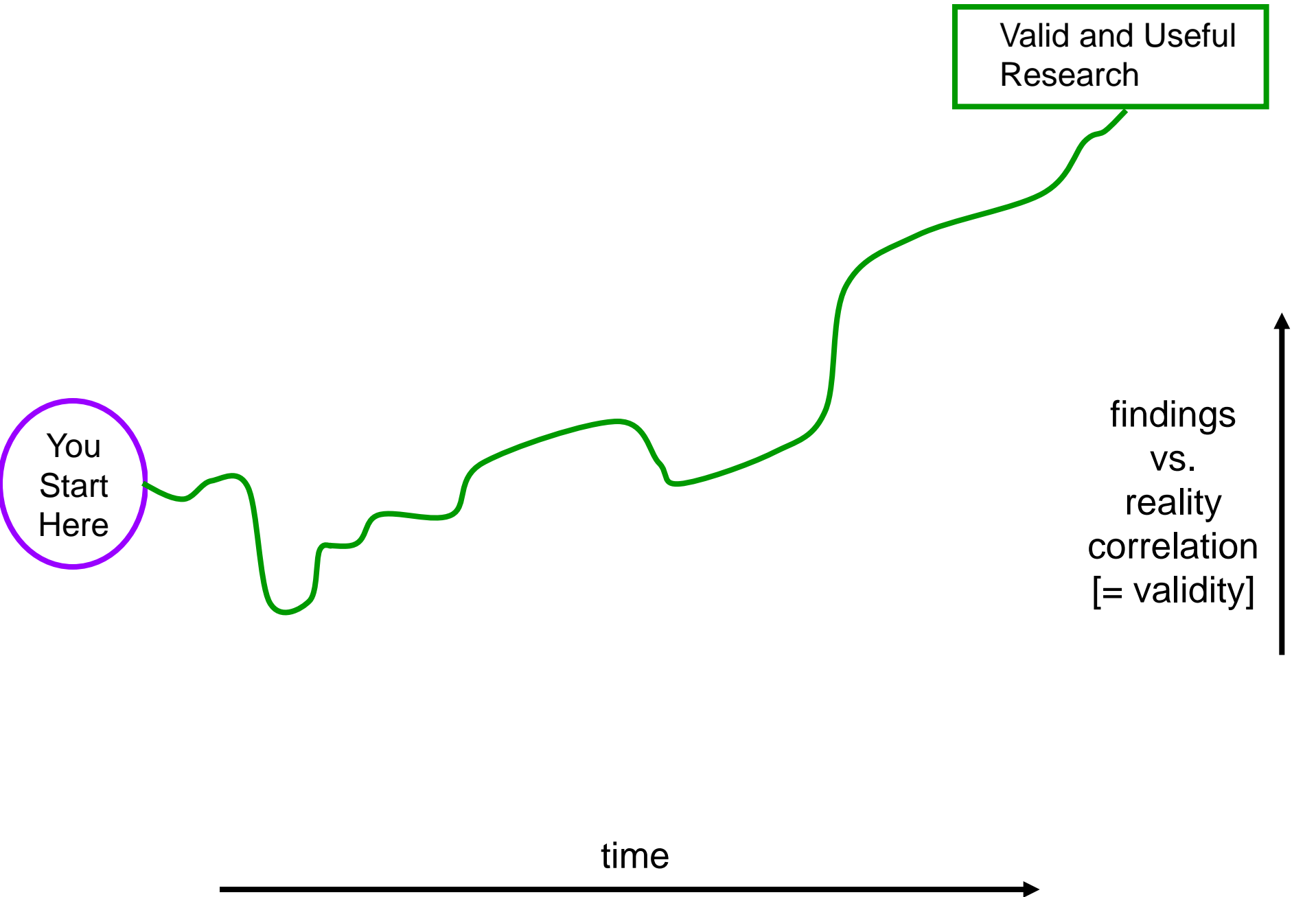


# **Avoiding the pitfalls on the research trail**

**Robert J. Kay, PhD**

Professor, Dept. of Medical Genetics

University of British Columbia

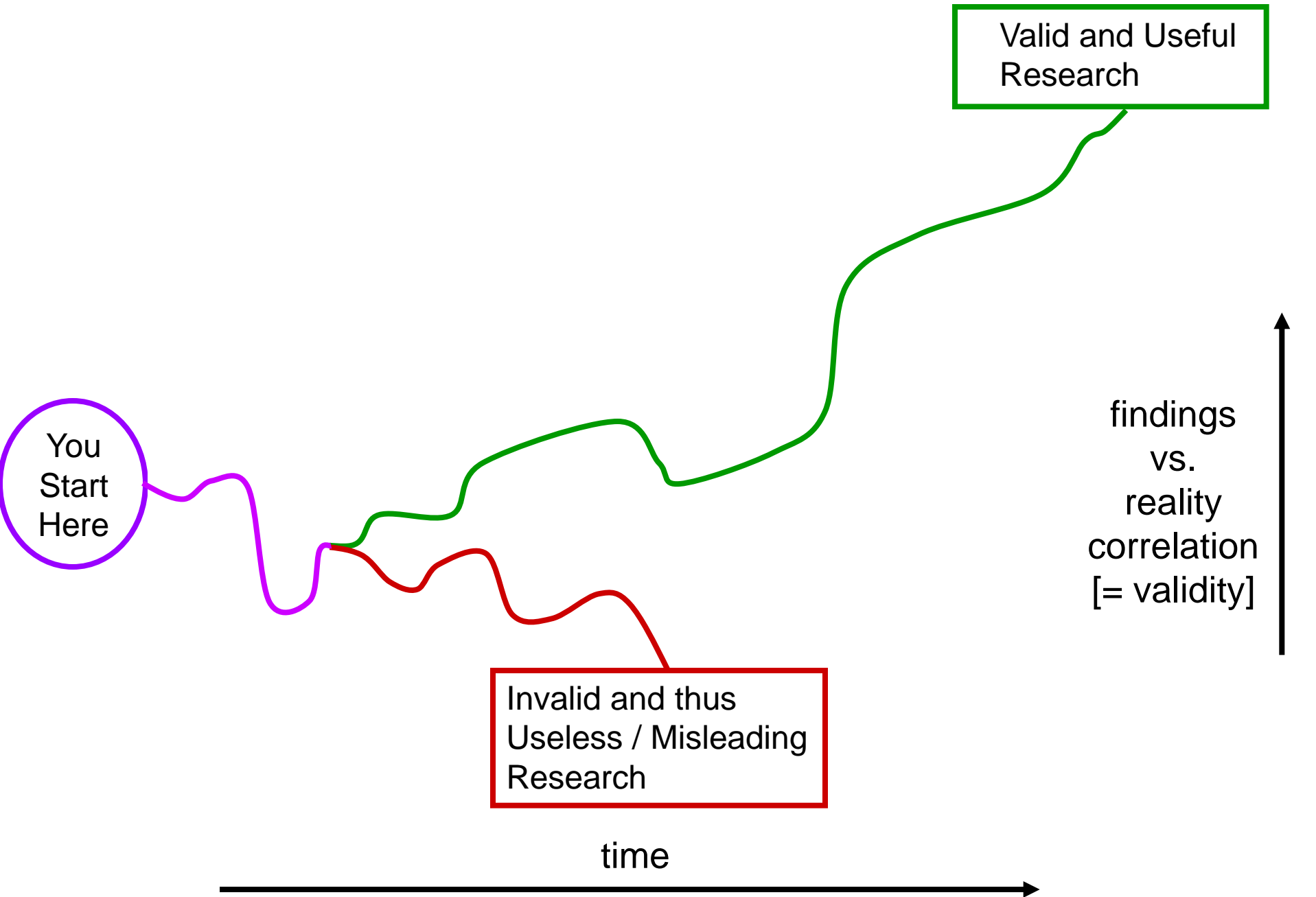


You Start Here

Valid and Useful Research

findings vs. reality correlation [= validity]

time



You Start Here

Valid and Useful Research

Invalid and thus Useless / Misleading Research

findings vs. reality correlation [= validity]

time

*Does low validity research  
get published often enough  
to cause a problem?*

**Deming, data and observational studies. A process out of control and needing fixing.** S. Stanley Young, Alan Karr. Significance 8: 116, 2011

**12 randomized clinical trials tested 52 claims made by observational studies**

<i>RCT ID #</i>	<i>signif. results of RCT</i>		<i># of observ'al studies tested by the RCT</i>	<i>Treatment(s)</i>
	<i>Pos.</i>	<i>Neg.</i>		
1				Vit E, beta-carotene
2				Hormone Replacement Ther.
3				Vit E, beta-carotene
4				Vit E
5				Low Fat
6				Vit D, Calcium
7				Folic acid, Vit B6, B12
8				Low Fat
9				Vit C, Vit E, beta-carotene
10				Vit C, Vit E
11				Vit E, Selenium
12				HRT + Vitamins
<hr/>				
<b>Totals</b>				

**How many of the 52 claims were confirmed by RCT???**

**Deming, data and observational studies. A process out of control and needing fixing.** S. Stanley Young, Alan Karr. Significance 8: 116, 2011

**12 randomized clinical trials tested 52 claims made by observational studies**

<i>RCT ID #</i>	<i>signif. results of RCT</i>		<i># of observ'al studies tested by the RCT</i>	<i>Treatment(s)</i>
	<i>Pos.</i>	<i>Neg.</i>		
1	0	1	3	Vit E, beta-carotene
2	0	3	4	Hormone Replacement Ther.
3	0	1	2	Vit E, beta-carotene
4	0	0	3	Vit E
5	0	0	3	Low Fat
6	0	0	3	Vit D, Calcium
7	0	0	2	Folic acid, Vit B6, B12
8	0	0	2	Low Fat
9	0	0	12	Vit C, Vit E, beta-carotene
10	0	0	12	Vit C, Vit E
11	0	0	3	Vit E, Selenium
12	0	0	3	HRT + Vitamins
<b>Totals</b>	<b>0</b>	<b>5</b>	<b>52</b>	

**0 of 52 claims were confirmed by RCT.  
5 of the claims had significant RCT evidence for an opposite effect.**

# A comprehensive review of genetic association studies

*Joel N. Hirschhorn, MD, PhD<sup>1-3</sup>, Kirk Lohmueller<sup>1</sup>, Edward Byrne<sup>1</sup>, and Kurt Hirschhorn, MD<sup>4</sup>*

Genetics in Medicine 4: 45, 2002

Over 600 associations between common gene variants and disease have been reported. These associations, if correct, would have tremendous importance for the prevention, prediction, and treatment of most common diseases.

Of the 166 putative genetic associations which have been studied three or more times, only 6 have been consistently replicated.

# **Almost all articles on cancer prognostic markers report statistically significant results**

*Panayiotis A. Kyzas<sup>a</sup>, Despina Denaxa-Kyza<sup>a</sup>, John P.A. Ioannidis<sup>a,b,c,\*</sup>*

Of 1915 published papers on cancer biomarkers, only 22 did not make any positive claims about a prognostic value of a biomarker.

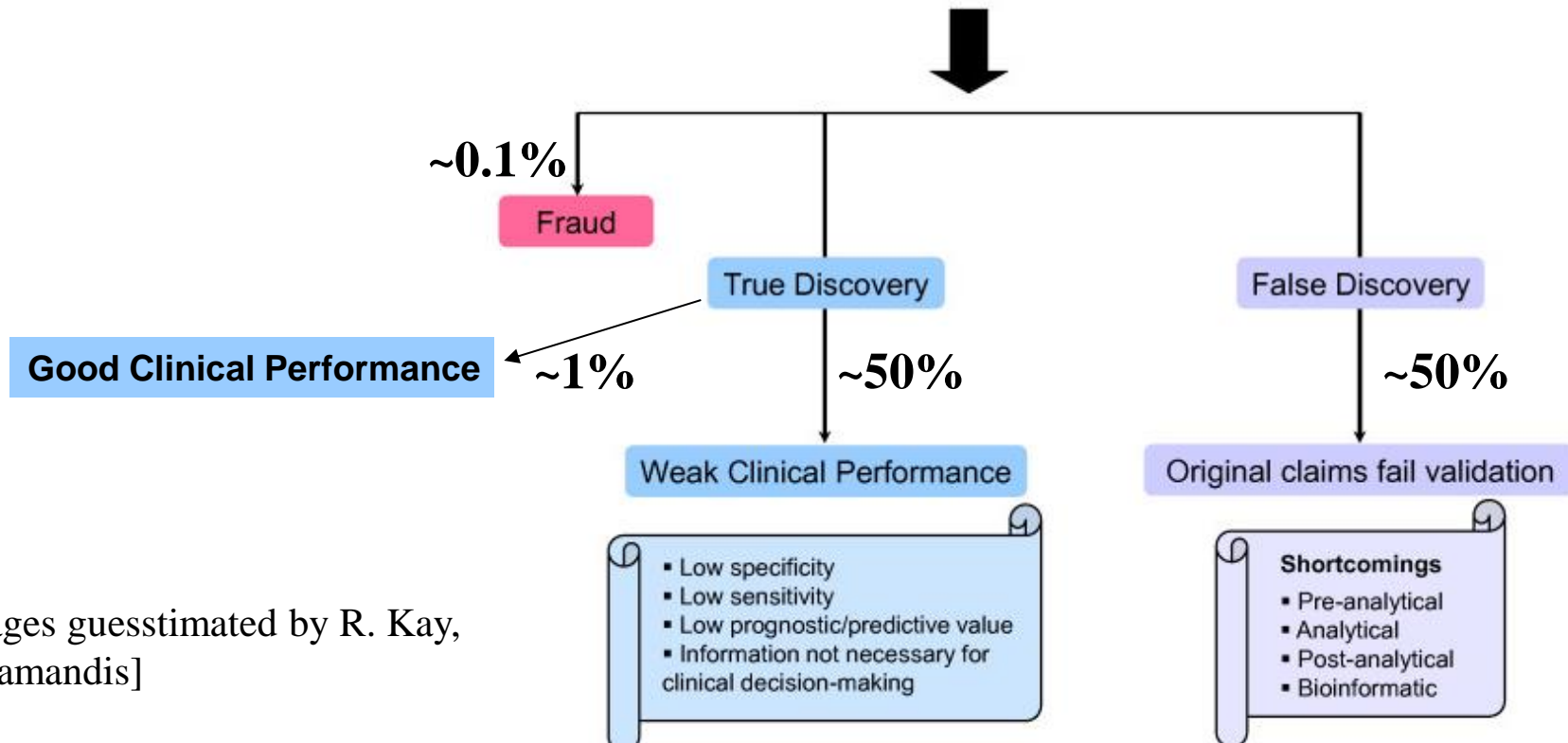


EP Diamandis, The failure of protein cancer biomarkers to reach the clinic, BMC Medicine 10: 87, 2012 [see also J Nat Cancer Inst 102:1462, 2010]

“There is a plethora of published cancer biomarkers but the reality is that very few, if any, new circulating cancer biomarkers have entered the clinic in the last 30 years.

“...major investments by both academia and industry have been made in this area of investigation but with very little return.

## Why Biomarkers Fail to Reach the Clinic



[percentages guesstimated by R. Kay, not in Diamandis]

# Puzzlingly High Correlations in fMRI Studies of Emotion, Personality, and Social Cognition

Edward Vul et al., Perspectives on Psychological Science May 2009 4: 274-290

Functional magnetic resonance imaging (fMRI) studies of emotion, personality, and social cognition have drawn much attention in recent years, with high-profile studies frequently reporting extremely high (e.g.,  $>.8$ ) correlations between brain activation and personality measures.

Measures of personality and emotion do not often have reliabilities greater than  $.8$ , and neuroimaging measures seem typically to be reliable at  $.7$  or less.

So the upper bound for the highest expected correlations would be  $(.8 \times .7)^{0.5} = .74$ .

Correlations exceeding this upper bound are often reported in recent fMRI studies on emotion, personality, and social cognition.

# Believe it or not: how much can we rely on published data on potential drug targets?

Florian Prinz, Thomas Schlange & Khusru Asadullah. Nature Reviews Drug Discovery 10: 712, 2011

An evaluation of the validity of basic science validity based on experience at the pharmaceutical company Bayer

**In almost two-thirds of [67 drug development] projects, there were inconsistencies between published data and in-house data**

*that, in most cases, resulted in termination of the projects because the evidence that was generated for the therapeutic hypothesis was insufficient to justify further investments into these projects.*

*The reproducibility of published data did not significantly correlate with:*

- journal impact factors*
- the number of publications on the respective target*
- the number of independent groups that authored the publications.*

# **Drug development: Raise standards for preclinical cancer research**

C. Glenn Begley & Lee M. Ellis. Nature 483: 531–533, 2012

An evaluation of the validity of basic science validity based on experience at the pharmaceutical company Amgen

**Prior scientific findings were confirmed in only 6 of 53 oncology drug development projects**

The  
Economist

OCTOBER 10TH-20TH 2015

Economist.com

Academic scientists acknowledge that they often get things wrong, but they also hold the idea that these errors get corrected over time as other scientists try to take the work further. **Evidence that many more dodgy results are published than are subsequently corrected calls that much-vaunted capacity for self-correction into question.**

HOW  
SCIENCE  
GOES  
WRONG

A career structure which lays great stress on publishing copious papers exacerbates all these problems [causing irreproducibility]. **“There is no cost to getting things wrong,” says Brian Nosek, “The cost is not getting them published.”**

The governments of the OECD spent \$59 billion on biomedical research in 2012. One of the justifications for this is that **basic-science results form the basis for private drug-development work. If companies cannot rely on academic research, that reasoning breaks down.**

*What allows low validity research  
to be performed,  
and to get published?*

*How to minimize this?*

# *Ambushed by Hidden Variables*

Observational studies repeatedly indicate that  $\beta$ -carotene lowers the risk of lung cancer

Randomized clinical trials repeatedly indicate that  $\beta$ -carotene raises the risk of lung cancer

Cochrane Database Syst Rev. 2012 Oct 17;10:CD002141

**Confounding by hidden [or ignored] variables?**



Observational studies repeatedly indicate that  $\beta$ -carotene lowers the risk of lung cancer

Randomized clinical trials repeatedly indicate that treatment with  $\beta$ -carotene raises the risk of lung cancer

Cochrane Database Syst Rev. 2012 Oct 17;10:CD002141

### **Confounding by hidden [or ignored] variables:**

People that eat lots of carrots smoke less than those who don't

People that eat lots of carrots now are nutritionally repenting for their sinful youths, during which they started smoking

People that eat lots of carrots are generally more compliant with healthy living recommendations than those who don't

People that metabolize  $\beta$ -carotene rapidly are particularly bad at metabolically-disabling carcinogens

Etcetera ad infinitum

## **Defenses against confounding:**

Assess all known or suspected confounding variables

Look for statistical indications of hidden confounders

Note the potential for hidden confounders in your conclusions

Do experiments, using a single controlled variable **????**

## **Hidden variables in**

### **experiments:**

Are you sure that your experimental intervention is the only variation between your test and control groups?

Cell lines, mice etc. are inherently heterogeneous

Experimenters can introduce heterogeneity by selective choosing of test vs control cells, mice etc

Experimenters can introduce heterogeneity by selective exclusion of data from tests vs controls

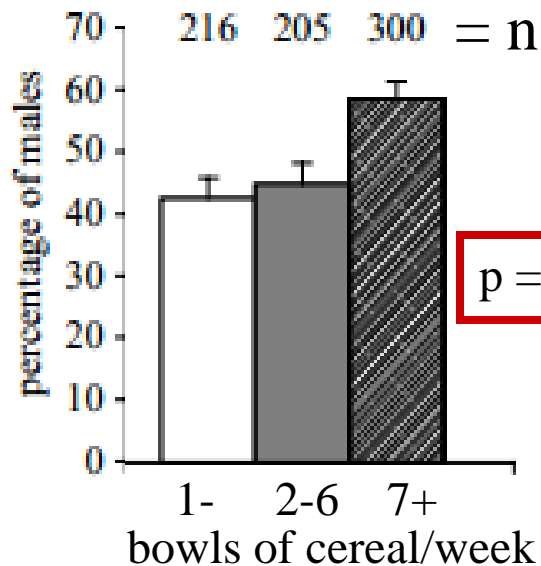
Etcetera

### **Defenses:**

- Blinding
- Randomization [but works only if n is large]
- pre-specification of experiment protocols etc

# *Too Much Choice*

## **You are what your mother eats: evidence for maternal preconception diet influencing foetal sex in humans**



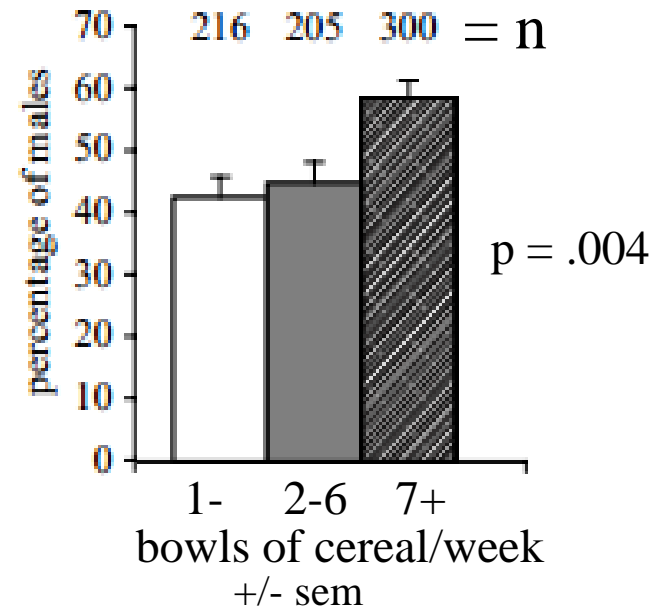
+/- sem

The odds ratio for a male infant was 1.87 (95% CI 1.31, 2.65) for women who consumed at least 7 bowls of breakfast cereal weekly compared with those who ate less than or equal to one bowlful per week. No other foods were significantly associated with infant sex.

Dietary changes may therefore explain the falling proportion of male births in industrialized countries.

**Too much choice in the supermarket,  
or too much choice in the research?**

**You are what your mother eats:  
evidence for maternal preconception diet  
influencing foetal sex in humans**



*Comment*

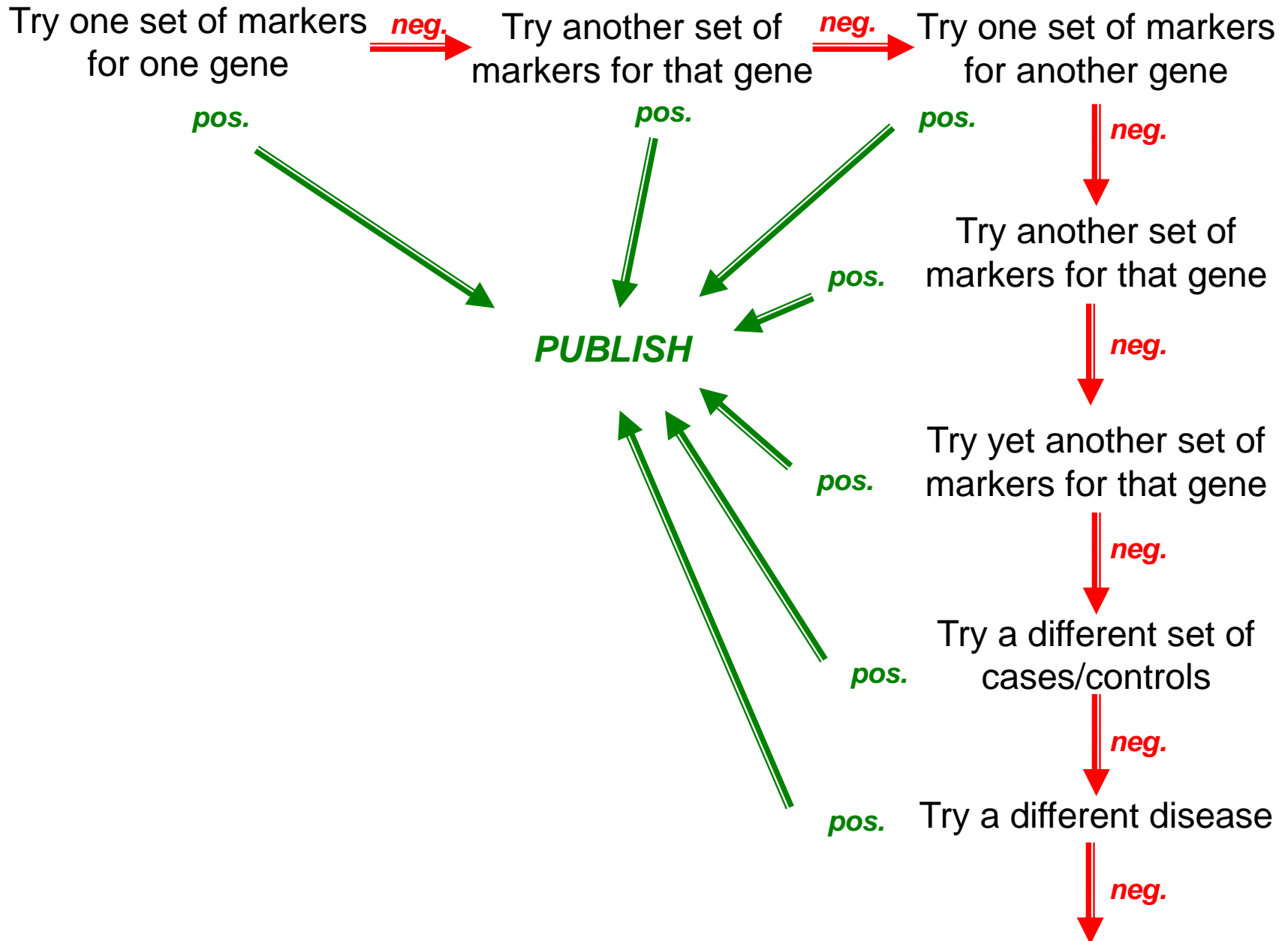
**Cereal-induced gender selection? Most likely a  
multiple testing false positive**

132 different diet variables x 3 time intervals = 396 tests

The probability of getting at least 1 test with  $p \leq .004$  is  $\sim 0.4$

- What other non-reported tests were performed?
- What is the prior probability that cereal consumption can radically alter the sex ratio?

# Dredging for hypotheses and data in a genetic association study



*Rampant*

*Heuristics*



# *Confirmation Bias*

I already expect that  $\beta$ -carotene will prevent lung cancer because:

-I like the notion that  $\beta$ -carotene will prevent a horrible disease

or

- my mom says carrots are good for you

or

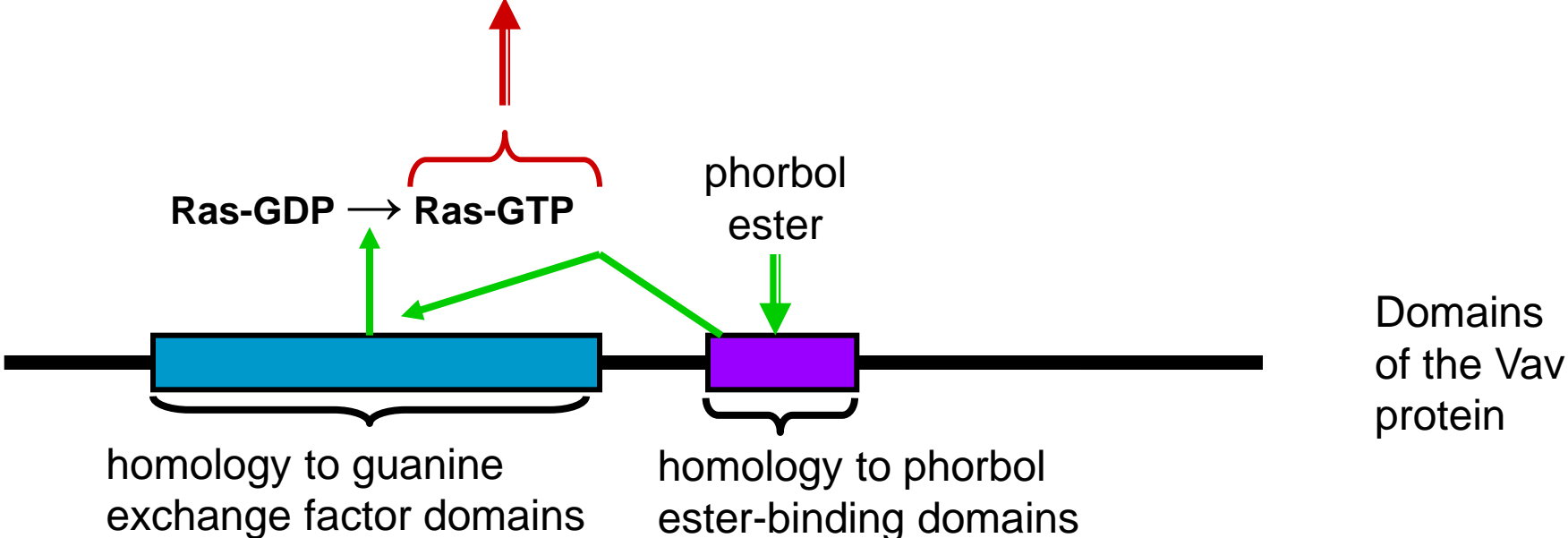
-because that's my hypothesis

## **Hazards of marrying a hypothesis:**

- In complex systems [high noise, many hidden variables], most non-obvious hypotheses will be wrong
- So hypotheses should be approached as being probably wrong, albeit worth investigating
- Love and scepticism don't mix

Direct stimulation of Vav guanine nucleotide exchange factor activity for Ras by phorbol esters and diglycerides

Causes oncogenic transformation

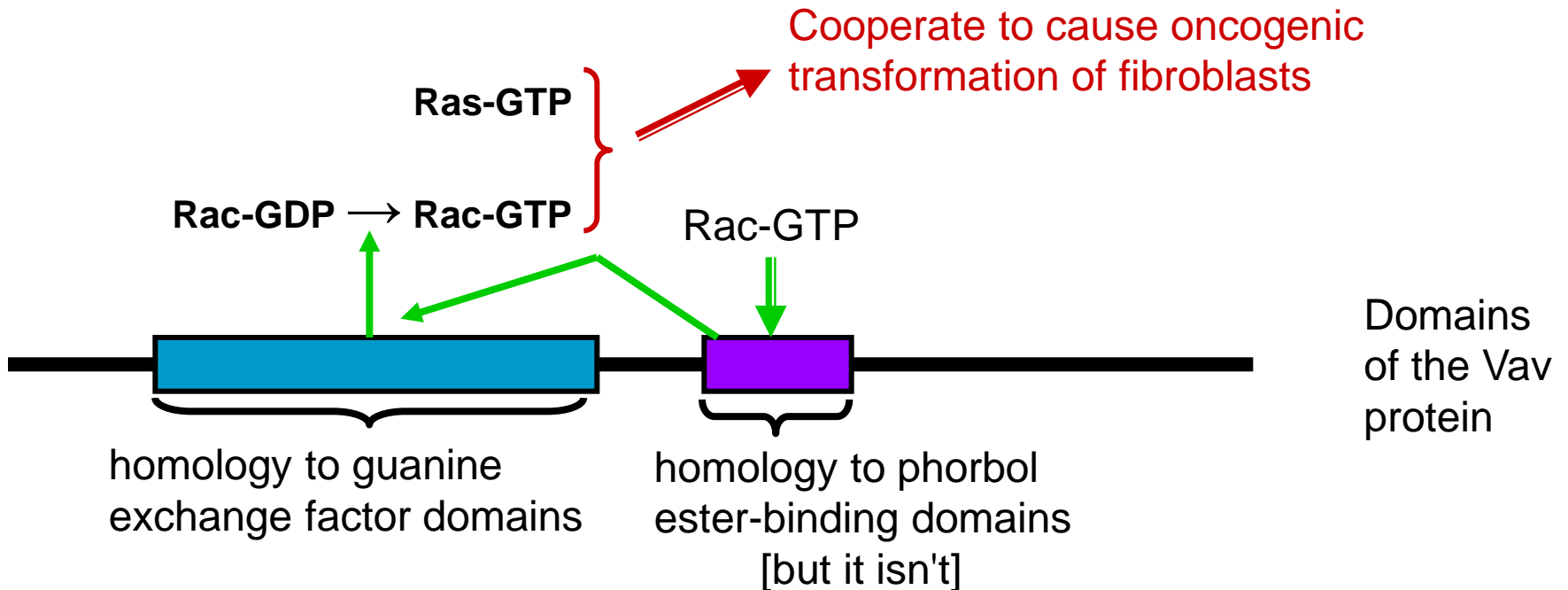


**Molecular and Cellular Biology** July 1994 p. 4749-4758

**Direct stimulation of Vav guanine nucleotide exchange factor activity for Ras by phorbol esters and diglycerides**

**Oncogene** August 1994 p. 2405-2413.

**Vav cooperates with Ras to transform rodent fibroblasts but is not a Ras GDP/GTP exchange factor**



# Commitment bias:

- My lab specializes in  $\beta$ -carotene
- I've already published on how  $\beta$ -carotene reduces cancer transplantation in mice
- I promised my mom that I'll cure lung cancer

# What confirmation and/or commitment bias causes:

-reluctance to search for possible problems, or alternative interpretations, of your hypothesis-supporting data or conclusions

-vigorous and non-objective searching for possible problems with data that contradict your hypothesis

-diligent searching for the magic combination of experimental conditions and data analyses that will make your experiments work

## **How to help your data confirm your expectations and meet your commitments:**

- exclude experiments that didn't work
- change your variables, experimental system etc. until it does work
- dredge through your data  
until you find some that works
- dredge for a stats tool  
that makes your data work
- change your hypothesis  
to fit your data
- once you find  $p < 0.05$ , don't acknowledge to yourself or others that  
you've performed a zillion different tests
- don't try to reproduce your results: the house of cards might collapse

**Defenses against confirmation and commitment biases?**



## **Defenses against confirmation and commitment biases:**

1. After exploring and refining your hypothesis and your experimental system, use it to test a pre-specified hypothesis, using pre-specified experimental conditions, data exclusion criteria, statistical analyses etc.

Do this blind and randomized if possible

2. Then see if your results are consistently reproducible under these conditions

3. Then publish, whether your results are positive or negative for your hypothesis

## **Drug development: Raise standards for preclinical cancer research**

C. Glenn Begley & Lee M. Ellis. Nature 483: 531–533, 2012

### **Prior scientific findings were confirmed in only 6 of 53 oncology drug development projects**

#### **Features of the small number of publications that were reproduced:**

Experiments were performed by blinded investigators

Experiments were repeated

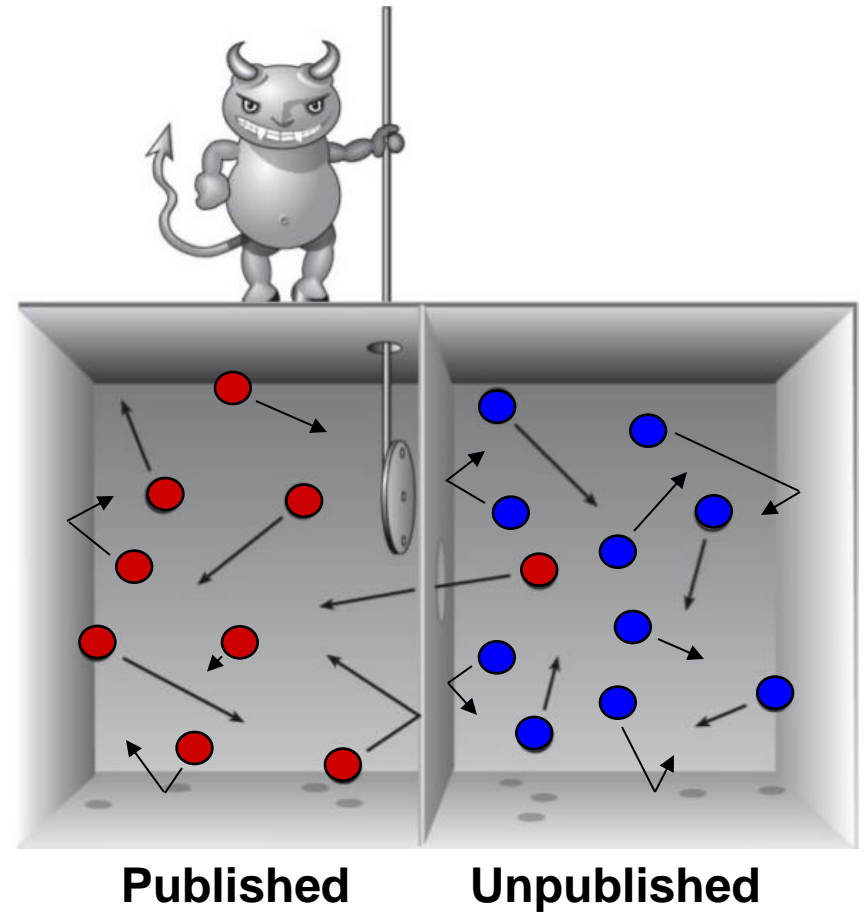
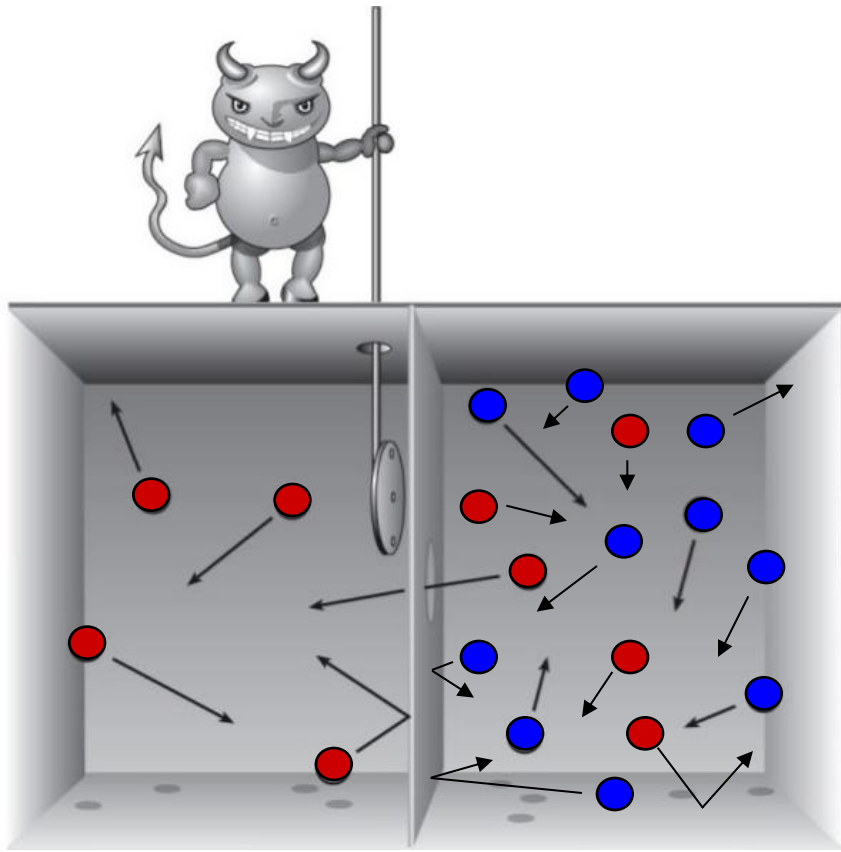
Experiments included positive and negative controls

Reagents were validated

Complete data sets were shown

Statistical tests were appropriately used

# *Publication Bias via Maxwell's Editor*



## Publication Bias

Journals are impressed by studies that confirm a hypothesis, especially if the hypothesis is exciting and novel

As a result, its much easier to publish studies that are positive than negative, even though:

- negative studies are often very useful
- positive studies are often very exaggerated
- in complex systems research, .....

# **Almost all articles on cancer prognostic markers report statistically significant results**

*Panayiotis A. Kyzas<sup>a</sup>, Despina Denaxa-Kyza<sup>a</sup>, John P.A. Ioannidis<sup>a,b,c,\*</sup>*

Of 1915 published papers on cancer biomarkers, only 22 did not make any positive claims about a prognostic value of a biomarker.

# Publication Bias in Reports of Animal Stroke Studies Leads to Major Overstatement of Efficacy

Emily S. Sena et al., PLoS Biol. 8(3): e1000344, 2010

We identified 525 publications of interventions tested in animal studies of acute ischaemic stroke.

Only ten publications (2%) reported no significant effects on infarct volume and only six (1.2%) did not report at least one significant finding.

Egger regression and trim-and-fill analysis suggested that publication bias was highly prevalent in animal studies modelling stroke.

Reported efficacy [would fall] from 31.3% to 23.8% after adjustment for publication bias.

## Publication Bias

Journals are impressed by studies that confirm a hypothesis, especially if the hypothesis is exciting and novel

As a result, its much easier to publish studies that are positive than negative, even though:

- negative studies are often very useful
- positive studies are often very exaggerated
- in complex systems research, .....

**How does Publication Bias affect you???**



## **How does Publication Bias affect you???**

The literature relevant to your research is incomplete, and highly biased due to depletion of negative results.

You should always consider the potential for a published study to be exaggerated, or even completely wrong.

You should try to avoid contributing to publication bias

# ***Counterproductive Science Culture***

*Don't blame me, my culture made me do it!*

C. Glenn Begley. Reproducibility: Six red flags for suspect work  
Nature 497:433–434, 2013.

**“Why do we repeatedly see these poor-quality papers in basic science?”**

- There are insufficient consequences for investigators or journals
- Many reviewers and co-authors do not actually read the papers
- Journals want to fill their pages with simple, complete, positive 'stories'
- Failure to recognize authors' competing interests that may interfere with their judgement

**What can science society do to improve research validity?**

# Consortium guidelines for improving the reporting [and hopefully conduct] of observational and experimental studies

STROBE = Strengthening the Reporting of Observational Studies in Epidemiology

STREGA = STrengthening the REporting of Genetic Association studies

Provides specific guidelines for:

- selection and participation of study participants
- rationale for choice of genes and variants investigated
- genotyping errors
- methods for inferring haplotypes
- population stratification
- assessment of Hardy–Weinberg equilibrium
- multiple testing
- reporting of quantitative outcomes
- selective reporting of study results

# Raising Standards: Nature journals' updated editorial policies aim to improve transparency and reproducibility.

Nature Neuroscience 16: 517, 2013

Ensuring systematic attention to reporting and transparency is only a small step toward solving the issues of reproducibility that have been highlighted across the life sciences and particularly in biomedical research. Much bigger underlying issues contribute to the problem.

- Too many biologists still do not receive adequate training in statistics and other quantitative aspects of their subject.
- Mentoring of young scientists on matters of rigor and transparency is inconsistent at best.
- In academia, the ever-increasing pressures to publish and obtain the next level of funding provide little incentive to pursue and publish studies that contradict or confirm previously published results.
- Those who would put effort into documenting the validity or irreproducibility of a published piece of work have little prospect of seeing their efforts valued by journals and funders; meanwhile, funding and efforts are wasted on false assumptions.

## Editorial: Reproducibility Science 343: 229, 2014

Recently, the scientific community was shaken by reports that a troubling proportion of peer-reviewed preclinical studies are not reproducible.

Because confidence in results is of paramount importance to the broad scientific community, we are announcing new initiatives to increase confidence in the studies published in Science.

For preclinical studies, we will be adopting recommendations of the U.S. National Institute of Neurological Disorders and Stroke (NINDS) for increasing transparency. Authors will indicate whether:

- there was a pre-experimental plan for data handling
- whether they conducted a sample size estimation
- whether samples were treated randomly
- whether the experimenter was blind to the conduct of the experiment

We are adding new members to our Board of Reviewing Editors from the statistics community to ensure that manuscripts receive appropriate scrutiny in their methods of data analysis.

# NIH plans to enhance reproducibility

Nature 505: 612, 2014

**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.

The recent evidence showing the irreproducibility of significant numbers of biomedical-research publications demands immediate and substantive action. The NIH is firmly committed to making systematic changes that should reduce the frequency and severity of this problem.

Here are some of the significant interventions that we are planning.

- Developing a training module on enhancing reproducibility and transparency of research findings, with an emphasis on good experimental design
- To counteract the over-emphasis on publishing in high-profile journals, which encourages rapid submission of research findings to the detriment of careful replication, we may modify the 'biographical sketch' form to emphasize the significance of advances resulting from work in which the applicant participated





## Psychology's Bold Initiative

**In an unusual attempt at scientific self-examination, psychology researchers are scrutinizing their field's reproducibility**

One project begun this year offers a site where psychologists can quickly and easily post the results of replications of experiments—whether they succeed or fail.

The Open Science Collaboration has begun a large-scale project to systematically replicate psychological experiments recently published in leading journals.