MiRoR
Methods in
Research on Research

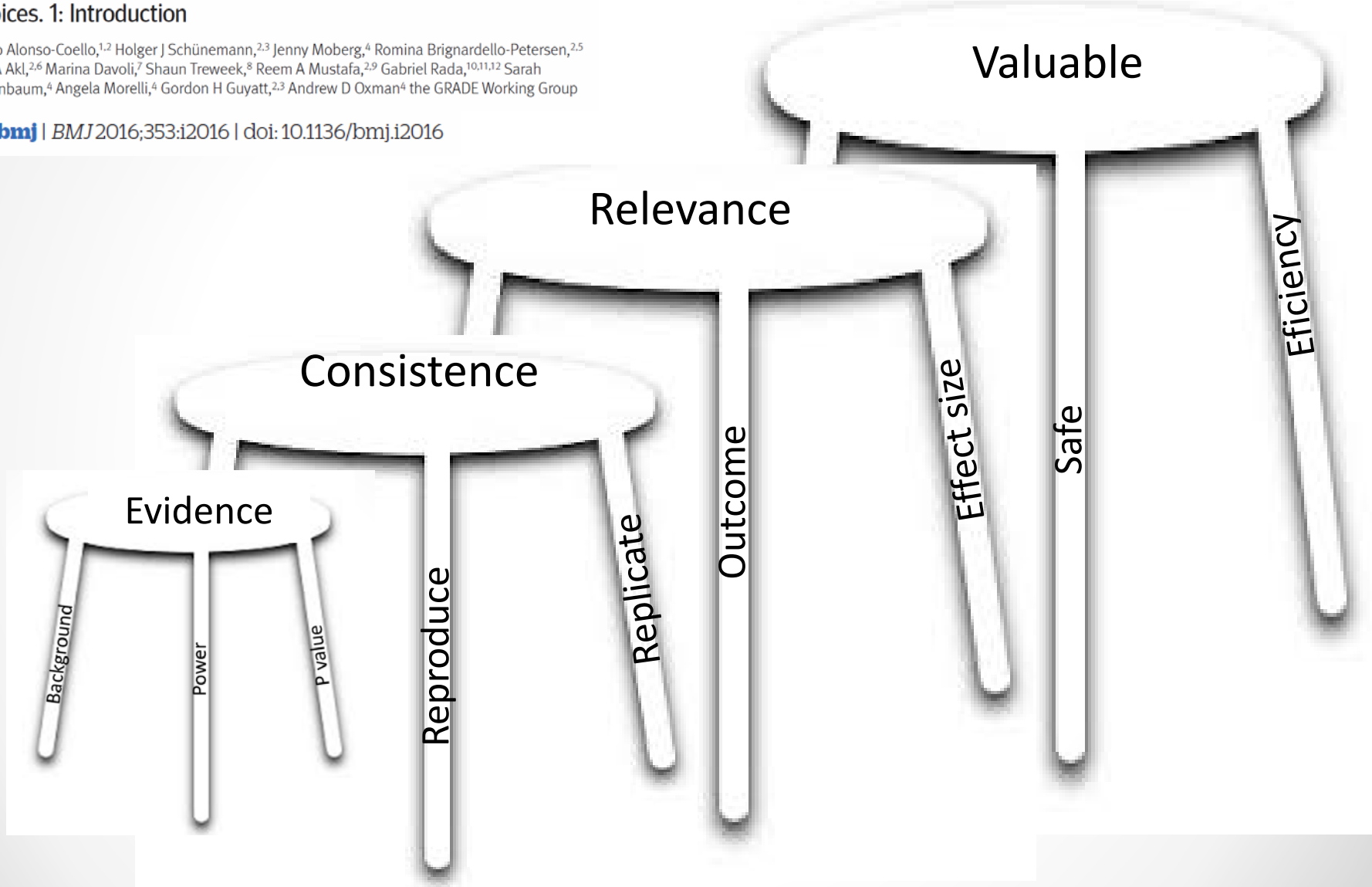
P value
(by the ASA statement)

UNIVERSITAT POLITÈCNICA DE CATALUNYA
BARCELONATECH
Departament d'Estadística
i Investigació Operativa

GRADE Evidence to Decision (EtD) frameworks: a systematic and transparent approach to making well informed healthcare choices. 1: Introduction

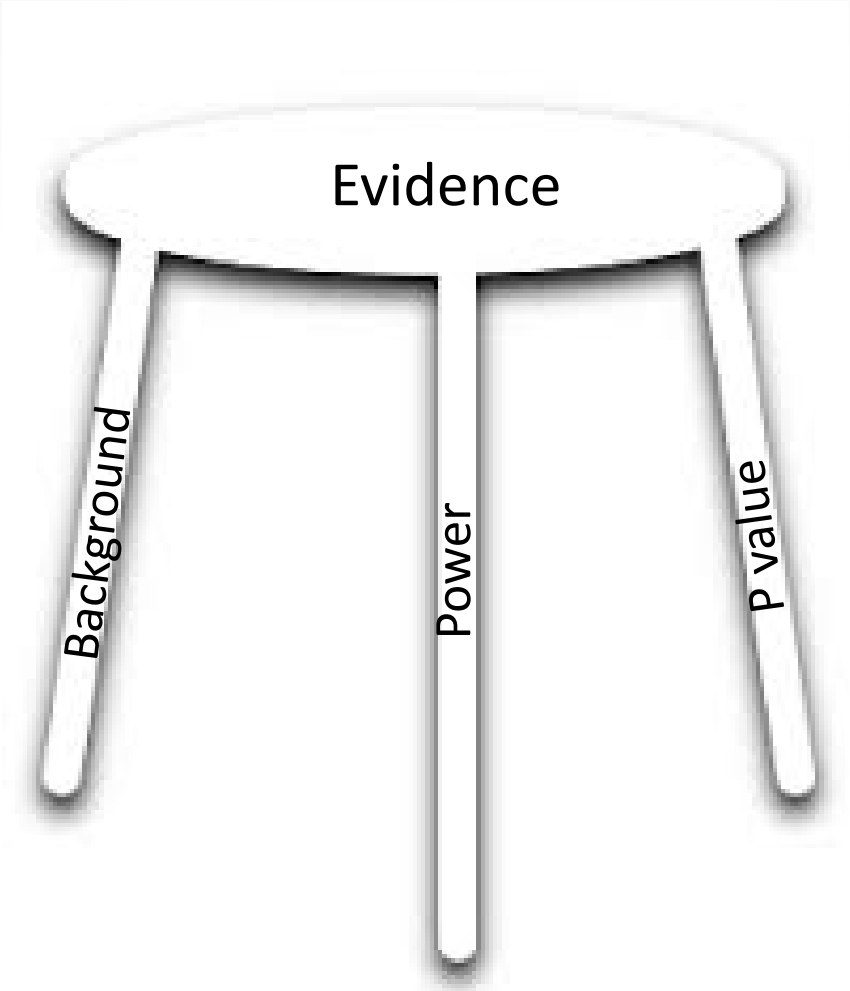
Pablo Alonso-Coello,[1,2] Holger J Schünemann,[2,3] Jenny Moberg,[4] Romina Brignardello-Petersen,[2,5] Elie A Akl,[2,6] Marina Davoli,[7] Shaun Treweek,[8] Reem A Mustafa,[2,9] Gabriel Rada,[10,11,12] Sarah Rosenbaum,[4] Angela Morelli,[4] Gordon H Guyatt,[2,3] Andrew D Oxman[4] the GRADE Working Group

thebmj | *BMJ* 2016;353:i2016 | doi: 10.1136/bmj.i2016

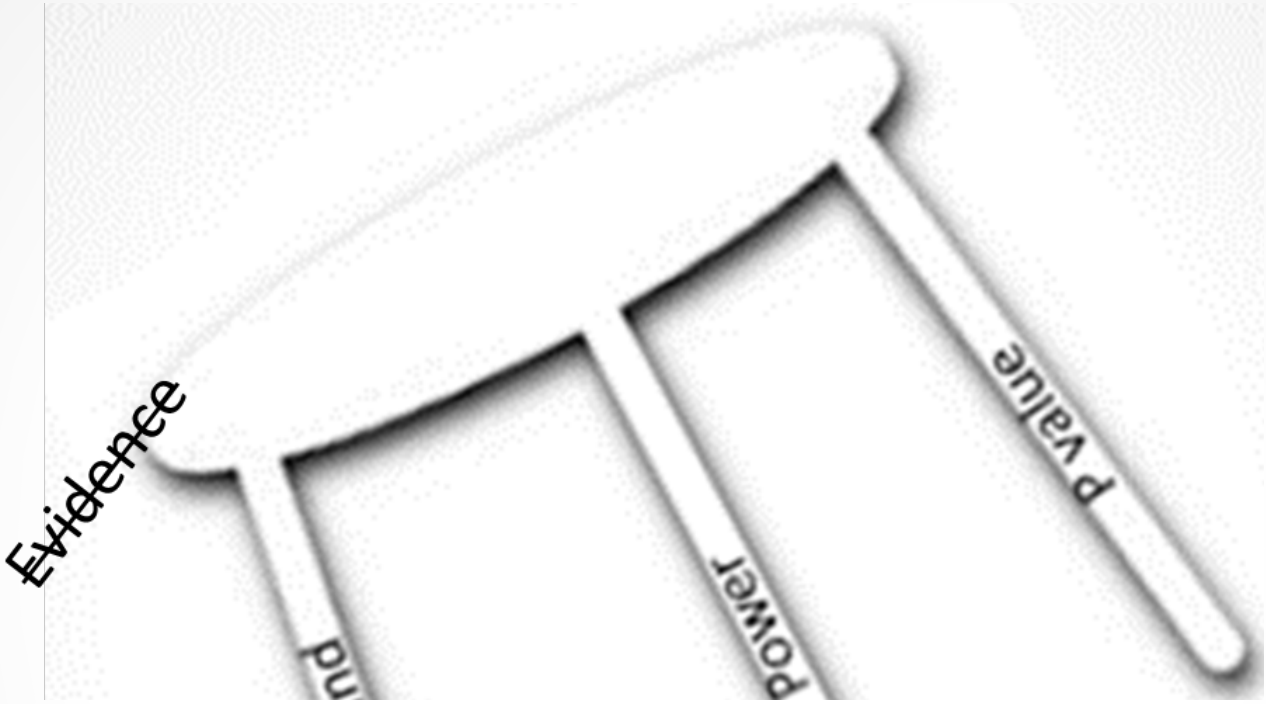


V A L U E:

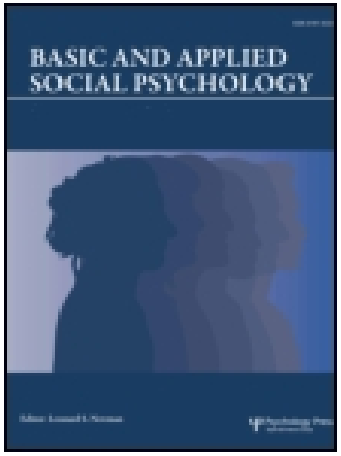
S T A T I S T I C S H E A L T H



Scientifically tested ⟷ Background, Power, P<0.05



Scientifically tested ≠ P<0.05



Psychology journal bans *P* values

Test for reliability of results 'too easy to pass', say editors.

nature

ScienceNews
MAGAZINE OF THE SOCIETY FOR SCIENCE & THE PUBLIC

P value ban: small step for a journal, giant leap for science

False-Positives, p-Hacking, Statistical Power, and Evidential Value

Leif D. Nelson
University of California, Berkeley
Haas School of Business

Summer Institute
June 2014

BITSS
Berkeley Initiative for Transparency in the Social Sciences

Six Ways to p-Hack

1. Stop collecting data once $p<.05$
2. Analyze many measures, but report only those with $p<.05$.
3. Collect and analyze many conditions, but only report those with $p<.05$.
4. Use covariates to get $p<.05$.
5. Exclude participants to get $p<.05$.
6. Transform the data to get $p<.05$.

THE LANCET

Perspective

Improving Patient Safety through Transparency

The NEW ENGLAND JOURNAL of MEDICINE

Research: increasing value, reducing waste

Research: increasing value, reducing waste 1

How to increase value and reduce waste when research priorities are set

Iain Chalmers, Michael B Bracken, Ben Djulbegovic, Silvio Garattini, Jonathan Grant, A Metin Gülmezoglu, David W Howells, John P A Ioannidis, Sandy Oliver

Research: increasing value, reducing waste 2

Increasing value and reducing waste in research design, conduct, and analysis

Research: increasing value, reducing waste 3

Increasing value and reducing waste in biomedical research regulation and management

Research: increasing value, reducing waste 4

Increasing value and reducing waste: addressing inaccessible research

Research: increasing value, reducing waste 5

Reducing waste from incomplete or unusable reports of biomedical research

BMJ

EDITORIAL

The scandal of poor medical research

BMJ 1994; 308 doi: http://dx.doi.org/10.1136/bmj.308.6924.283 (Published 29 January 1994)
Cite this as: *BMJ* 1994;308:283

EDITORIAL

The scandal of poor epidemiological research

BMJ 2004; 329 doi: http://dx.doi.org/10.1136/bmj.329.7471.868 (Published 14 October 2004
Cite this as: *BMJ* 2004;329:868

Is animal research sufficiently evidence based to be a cornerstone of biomedical research?

nature

NIH plans to enhance reproducibility

Francis S. Collins and **Lawrence A. Tabak** discuss initiatives that the US National Institutes of Health is exploring to restore the self-correcting nature of preclinical research.

MEDICINA CLINICA

Declaración de transparencia para las publicaciones científicas
Declaration of transparency for scientific publications

Why Most Published Research Findings Are False

John P. A. Ioannidis

Table 4. PPV of Research Findings for Various Combinations of Power (1 – β), Ratio of True to Not-True Relationships *(R)*, and Bias *(u)*

1 – β	*R*	*u*	Practical Example	PPV
0.80	1:1	0.10	Adequately powered RCT with little bias and 1:1 pre-study odds	0.85
0.95	2:1	0.30	Confirmatory meta-analysis of good-quality RCTs	0.85
0.80	1:3	0.40	Meta-analysis of small inconclusive studies	0.41
0.20	1:5	0.20	Underpowered, but well-performed phase I/II RCT	0.23
0.20	1:5	0.80	Underpowered, poorly performed phase I/II RCT	0.17
0.80	1:10	0.30	Adequately powered exploratory epidemiological study	0.20
0.20	1:10	0.30	Underpowered exploratory epidemiological study	0.12
0.20	1:1,000	0.80	Discovery-oriented exploratory research with massive testing	0.0010
0.20	1:1,000	0.20	As in previous example, but with more limited bias (more standardized)	0.0015

AMERICAN STATISTICAL ASSOCIATION
Promoting the Practice and Profession of Statistics

The ASA's statement on p-values: context, process, and purpose

Ronald L. Wasserstein and Nicole A. Lazar

1. P-values can indicate how incompatible the data are with a specified statistical model.

2. P-values do NOT measure the probability that the studied hypothesis is true, or the probability that the data were produced by random chance alone.

3. Scientific conclusions and business or policy decisions should NOT be based only on whether a p-value passes a specific threshold.

4. Proper inference requires full reporting and transparency

5. A p-value, or statistical significance, does NOT measure the size of an effect or the importance of a result.

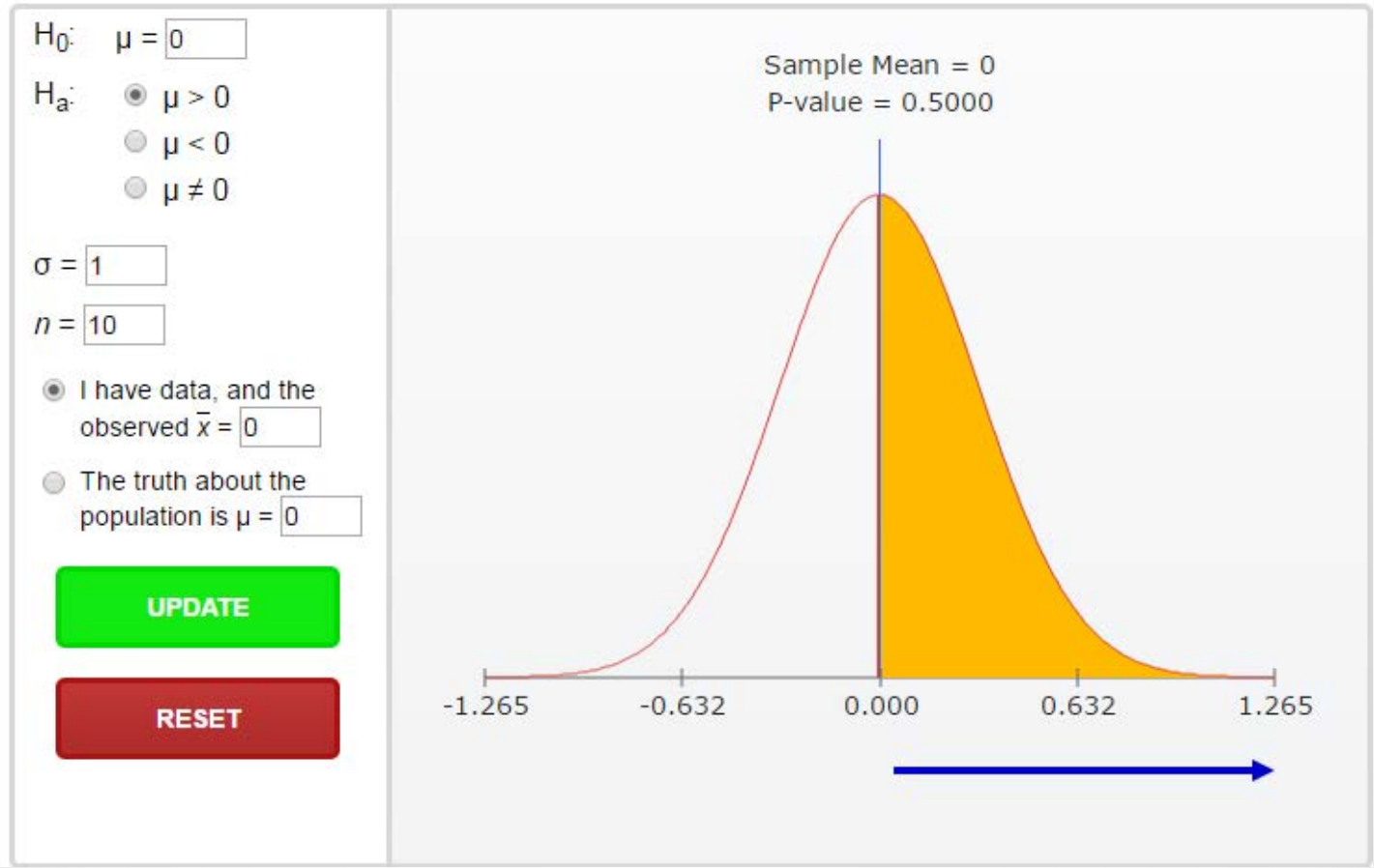
6. By itself, a p-value does NOT provide a good measure of evidence regarding a model or hypothesis.

Declaración de la ASA sobre la significación estadística y los p-valores

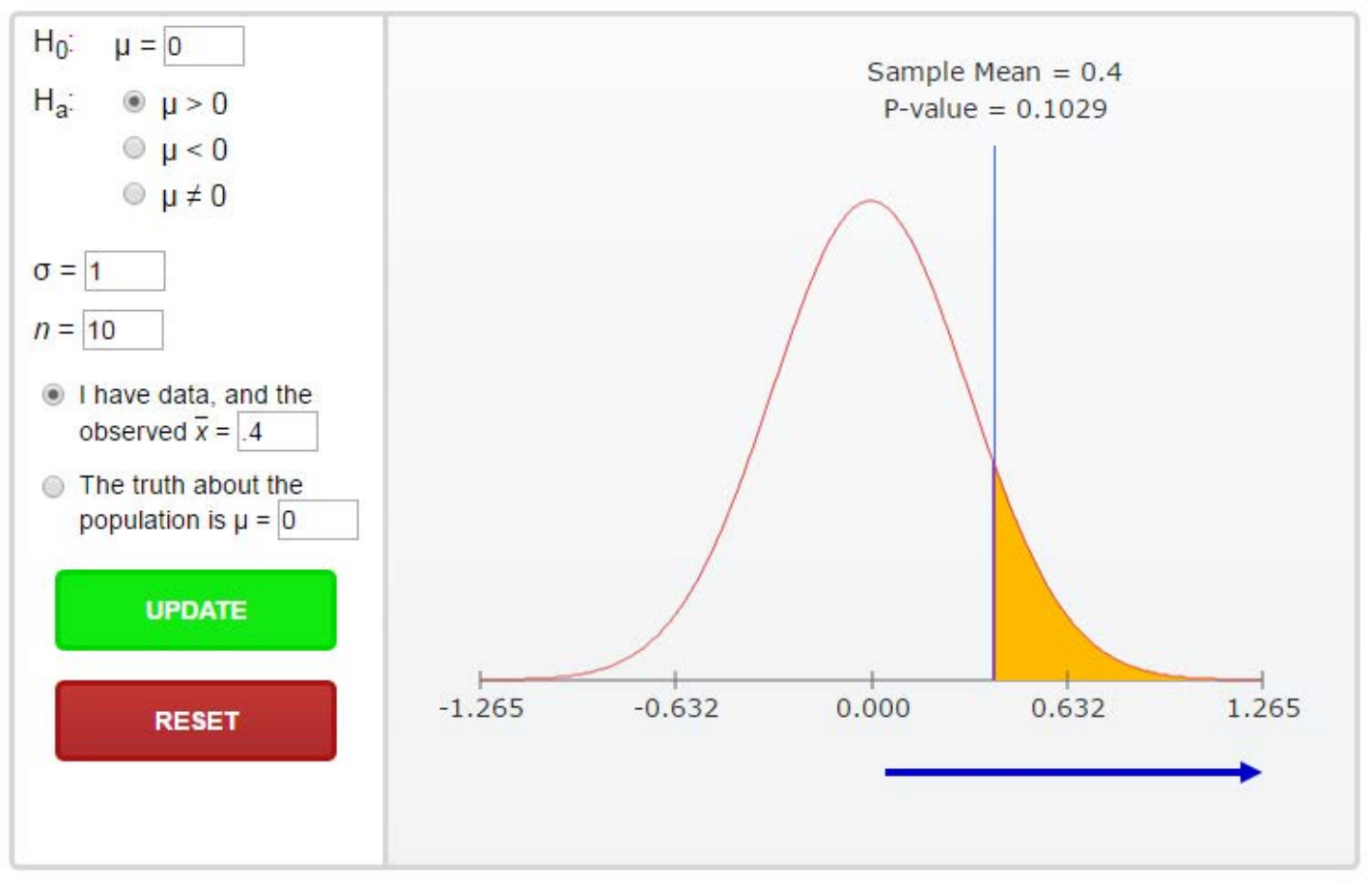
SOCIETAT CATALANA D'ESTADÍSTICA

http://soce.iec.cat/

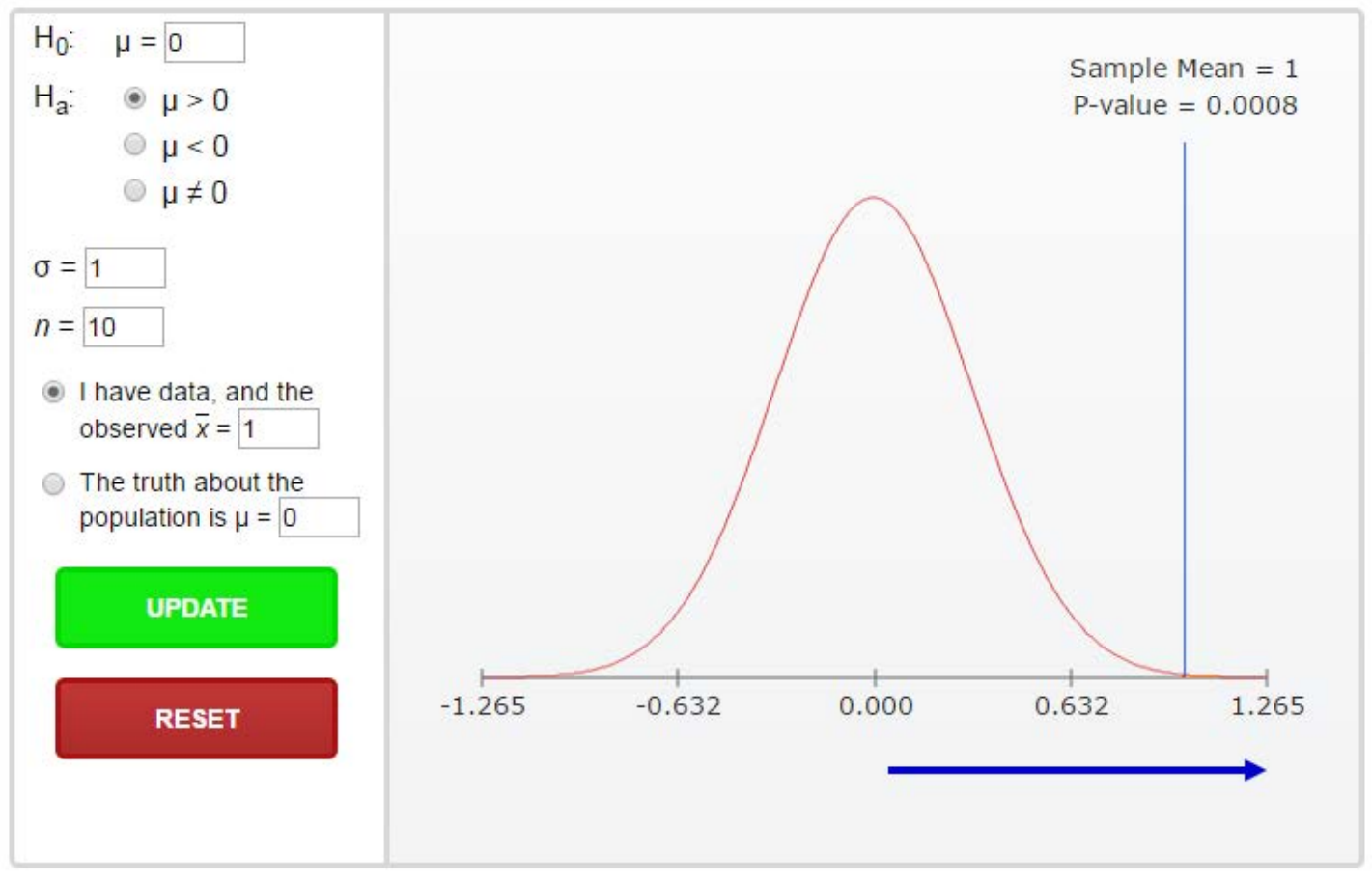
1. *P-values can indicate how incompatible the data are with a specified statistical model*



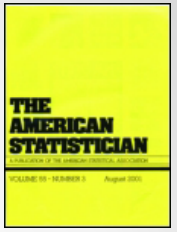
1. *P-values can indicate how incompatible the data are with a specified statistical model*



1. *P-values can indicate how incompatible the data are with a specified statistical model*



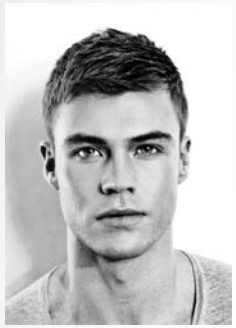
ASA
AMERICAN STATISTICAL ASSOCIATION
Promoting the Practice and Profession of Statistics

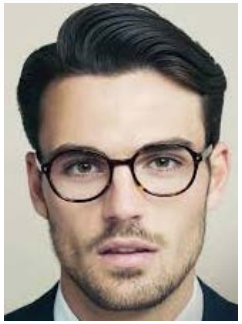


The ASA's statement on p-values: context, process, and purpose

Ronald L. Wasserstein and Nicole A. Lazar

1. *P-values can indicate how incompatible the data are with a specified statistical model*
2. *P-values do NOT measure the probability that the studied hypothesis is true*



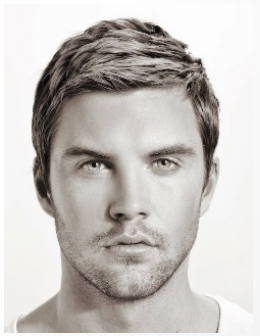


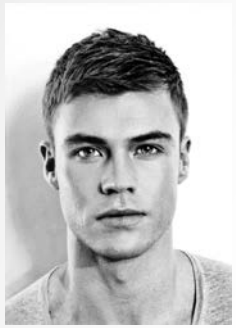


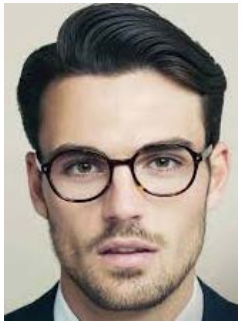










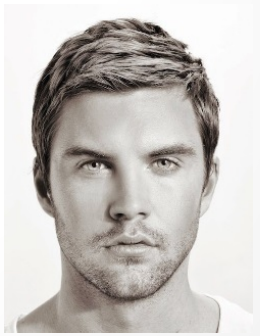




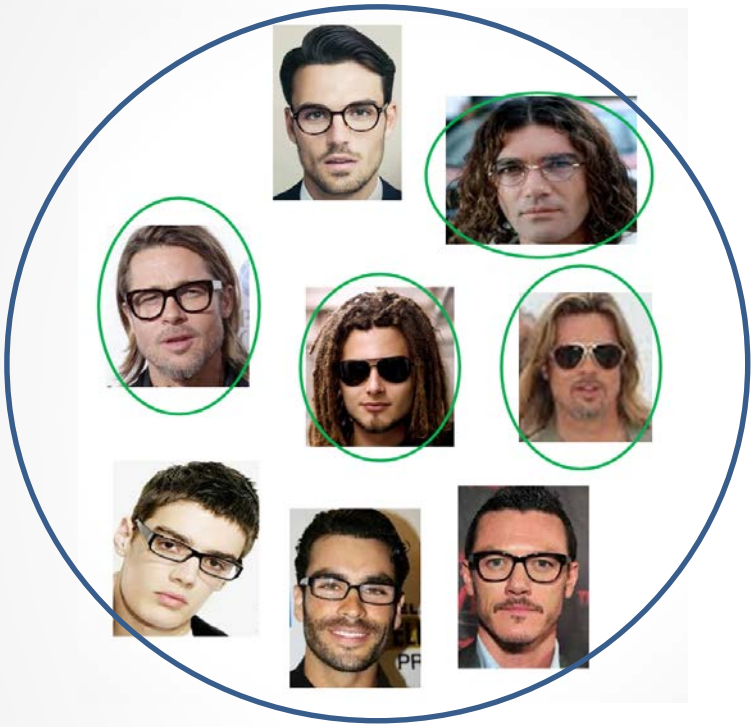


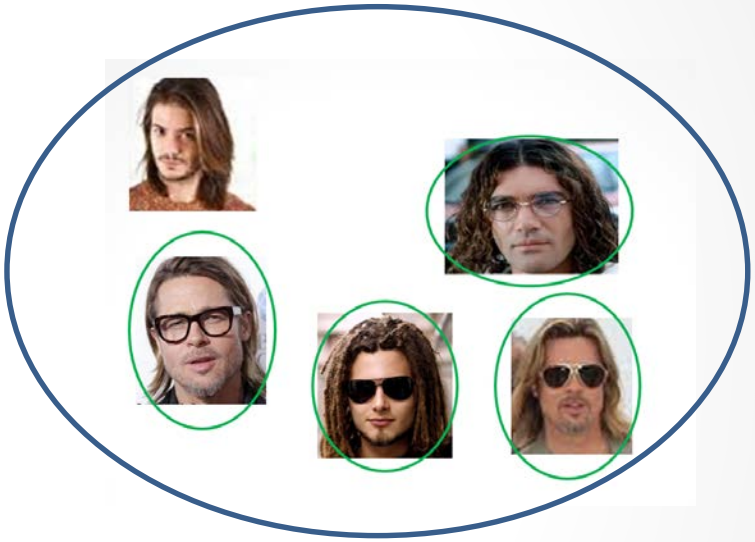






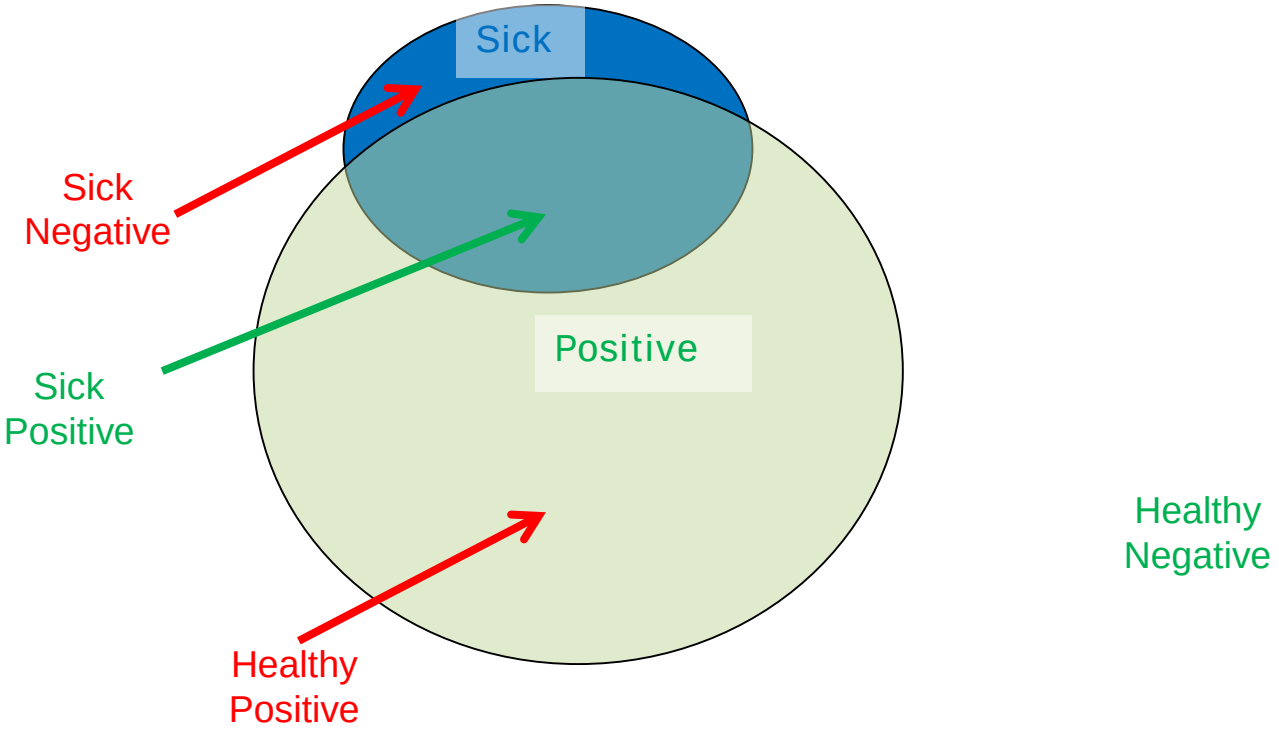
Denominator changed

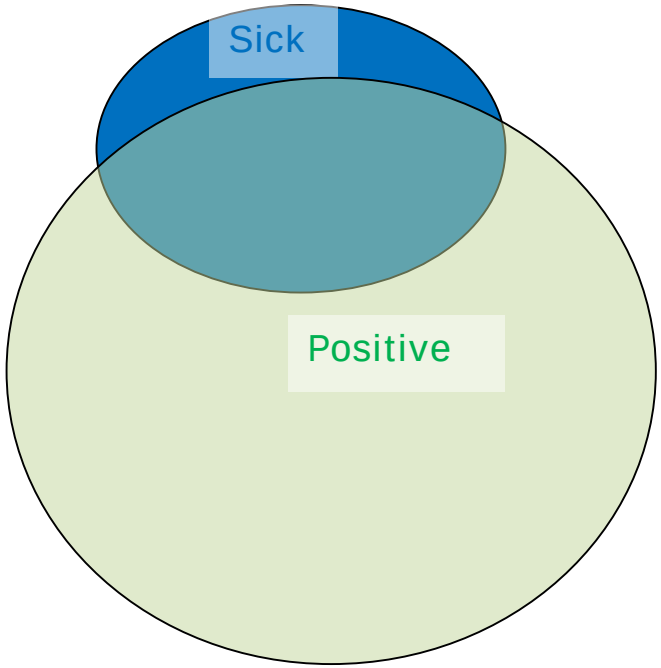




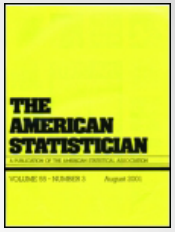
$$\frac{4}{8} = 50\%$$

$$\frac{4}{5} = 80\%$$


Sick
Sick
Negative
Positive
Sick
Positive
Healthy
Negative
Healthy
Positive


Sick
Positive

ASA
AMERICAN STATISTICAL ASSOCIATION
Promoting the Practice and Profession of Statistics



1. *P-values can indicate how incompatible the data are with a specified statistical model*
2. *P-values do NOT measure the probability that the studied hypothesis is true*

Decision errors and risks		Decision	
		A_0	A_1
Reality	H_0	$1-\alpha$	Type **I** risk α
	H_1	Type **II** risk β	Power = $1-\beta$

Confusion Over Measures of Evidence (p's) Versus Errors (α's) in Classical Statistical Testing

Raymond HUBBARD and M. J. BAYARRI

 DOI: 10.1198/0003130031856 The American Statistician, August 2003, Vol. 57, No. 3

ELSEVIER

Seminars in HEMATOLOGY

A Dirty Dozen: Twelve *P*-Value Misconceptions

Steven Goodman

The *P* value is a measure of statistical evidence that appears in virtually all medical research papers. Its interpretation is made extraordinarily difficult because it is not part of any formal system of statistical inference. As a result, the *P* value's inferential meaning is widely and often wildly misconstrued, a fact that has been pointed out in innumerable papers and books appearing since at least the 1940s. This commentary reviews a dozen of these common misinterpretations and explains why each is wrong. It also reviews the possible consequences of these improper understandings or representations of its meaning. Finally, it contrasts the *P* value with its Bayesian counterpart, the Bayes' factor, which has virtually all of the desirable properties of an evidential measure that the *P* value lacks, most notably interpretability. The most serious consequence of this array of *P*-value misconceptions is the false belief that the probability of a conclusion being in error can be calculated from the data in a single experiment without reference to external evidence or the plausibility of the underlying mechanism.
Semin Hematol 45:135-140

Eur J Epidemiol (2016) 31:337–350
DOI 10.1007/s10654-016-0149-3

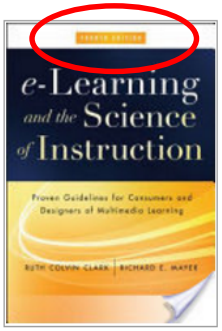
CrossMark

ESSAY

Statistical tests, *P* values, confidence intervals, and power: a guide to misinterpretations

Sander Greenland[1] · Stephen J. Senn[2] · Kenneth J. Rothman[3] · John B. Carlin[4] · Charles Poole[5] · Steven N. Goodman[6] · Douglas G. Altman[7]

e-Learning and the Science of Instruction:



Ruth C. Clark, Richard E. Mayer

John Wiley & Sons, 19 feb. 2016 - 528 páginas

★★★★½

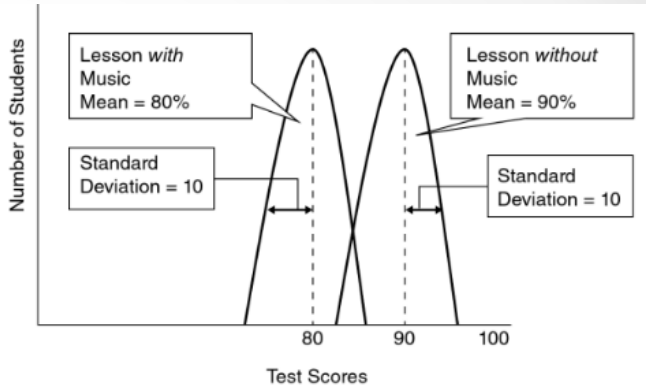
2 Reseñas

G+1 0

The essential e-learning design manual, updated witl

e-Learning and the Science of Instruction is the ultimate e-learning has grown to account for at least 40% of all tra learning effectiveness and efficiency. This guide provides

Más »



Statistical Significance: Probability Less Than .05

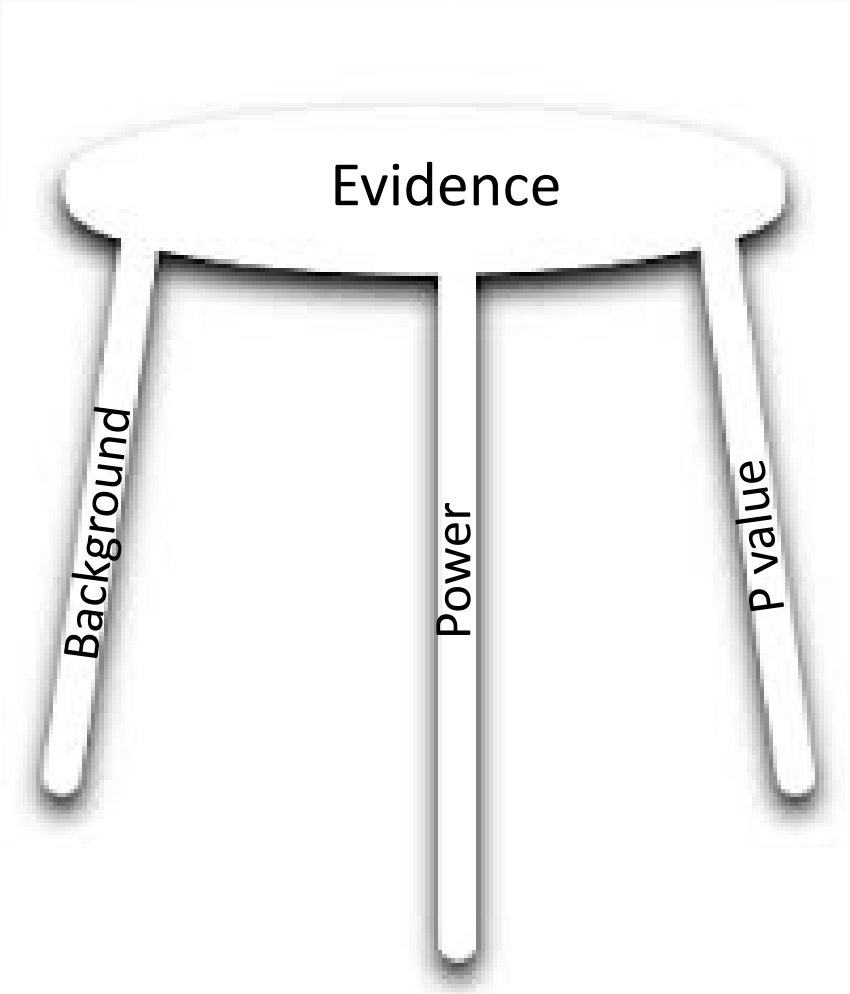
In other words, if you concluded there is a difference in test performance between the groups, there is less than a 5 percent chance that you are wrong and more than a 95 percent chance that you are right.

Why Most Published Research Findings Are False

John P. A. Ioannidis

Table 4. PPV of Research Findings for Various Combinations of Power (1 – β), Ratio of True to Not-True Relationships *(R)*, and Bias *(u)*

1 – β	*R*	*u*	Practical Example	PPV
0.80	1:1	0.10	Adequately powered RCT with little bias and 1:1 pre-study odds	0.85
0.95	2:1	0.30	Confirmatory meta-analysis of good-quality RCTs	0.85
0.80	1:3	0.40	Meta-analysis of small inconclusive studies	0.41
0.20	1:5	0.20	Underpowered, but well-performed phase I/II RCT	0.23
0.20	1:5	0.80	Underpowered, poorly performed phase I/II RCT	0.17
0.80	1:10	0.30	Adequately powered exploratory epidemiological study	0.20
0.20	1:10	0.30	Underpowered exploratory epidemiological study	0.12
0.20	1:1,000	0.80	Discovery-oriented exploratory research with massive testing	0.0010
0.20	1:1,000	0.20	As in previous example, but with more limited bias (more standardized)	0.0015

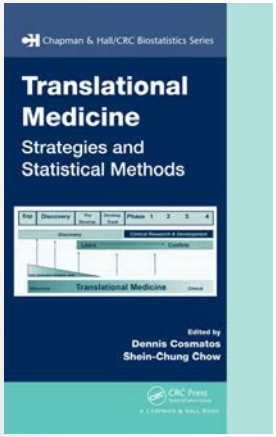

Evidence
Background
Power
P value

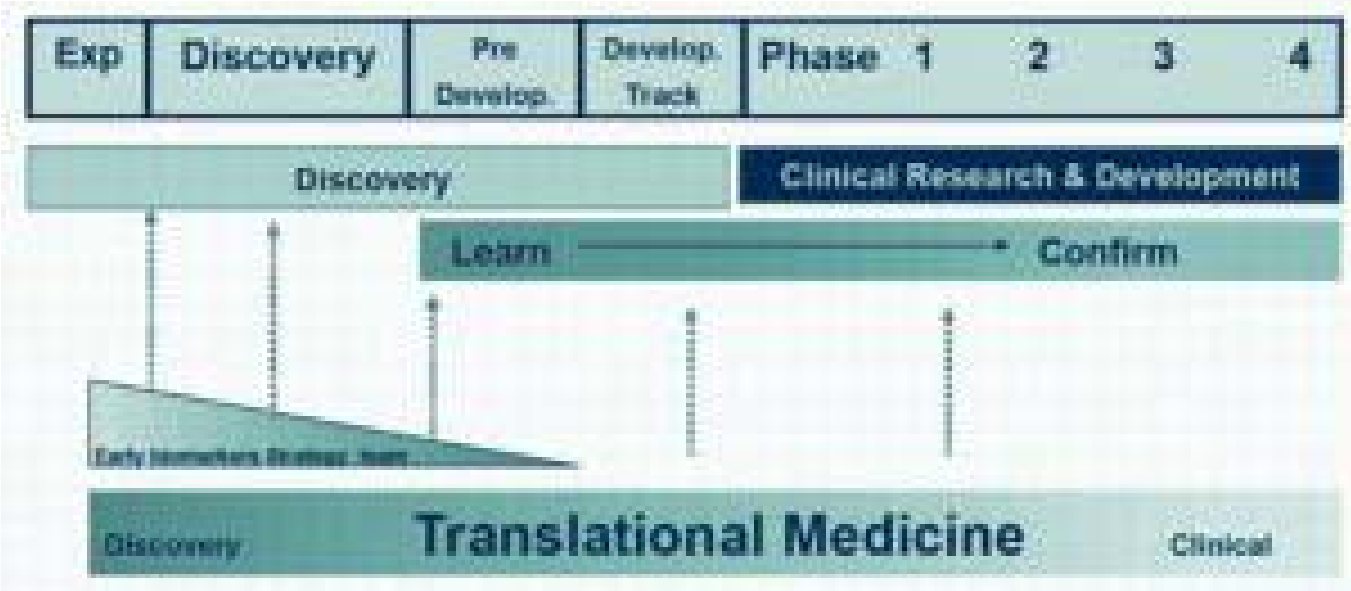
1.- Background

Steps within R & D

R & D

Discovery → Evaluation → Validation

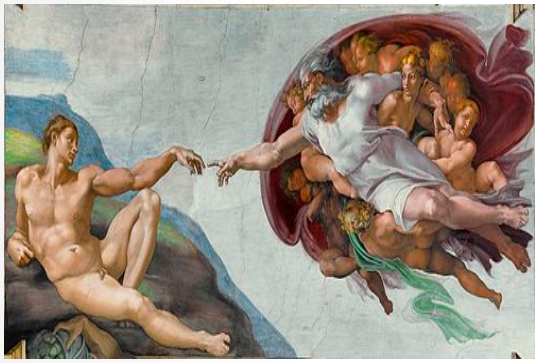


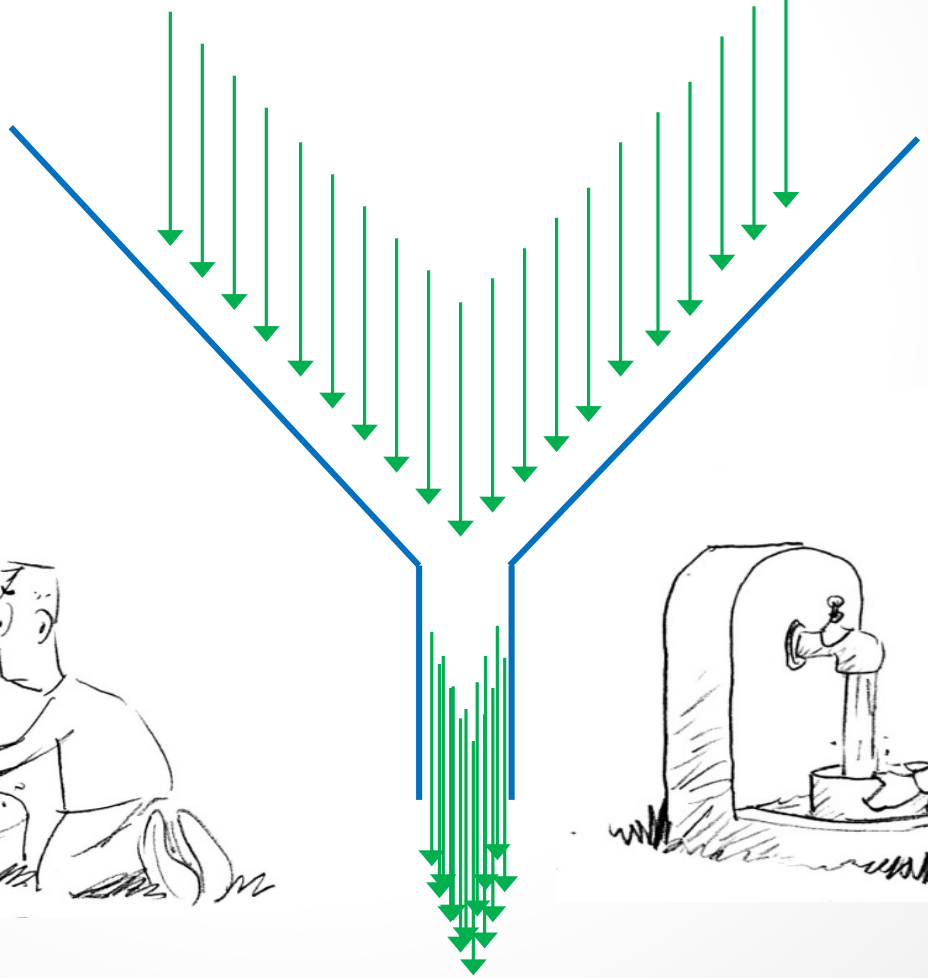


2.- Power

Filter capacity to select good ideas

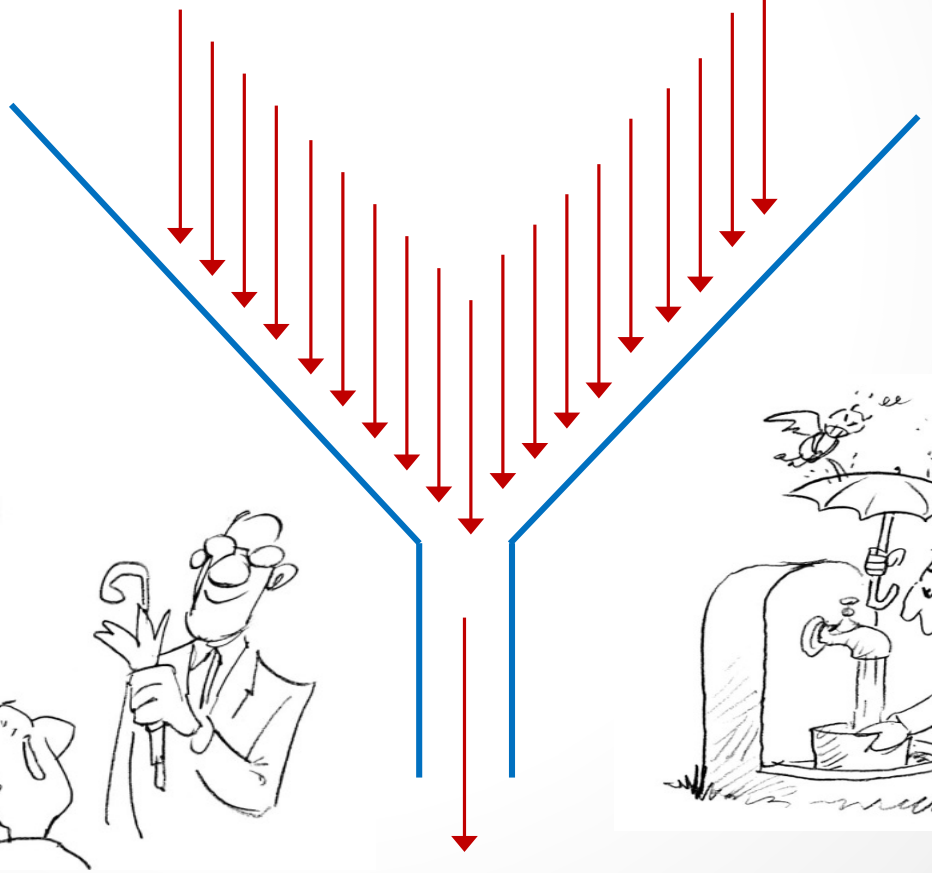
Good ideas

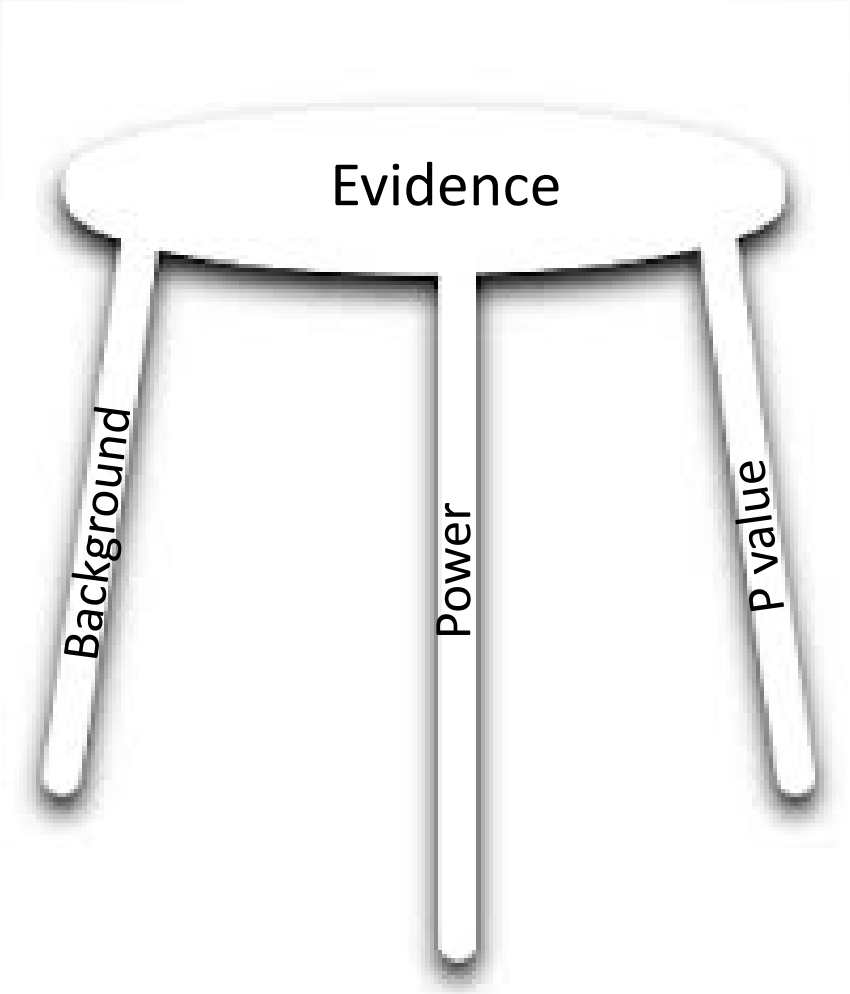




3.- P value

Filter capacity to discard bad ideas




Evidence
Background
Power
P value

Credibilidad de la ciencia

Calculadora | Curva de credibilidad | Ayuda

Esta aplicación pretende mostrarle que la posibilidad de repetir un resultado depende de:

1. la base científica de la hipótesis (mayor justificación previa, mayor credibilidad);
2. la capacidad del estudio para alcanzar sus objetivos según su diseño, tamaño y precisión (mayor potencia, mayor credibilidad);
3. el control del riesgo de error por falsas señales positivas (mayor control, mayor credibilidad).

Vd. debe jugar con estos 3 parámetros, y el gráfico adjunto le mostrará cómo cambia el escenario: un efecto real o no, combinado con resultado del análisis significativo o no. Además, también obtendrá la probabilidad de que un resultado "estadísticamente significativo" se repita en el futuro.

Expectativa del efecto

0 — 75 — 100

0 10 20 30 40 50 60 70 80 90 100

Potencia

0 — 80 — 100

0 10 20 30 40 50 60 70 80 90 100

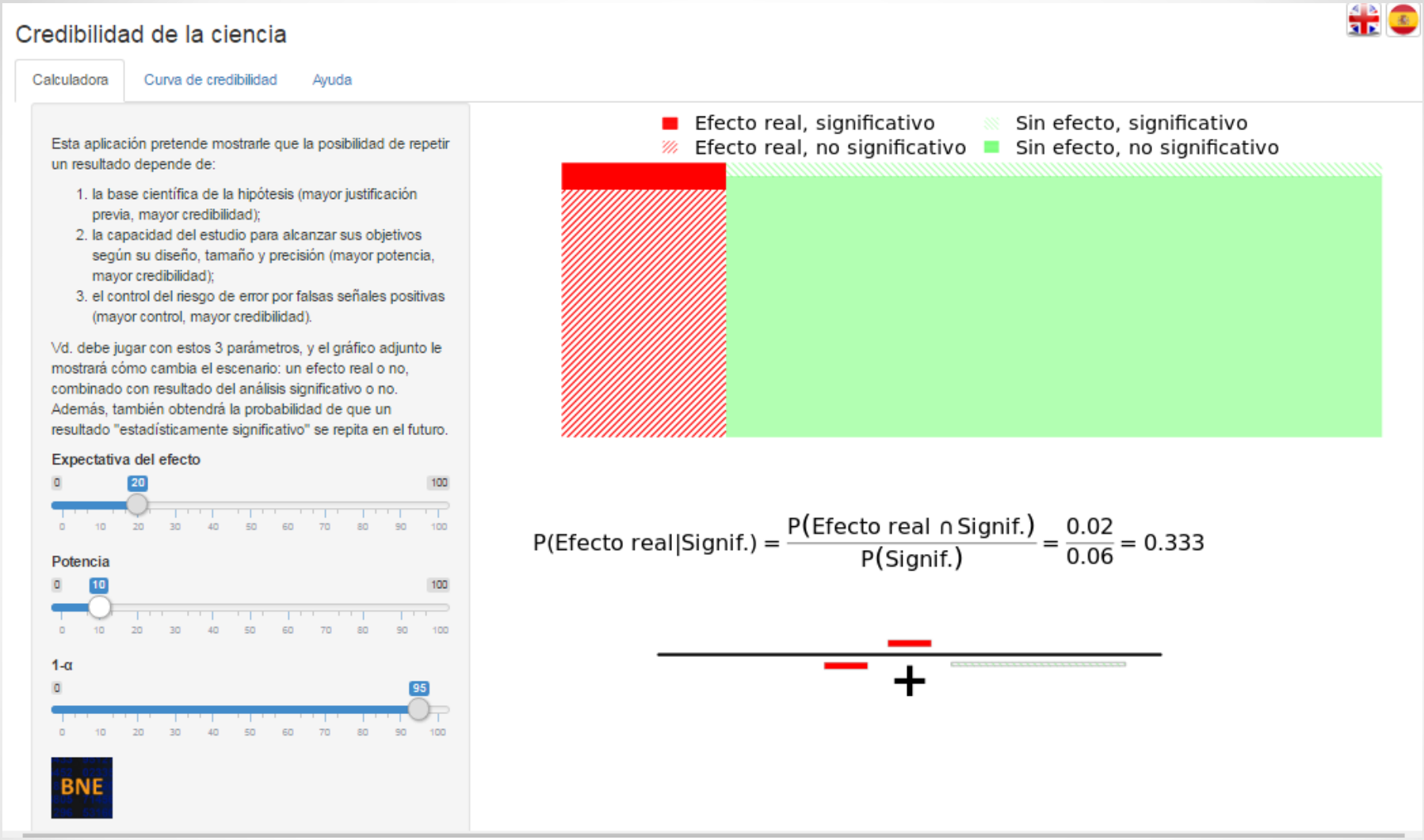
1-α

0 — 95

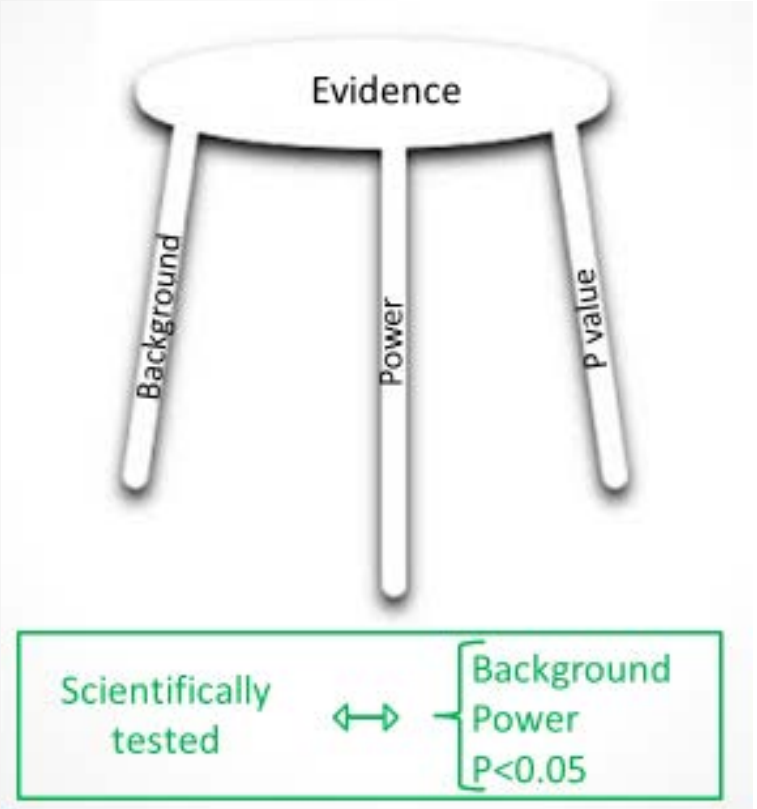
0 10 20 30 40 50 60 70 80 90 100

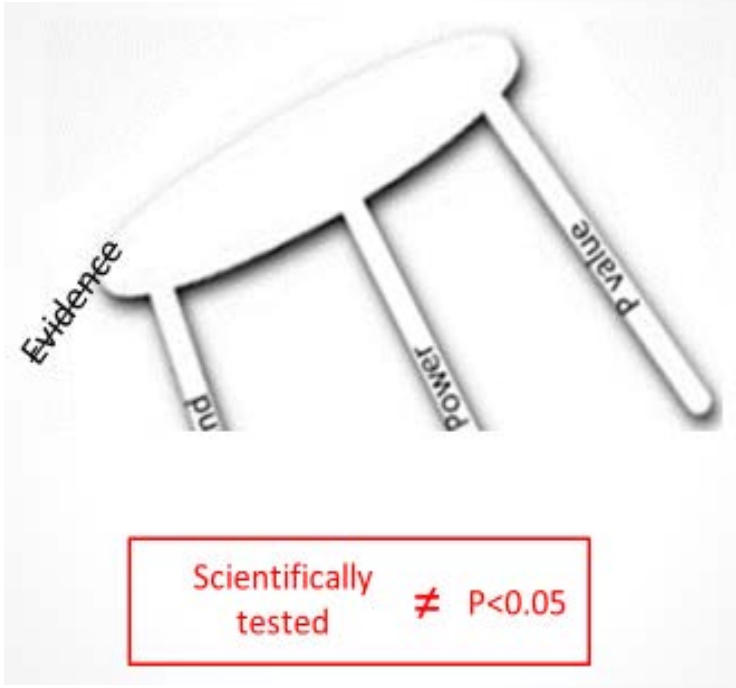
BNE

$$P(\text{Efecto real}|\text{Signif.}) = \frac{P(\text{Efecto real} \cap \text{Signif.})}{P(\text{Signif.})} = \frac{0.6}{0.6125} = 0.98$$



Link to applet: http://shiny-eio.upc.edu/bne/efectos2/


Evidence
Background
Power
P value
Scientifically tested
Background
Power
P<0.05


Evidence
Power
P value
Scientifically tested
≠
P<0.05

The NEW ENGLAND JOURNAL of MEDICINE

ESTABLISHED IN 1812 JUNE 23, 2016 VOL. 374 NO. 25

Adapted Treatment Guided by Interim PET-CT Scan in Advanced Hodgkin's Lymphoma

Peter Johnson, M.D., Massimo Federico, M.D., Amy Kirkwood, M.Sc., Alexander Fosså, M.D., Leanne Berkahn, M.D., Angelo Carella, M.D., Francesco d'Amore, M.D., Gunilla Enblad, M.D., Antonella Franceschetto, M.D., Michael Fulham, M.D., Stefano Luminari, M.D., Michael O'Doherty, M.D., Pip Patrick, Ph.D., Thomas Roberts, B.Sc., Gamal Sidra, M.D., Lindsey Stevens, Paul Smith, M.Sc., Judith Trotman, M.D., Zaid Viney, M.D., John Radford, M.D., and Sally Barrington, M.D.

ABSTRACT

BACKGROUND

We tested interim positron-emission tomography–computed tomography (PET-CT) as a measure of early response to chemotherapy in order to guide treatment for patients with advanced Hodgkin's lymphoma.

METHODS

Patients with newly diagnosed advanced classic Hodgkin's lymphoma underwent a baseline PET-CT scan, received two cycles of ABVD (doxorubicin, bleomycin, vinblastine, and dacarbazine) chemotherapy, and then underwent an interim PET-CT scan. Images were centrally reviewed with the use of a 5-point scale for PET findings. Patients with negative PET findings after two cycles were randomly assigned to continue ABVD (ABVD group) or omit bleomycin (AVD group) in cycles 3 through 6. Those with positive PET findings after two cycles received BEACOPP (bleomycin, etoposide, doxorubicin, cyclophosphamide, vincristine, procarbazine, and prednisone). Radiotherapy was not recommended for patients with negative findings on interim scans. The primary outcome was the difference in the 3-year progression-free survival rate between randomized groups, a noninferiority comparison to exclude a difference of 5 or more percentage points.

RESULTS

A total of 1214 patients were registered; 937 of the 1119 patients (83.7%) who underwent an interim PET-CT scan according to protocol had negative findings. With a median follow-up of 41 months, the 3-year progression-free survival rate and overall survival rate in the ABVD group were 85.7% (95% confidence interval [CI], 82.1 to 88.6) and 97.2% (95% CI, 95.1 to 98.4), respectively; the corresponding rates in the AVD group were 84.4% (95% CI, 80.7 to 87.5) and 97.6% (95% CI, 95.6 to 98.7). The absolute difference in the 3-year progression-free survival rate (ABVD minus AVD) was 1.6 percentage points (95% CI, −3.2 to 5.3). Respiratory adverse events were more severe in the ABVD group than in the AVD group. BEACOPP was given to the 172 patients with positive findings on the interim scan, and 74.4% had negative findings on a third PET-CT scan; the 3-year progression-free survival rate was 67.5% and the overall survival rate 87.8%. A total of 62 patients died during the trial (24 from Hodgkin's lymphoma), for a 3-year progression-free survival rate of 82.6% and an overall survival rate of 95.8%.

CONCLUSIONS

Although the results fall just short of the specified noninferiority margin, the omission of bleomycin from the ABVD regimen after negative findings on interim PET resulted in a lower incidence of pulmonary toxic effects than with continued ABVD but not significantly lower efficacy. (Funded by Cancer Research UK and Others; ClinicalTrials.gov number, NCT00678327.)

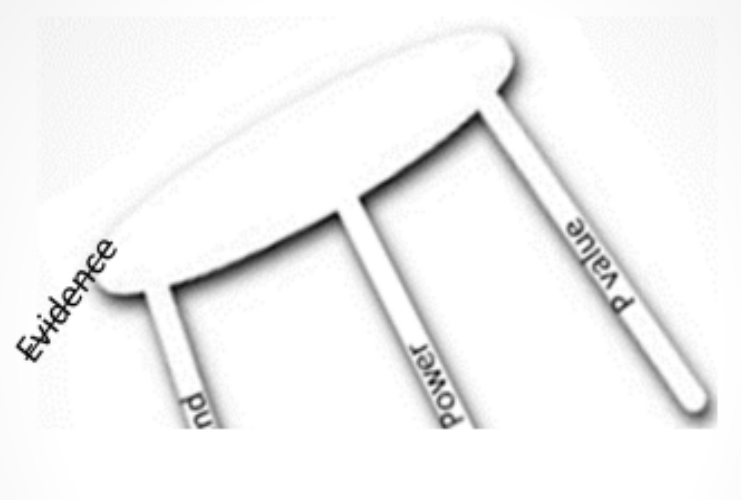
From the Cancer Research UK Centre, University of Southampton, Southampton (P.J.), Cancer Research UK and University College London Cancer Trials Centre (A.K., P.P., T.R., L.S., P.S.) and the PET Imaging Centre, King's College London, King's Health Partners, St. Thomas' Hospital (M.O., Z.V., S.B.), London, the Department of Haematology, Lincoln County Hospital, Lincoln (G.S.), and the Department of Medical Oncology, Christie Hospital, Manchester (J.R.) — all in the United Kingdom; the Department of Diagnostic, Clinical, and Public Health Medicine, University of Modena and Reggio Emilia, Modena (M. Federico, A. Franceschetto, S.L.), the Department of Hematology, San Martino University Hospital, Genoa (A.C.), and Arcispedale Santa Maria Nuova–Istituti di Ricovero e Cura a Carattere Scientifico, Reggio Emilia (S.L.) — all in Italy; the Department of Medical Oncology, Oslo University Hospital, Oslo (A. Fosså); the Department of Haematology, Auckland City Hospital, Auckland, New Zealand (L.B.); the Department of Hematology, Aarhus University Hospital, Aarhus, Denmark (F.A.); the Department of Immunology, Genetics, and Pathology, Uppsala University, Uppsala, Sweden (G.E.); and the Department of Molecular Imaging, Royal Prince Alfred Hospital (M. Fulham), and Concord Repatriation General Hospital, University of Sydney (J.T.), Sydney. Address reprint requests to Dr. Johnson at the Cancer Research UK Centre, University of Southampton, Somers Cancer Research Bldg., Southampton General Hospital, Southampton SO16 6YD, United Kingdom, or at johnsonp@soton.ac.uk.

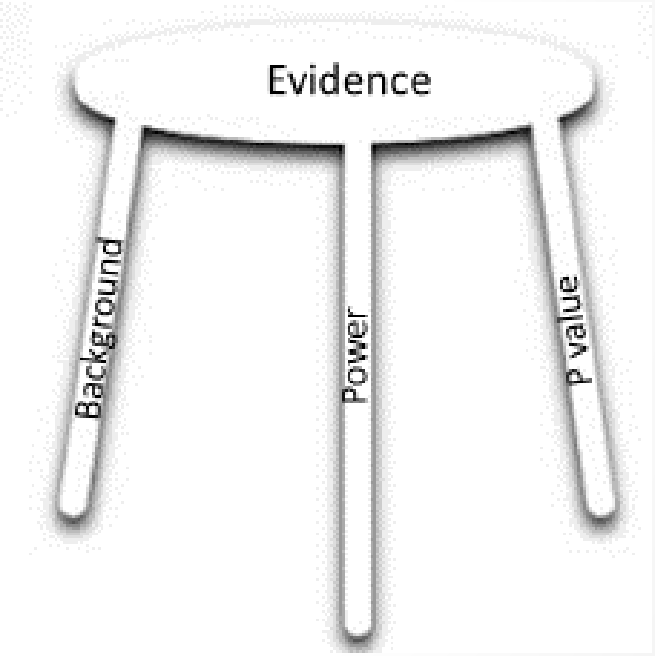
N Engl J Med 2016;374:2419-29.
DOI: 10.1056/NEJMoa1510093

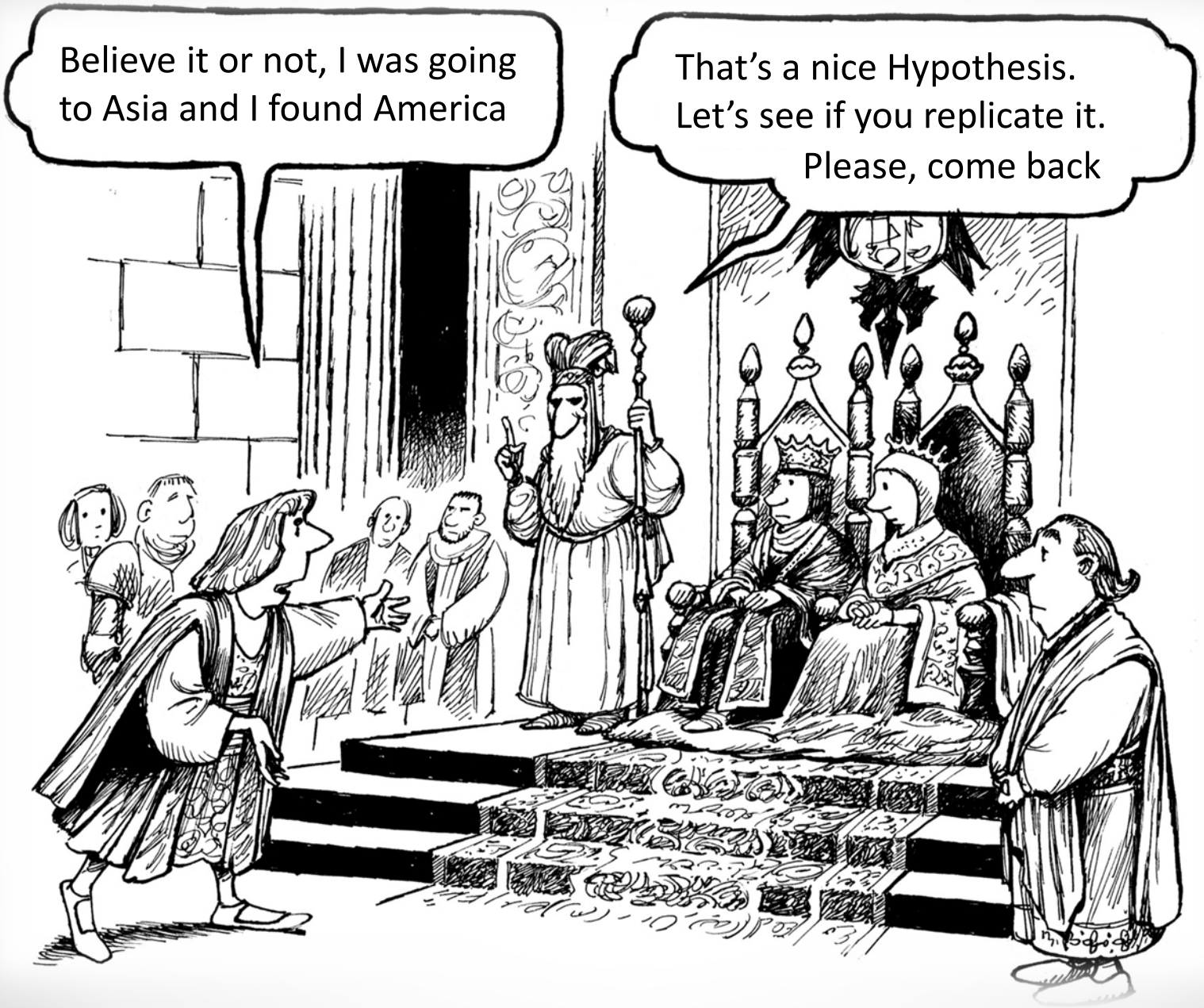
R & D

To explore

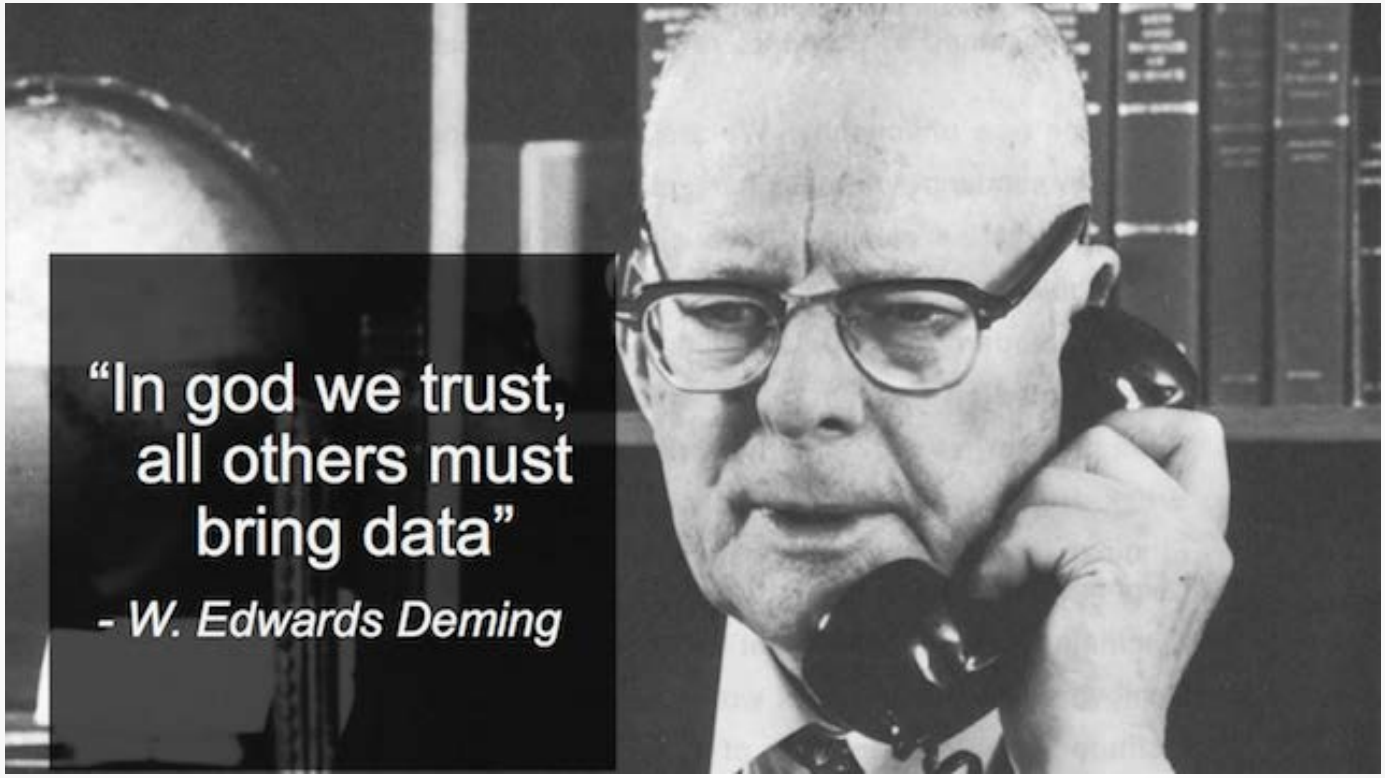
To confirm






Believe it or not, I was going to Asia and I found America
That's a nice Hypothesis. Let's see if you replicate it. Please, come back

Thanks



October, the 20 of 2016

UNIVERSITAT POLITÈCNICA DE CATALUNYA
BARCELONATECH
Departament d'Estadística
i Investigació Operativa