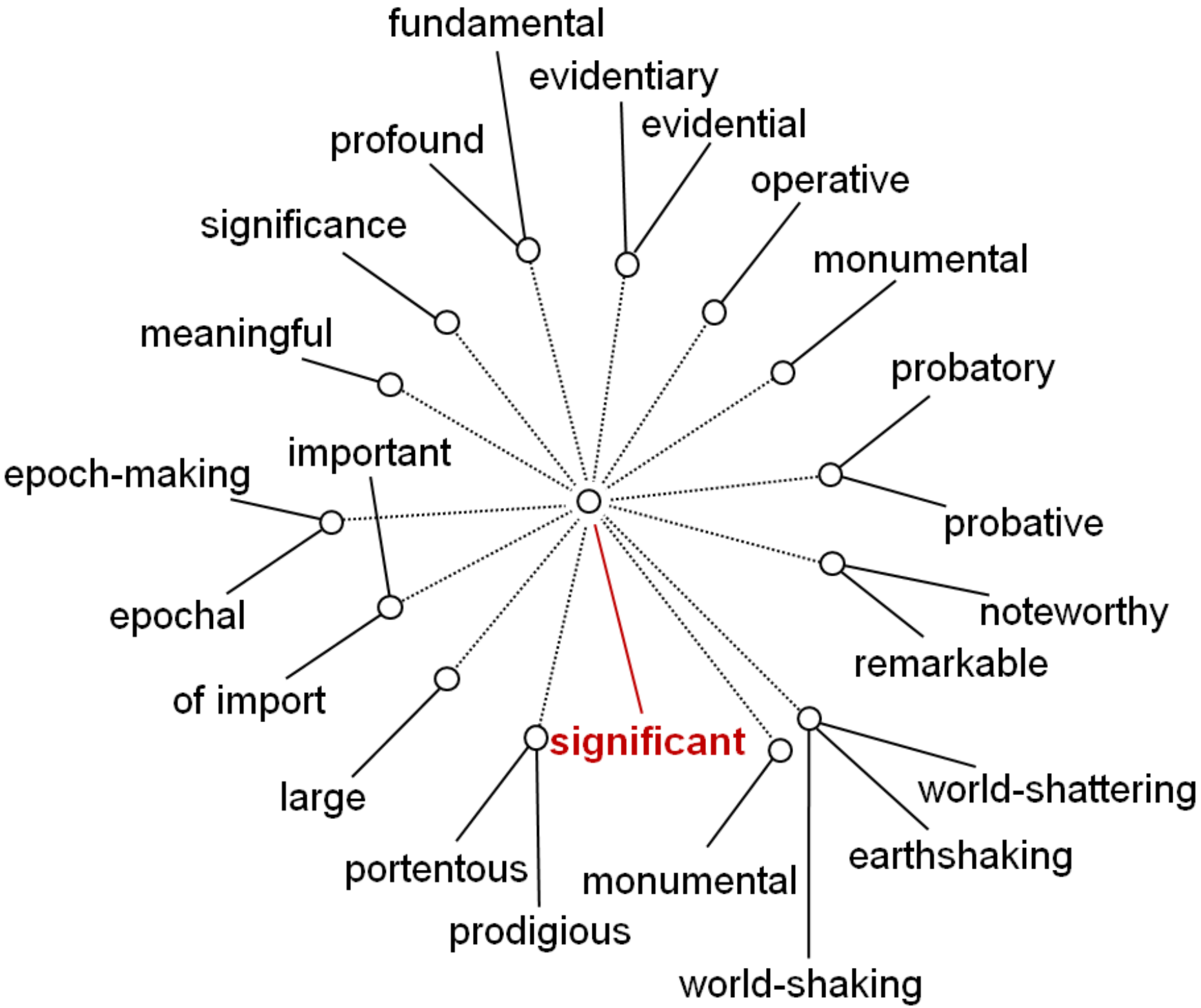
Broken * tests

Incorrectly conducted

Fail to estimate power

Arbitrary level of α (.05, .01)


Broken * tests
Unqualified “significant”
Legitimizes trivial topics
Analysis as camouflage


fundamental
evidentiary
evidential
profound
operative
significance
monumental
meaningful
probatory
important
epoch-making
probative
noteworthy
epochal
remarkable
of import
significant
large
world-shattering
earthshaking
portentous
monumental
prodigious
world-shaking


Broken*tests
Inhibits new learning
Great p value blank-out
Tunnel vision
ROAD CLOSED

Broken * tests

Researcher *df*

False-positive psychology

Junkyard of false positives

Simmons et al. (2011)

Broken* tests

Widely misunderstood

Students ≈ professors

Cycle of misinformation

The textbooks are wrong. The teaching is wrong. The seminar you just attended is wrong. The most prestigious journal in your scientific field is wrong.

Ziliak & McCloskey (2008)


5

Odds against chance

Inverse probability

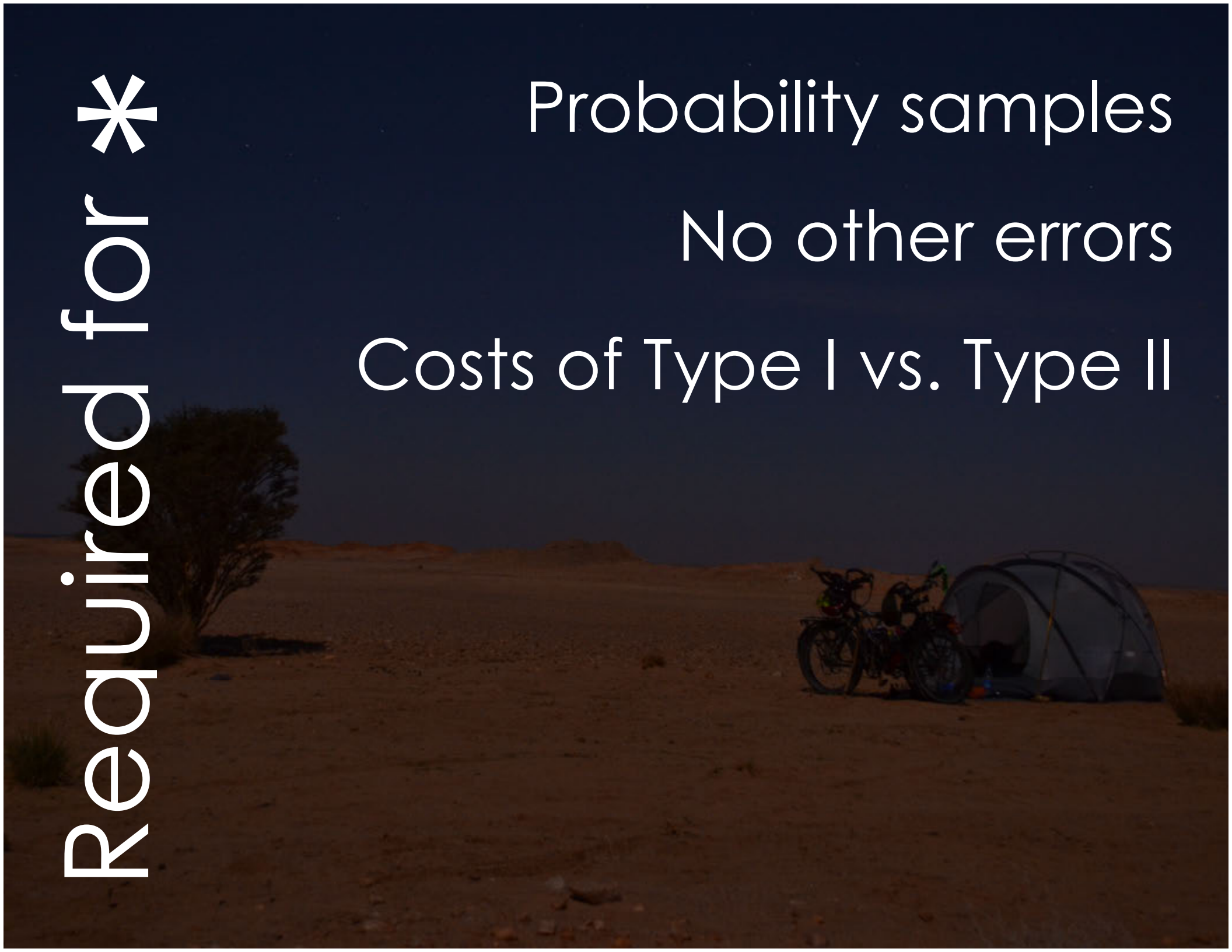
Local Type I error

Replicability

Validity

Fallacy	Ψ profs[a]	Ugrads
Odds against chance	—	72%
Inverse probability	17–36%	35
Local Type I error	67–73	45
Replicability	37–60	42
Validity	33–66	15

[a]Haller & Krauss (2002), Oakes (1986)


Required for *
Probability samples
No other errors
Costs of Type I vs. Type II

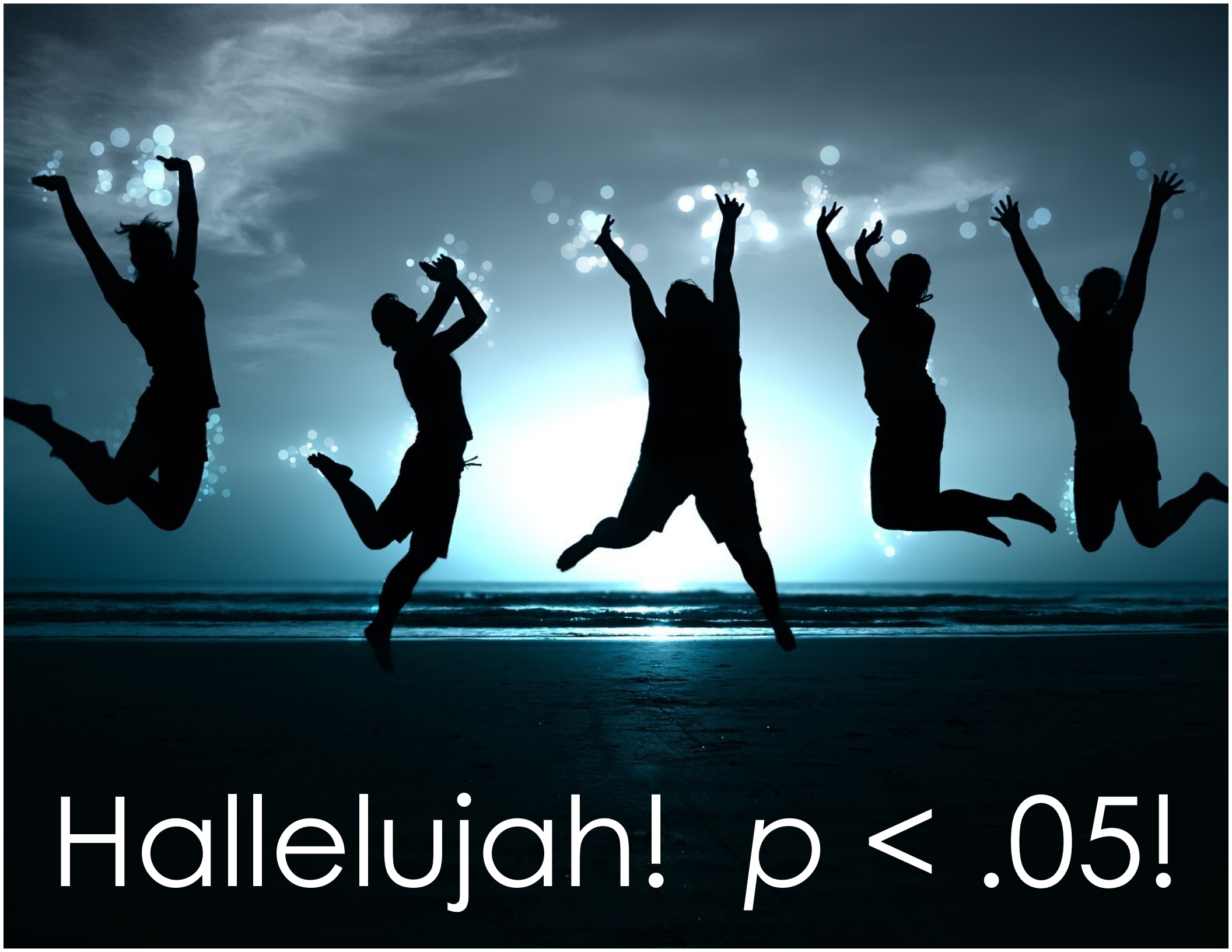
Required for *

Set α intelligently

Estimate power

Verify assumptions

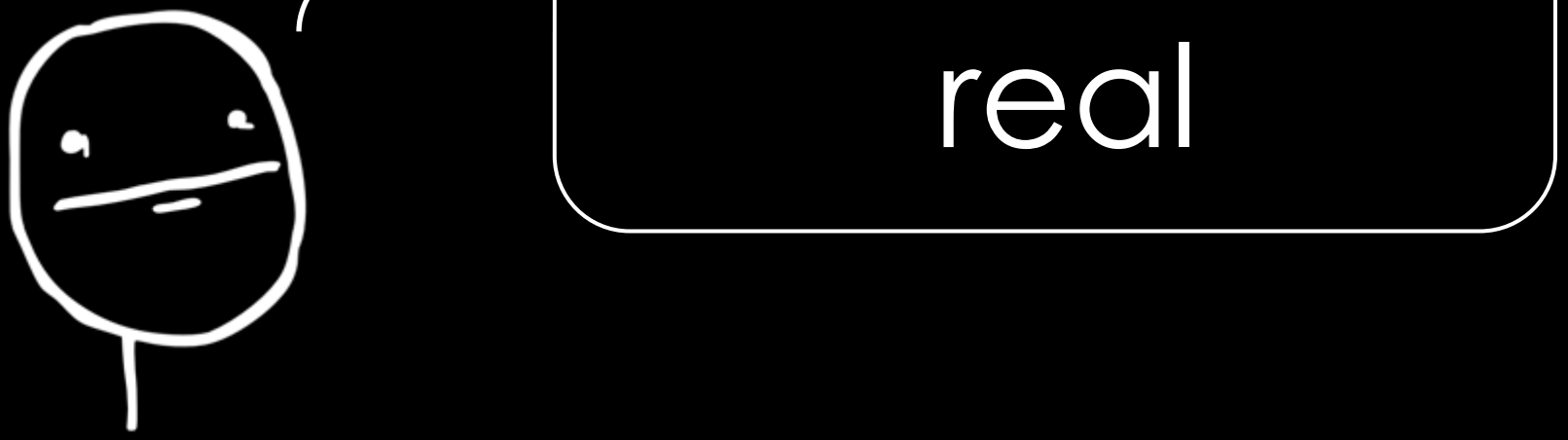

Required for *
Plausible nil H_0
Not misinterpreted
Replicate


Hallelujah! $p < .05$!


ERROR

$t\ (28) = 2.37,$

$p = .025113332794$




AN INCONVENIENT TRUTH
A REASSURING LIE
THE CHRISTIAN SCIENCE MONITOR
Bennett

Statistical buffoonery

Sorcery, shamanism

Intellectually dishonest

Lambdin (2012)

Our obsession with statistical tests of significance has made much of our research blatantly unscientific.

Lambdin (2012)

Empirical goals

Existence? (relation)

How much? (oomph)

So what? (substantive)

War is peace, you must report *p*

When ya gotta

Just report *p*

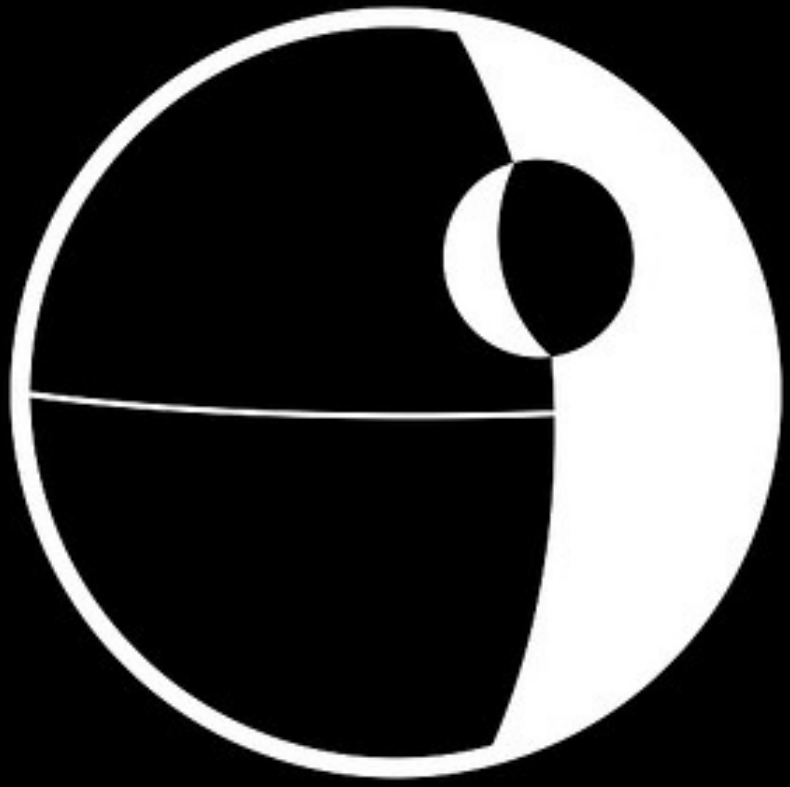
Do not dichotomize

No "significant" or *

Hurlbert & Lombardi (2009)

Source	*SS*	*df*	*MS*	*F*	R^2
Between (*A*)	40.00	2	20.00	3.64[a]	.37 (0–.60)[b]
Within (error)	66.00	12	5.50		
Total	106.00	14			

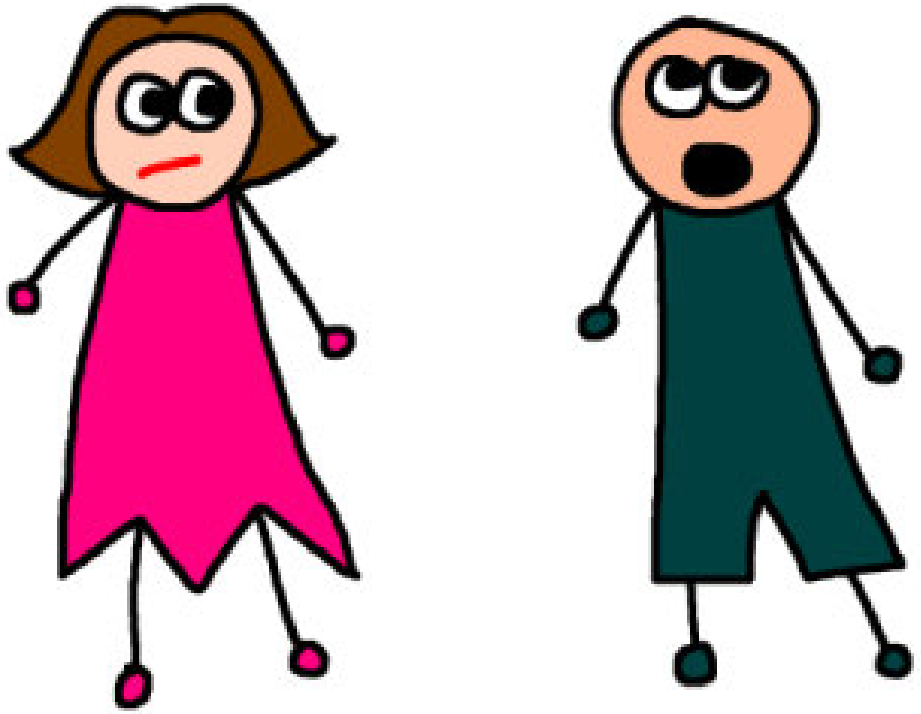
[a]*p* = .058 [b]95% confidence interval



TOO BIG TO FAIL ?


Change

BLAH* BLAH** BLAH***



* $p < .05$ ** $p < .01$ *** $p < .001$


thank you
come again