

Reducing scientific error and ensuring innovation

CUOG lecture for CUA 2019

Laurence Klotz
Professor of Surgery
University of Toronto

Laurence Klotz disclosures

- **Chief Medical Officer, miR Scientific**
- **Research support from Profound Medical, Exact Imaging, Prostate Cancer Canada, Ontario Institute of Cancer Research, Ferring, Astellas, Sanofi Genzyme**

THE LANCET

2002 Dec 7;360(9348):1869.

Uses of error

Reactions

Laurence Klotz

During my residency years, it was said that every surgical resident was responsible for a couple of cadavers by the time he finished his training. This observation both emphasised the seriousness of the endeavour, and relieved some of the anxiety associated with the inevitable errors in judgment to which we were vulnerable. Having made errors that resulted in unnecessary death or morbidity, how did we handle this knowledge? An understandable mixture of emotions; guilt, relief (at having avoided serious sanction), and anxiety. In

The history of science, like the history of all ideas, is a history of irresponsible dreams, of obstinacy, and of error.

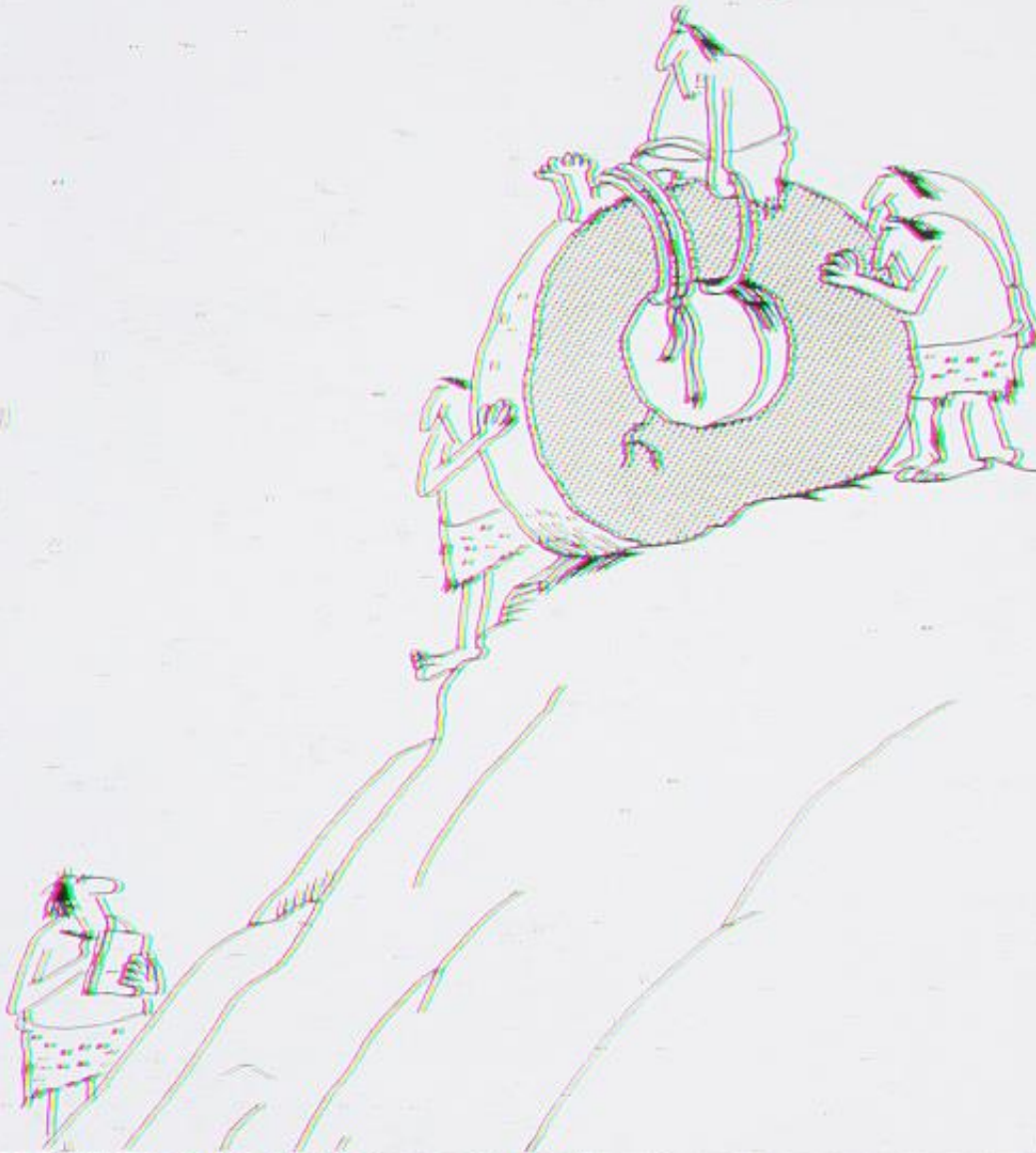
But science is one of the very few human activities—perhaps the only one—in which errors are systematically criticized and fairly often corrected.

In science, we often learn from our mistakes, and can speak clearly and sensibly about making progress.

Karl Popper

Conjectures and Refutations: The Growth of Scientific Knowledge (1963)

Early experiments in transportation



Bias:

If a man is offered a fact which goes against his instincts, he will scrutinize it closely, and unless the evidence is overwhelming, he will refuse to believe it.

If, conversely, he is offered something which affords a reason for acting in accordance with his instincts, he will accept it even on the slightest evidence. This is the origin of myths.

Bertrand Russell

What is scientific error?

- Errors that may occur in the execution or analysis of an experiment.
- Types:
 - Human error, or mistakes in data collection
 - Systematic error, or flaws in experimental design
 - Random error, caused by environmental conditions or other unpredictable factors

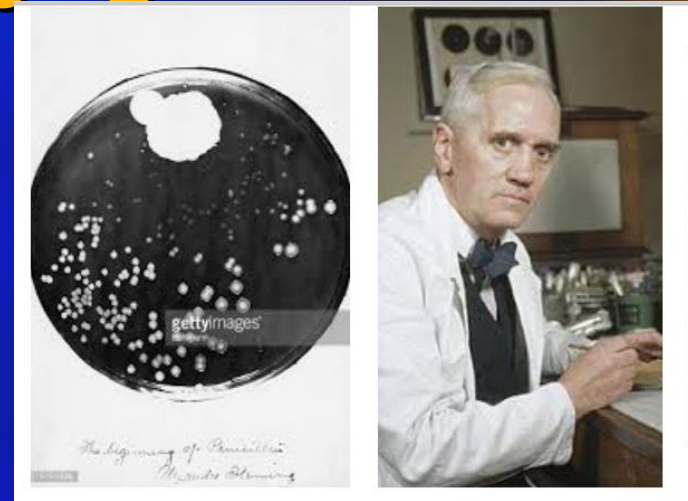
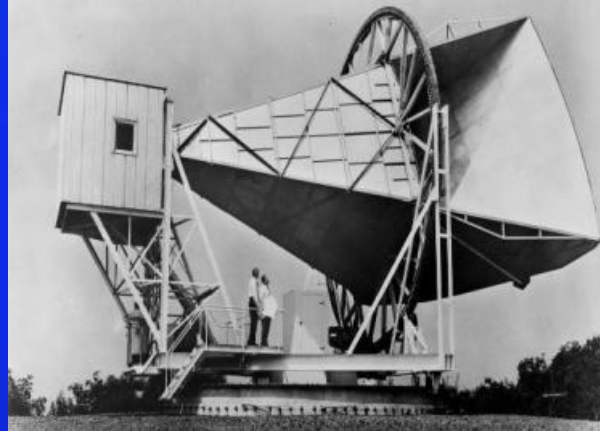
Errors that led to world changing discoveries

Penicillin: Alexander Fleming:

The Big Bang: Penzias and Wilson



Safety Glass



Saccharine:
C Fahlber

Microwave
Oven



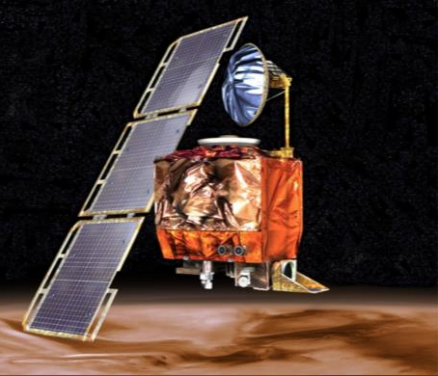
Error due to technical error/miscalculation/incompetence



Challenger disaster



Africanized killer bees



Mars Climate Orbiter



Thalidomide



Faster than light Neutrino—'Opera project'



Fraud/Scientific Misconduct



The New York Times | <https://nyti.ms/29JMVXf>

ARCHIVES | 1983

NOTORIOUS DARSEE CASE SHAKES ASSUMPTIONS ABOUT SCIENCE

By WILLIAM J. BROAD

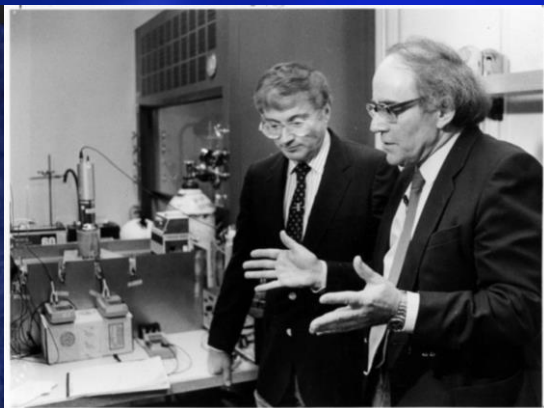
A YOUNG researcher fakes the bulk of his 100 publications, gets caught red-handed, and publicly apologizes after Federal and university investigators expose his deception. The case would appear to be closed. Yet the celebrated fakery of Dr.

William McBride and the Debendox debacle



Piltdown Chicken

Miscalculation/error with agendas



Cold Fusion (Pons and Fleischmann)



Breast implants and autoimmune disorders



Health effects of electromagnetic radiation



Junk Science leading to bad public policy

Adding Fluoride to Water is Dangerous



Vaccines cause Autism



Dietary supplements are necessary, effective, and perfectly safe



Genetically modified foods are inherently dangerous



State of The Climate Report Reveals 23-Year Temperature Pause in the Stratosphere

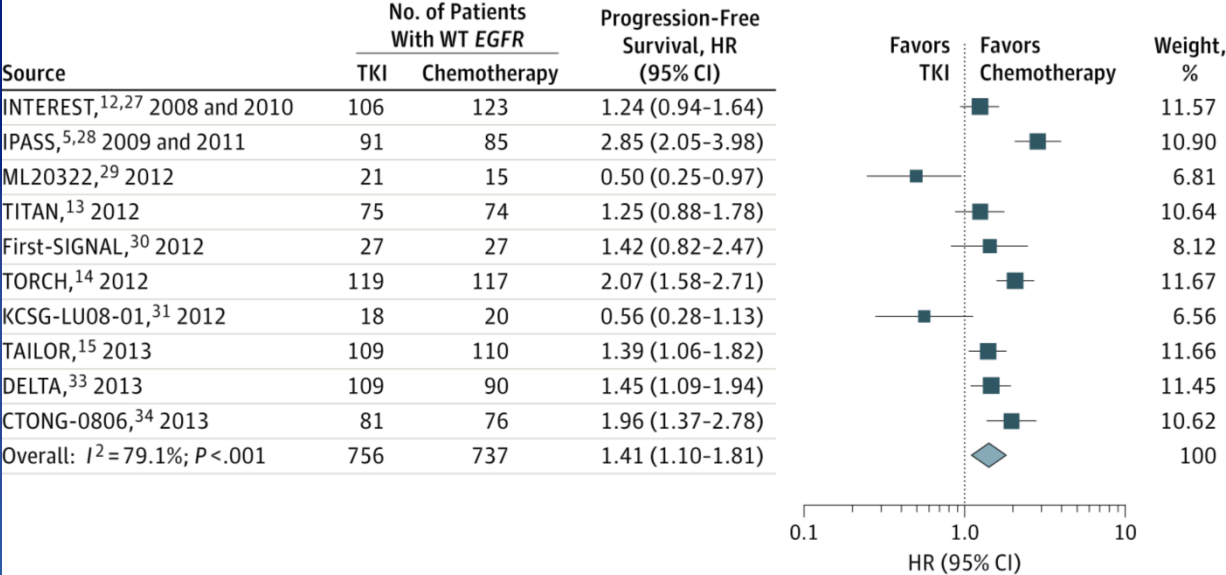
- **London 27 March 2018: Report from the Global Warming Policy Foundation (GWPF) counters media hype over recent warm global temperatures**
- **Most of the increase in temperatures in the last couple of years caused by a record strong natural El Nino phenomenon rather than global warming.**
- **Ole Humlum (Emeritus Prof, University of Oslo) ‘The atmosphere is still not behaving the way most climatologists say it should’ (models have not been predictive)**
- **Polar bear population is at its highest in 30 years**

Missing subgroup effects

- Gefinitib and EGF receptor positive lung cancer
- SELECT trial
- PCPT

Gefitinib and lung cancer: Example of type 2 error—failing to recognize a benefit when it is confined to a subgroup.

Phase 3 trials negative vs chemotherapy



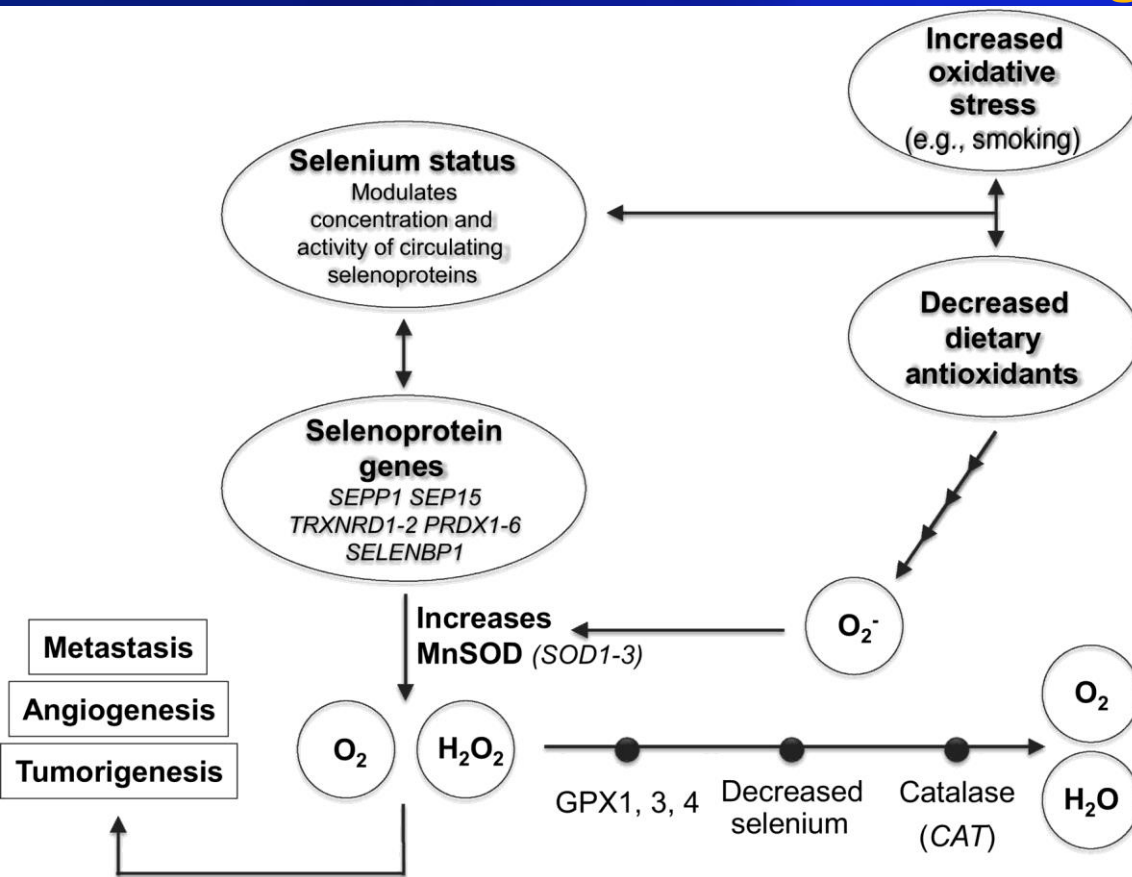
PFS for Gefitinib vs chemo in EGFR mutation positive patients

Study or Subgroup	log[Hazard Ratio]	SE	Weight	Hazard Ratio	
				IV, Random, 95% CI	IV, Random, 95% CI
6.3.1 Biomarker driven selection					
Maemondo 2010 NEJ002	-1.1332037	0.1561584	26.6%	0.32 [0.24, 0.44]	
Mitsudomi 2010 WJTOG3405	-0.7153928	0.1889277	23.0%	0.49 [0.34, 0.71]	
Subtotal (95% CI)			49.6%	0.39 [0.26, 0.59]	
Heterogeneity: $\text{Tau}^2 = 0.06$; $\text{Chi}^2 = 2.91$, $\text{df} = 1$ ($P = 0.09$); $I^2 = 66\%$					
Test for overall effect: $Z = 4.50$ ($P < 0.00001$)					

Gene variant in the antioxidant pathways modify the effect of selenium on the risk of aggressive PCa.

J. Chan et al. *Cancer Epidemiol Biomarkers Prev* 2016;25:1050-1058

MnSOD-Manganese Superoxide dismutase detoxifies mitochondrial free oxygen radicals



SOD2 gene rs7855	Genotype frequency	HR high grade cancer	
		Placebo	Selenium
AA	90%	1.0	1.46. p=.02
Any G	10%	2.14	0.77

Androgen metabolism gene SRD5A1 SNPs and Pca risk from PCPT. Price DK, Cancer. 2016 Aug 1;122(15)

- PCPT: No difference in Gleason ≥ 7 with Finasteride
- But subgroup analysis shows benefit in some SNP groups

Risk of high grade cancer by SNP genotype

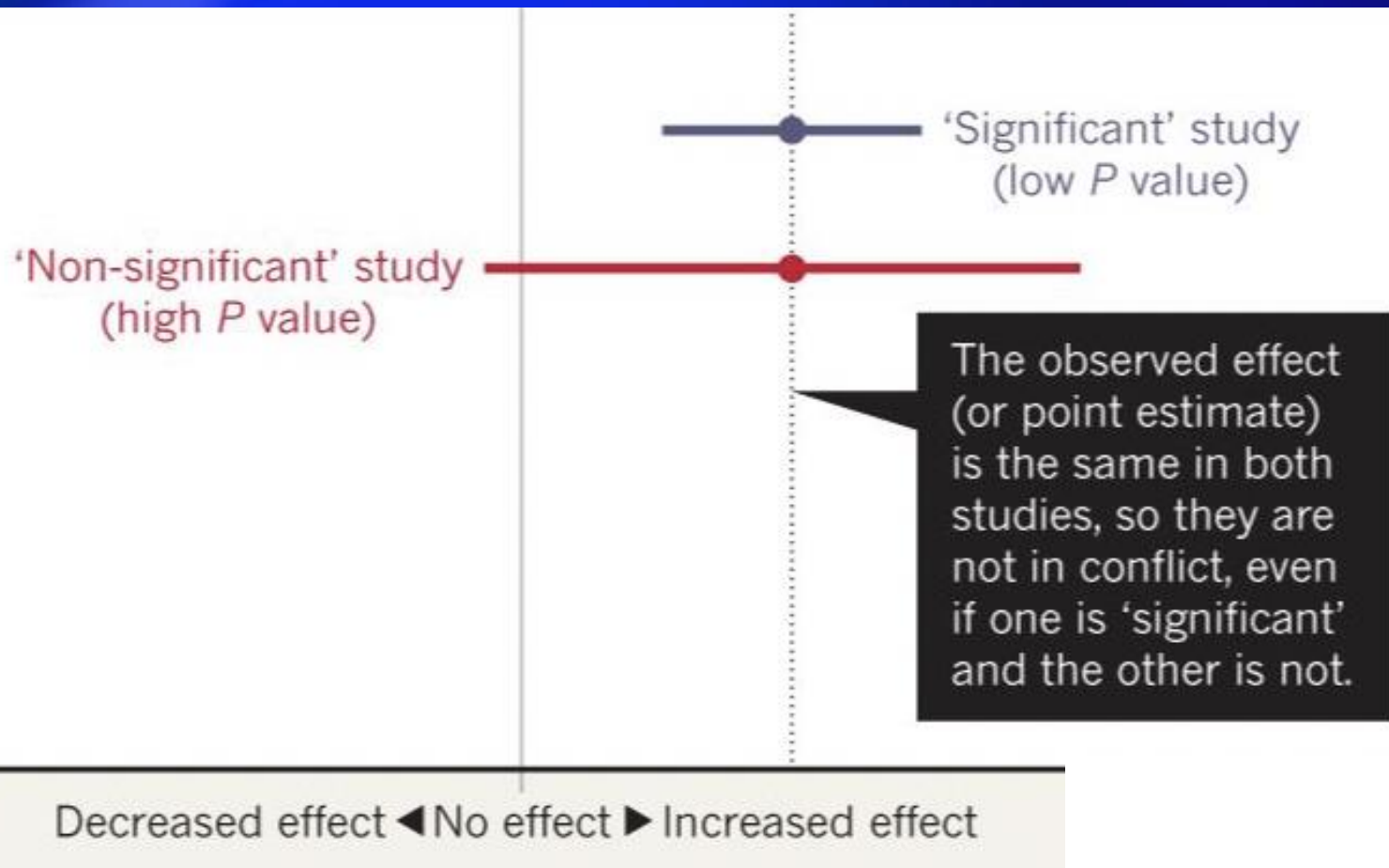
SNP	Genotype	Cases	OR	P value
rs3736316	GG	79	1	0.02
	AG	76	0.8	
	AA	10	0.43	
rs3822430	AA	76	1	0.02
	AG	75	0.81	
	AA	15	0.50	
rs1560149	GG	84	1	0.07
	GG	80	1.82	
	CC	5	0.63	
rs248797	CC	53	1	0.03
	CT	88	0.92	
	TT	27	0.56	

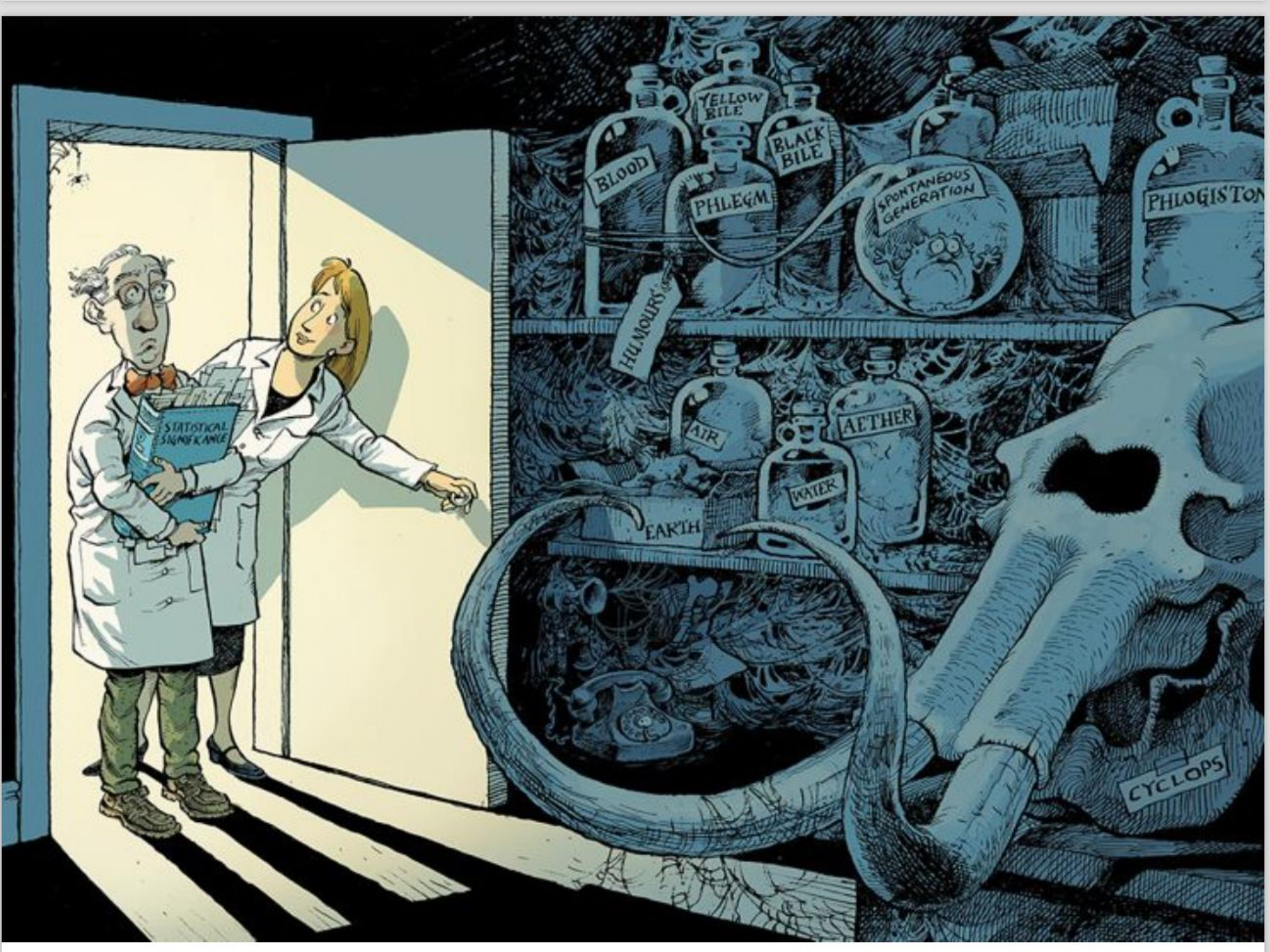
Scientists rise up against statistical significance

Valentin Amrhein, Sander Greenland, Blake McShane and more than 800 signatories call for an end to hyped claims and the dismissal of possibly crucial effects.

Beware false conclusions: Studies dubbed 'statistically significant' and 'statistically non-significant' may not be contradictory

Such designations may cause genuine effects to be dismissed.





COMMENT · 20 MARCH 2019

Scientists rise up against statistical significance

Valentin Amrhein, Sander Greenland, Blake McShane and more than 800 signatories call for an end to hyped claims and the dismissal of possibly crucial effects.

Valentin Amrhein , Sander Greenland & Blake McShane

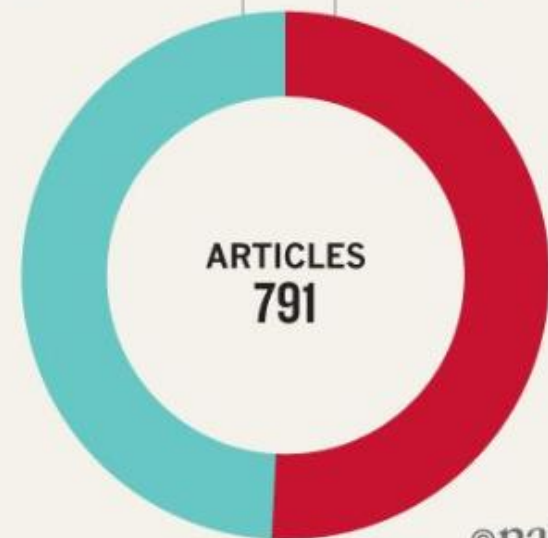
WRONG INTERPRETATIONS

An analysis of 791 articles across 5 journals* found that around half mistakenly assume non-significance means no effect.

*Data taken from: P. Schatz *et al.* *Arch. Clin. Neuropsychol.* **20**, 1053–1059 (2005); F. Fidler *et al.* *Conserv. Biol.* **20**, 1539–1544 (2006); R. Hoekstra *et al.* *Psychon. Bull. Rev.* **13**, 1033–1037 (2006); F. Bernardi *et al.* *Eur. Sociol. Rev.* **33**, 1–15 (2017).

Appropriately interpreted
49%

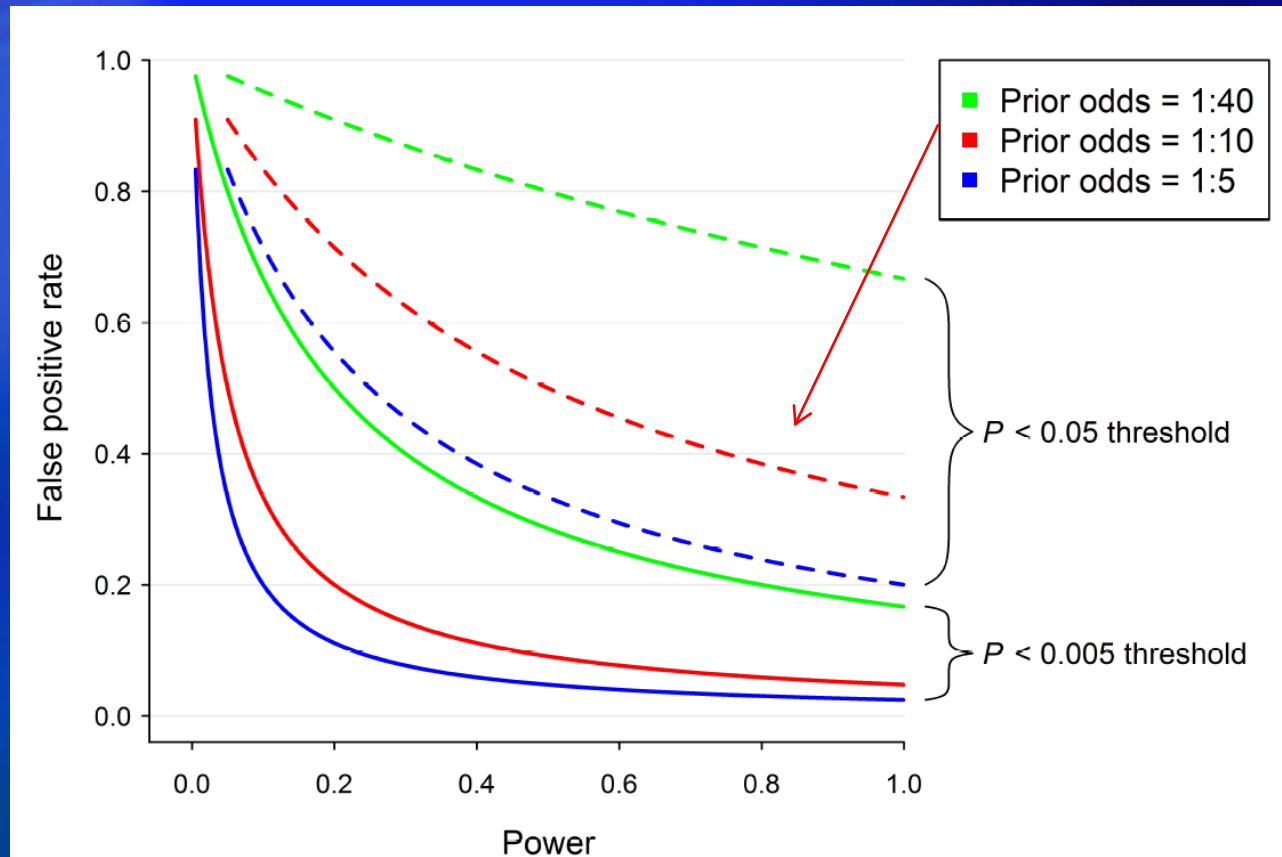
Wrongly interpreted
51%



Benjamin DJ, Berger JO, Johnson VE, et al. Redefine statistical significance. *Nat Hum Behav.* 2018;2:6-10.

Relationship between the P -value threshold, power, and the false positive rate.

Bayesian Factor



The Proposal to Lower P Value Thresholds to < 0.005

Ioannidis JPA. JAMA. 2018 Apr 10;319(14):1429-1430.

'P values are misinterpreted, overtrusted, and misused'

	Apply to past studies	Future research
Lower P value threshold	Simple temporizing solution	Collateral harms, requires adoption/enforcement
Abandon P value thresholds; use exact P val.	Many P values only reported with thresholds	Success depends on extent of adoption
Abandon P values	May not be any other stat, ie effect sizes/CI lacking	Better success in some areas (Dx test performance)
Use alternative inference (Bayesian stats)	Needs sophisticated training	Suitable for most studies
Focus on effect sizes	Often not reported (more common recently)	Widely relevant and readily understood
Better training	Takes time, commitment	More appropriate use of statistics and inference tools
Address biases leading to inflated results	Requires major training; biases may be subtle	Needs commitment of multiple stakeholders

Steps in the scientific method (Wikipedia)

- Define a question
- Gather information and resources (observe)
- Form an explanatory hypothesis
- Test the hypothesis by performing an experiment and collecting data in a reproducible manner
- Analyze the data
- Interpret the data and draw conclusions that serve as a starting point for new hypothesis
- Publish results
- Reproduce results (frequently done by other scientists)

The first-ever English language retraction (1756)?

The notice appeared in the *Philosophical Transactions* of the Royal Society on June 24, 1756. It reads:

Gentlemen,

I think it necessary to retract an opinion concerning the explication of the Leyden experiment, which I troubled this Society with in the year 1746, and afterwards published more at large in a Treatise upon Electicity, in the year 1750; as I have lately made some farther discoveries relative to that experiment, and the minus electricity of Mr. Franklin, which shew I was then mistaken in my notions about it.



Benjamin Wilson self-portrait, via Wikimedia
<http://bit.ly/zkWs5C>

Reproducibility

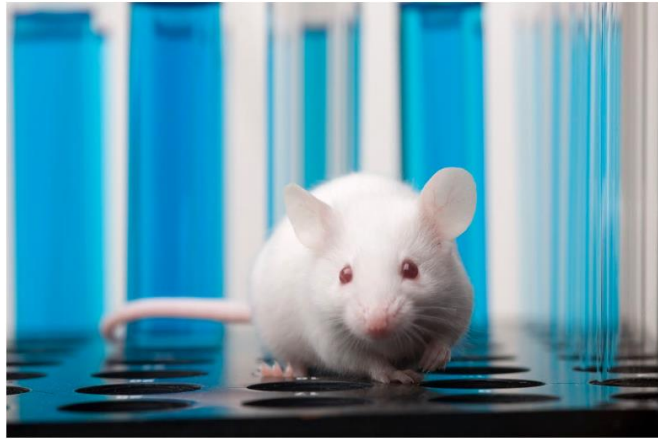
Aren't published results always reproducible?

Not always.

Doesn't the literature correct itself?

Usually not!

More science than you think is retracted. Even more should be.



(dra_schwartz/Getty Images)

By **Adam Marcus and Ivan Oransky**
December 26 at 6:01 PM

Ad closed by Google

Report this ad

Why this ad?

STAT [Sections](#) [Topics](#) [Multimedia](#) [Newsletters](#) [More](#)

[Log In](#) [Subscribe](#)

[TRY STAT PLUS](#)

EXCLUSIVE

Harvard and the Brigham call for more than 30 retractions of cardiac stem cell research

By [IVAN ORANSKY](#) @ivanoransky and [ADAM MARCUS](#) @armarcus / OCTOBER 14, 2018



Lack of reproducibility



Re-tested 70+ drugs from 221 independent studies¹

→ 0 reproduced

→ Minocycline: effective in four separate ALS mouse studies worsened symptoms in a clinical trial of more than 400 patients²



Sponsored replication of 12 spinal cord injury studies

→ 2/12 fully reproduced³



Conducted in-house target validation studies

→ 14/67 reproduced⁴



Attempted to reproduce 53 “landmark” oncology publications

→ 6/53 reproduced⁵

1. Scott et al. Amyotroph Lateral Scler. 9, 4-15 (2008).

2. Gordon et al. Lancet Neurol. 6, 1045-1053 (2007).

3. Stuart et al. Experimental Neurology 233, 597-605 (2012).

4. Prinz et al. Nat Rev Drug Discov. 10, 712 (2011).

5. Begley and Ellis. Nature. 483, 531-3 (2012).

Reproducibility

- 0.2% of the literature retracted (vs >50% irreproducibility)
- 55% of researchers in one survey tried and failed to reproduce published results. ¹
 - Less than 30% published their failure
 - 44% had difficulty publishing their contradictory results
- Likelihood of reproducibility unrelated to number of publications³ or citations⁴

1. Mobley et al. PLOS ONE. 8, e63221 (2013) 2. Fanelli. Scientometrics. 90, 891 (2012) 3. Prinz et al. Nat Rev Drug Discov. 10, 712 (2011) 4. Begley and Ellis. Nature. 483, 531-3 (2012).

The Retraction Watch Leaderboard

Who has the most retractions? Here's our unofficial list (see notes on methodology), which we'll update as more information comes to light:

1. [Yoshitaka Fujii](#) (total retractions: 183) See also: [Final report of investigating committee](#), [our reporting](#), [additional coverage](#)
2. [Joachim Boldt](#) (96) See also: [Editors-in-chief statement](#), [our coverage](#)
3. [Diederik Stapel](#) (58) See also: [our coverage](#)
4. [Adrian Maxim](#) (48) See also: [our coverage](#)
5. [Yuhji Saitoh](#) (48) See also: [our coverage](#)
6. [Yoshihiro Sato](#) (48) See also: [our coverage](#)
7. [Jun Iwamoto](#) (44) See also: [our coverage](#)
8. [Chen-Yuan \(Peter\) Chen](#) (43) See also: [SAGE](#), [our coverage](#)
9. [Fazlul Sarkar](#) (41) See also: [our coverage](#)
10. [Hua Zhong](#) (41) See also: [journal notice](#)
11. [Shigeaki Kato](#) (40) See also: [our coverage](#)
12. [James Hunton](#) (37) See also: [our coverage](#)
13. [Hyung-In Moon](#) (35) See also: [our coverage](#)
14. [Naoki Mori](#) (32) See also: [our coverage](#)
15. [Jan Hendrik Schön](#) (32) See also: [our coverage](#)
16. [Soon-Gi Shin](#) (30) See also: [our coverage](#)
17. [Tao Liu](#) (29) See also: [our coverage](#)
18. [Bharat Aggarwal](#) (28) See also: [our coverage](#)
19. [Cheng-Wu Chen](#) (28) See also: [our coverage](#)
20. [A Salar Elahi](#) (27) See also: [our coverage](#)
21. [Richard L E Barnett](#) (26) See also: [our coverage](#)
22. [Antonio Orlandi](#) (26) See also: [our coverage](#)
23. [Prashant K Sharma](#) (26) See also: [our coverage](#)
24. [Rashmi Madhuri](#) (24) See also: [our coverage](#)
25. [Scott Reuben](#) (24) See also: [our coverage](#)
26. [M Ghoranneviss](#) (22) See also: [our coverage](#)
27. [Anil K Jaiswal](#) (22) See also: [our coverage](#)
28. [Gilson Khang](#) (22) See also: [our coverage](#)
29. [Noel Chia](#) (21) See also: [our coverage](#)
30. [Alfredo Fusco](#) (21) See also: [our coverage](#)
31. [Friedhelm Herrmann](#) (21) See also: [our coverage](#)
32. [Stanley Rapoport](#) (21) See also: [our coverage](#)

**No mechanism exists to confirm
reproducibility.**

We have a problem!

Incentives to replicate research are poor

“

Pressure-testing academic theories remains an important...and profoundly undervalued activity.

Forbes

“

**The
Economist**

Nowadays verification studies do little to advance a researcher's career. And without verification, dubious findings live on to mislead.

“

those who document the invalidity of a published piece of work seldom get a welcome from journals, funding agencies, conference organizers

WILEY

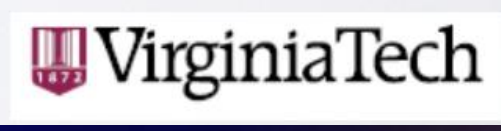
A solution: Science Exchange Cancer Biology Reproducibility Project

- Goal: Identify key studies in the literature and reproduce them
- Top 50 cited cancer biology studies from 2010-2012
- Perform rapid, cost effective replication by expert independent labs
- Budget: \$1.3M

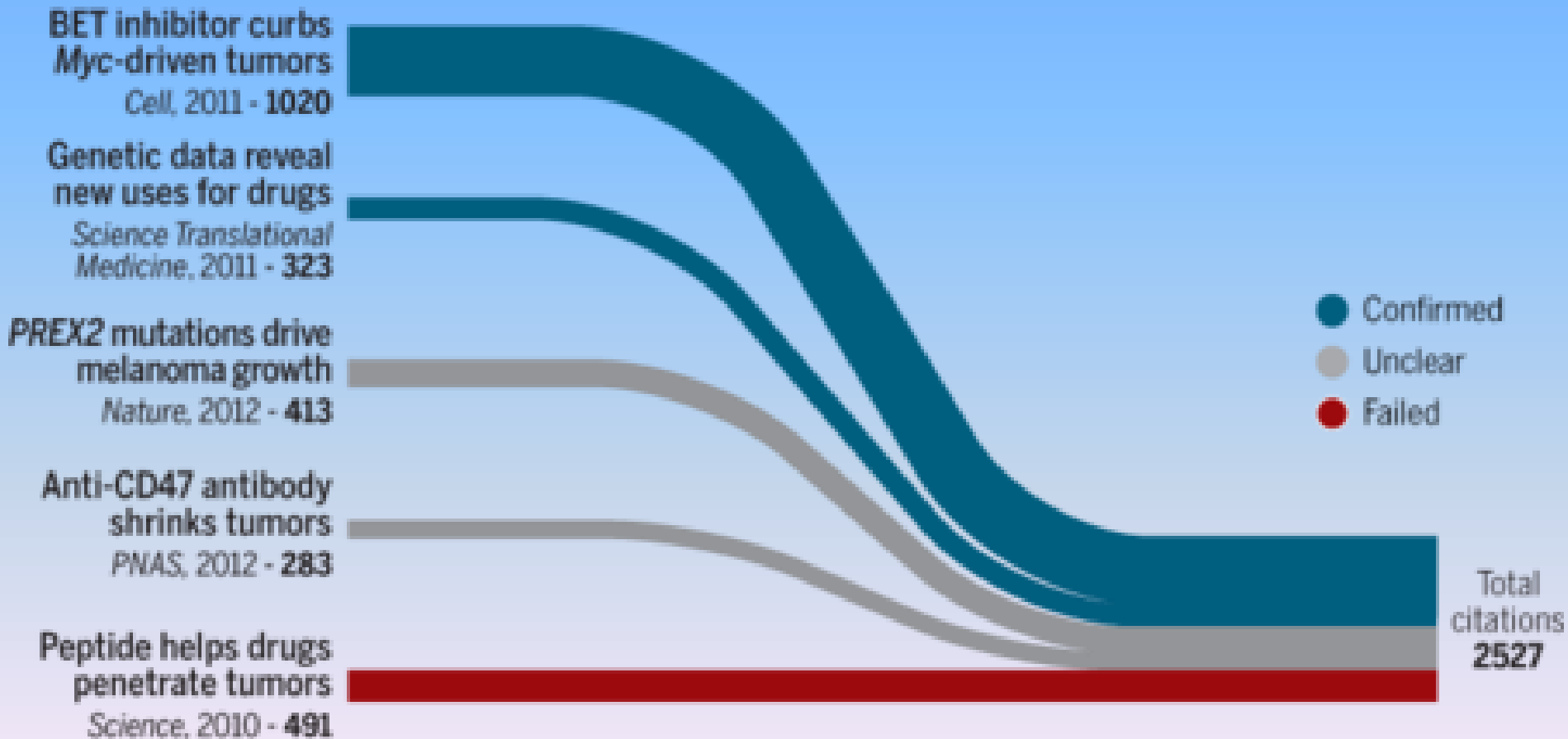
THE SCIENCE EXCHANGE NETWORK

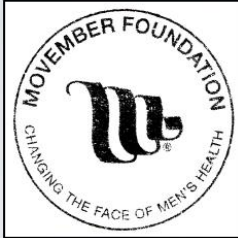


Labs from 400+ research institutions:
75 of top 100 U.S. research universities



Replication effort succeeds for two of five cancer papers. Kaiser J, *Science*, Jan. 2017





**PROSTATE
CANCER
FOUNDATION**



- **Movember Foundation-Prostate Cancer Foundation-Scientific Exchange Collaboration**
- **Pilot study to assess the reproducibility of research findings with implications for prostate cancer patients**
- **Goal: Replicate four major studies**

The 'Zeitgeist' has increased the risk of scientific misconduct

- **'Truthiness'**
- **Fake news**
- **'Relativism' (no objective truth)**
- **Intense academic pressure to publish**
- **Long distance collaboration, large teams means less direct individual oversight**
- **Undisclosed conflict of interest (eg, Dr. Basalga)**
- **Limitations of journal editorial review process**
- **Media complicity, desire to enhance the appeal of the story**
- **Result: Objectivity falls by the wayside**

Forbes, Sep 10, 2018

How To Really Take Medical Conflicts Of Interest Seriously

The *New York Times* reported on its front page that Dr. Jose Baselga, one of the world's top oncologists and the chief medical officer of Memorial Sloan Kettering Cancer Center, failed to disclose financial conflicts of interest when he wrote articles for medical journals including the *New England Journal of Medicine*. That includes more than \$3 million dollars Baselga made when Roche bought Seragon, a startup that had paid him in stock.



Dr. José Baselga Memorial Sloan Kettering Cancer Center via Creative Commons

How to reduce scientific error in the future

- Recognize the key role of 'error' in the scientific method (falsifiable hypothesis)
- Maintain and foster scientific tradition of scrupulous adherence to truth and objectivity
- Enhance awareness of pitfalls of human investment in established paradigms (Popper, Russell, et al)
- Recognize risks of statistical error, particularly type 2 error (underpowered studies, missing subgroups) and hazards of $P < 0.05$
- Support reproducibility studies for key research findings
- Be aware of the high prevalence of scientific misconduct

Innovation: introduce something new with hope that it will become commonplace

- **Percutaneous nephrolithotomy**
 - **Clayman**
- **Anatomic radical prostatectomy**
 - **Walsh, Donker**
- **BCG--Morales**
- **RCC genetics—Linehan**
- **PSA—Murphy, Catalona**
- **Cryosurgery—Cohen, Onick**
- **Retroperitoneal lymphadenectomy—Donahue**
- **Continent Diversion—Koch, Studer**
- **Huggins—ADT**
- **Bruchovsky: Intermittent Rx**
- **PCA3—Schalken**
- **Da Vinci—Menon**
- **Focal therapy—Emberton**
- **AR targeted agents in CRPC: Sawyer, DeBono**
- **Active Surveillance—Klotz, Carter**

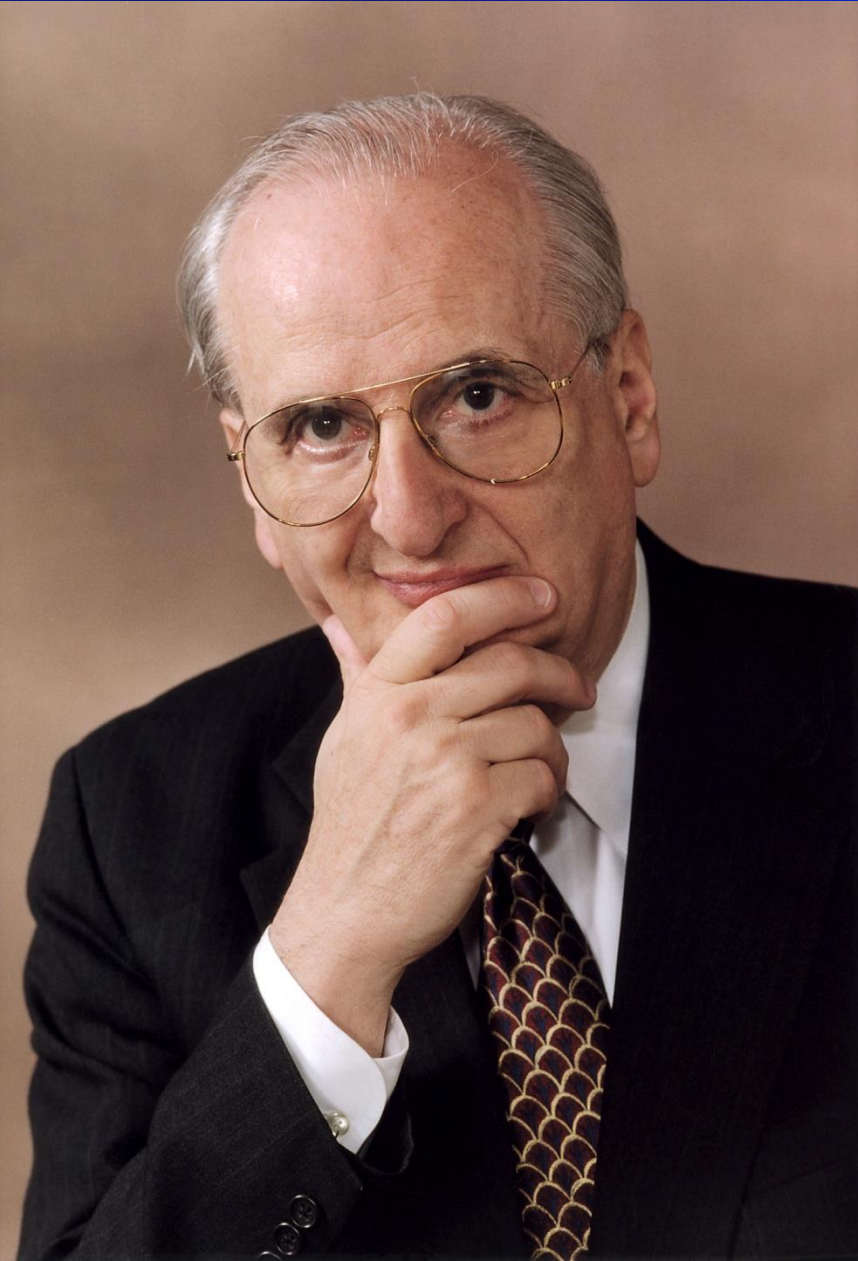
Change by evolution: Darwin's revolutionary proposal

- **Survival of fittest led to adaptation and evolution.**

BUT

- **Adaptation is a reaction to the environment—not enough.**
- **We must go beyond: to lead, anticipate, innovate.**

Judah Folkman: Angiogenesis



There is a fine line between persistence and obstinacy. I have come to realize the key is to choose a problem that is worth persistent effort.

Well, I always thought I was right because I knew something that none of them knew; I had been at the operating table. It wasn't the surgeons who were criticizing my theories of angiogenesis, it was the basic scientists, and I knew that many of them had never seen cancer except in a dish. I knew that they had not experienced what I had experienced. The idea of tumors growing in three dimensions and needing blood vessels in the eye, in the peritoneal cavity, in the thyroid, and many other places, and the concept of in situ cancers and tumors waiting dormant—I had seen all that. So I kept saying the ideas are right.

*** Judah Folkman: Discovery Channel, 2000**

Convergent advances: unprecedented opportunities

- **Biological revolution applied to medicine**
 - **‘Omics’, systems biology—mechanisms, biomarker signatures**
 - **Stem cells, regenerative biology**
 - **Robotic/virtual reality, imaging, AI**
 - **Information technology, big data**
 - **Digitalization of biology, drug/biomarker discovery**
- **Increasing emphasis on translational research and team science**
 - **Bedside-Bench-Bedside and marketplace “traffic” has never been greater**
- **Grow clinical trials consortia for surgery**
- **Train clinician scientists**

Innovation



- **There is nothing more penetrating than the human imagination and the human capacity for reason**
- **The most dazzling scientific and medical breakthroughs first emerge from the interior worlds of individual minds**
- **That individual may be unknown, unrecognized, under-resourced, and lacking in confidence**
- **That individual may be you!**

We are in an era of profound change and great challenges

- **Someone will lead this change**
- **Think about how to do things differently (better)**
 - **Write these down; discuss them**
- **Try these ideas out; propose a trial**
- **Remember pharma is interested in investigator initiated trials and will fund these**

Fildes, Doctor and patient



The painting is a reminder that all the marvels of science and medicine don't replace what patients want of their doctors and what most of us wanted when we felt the calling to medicine: the opportunity to be fully present at the bedside, to bring the human comfort that only the presence of an attentive and knowledgeable physician can bring, and to convey to patient and family the unspoken promise, "I will stay with you through thick and thin." That's a privilege I won't trade for anything. **A Verghese: The Myth of Prevention 2009**

Our challenge:

- **Continue the remarkable scientific progress of the last 30 years**
- **Maintain our role as caring physicians and human beings**
- **Maintain our trusted and privileged position in society**