# Reporting and reproducible research: salvaging the self-correction principle of science

John P.A. Ioannidis

Stanford University

# Self-correction in science

- The ability of self-correction is considered one of the main features of science.
- In a cumulative meta-analysis framework, if sufficient time elapses, the accumulation of replication effects should gravitate towards the "truth".
- However, self-correction is often not happening.
- Self-correction may be impeded by destruction of evidence, production of wrong evidence, and/or distortion of evidence.

# Ancient scholarship: >1,000,000 volumes of information

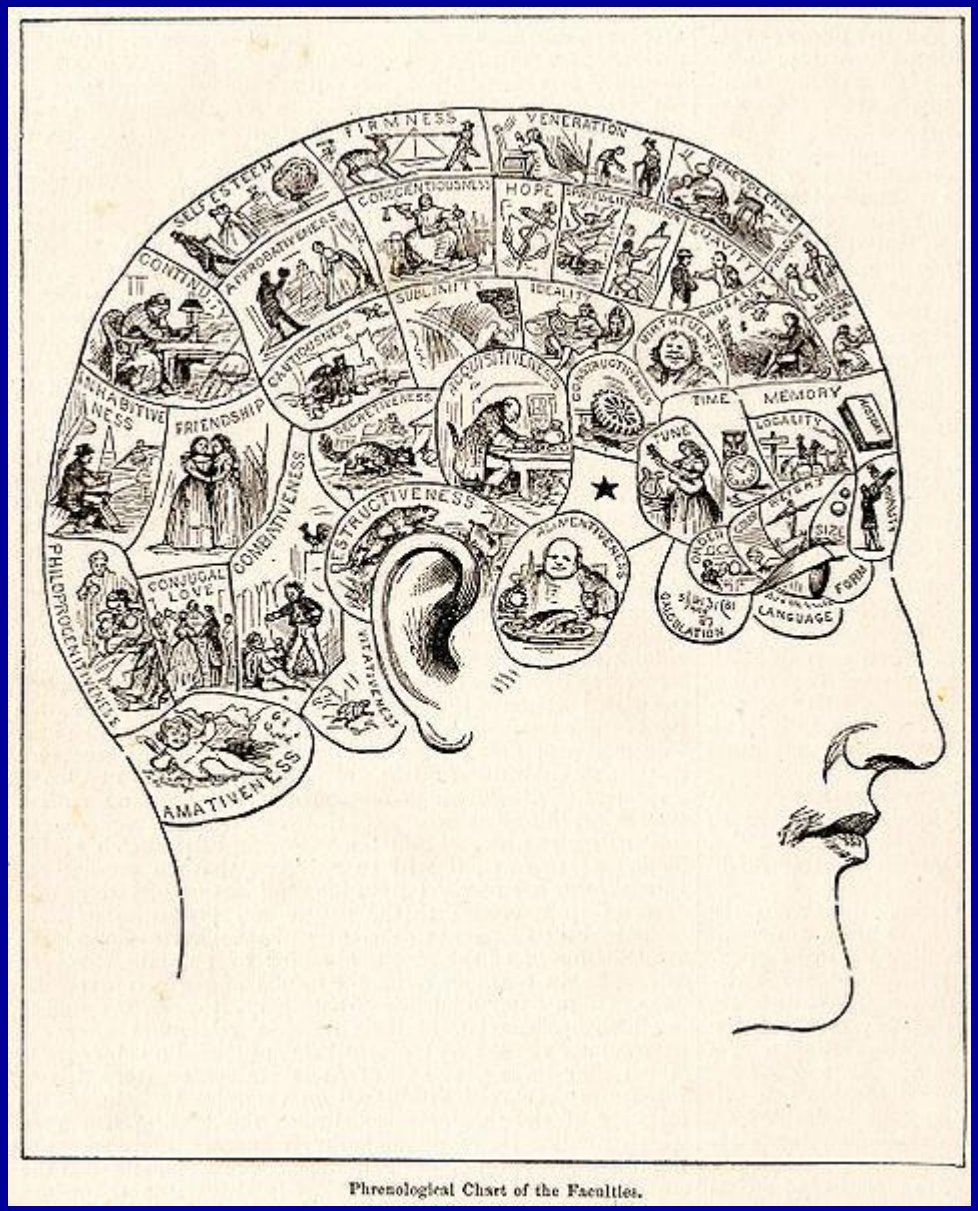
# Destroyed four times until final extinction, <1% surviving

# Destruction of evidence

- Julius Caesar defending Roman interests and his mistress Cleopatra
- War between Emperor Aurelian and Queen Zenobia of Palmyra
- Christian mobs licensed by the zealot Pope of Alexandria to destroy hallmarks of Hellenic civilization
- Arab conquest – we have no need of replication: “Quod ad libros quorum mentionem fecisti: si in illis contineatur, quod cum libro Dei conveniat, in libro Dei [est] quod sufficiat absque illo; quod si in illis fuerit quod libro Dei repugnet, neutiquam est eo [nobis] opus, jube igitur e medio tolli.”

# Destruction of scientists

Hypatia of Alexandria (according to Raphael)




Phrenological Chart of the Faculties.

# Production of wrong evidence or distortion of evidence

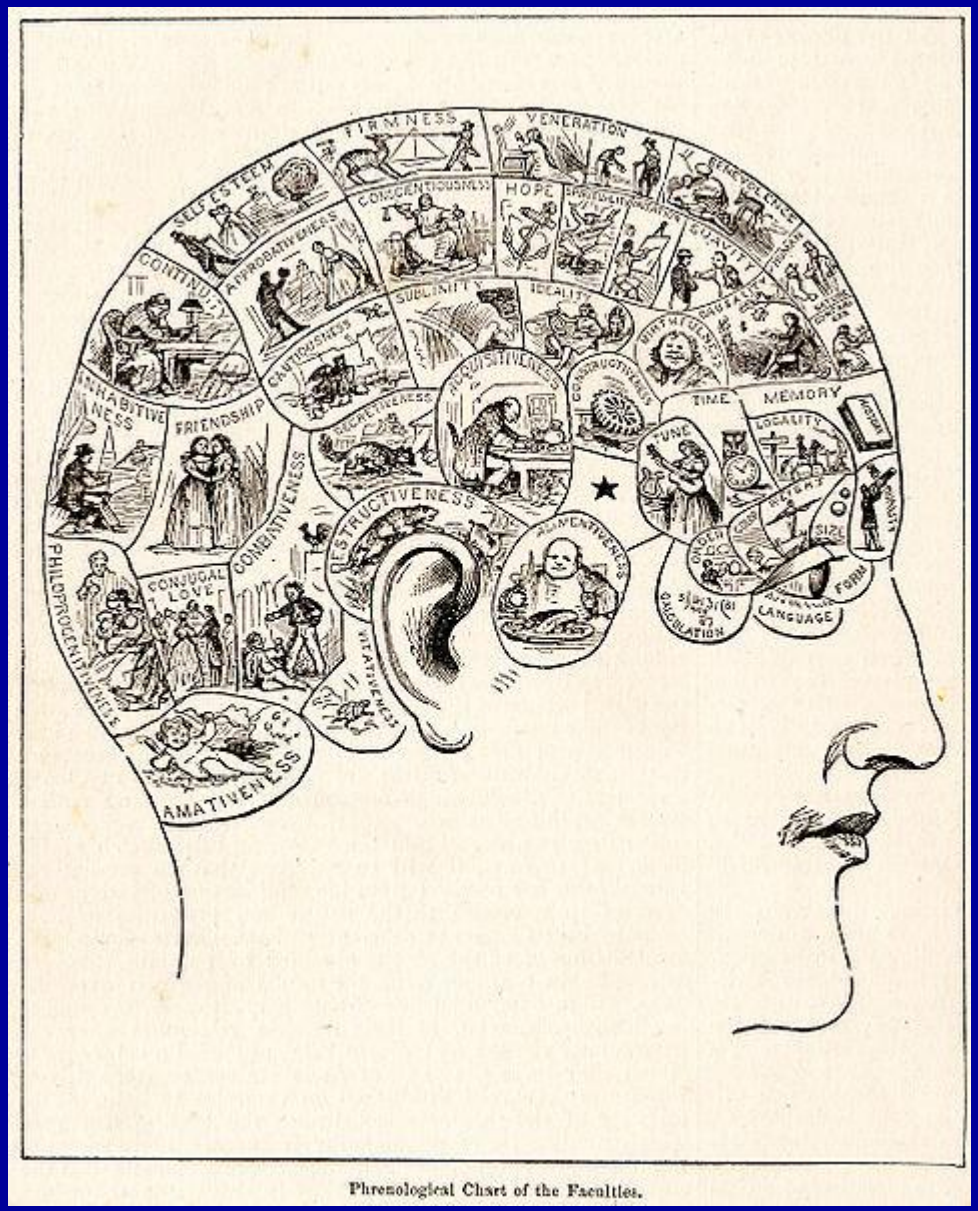
# From wikipedia

- In the context of Victorian society phrenology was a respectable scientific theory. The Phrenological Society of Edinburgh founded by George and Andrew Combe was an example of the credibility of phrenology at the time…In 1826, out of the 120 members of the Edinburgh society an estimated one third were from a medical background and by the 1840s there were over twenty-eight phrenological societies in London with over 1000 members.

# Could it be that several Libraries of Alexandria disappear daily?

# People who seek knowledge continue to be shot

Malala Yousafzai, 14 years old, shot for wanting to get educated




Phrenological Chart of the Faculties.

Could it be that many phrenologies circulate in biomedical journals nowadays?

# Possibilities for discovery and replication

- Optimal: Discovery correct – Replication correct
- Self-correcting: Discovery wrong – Replication correct
- False non-replication: Discovery correct – Replication wrong
- Perpetuated fallacy: Discovery wrong – Replication wrong
- Unchallenged fallacy: Discovery wrong – Replication not done
- Unconfirmed (genuine) discovery: Discovery correct – Replication not done

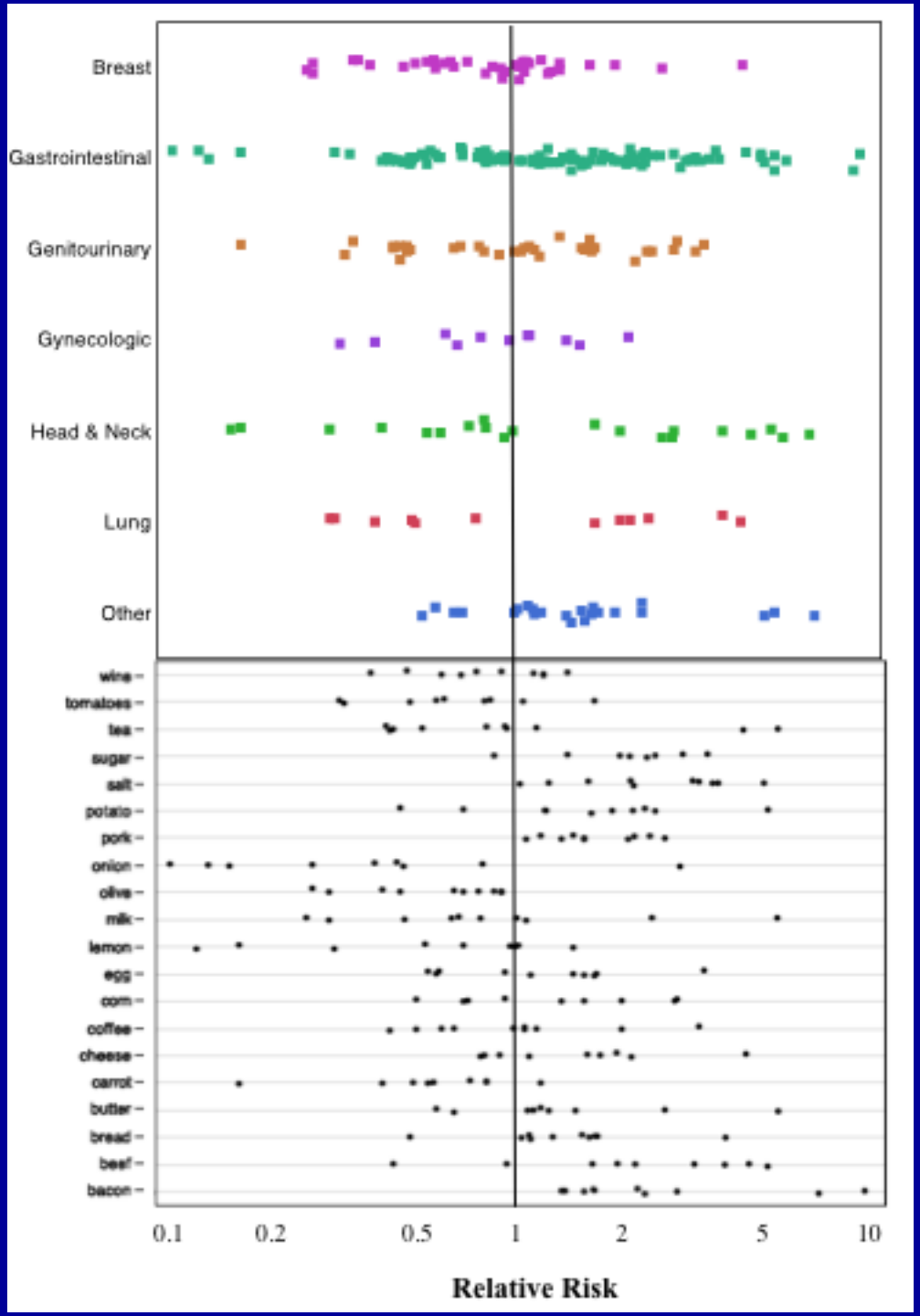
# Diet causes cancer

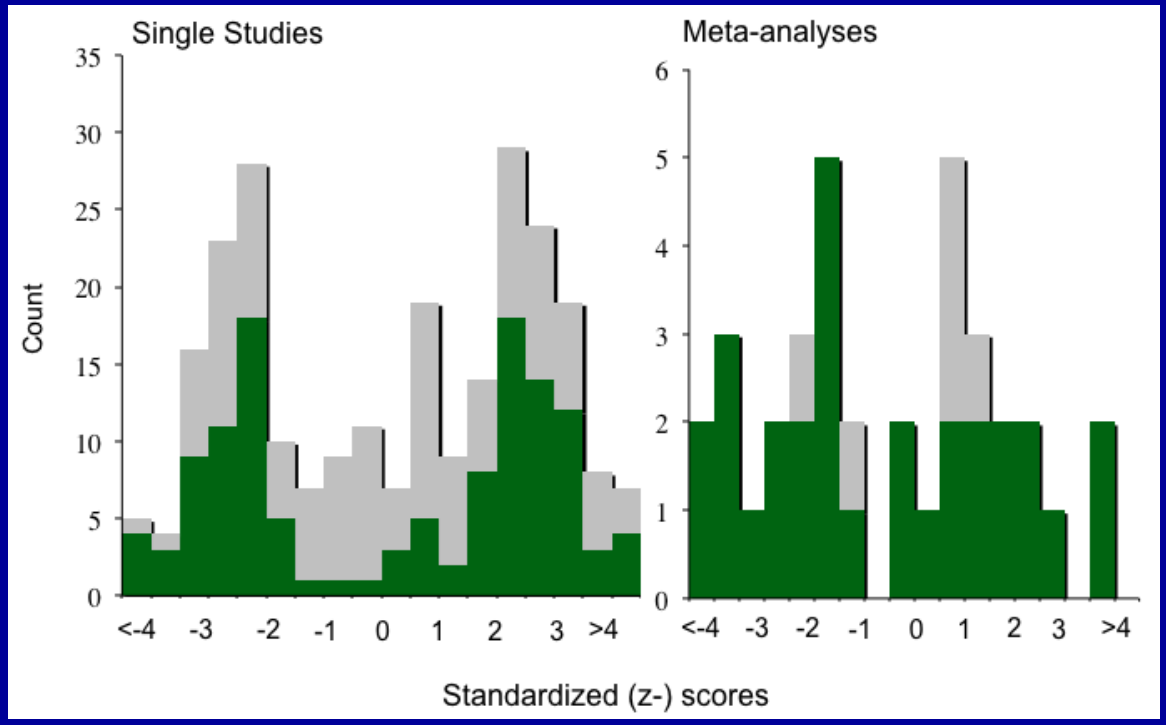
- Open a popular cookbook
- Randomly check 50 ingredients
- How many of those are associated with increased or significantly decreased cancer risk in the scientific literature?

# Associated with cancer risk

- veal, salt, pepper spice, flour, egg, bread, pork, butter, tomato, lemon, duck, onion, celery, carrot, parsley, mace, sherry, olive, mushroom, tripe, milk, cheese, coffee, bacon, sugar, lobster, potato, beef, lamb, mustard, nuts, wine, peas, corn, cinnamon, cayenne, orange, tea, rum, raisin

Schoenfeld and Ioannidis, Am J Clin Nutrition 2012 in press


Breast
Gastrointestinal
Genitourinary
Gynecologic
Head & Neck
Lung
Other
wine
tomatoes
tea
sugar
salt
potato
pork
onion
olive
milk
lemon
egg
corn
coffee
cheese
carrot
butter
bread
beef
bacon
0.1
0.2
0.5
1
2
5
10
Relative Risk

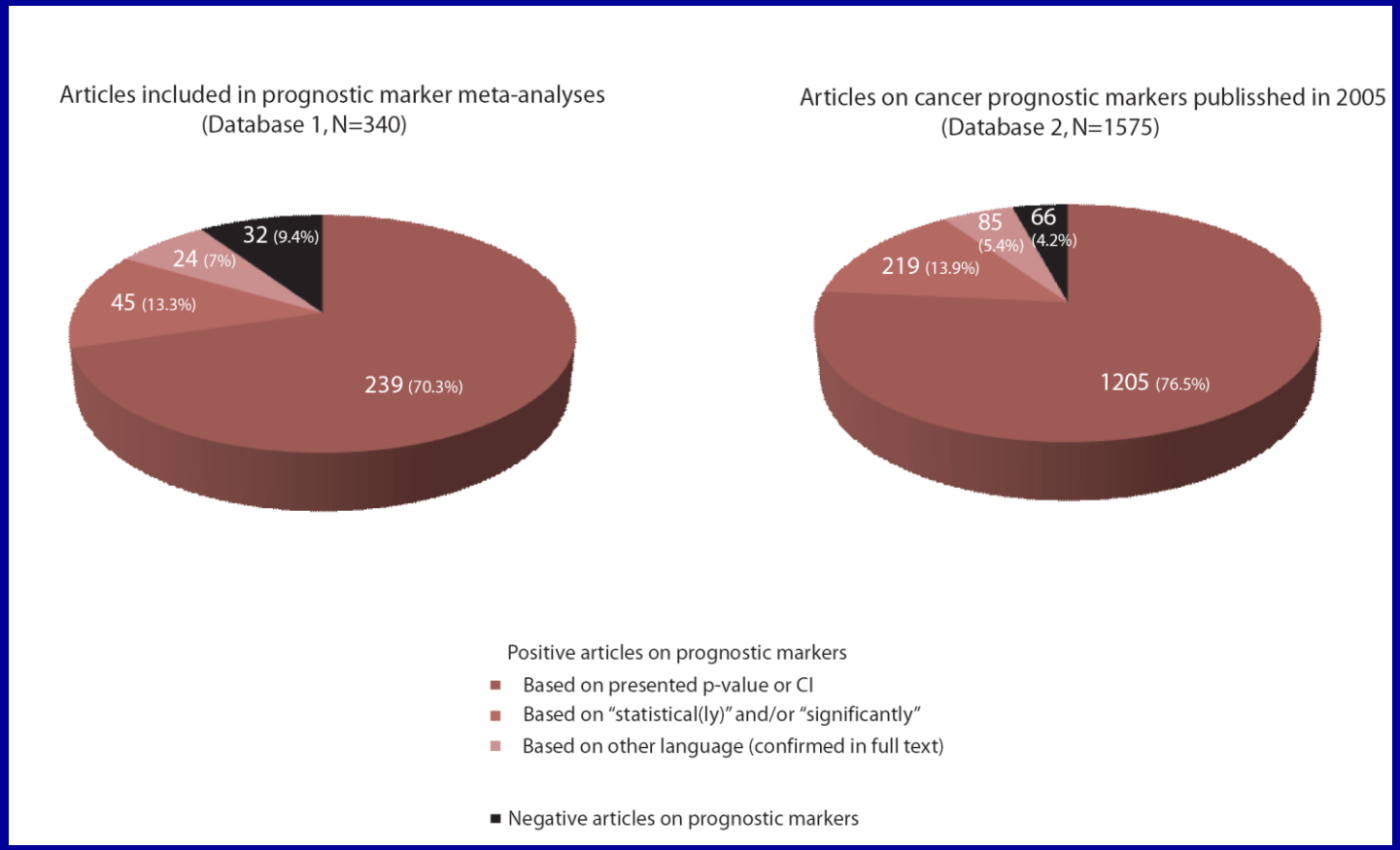

Single Studies
Meta-analyses
Count
35
30
25
20
15
10
5
0
6
5
4
3
2
1
0
<-4
-3
-2
-1
0
1
2
3
>4
Standardized (z-) scores

# Almost all biomedical studies currently are “statistically significant”

- In a survey of 389 epidemiological studies published in 2004-2005 in major general and specialty medical journals, 89% highlighted some statistically significant relative risk

Kavvoura et al. PLoS Medicine (March 2007)

# How can it all be significant?

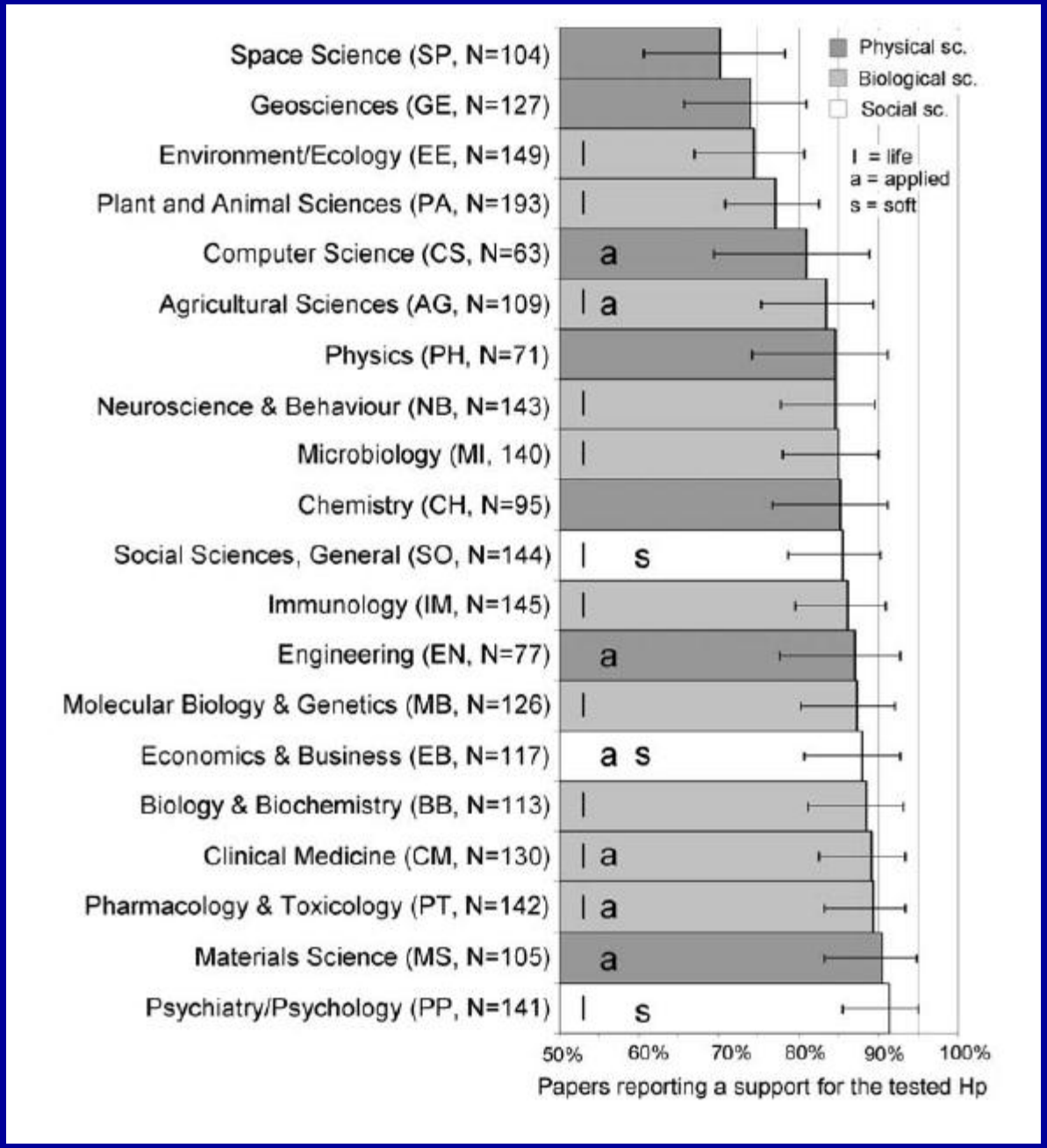



Kyzas et al. Eur J Cancer 2007

# Further analysis of claims in "negative" prognostic studies

| | **Database 1** N (%) | **Database 2** N (%) |
|---|---|---|
| Not admitted to be fully "negative" | 27 (7.9) | 45 (2.8) |
| Significance for other (non-prognostic) analyses | 6 (1.7) | 11 (0.6) |
| Discussion of non-significant trends | 2 (0.6) | 5 (0.3) |
| Offered apologies | 9 (2.8) | 13 (0.8) |
| Significance for other analyses + Discussion of non-significant trends | 1 (0.3) | 3 (0.2) |
| Significance for other analyses + Offered apologies | 6 (1.7) | 7 (0.5) |
| Discussion of non-significant trends + Offered apologies | 3 (0.8) | 4 (0.3) |
| All three mechanisms | - | 2 (0.1) |
| Admitted to be fully "negative" | 5 (1.5) | 21 (1.3) |

Kyzas et al. Eur J Cancer 2007




Fanelli, PLoS ONE 2010

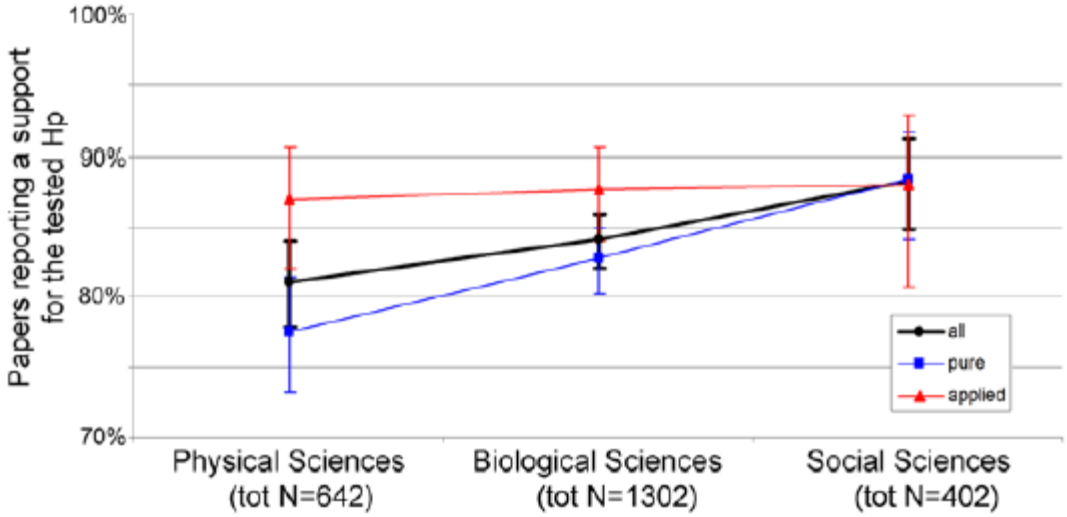



**Figure 3. Positive Results by Disciplinary Domain.** Percentage of papers that supported a tested hypothesis, classified by disciplinary domain. Blue = including only pure disciplines, Red = including only applied disciplines, Black = all disciplines included. Error bars represent

Empirical studies suggest that most of the claimed statistically significant effects in traditional medical research are false positives or substantially exaggerated

# Estimates from psychological science



**Table 1.** Possibilities of Discovery and Replication: Six Possible Paradigms

| | Replication results | | |
|---|---|---|---|
| Discovery results | Correct | Wrong | Not obtained |
| Correct (true positive) | Optimal: ≤ 1%* | False nonreplication: <<1%* | Unconfirmed genuine discovery: 43%** |
| Wrong (false positive) | Self-correcting: ≤1%* | Perpetuated fallacy: 2%* | Unchallenged fallacy: 53%** |

*The sum of the items in the first two columns is assumed to be close to 4%, a probably generous estimate vis-à-vis the data by Makel et al. (2012).

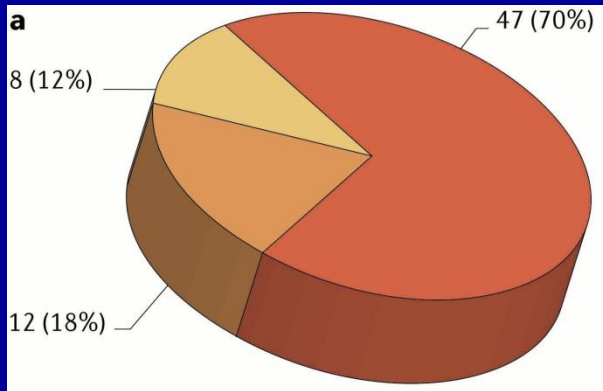
Ioannidis, Perspectives on Psychological Science in press

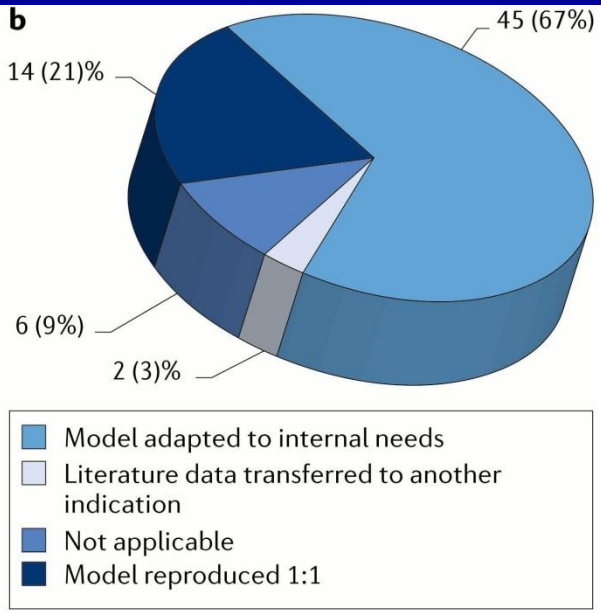
# Candidate genes replicated through GWAS: in search for survivors

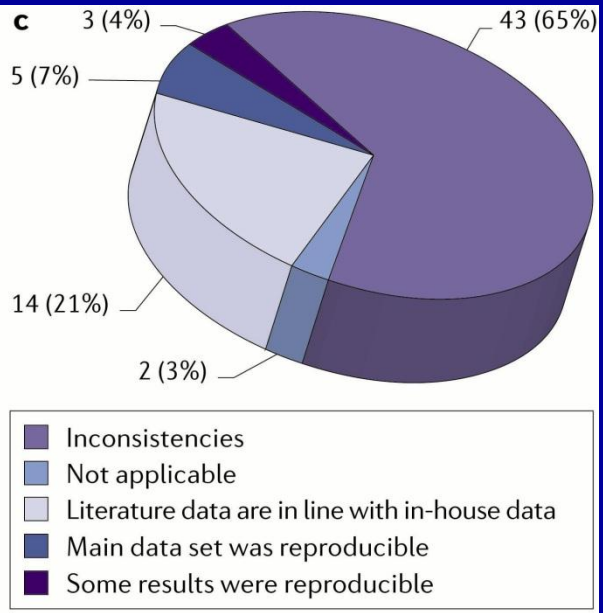
Table. Large-scale efforts to massively replicate reported candidate gene associations

| First author | Disease/phenotype | Gene loci tested | Sample size (design) | Replicated gene loci |
|---|---|---|---|---|
| Bosker (16) | Major depressive disorder | 57 | 3540 (Case-control) | 1 |
| Caporaso (17) | Smoking (7 phenotypes) | 359 | 4611 (Cohort) | 1 |
| Morgan (18) | Acute coronary syndrome | 70 | 1461 (Case-control) | 0 |
| Richards (19) | Osteoporosis (2 phenotypes) | 150 | 19,195 (Cohort) | 9 |
| Samani (20) | Coronary artery disease | 55 | 4864; 2519 (Case-control) | 1 |
| Scuteri (21) | Obesity (3 phenotypes) | 74 | 6148 (Cohort) | 0 |
| Soeber (22) | Blood pressure | 149 | 1644; 8023 (Cohort) | 0 |
| Wu (23) | Childhood asthma | 237 | 1476 (Triads) | 1 |

Ioannidis, Tarone, McLaughlin, Epidemiology 2011




**d**

| | Model reproduced 1:1 | Model adapted to internal needs (cell line, assays) | Literature data transferred to another indication | Not applicable |
|---|---|---|---|---|
| In-house data in line with published results | 1 (7%) | 12 (86%) | 0 | 1 (7%) |
| Inconsistencies that led to project termination | 11 (26%) | 26 (60%) | 2 (5%) | 4 (9%) |

Nature Reviews | Drug Discovery

Prinz et al., Nature Reviews Drug Discovery 2011

# Replicated: only 6 of 53 landmark studies for Amgen oncology drug target projects

- “The failure to win “the war on cancer” has been blamed on many factors, … But recently a new culprit has emerged: too many basic scientific discoveries… are wrong.”

Begley et al. Nature 2012

# Non-reproducible research is highly-cited

| Journal impact factor | Number of articles | Mean number of citations of non-reproduced articles | Mean number of citations of reproduced articles |
|---|---|---|---|
| >20 | 21 | 248 (range 3– 800) | 231 (range 82–519) |
| 5–19 | 32 | 169 (range 6–1,909) | 13 (range 3–24) |

*Source of citations: Google Scholar, May 2011.

# Hedge funds don’t trust science

- “at least 50% of published studies, even those in top-tier academic journals, can't be repeated with the same conclusions by an industrial lab”
- “The potential for not being able to reproduce academic data is a disincentive to early stage investors.”

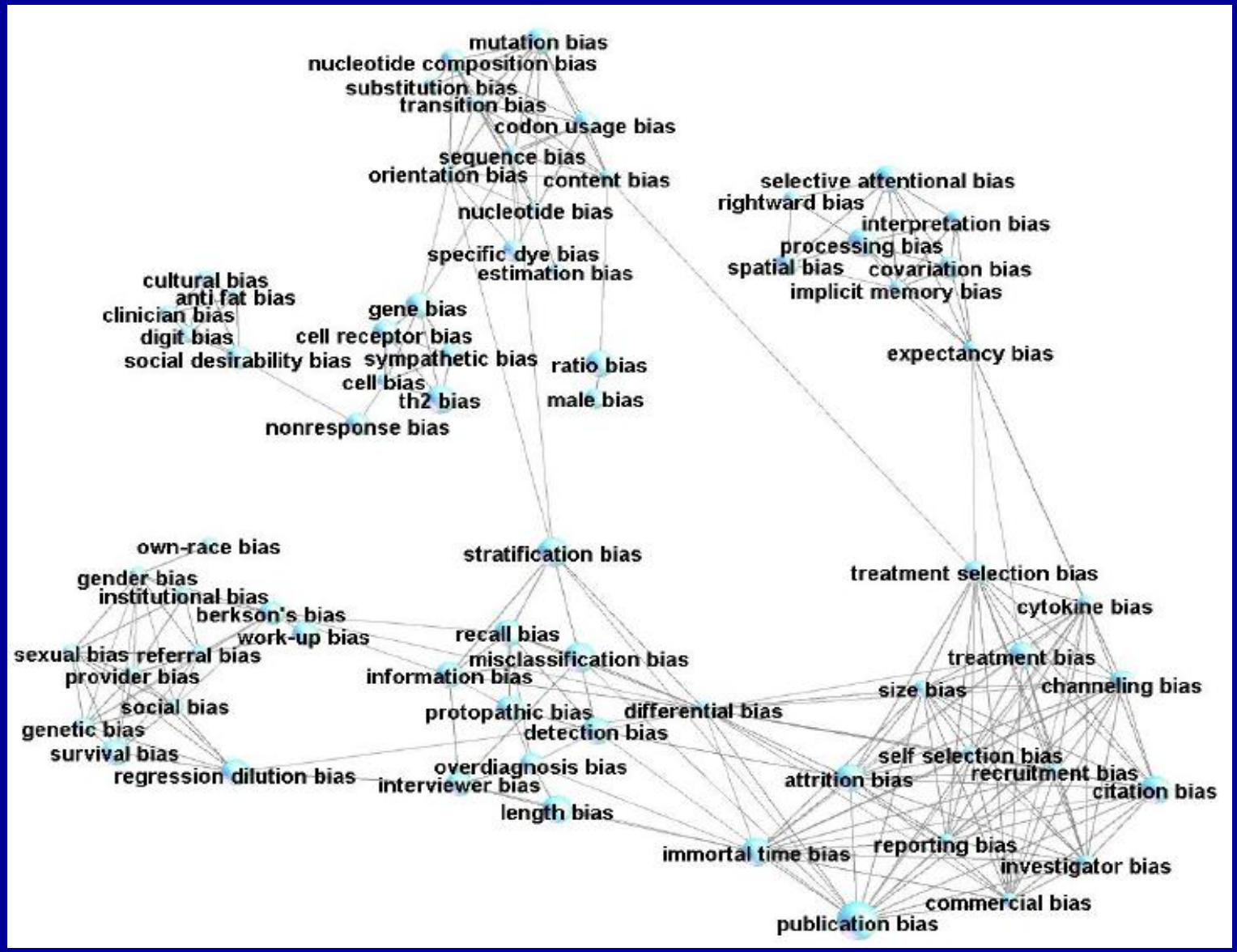
Osherovich L. Hedging against academic risk. SciBX 4(15);doi:10.1038/scibx.2011.416. Published online April 14 2011

# Why research findings may not be credible?

- There is bias
- There is random error (see multiple comparisons)

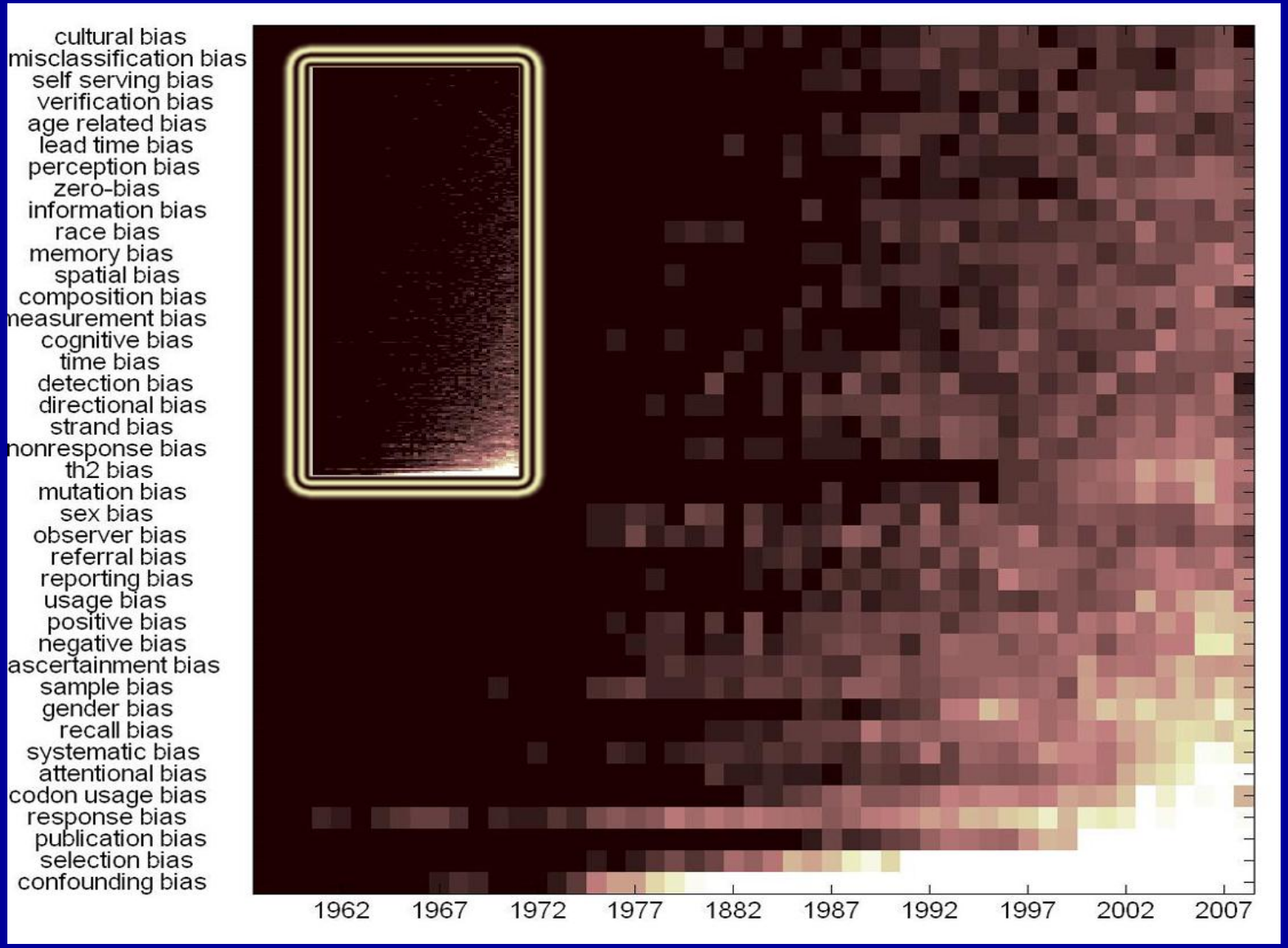
- Usually there is plenty of both

# Mapping 235 biases in 17 million Pub Med papers

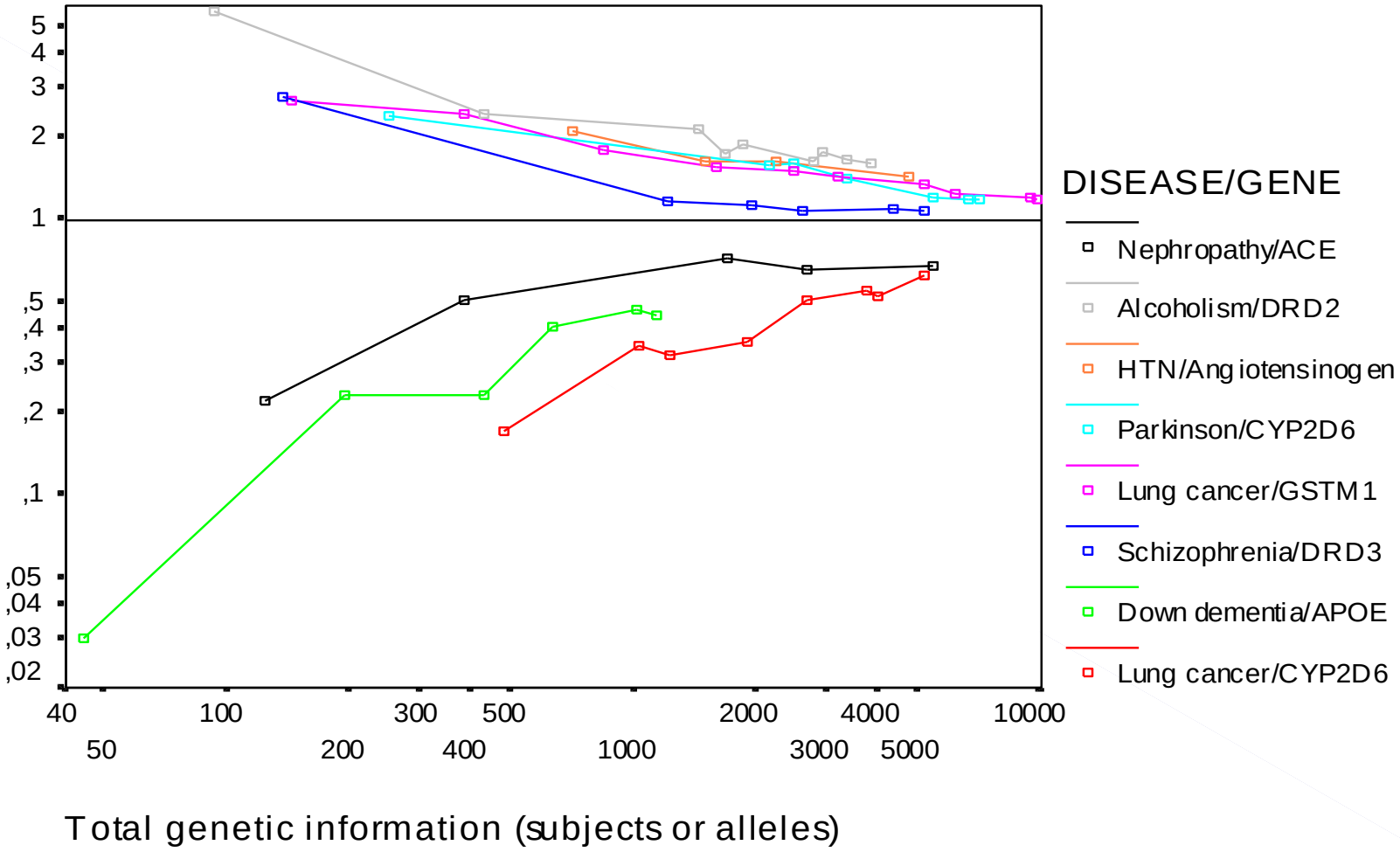



Chavalarias and Ioannidis, JCE 2010

# A time array for biases



# Corrected bias: non-replication




Ioannidis et al, Nature Genetics 2001

# Contradicted and Initially Stronger Effects in Highly Cited Clinical Research

John P. A. Ioannidis, MD

**Context** Controversy and uncertainty ensue when the results of clinical research on the effectiveness of interventions are subsequently contradicted. Controversies are most prominent when high-impact research is involved.

**Objectives** To understand how frequently highly cited studies are contradicted or find effects that are stronger than in other similar studies and to discern whether specific characteristics are associated with such refutation over time.

**Design** All original clinical research studies published in 3 major general clinical journals or high-impact-factor specialty journals in 1990-2003 and cited more than 1000 times in the literature were examined.

**Main Outcome Measure** The results of highly cited articles were compared against subsequent studies of comparable or larger sample size and similar or better controlled designs. The same analysis was also performed comparatively for matched studies that were not so highly cited.

**Results** Of 49 highly cited original clinical research studies, 45 claimed that the intervention was effective. Of these, 7 (16%) were contradicted by subsequent studies, 7 others (16%) had found effects that were stronger than those of subsequent studies, 20 (44%) were replicated, and 11 (24%) remained largely unchallenged. Five of 6 highly-cited nonrandomized studies had been contradicted or had found stronger effects vs 9 of 39 randomized controlled trials ($P=.008$). Among randomized trials, studies with contradicted or stronger effects were smaller ($P=.009$) than replicated or unchallenged studies although there was no statistically significant difference in their early or overall citation impact. Matched control studies did not have a significantly different share of refuted results than highly cited studies, but they included more studies with "negative" results.

**Conclusions** Contradiction and initially stronger effects are not unusual in highly cited research of clinical interventions and their outcomes. The extent to which high citations may provoke contradictions and vice versa needs more study. Controversies are most common with highly cited nonrandomized studies, but even the most highly cited randomized trials may be challenged and refuted over time, especially small ones.

*JAMA. 2005;294:218-228* www.jama.com

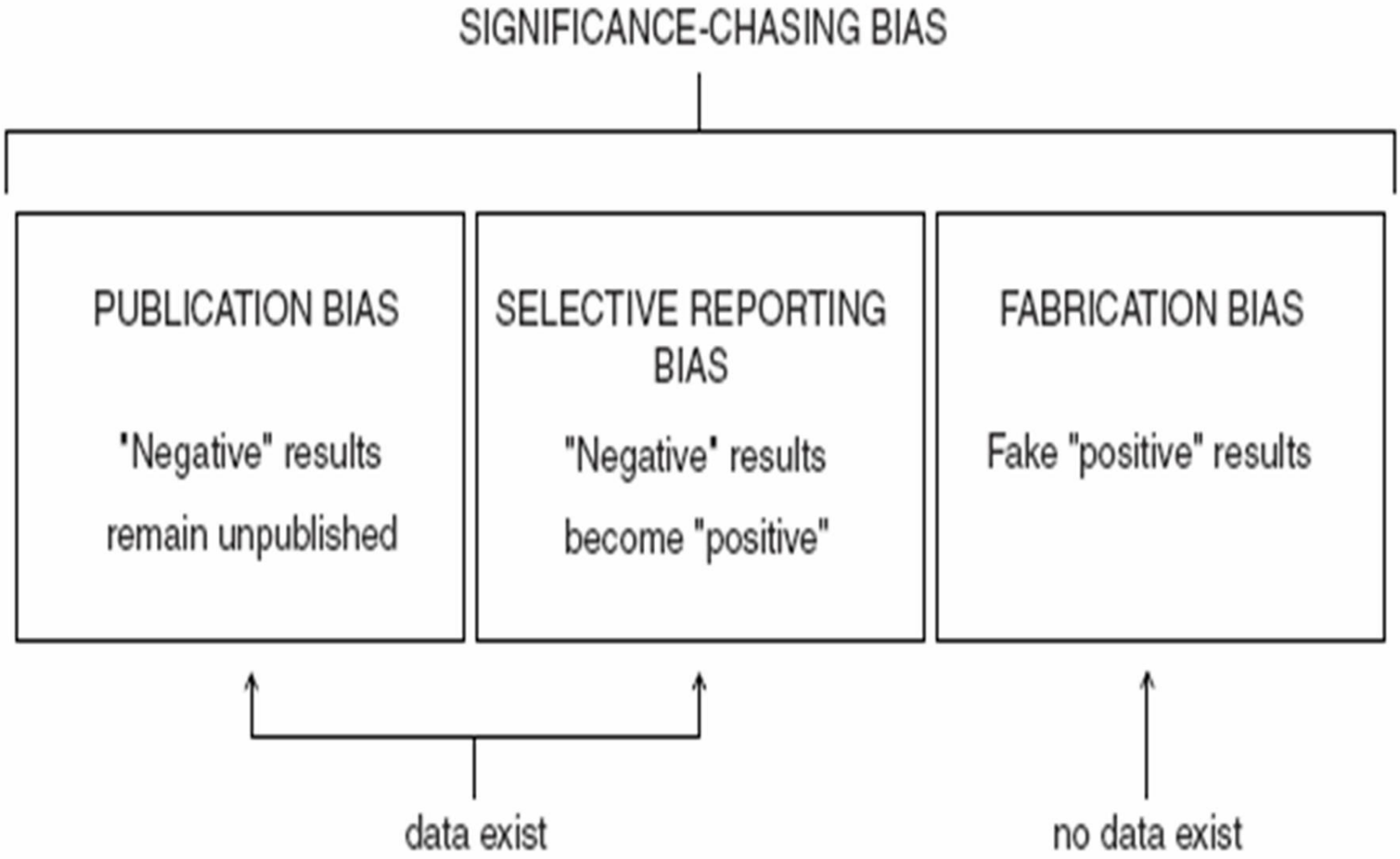
CLINICAL RESEARCH ON IMPORtant questions about the efficacy of medical interventions is sometimes followed by subsequent studies that either reach opposite conclusions or suggest that the original claims were too strong. Such disagreements may upset clinical practice and acquire publicity in both scientific circles and in the lay press. Several empirical investigations have tried to address whether specific types of studies are more likely to be contradicted and to explain observed controversies. For example, evidence exists that small studies may sometimes be refuted by larger ones.[1,2]

Similarly, there is some evidence on disagreements between epidemiological studies and randomized trials.[3-5] Prior investigations have focused on a variety of studies without any particular attention to their relative importance and scientific impact. Yet, most research publications have little impact while a small minority receives most attention and dominates scien-

# Excess significance bias

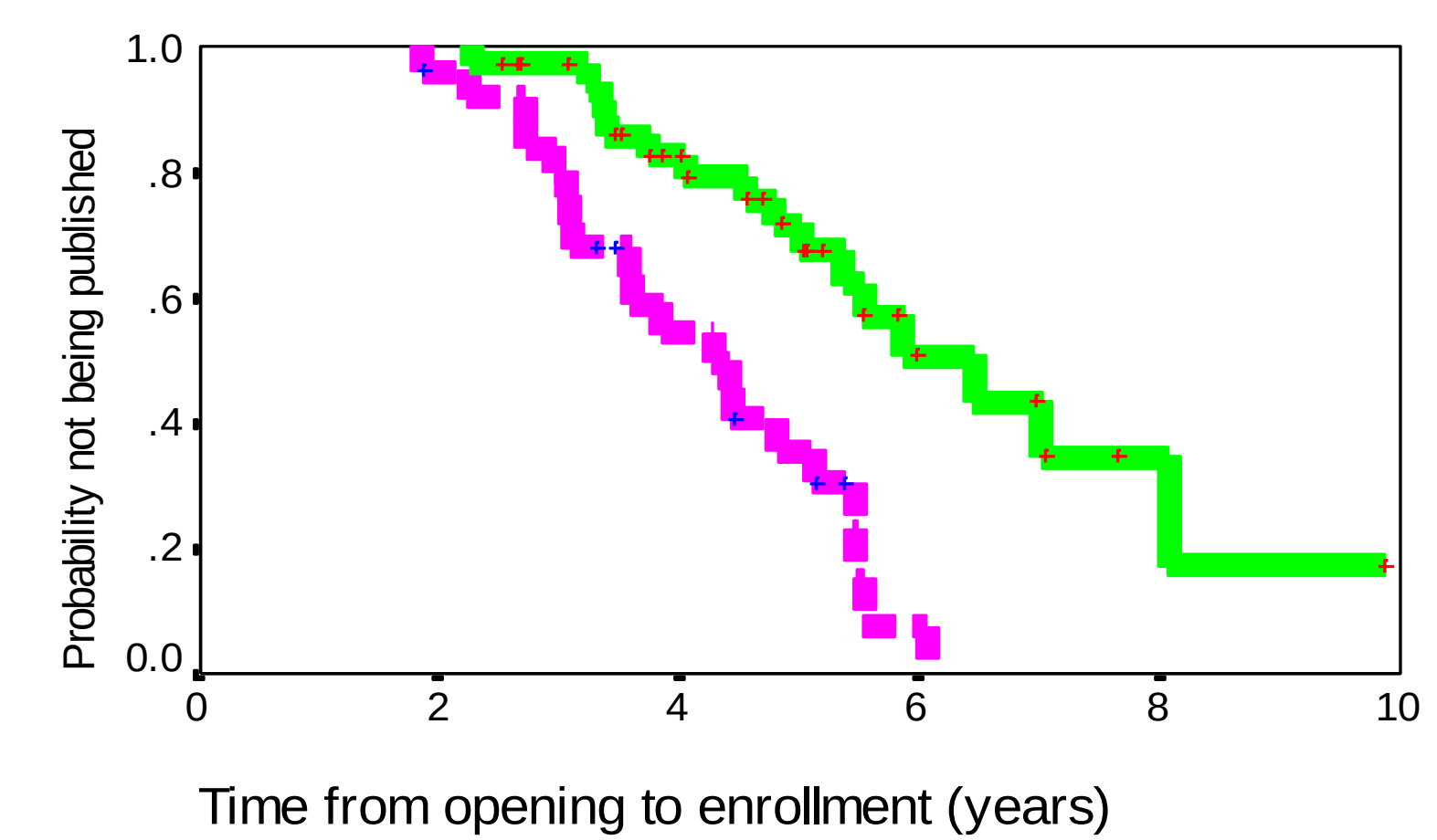
- Results that become “positive” while they should have been “negative”
- Results that are “negative” are suppressed
- Fake “positive” results are created

- The common consequence of all these practices is an inflation in the proportion of observed “positive” results



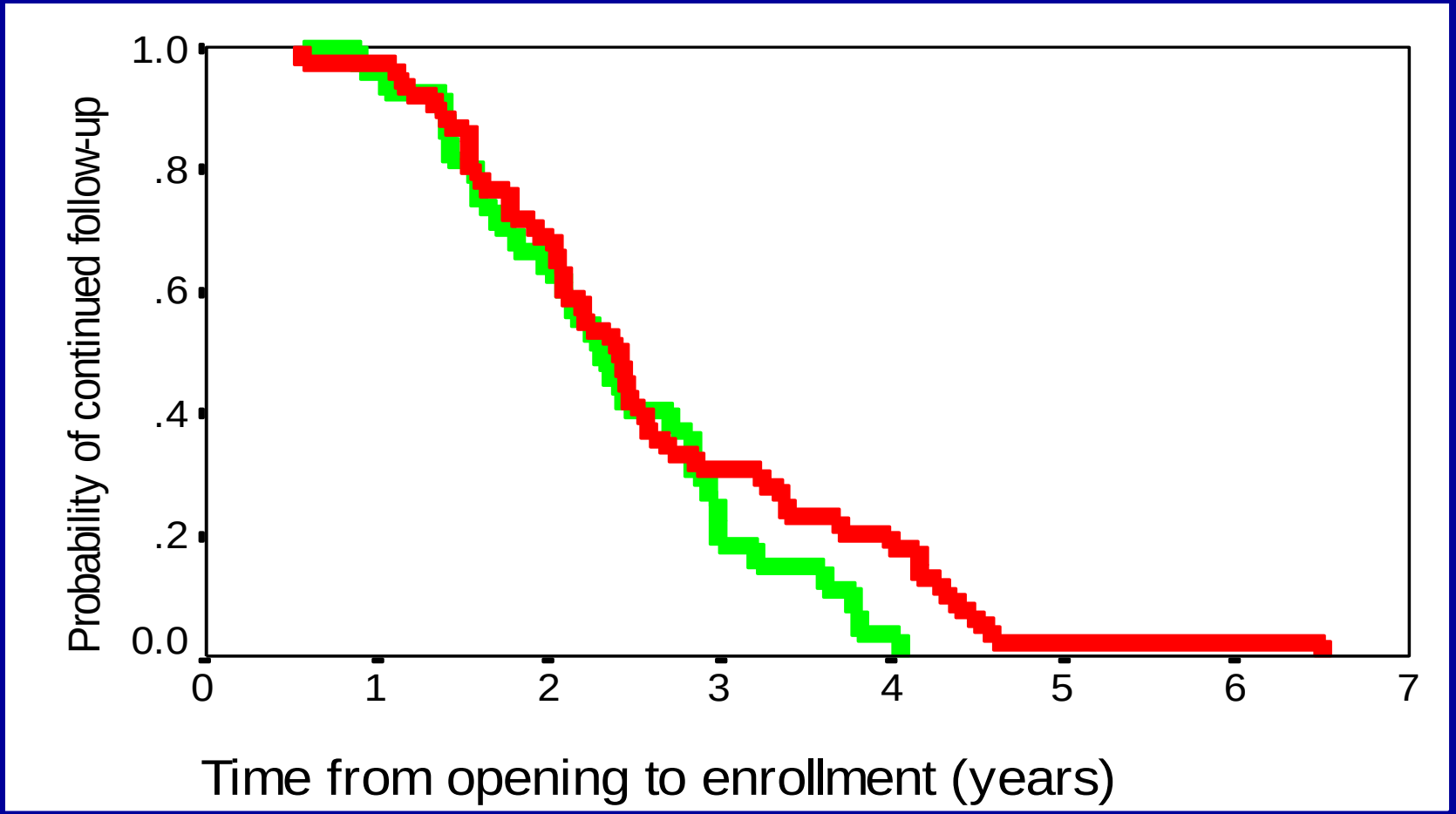

Ioannidis PLoS Clinical Trials 2006 and Clinical Trials 2007

# Time lag: bad news take longer to appear

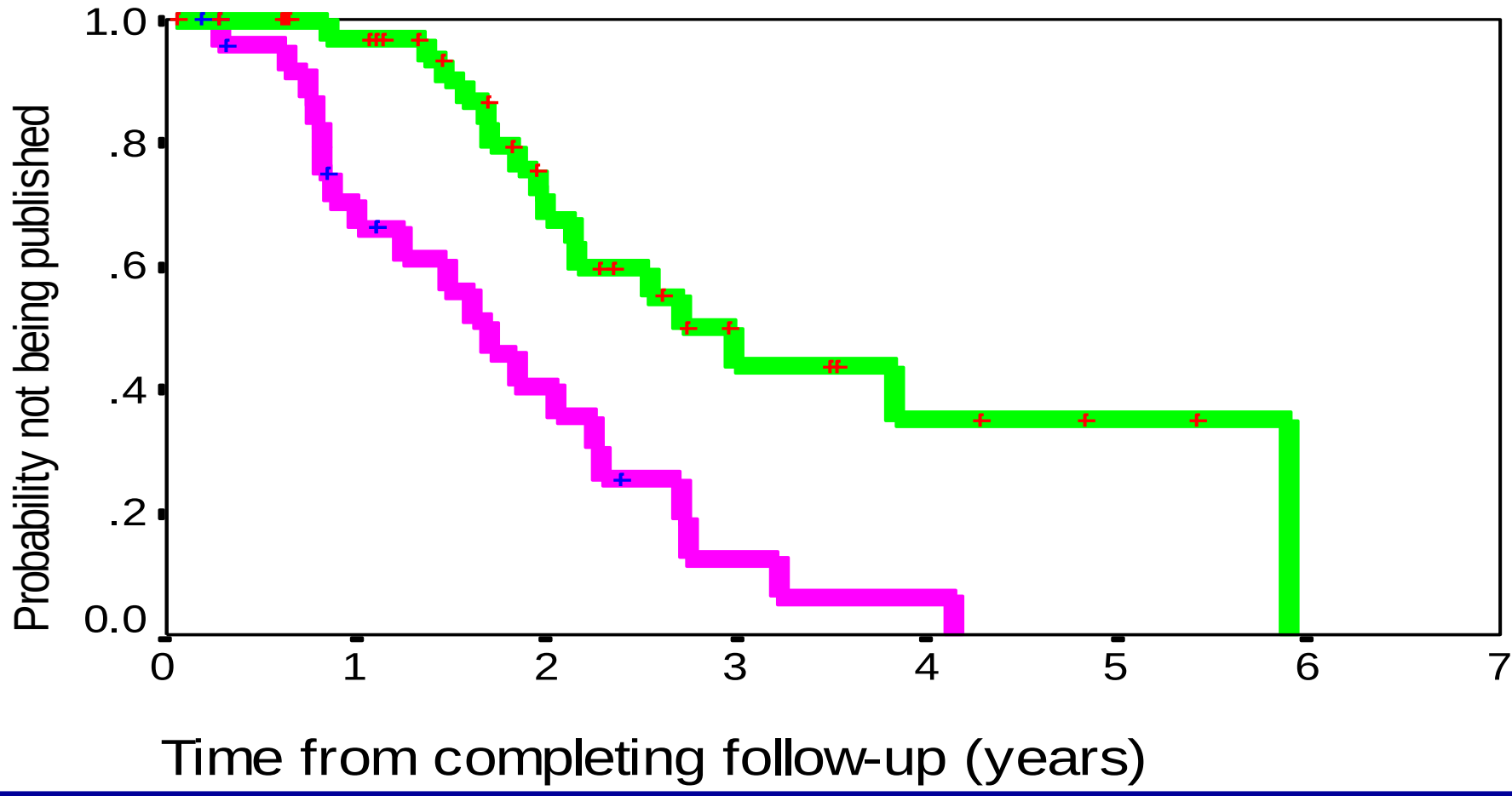



Ioannidis JP. JAMA 1998

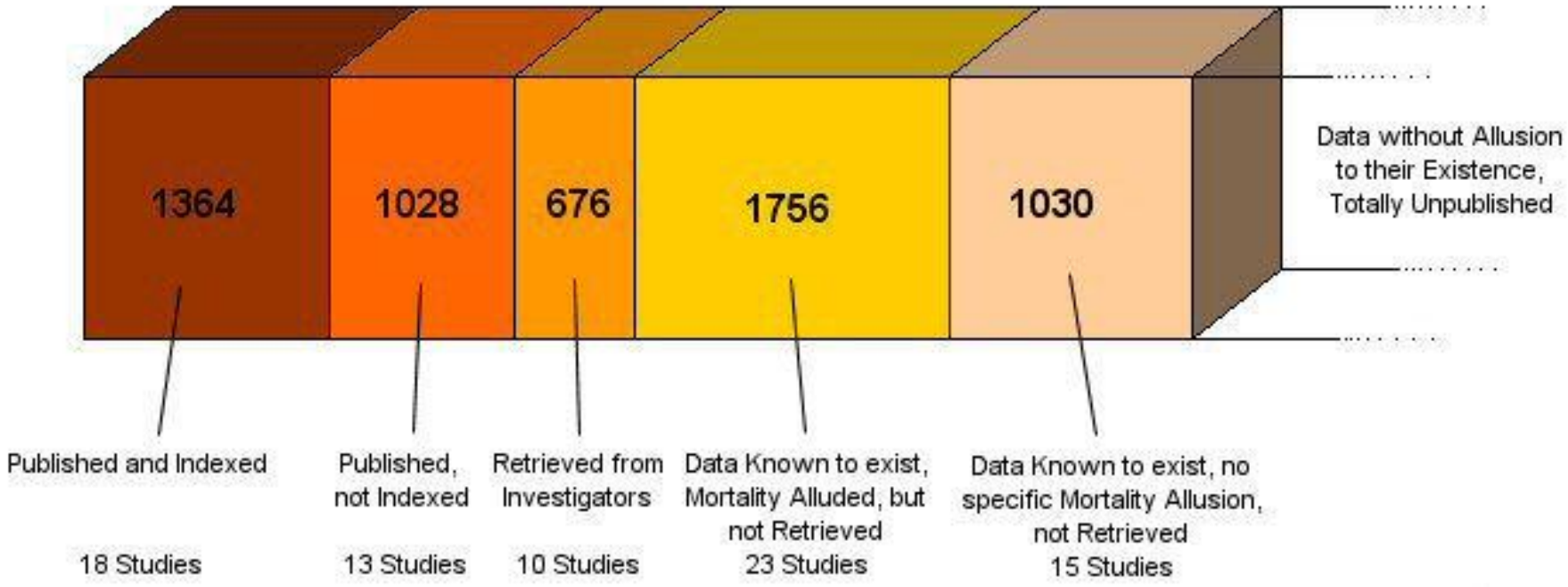
# … even though they are obtained as fast..



# …but publication is delayed

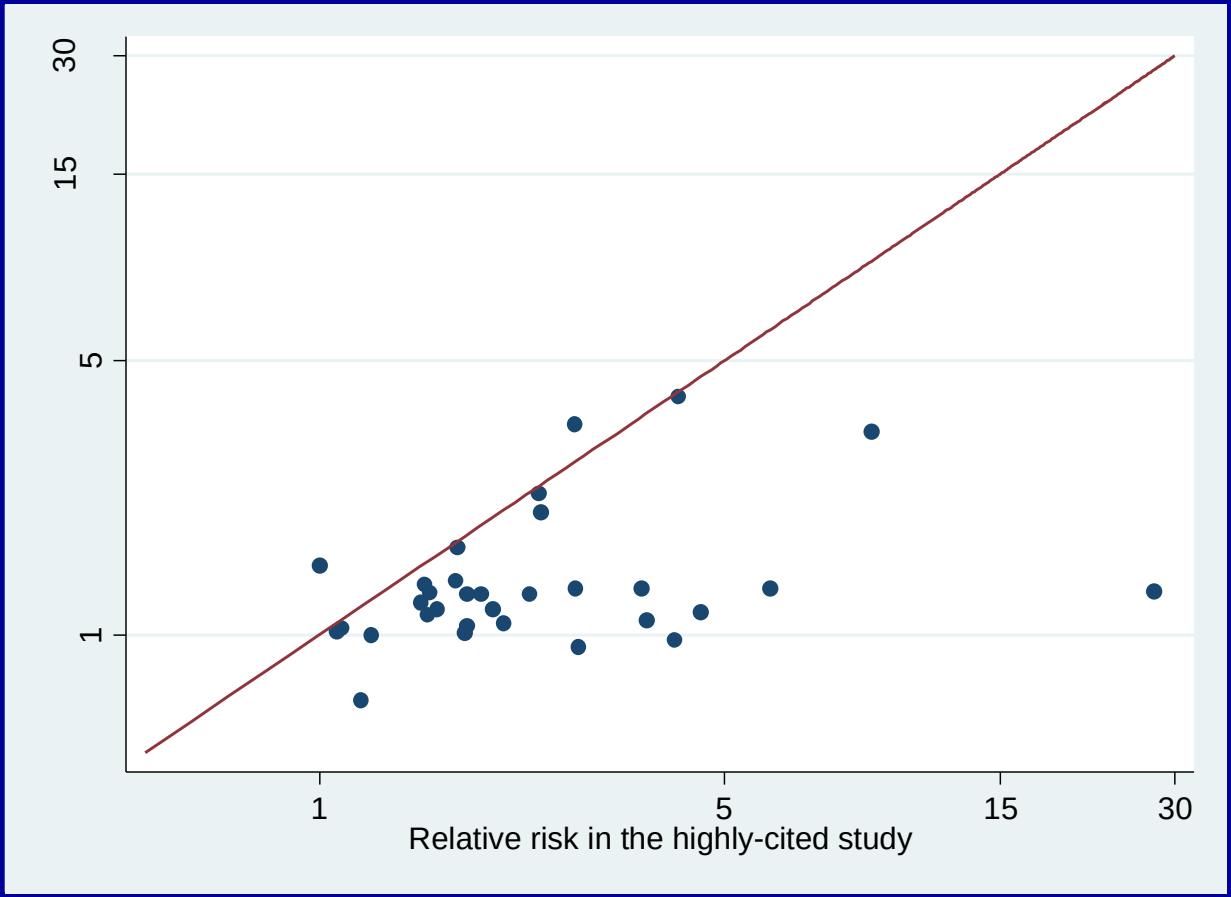


# Readily available, available, hidden, and very well hidden data




Kyzas, Loizou, Ioannidis. JNCI 2005

# Effect sizes for the top-cited biomarkers in the biomedical literature




Ioannidis and Panagiotou, JAMA 2011

# Minimal and null effects for the most popular cardiovascular biomarkers



**Table 1. Highly Popular Blood Biomarkers for Cardiovascular Disease**

| Biomarker | No. of Items in PubMed | No. With Cardiovascular Focus | Percentage |
|---|---|---|---|
| Triglycerides | 95 058 | 35 700 | 37.6 |
| C-reactive protein | 42 862 | 16 509 | 38.5 |
| Fibrinogen | 48 853 | 13 095 | 26.8 |
| Interleukin 6 | 67 690 | 10 151 | 15 |
| B-type natriuretic peptide | 10 507 | 9 409 | 89.5 |
| Serum albumin | 83 278 | 9 315 | 11.2 |
| Myeloperoxidase | 88 039 | 8 598 | 9.8 |
| ICAM-1 | 20 033 | 6 844 | 34.1 |
| Homocysteine | 17 325 | 6 754 | 39 |
| Uric acid | 27 398 | 6 538 | 23.4 |

B-type indicates brain-type; ICAM-1, intercellular adhesion molecule-1.

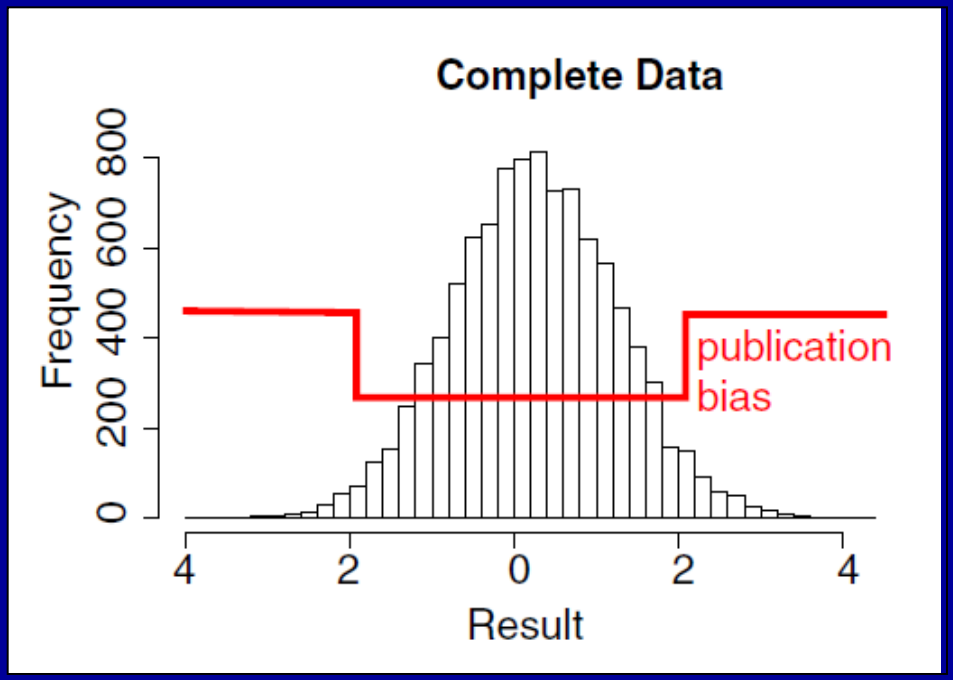
**Table 2. Predictive Effects of Popular Blood Biomarkers for Coronary Heart Disease in General Populations**

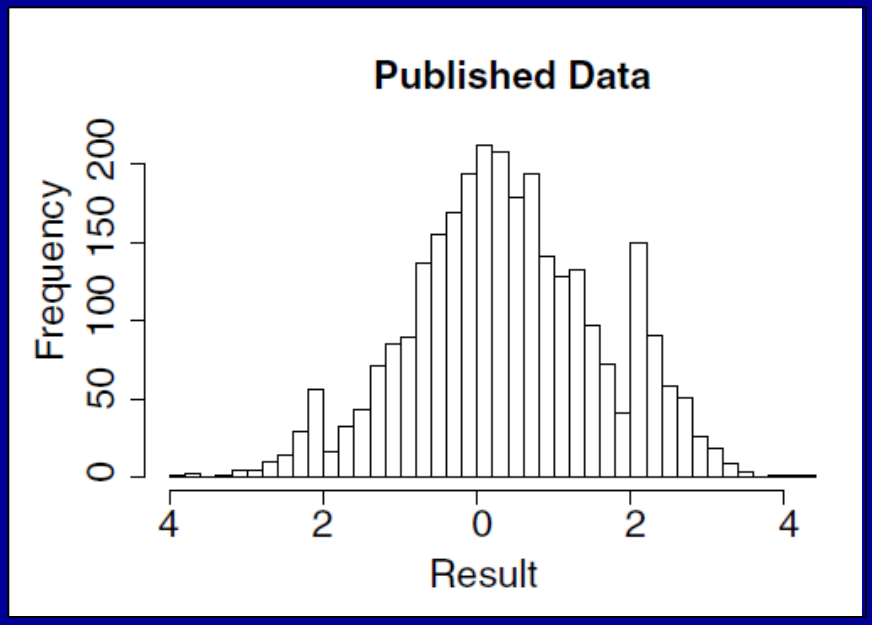
| Biomarker (Reference) | Unadjusted Relative Risk (95% CI)* | Adjusted Relative Risk (95% CI)† |
|---|---|---|
| Triglycerides[44] | | 0.99 (0.94–1.05) |
| CRP[45] | | 1.39 (1.32–1.47) |
| Fibrinogen[46] | 1.74 (1.66–1.82) | 1.45 (1.34–1.57) |
| Interleukin-6[47] | | 1.27 (1.19–1.35) |
| BNP or NT-proBNP[48] | | 1.42 (1.24–1.63) |
| Serum albumin[49] | | 1.2 (1.1–1.3) |
| ICAM-1[50] | 1.68 (1.32–2.14) | 1.11 (0.75–1.64) |
| Homocysteine[51] | | 1.05 (1.03–1.07) |
| Uric acid[12] | 1.34 (1.19, 1.49) | 1.09 (1.03, 1.16) |

CI indicates confidence interval; BNP, brain-type natriuretic peptide; NT-proBNP, N-terminal prohormone of BNP; and ICAM-1, intercellular adhesion molecule 1.

Meta-analyses of individual level data were available for triglycerides, CRP, and fibrinogen, and meta-analyses of published group-level data were available for all other markers among the most popular ones, except for myeloperoxidase (not shown in the Table). All relative risks are expressed

# Modeling the publication selection process




Pfeiffer, Bertram, Ioannidis, PLoS ONE 2011

# Missing negative studies or vibration of effects? Barely scratching the significance threshold



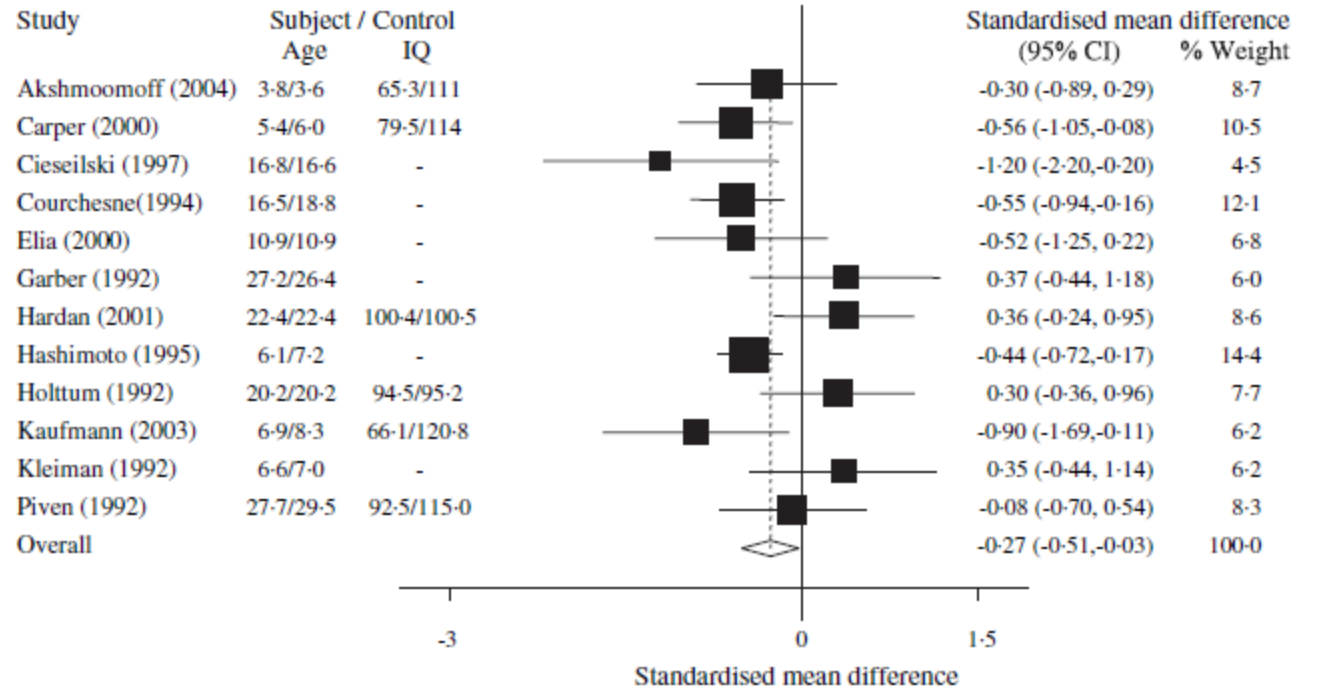
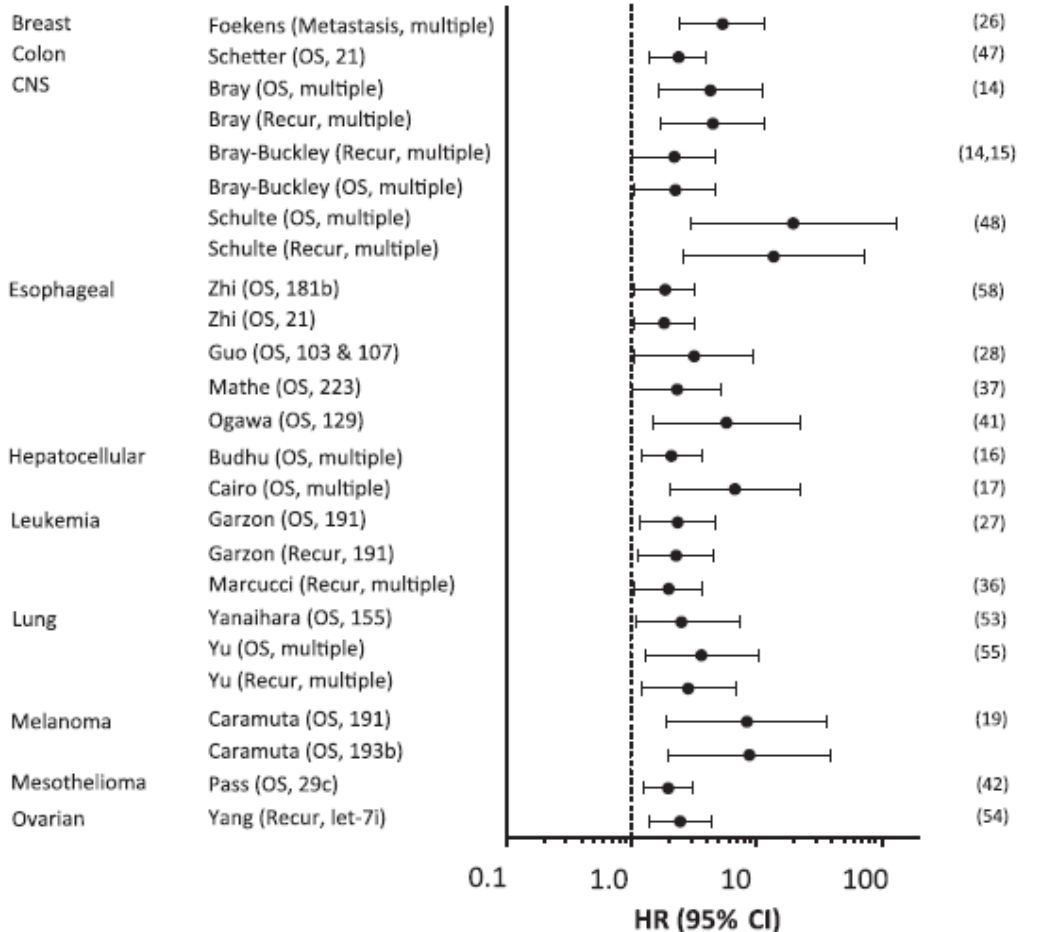



Fig. 2. Forest plot for vermal lobules VI–VII, showing the mean/median age and IQ of the subject and control groups.

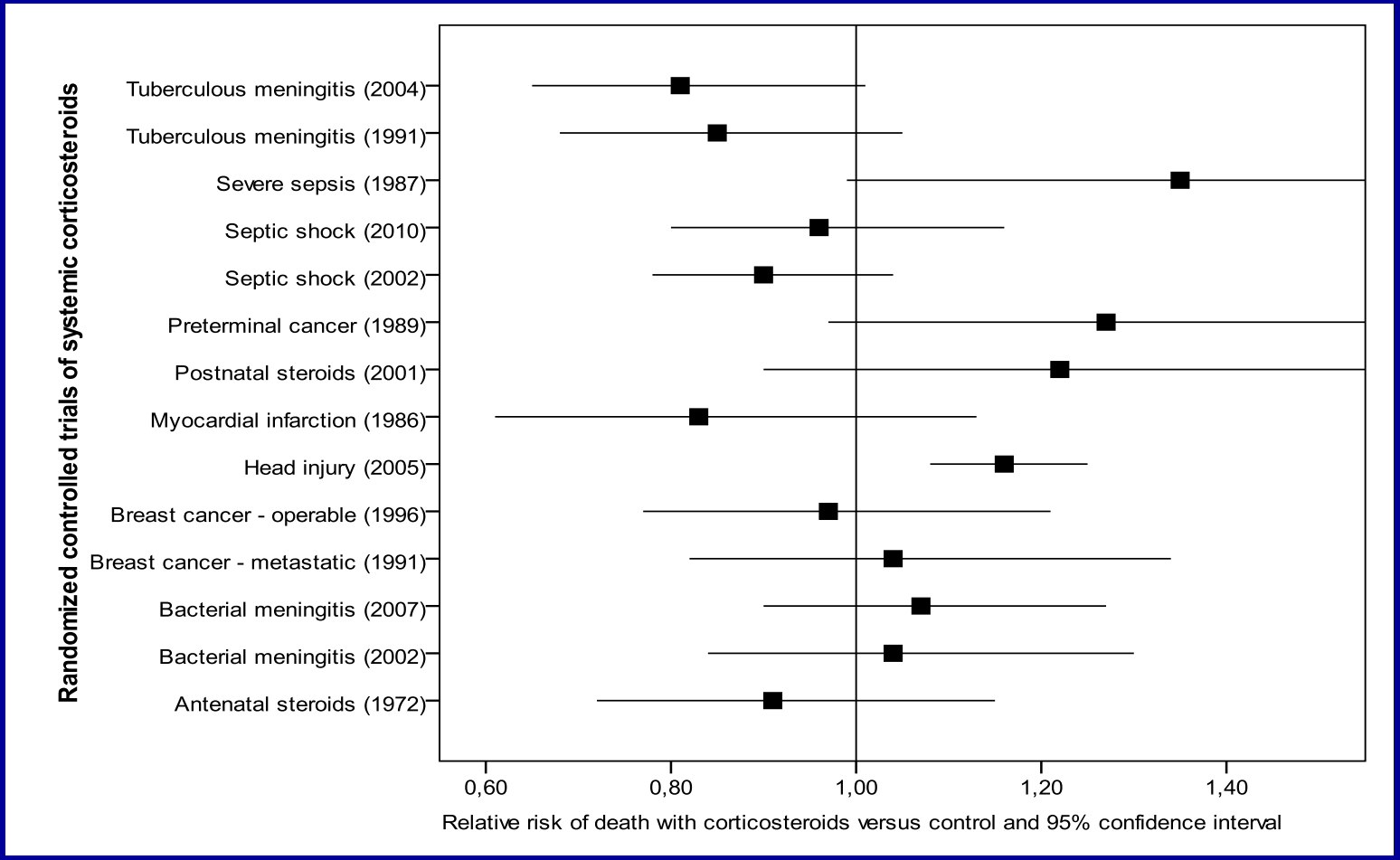
# Clinical Outcome Prediction by MicroRNAs in Human Cancer: A Systematic Review

Viswam S. Nair, Lauren S. Maeda, John P.A. Ioannidis

Manuscript received June 7, 2011; revised January 1, 2012; accepted January 10, 2012.



# Nine of the 14 largest RCTs on steroids claim significant mortality benefits




Contopoulos-Ioannidis and Ioannidis EJCI 2011

**Table.** Selected meta-analyses of brain volume abnormalities

| Topic and brain structure | Study Datasets | Cases/ Controls | Effect size (95% CI) |
|---|---|---|---|
| *Major depressive disorder* | | | |
| Anterior cingulate cortex | 8 | 181/170 | -0.77 (-1.32, -0.22) |
| Orbitofrontal cortex | 7 | 373/204 | -0.43 (-0.78, -0.09) |
| Prefrontal cortex | 7 | 242/181 | -0.34 (-0.52, -0.16) |
| Hippocampus | 31 | 1114/991 | -0.41 (-0.24, -0.28) |
| Putamen | 6 | 192/184 | -0.48 (-0.80, -0.16) |
| Caudate nucleus | 10 | 467/316 | -0.31 (-0.58, -0.04) |
| Left amygdala | 20 | 409/482 | 0.04 (-0.21, 0.28) |
| Right amygdala | 20 | 409/482 | -0.08 (-0.37, 0.21) |
| *Bipolar disorder* | | | |
| Lateral ventricles | 17 | 375/589 | 0.39 (0.24, 0.55) |
| Third ventricle | 12 | 208/271 | 0.27 (0.00, 0.53) |
| Gray matter | 14 | 257/310 | -0.18 (-0.50, 0.13) |
| White matter | 14 | 221/284 | -0.09 (-0.32, 0.15) |
| Left caudate nucleus | 11 | 273/273 | -0.03 (-0.21, 0.15) |
| Right caudate nucleus | 11 | 273/273 | -0.07 (-0.24, 0.10) |
| Left putamen | 7 | 197/183 | -0.02 (-0.22, 0.18) |
| Right putamen | 7 | 197/183 | 0.00 (-0.20, 0.21) |
| Globus pallidus | 6 | 135/106 | 0.50 (0.00, 1.01) |
| Thalamus | 10 | 235/207 | -0.02 (-0.32, 0.28) |
| Left temporal lobe | 12 | 258/277 | -0.08 (-0.35, 0.20) |
| Right temporal lobe | 12 | 258/277 | -0.16 (-0.44, 0.12) |
| Left hippocampus | 18 | 380/487 | 0.10 (-0.06, 0.26) |
| Right hippocampus | 18 | 380/487 | 0.02 (-0.13, 0.17) |
| Left amygdala | 11 | 236/354 | -0.07 (-0.47, 0.33) |
| Right amygdala | 11 | 236/354 | -0.04 (-0.45, 0.37) |
| *Obsessive compulsive disorder* | | | |
| Left caudate nucleus | 8 | 159/160 | -0.10 (-0.37, 0.17) |
| Right caudate nucleus | 8 | 159/160 | -0.08 (-0.40, 0.25) |
| *Post-traumatic stress disorder* | | | |
| Right hippocampus | 15 | 250/312 | -0.28 (-0.42, -0.13) |
| Left hippocampus | 15 | 250/312 | -0.29 (-0.43, -0.14) |
| Right amygdala | 7 | 131/188 | -0.07 (-0.21, 0.07) |
| Left amygdala | 7 | 131/188 | -0.14 (-0.26, -0.00) |
| *Autism* | | | |
| Left amygdala | 6 | 109/100 | 0.15 (-0.46, 0.76) |
| Vermal lobules I-IV | 10 | 290/310 | 0.10 (-0.28, 0.49) |
| Vermal lobules VI-VII | 12 | 348/337 | -0.27 (-0.51, -0.03) |
| *First episode schizophrenia* | | | |
| Left hippocampus | 11 | 300/287 | -0.53 (-0.74, -0.33) |
| Right hippocampus | 11 | 300/287 | -0.53 (-0.76, -0.31) |
| Left lateral ventricle | 9 | 262/248 | 0.60 (0.42, 0.78) |
| Right lateral ventricle | 9 | 262/248 | 0.46 (0.28, 0.64) |
| Third ventricle | 8 | 204/209 | 0.59 (0.39, 0.79) |
| *Schizophrenia relatives* | | | |
| Hippocampus | 9 | 421/603 | -0.31 (-0.49, -0.13) |
| Gray matter | 7 | 249/285 | -0.18 (-0.33, -0.02) |
| Third ventricle | 7 | 414/418 | 0.21 (0.03, 0.40) |

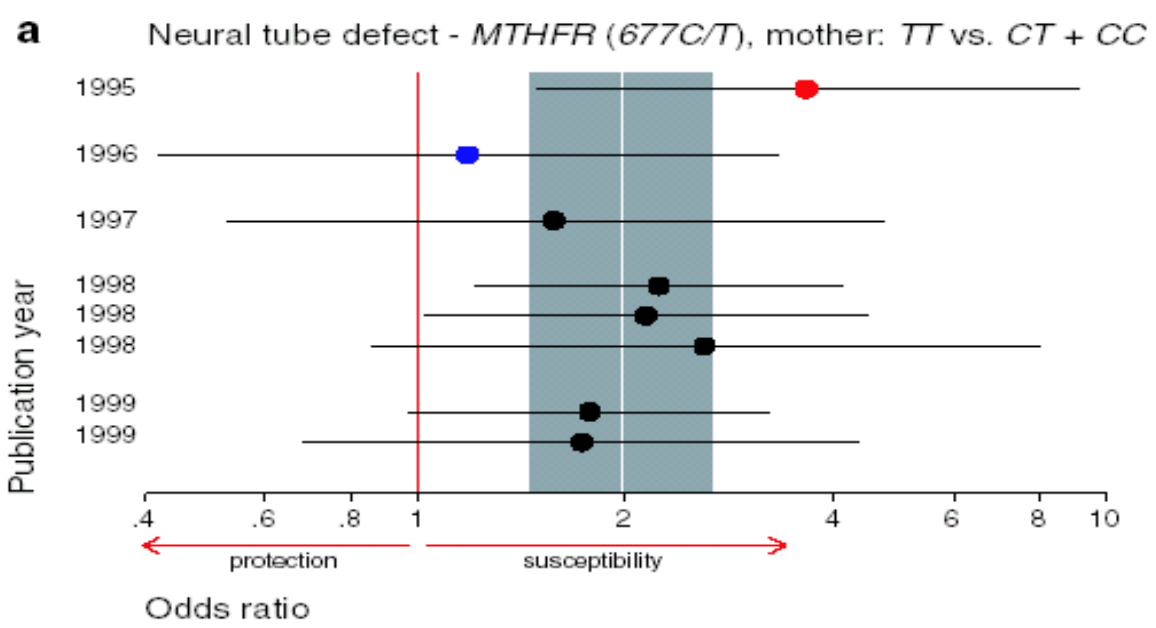
# Brain volumes (MRI, not phrenology)

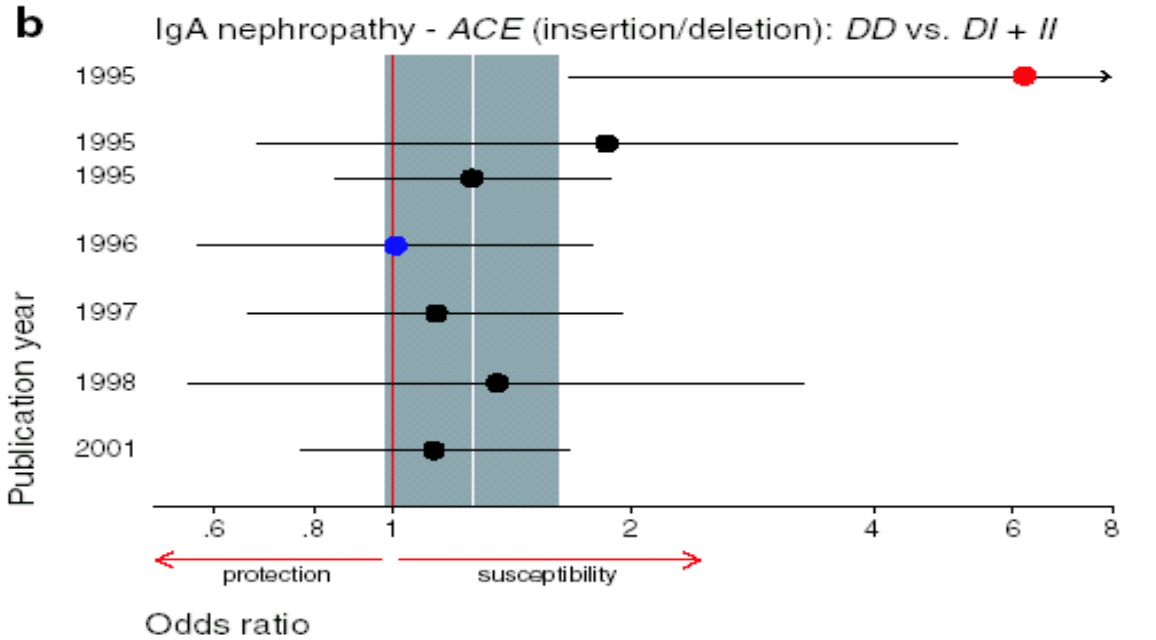
Ioannidis, Arch Gen Psych 2011

# Observed and expected studies with p<0.05

**Table.** Observed and expected number of "positive" study datasets across all meta-analyses for each condition and for each brain structure.

| | Study datasets | Observed "positive" | Expected "positive" Effect* | Half-effect* |
|---|---|---|---|---|
| *According to condition* | | | | |
| Major depressive disorder | 109 | 42 | 26.8** | 11.4** |
| Bipolar disorder | 191 | 36 | 15.6** | 11.2** |
| Obsessive compulsive disorder | 16 | 3 | 0.8** | 0.8** |
| Post-traumatic stress disorder | 44 | 17 | 4.8** | 2.8** |
| Autism | 28 | 12 | 3.1** | 1.7** |
| First episode schizophrenia | 48 | 26 | 22.3 | 7.2** |
| Schizophrenia relatives | 25 | 6 | 4.9 | 2.0** |
| *According to brain structure* | | | | |
| Anterior cingulate cortex | 8 | 4 | 4.9 | 1.7* |
| Orbitofrontal cortex | 7 | 4 | 2.9 | 1.2** |
| Prefrontal cortex | 7 | 1 | 1.7 | 0.9 |
| Hippocampus | 138 | 44 | 28.5** | 11.9** |
| Putamen | 20 | 3 | 3.4 | 1.7 |
| Caudate nucleus | 48 | 5 | 2.0** | 2.0** |
| Amygdala | 82 | 32 | 4.7** | 4.3** |
| Lateral ventricles | 35 | 15 | 14 | 3.5** |
| Third ventricle | 27 | 9 | 6.8 | 2.5** |
| Gray matter | 21 | 5 | 2.0** | 1.2** |
| White matter | 14 | 1 | 0.8 | 0.7 |
| Globus pallidus | 6 | 2 | 0.5** | 0.4** |
| Thalamus | 10 | 3 | 0.5** | 0.5** |
| Temporal lobe | 24 | 6 | 1.6** | 1.2** |
| Vermal lobules | 22 | 8 | 2.6** | 1.4** |

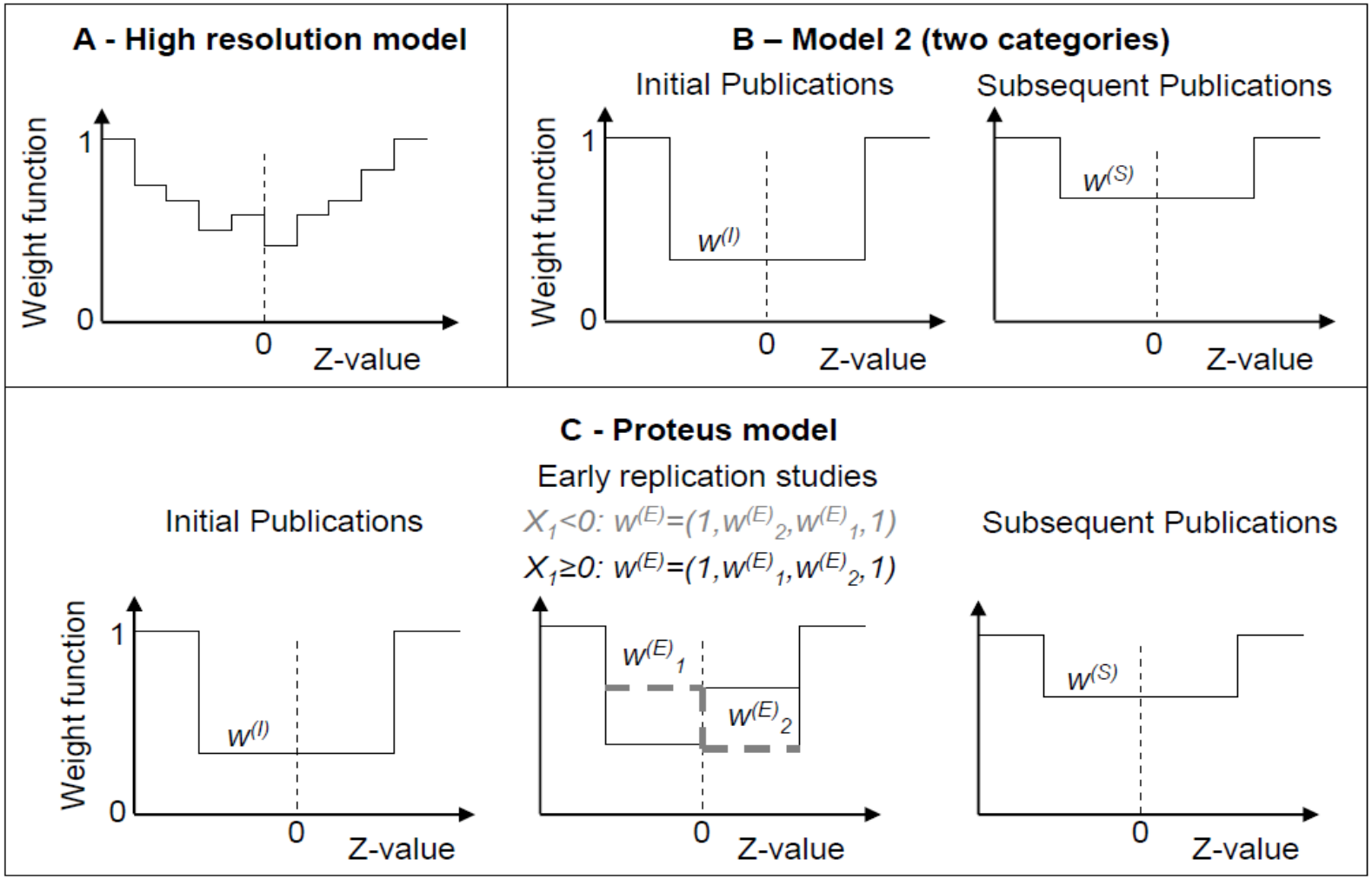



# Succession of early extremes: the Proteus phenomenon

J Clin Epidemiol 2005;58:543-8.

# Adding different selection processes for initial studies, early replications, late replications



# Selecting the selection model

| | Random-effects model | | | | |
|---|---|---|---|---|---|
| | **Unbiased** | **Model 1** | **Model 2** | **Model 3** | **Proteus** |
| $log\ w^{(I)}$ | - | - | -0.81 (0.17) | -0.82 (0.17) | -0.81 (0.17) |
| $log\ w^{(E)}{}_1$ | - | - | - | -0.33 (0.17) | -0.11 (0.17) |
| $log\ w^{(E)}{}_2$ | - | - | - | -0.24 (0.17) | -0.43 (0.17) |
| $log\ w^{(S)}$ | - | -0.33 (0.11) | -0.17 (0.12) | -0.08 (0.14) | -0.08 (0.14) |
| $\Delta_L$ | 0 | 4.4 | 10.6 | 11.4 | 13.9 |
| **Parameters** | 0 | 1 | 2 | 4 | 4 |
| $\Delta_{AIC}$ | 0 | -6.8 | -17.2 | -14.8 | -19.8 |

Pfeiffer, Bertram, Ioannidis, PLoS ONE 2011

# Any solutions?

- Learning to live with small/tiny effects
- Adjusting effects downwards
- Getting used to estimating credibility
- Large-scale collaborations
- Improving standards for reporting of research after the research is done
- Public registration and deposition of protocols, data, and analyses
- Improving validation practices
- Rewards (and penalties?) for reproducible research
- Prospective, open live streaming of research

# Learning to live with small/tiny effects

Published by Oxford University Press on behalf of the International Epidemiological Association
International Journal of Epidemiology 2011;40:1292–1307
doi:10.1093/ije/dyr099


## Risk factors and interventions with statistically significant tiny effects

George CM Siontis[1] and John PA Ioannidis[1,2]*

[1]Clinical Trials and Evidence-Based Medicine Unit and the Clinical and Molecular Epidemiology Unit, Department of Hygiene and Epidemiology, University of Ioannina School of Medicine, Ioannina, Greece and [2]Stanford Prevention Research Center, Department of Medicine, Stanford University School of Medicine, Stanford, USA

*Corresponding author. Stanford Prevention Research Center, Department of Medicine, Stanford University School of Medicine, Stanford, CA 94305, USA. E-mail: jioannid@stanford.edu

**Accepted** 19 May 2011

**Background** Large studies may identify postulated risk factors and interventions with very small effect sizes. We aimed to assess empirically a large number of statistically significant relative risks (RRs) of tiny magnitude and their interpretation by investigators.

**Methods** RRs in the range between 0.95 and 1.05 were identified in abstracts of articles of cohort studies; articles published in *NEJM*, *JAMA* or *Lancet*; and Cochrane reviews. For each eligible tiny effect and the respective study, we recorded information on study design, participants, risk factor/intervention, outcome, effect estimates, *P*-values and interpretation by study investigators. We also calculated the probability that each effect lies outside specific intervals around the null (RR interval 0.97–1.03, 0.95–1.05, 0.90–1.10).

**Results** We evaluated 51 eligible tiny effects (median sample size 112 786 for risk factors and 36 021 for interventions). Most (37/51) appeared in articles published in 2006–10. The effects pertained to nutrition ($n = 19$), genetic and other biomarkers ($n = 8$), correlates of health care ($n = 8$) and diverse other topics ($n = 16$) of clinical or public health importance and mostly referred to major clinical outcomes. A total of 15 of the 51 effects were >80% likely to lie outside the RR interval 0.97–1.03, but only 8 were >40% likely to lie outside the RR interval 0.95–1.05 and none was >1.7% likely to lie outside the RR interval 0.90–1.10. The authors discussed at least one concern for 23 effects (small magnitude $n = 19$, residual confounding $n = 11$, selection bias $n = 1$). No concerns were expressed for 28 effects.

**Conclusions** Statistically significant tiny effects for risk factors and interventions of clinical or public health importance become more common in the literature. Cautious interpretation is warranted, since most of these effects could be eliminated with even minimal biases and their importance is uncertain.

# Adjusting effects downwards

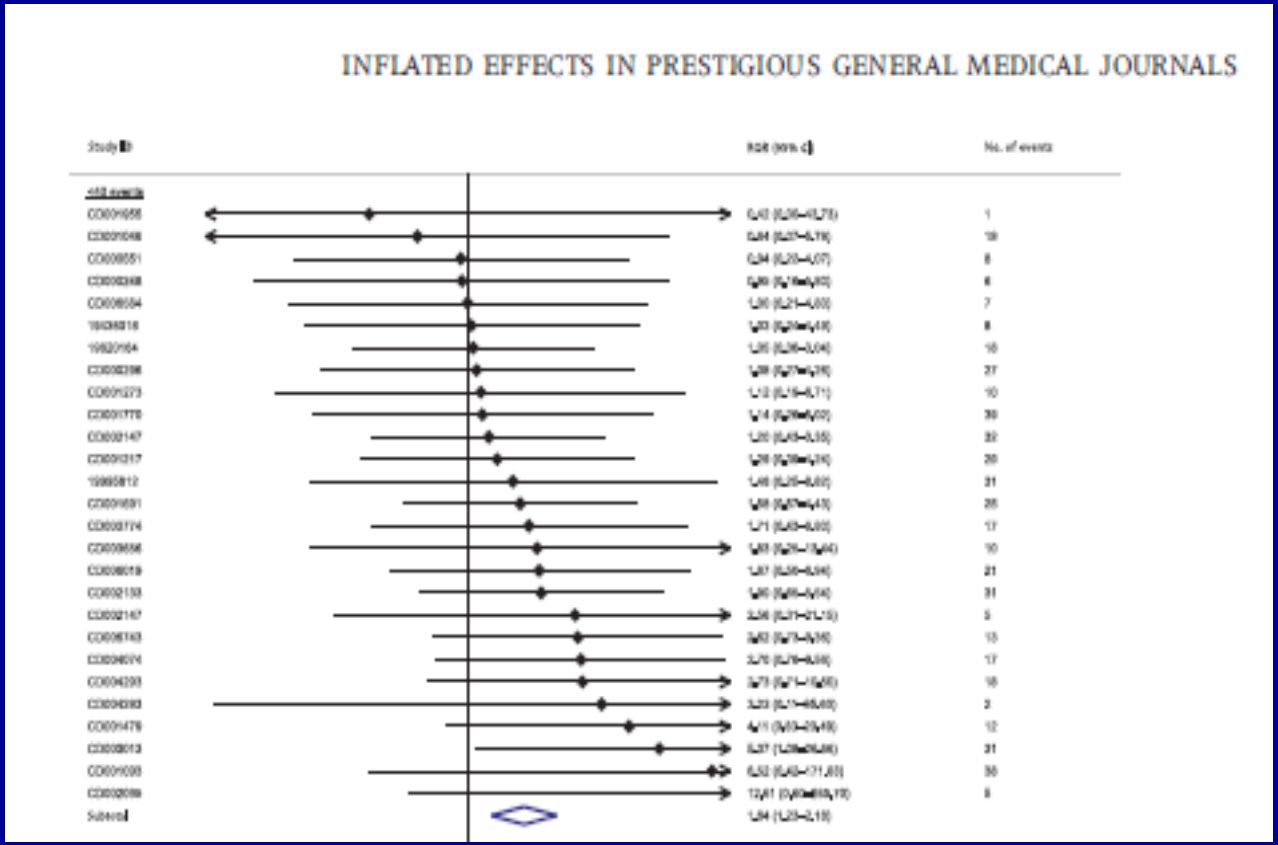
Published by Oxford University Press on behalf of the International Epidemiological Association

*International Journal of Epidemiology* 2011;**40**:1280–1291
doi:10.1093/ije/dyr095

METHODOLOGY

## Magnitude of effects in clinical trials published in high-impact general medical journals

Konstantinos CM Siontis,[1] Evangelos Evangelou[1] and John PA Ioannidis[1,2,3,4]*



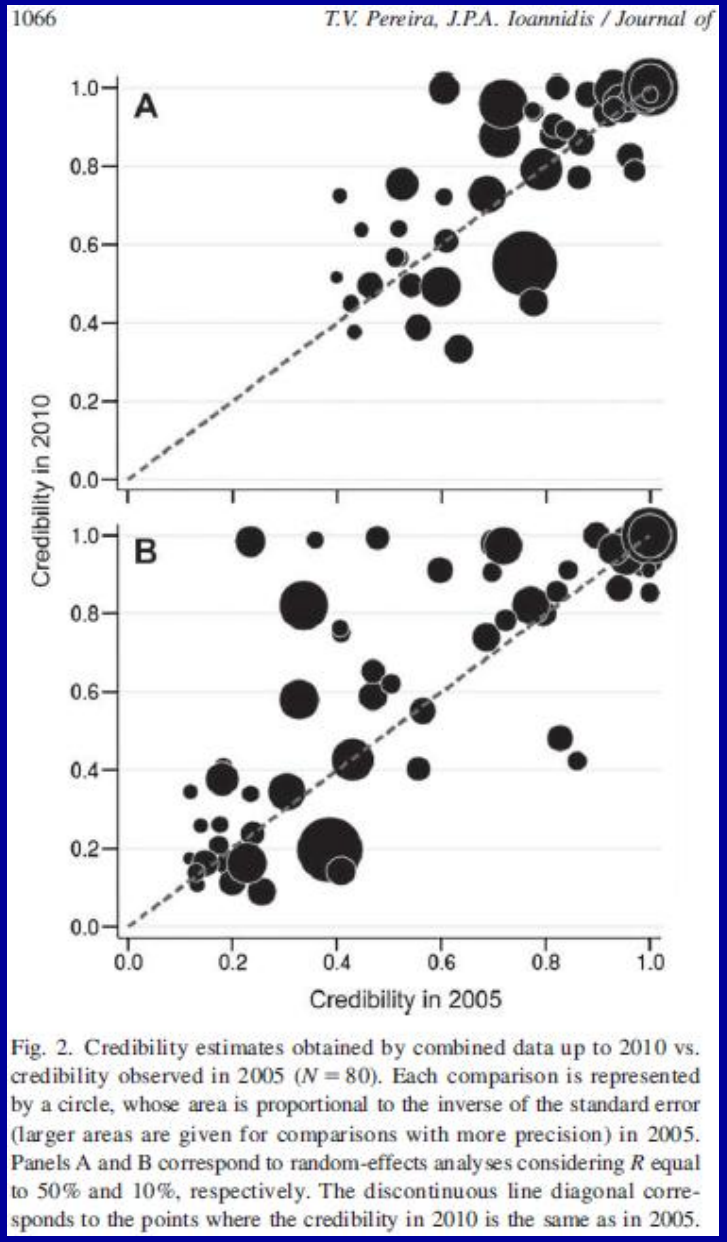
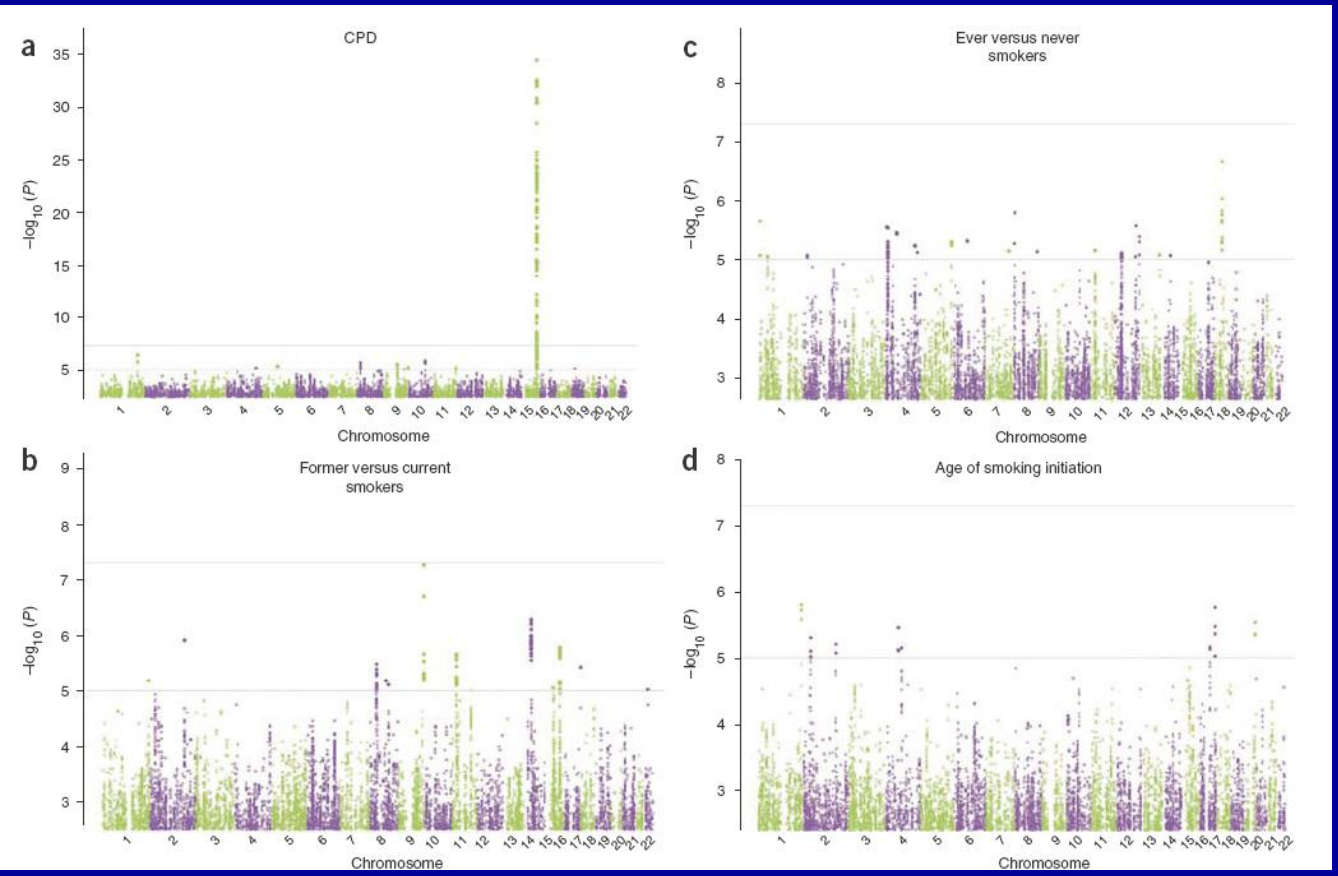



Fig. 2. Credibility estimates obtained by combined data up to 2010 vs. credibility observed in 2005 ($N = 80$). Each comparison is represented by a circle, whose area is proportional to the inverse of the standard error (larger areas are given for comparisons with more precision) in 2005. Panels A and B correspond to random-effects analyses considering $R$ equal to 50% and 10%, respectively. The discontinuous line diagonal corresponds to the points where the credibility in 2010 is the same as in 2005.

Get used to estimating credibility… even in meta-analyses of randomized trials

# Large-scale collaboration




nature genetics

Genome-wide meta-analyses identify multiple loci associated with smoking behavior

Nature Genetics, 2010

# Improving research reporting standards: EQUATOR

# Register (everything) and publish (everything)

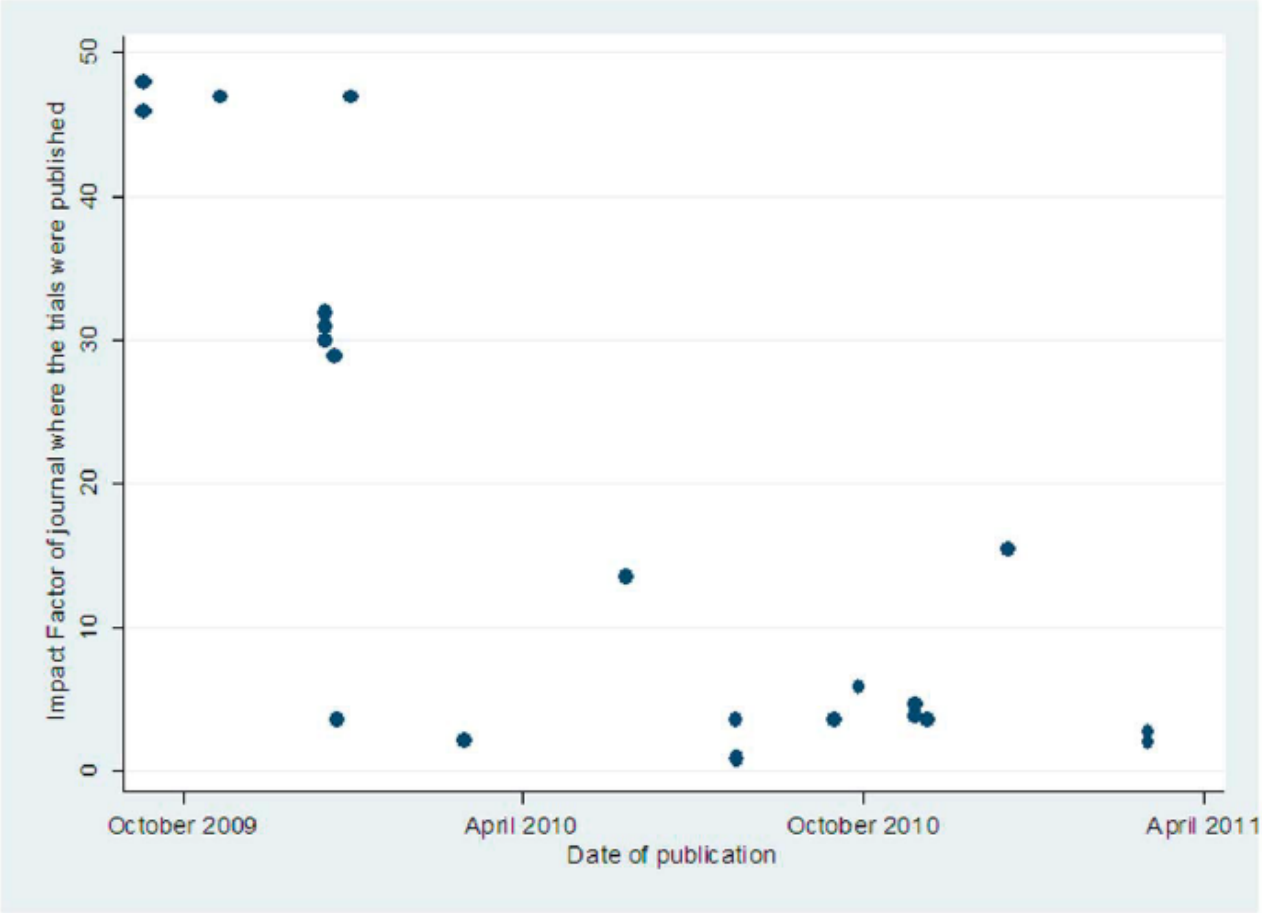



**Figure 4. Scatter plot showing the impact factor of published randomized trials by time of publication.**
doi:10.1371/journal.pone.0028346.g004

OPEN ACCESS Freely available online

PLoS one

## Publication Delay of Randomized Trials on 2009 Influenza A (H1N1) Vaccination

John P. A. Ioannidis[1], Lamberto Manzoli[2*], Corrado De Vito[3], Maddalena D'Addario[3], Paolo Villari[3]

# Public Availability of Published Research Data in High-Impact Journals

Alawi A. Alsheikh-Ali[1,2§], Waqas Qureshi[3§], Mouaz H. Al-Mallah[4,5], John P. A. Ioannidis[2,6,7,8,9*]

| Journal | Impact Factor | Policy of Required Public Deposition for Types of Data: Microarray | Nucleic Acid | Protein | Macromolecular | Policy of Provision of Materials and Methods: Materials upon request | Protocols upon request | Condition of publication | Full data deposited: Percentage of papers |
|---|---|---|---|---|---|---|---|---|---|
| New England Journal of Medicine | 52.589 | | | | | | | | 0 |
| Cell | 29.887 | | | | | | | | 1 |
| Nature | 28.751 | | | | | | | | 0 |
| Lancet | 28.638 | | | | | | | | 0 |
| Nature Medicine | 26.382 | | | | | | | | 0 |
| Science | 26.372 | | | | | | | | 1 |
| Nature Immunology | 26.218 | | | | | | | | 9 |
| Nature Genetics | 25.556 | | | | | | | | 0 |
| JAMA | 25.547 | | | | | | | | 1 |
| Nature Biotechonology | 22.848 | | | | | | | | 5 |
| Nature Materials | 19.782 | | | | | | | | 0 |
| Immunity | 19.266 | | | | | | | | 0 |
| Nature Cell Biology | 17.623 | | | | | | | | 0 |
| Journal of Clinical Investigation | 16.915 | | | | | | | | 0 |
| Archives of General Psychiatry | 15.976 | | | | | | | | 0 |
| Journal of the National Cancer Institute | 15.678 | | | | | | | | 0 |
| Nature Neuroscience | 15.664 | | | | | | | | 1 |
| Journal of Experimental Medicine | 15.612 | | | | | | | | 0 |
| Annals of Internal Medicine | 15.516 | | | | | | | | 0 |
| Journal of Clinical Oncology | 15.484 | | | | | | | | 0 |
| Nature Methods | 15.478 | | | | | | | | 6 |
| Genes and Development | 14.795 | | | | | | | | 3 |
| Nature Physics | 14.677 | | | | | | | | 0 |
| PLoS Biology | 13.501 | | | | | | | | 2 |
| Neuron | 13.41 | | | | | | | | 0 |
| Molecular Cell | 13.156 | | | | | | | | 0 |
| Circulation | 12.755 | | | | | | | | 0 |
| PLoS Medicine | 12.601 | | | | | | | | 0 |
| Developmental Cell | 12.436 | | | | | | | | 0 |
| Gastroenterology | 11.673 | | | | | | | | 0 |
| Genome Research | 11.224 | | | | | | | | 6 |
| American Journal of Human Genetics | 11.092 | | | | | | | | 3 |
| Nature Structural and Molecular Biology | 11.085 | | | | | | | | 0 |
| Journal of the American College of Cardiology | 11.054 | | | | | | | | 0 |
| Blood | 10.896 | | | | | | | | 0 |
| Hepatology | 10.734 | | | | | | | | 0 |
| Current Biology | 10.539 | | | | | | | | 0 |
| Gut | 10.015 | | | | | | | | 0 |
| British Medical Journal | 9.723 | | | | | | | | 0 |
| Circulation Research | 9.721 | | | | | | | | 1 |
| Plant Cell | 9.653 | | | | | | | | 0 |
| Nano Letters | 9.627 | | | | | | | | 0 |
| Journal of Cell Biology | 9.598 | | | | | | | | 0 |
| PNAS | 9.598 | | | | | | | | 1 |
| Molecular and Cellular Proteomics | 9.425 | | | | | | | | 7 |
| PLoS Pathogens | 9.336 | | | | | | | | 0 |
| American Journal of Psychiatry | 9.127 | | | | | | | | 0 |
| American Journal of Respiratory and Critical Care Medicine | 9.074 | | | | | | | | 0 |
| Annals of Neurology | 8.813 | | | | | | | | 0 |
| PLoS Genetics | 8.721 | | | | | | | | 0 |

# Repeatability

nature genetics

ANALYSIS

## Repeatability of published microarray gene expression analyses

John P A Ioannidis[1–3], David B Allison[4], Catherine A Ball[5], Issa Coulibaly[4], Xiangqin Cui[4], Aedín C Culhane[6,7], Mario Falchi[8,9], Cesare Furlanello[10], Laurence Game[11], Giuseppe Jurman[10], Jon Mangion[11], Tapan Mehta[4], Michael Nitzberg[5], Grier P Page[4,12], Enrico Petretto[11,13] & Vera van Noort[14]

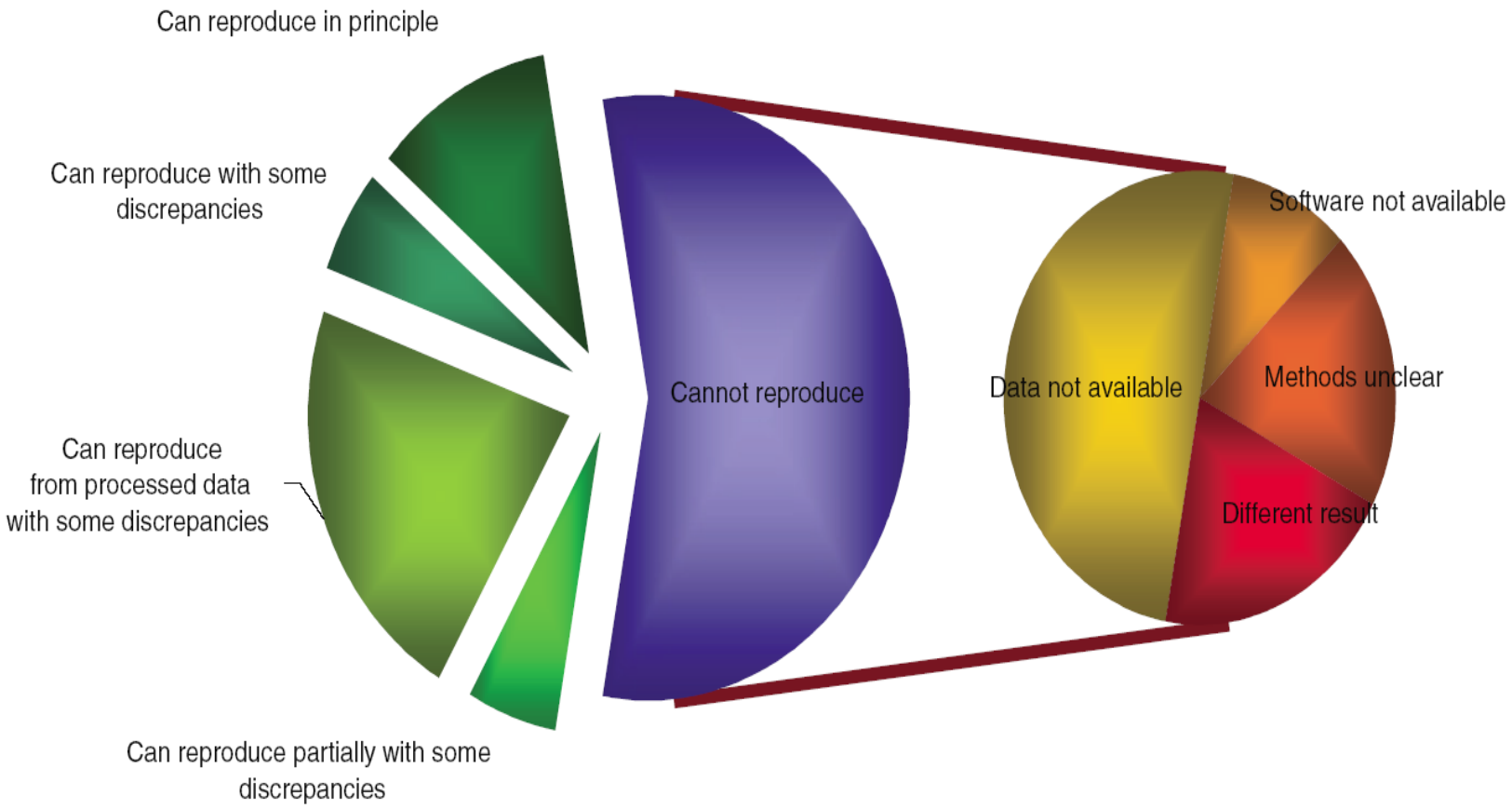



**Figure 1** Summary of the efforts to replicate the published analyses.

PERSPECTIVE

# Improving Validation Practices in "Omics" Research

John P. A. Ioannidis[1] and Muin J. Khoury[2]*

"Omics" research poses acute challenges regarding how to enhance validation practices and eventually the utility of this rich information. Several strategies may be useful, including routine replication, public data and protocol availability, funding incentives, reproducibility rewards or penalties, and targeted repeatability checks.

The exponential growth of the "omics" fields (genomics, transcriptomics, proteomics, metabolomics, and others) fuels expectations for a new era of personalized medicine. ation of the predictive value in real-practice populations, whereas clinical utility requires evaluation of the balance of benefits and harms associated with the adoption of these technologies issues (*8*, *9*), because information on available classifiers constantly changes and new classifiers are proposed. There is at least one recent unfortunate example, where gene signatures were moved into clinical trial experimentation with insufficient previous validation. Three trials of gene signatures to predict outcomes of chemotherapy in treating non–small-cell lung cancer and breast cancer were suspended in 2011 after the realization that their supporting published evidence was nonreproducible (*10*).

Many scientists now demand reproducible omics research (*11*). This requires access to the full data, protocols, and analysis codes for published studies so that other scientists can repeat analyses and verify results. Fortunately, several public data repositories exist, such as the Gene Expression Omnibus, ArrayExpress, and the Stanford Microarray Database. There have also

Science, December 2, 2011

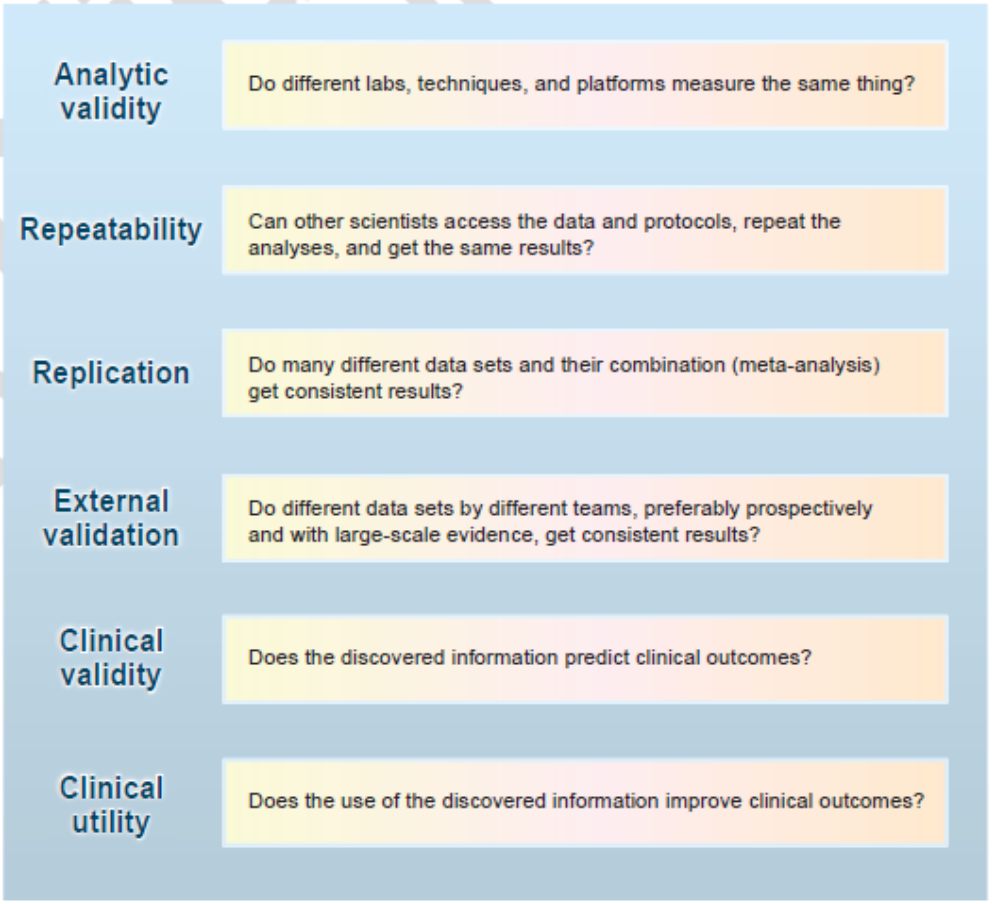



**Fig. 1.** The validation of omics research for use in medicine and public health requires fulfilling multiple steps. [Adapted from (*7*)]

# Fragmented information

- ΑΥΤΟ ΜΟΝΟΝ ΠΕΙΣΘΕΝΤΕΣ ΟΤΩ ΠΡΟΣΕΚΥΡΣΕΝ ΕΚΑΣΤΟΣ ΠΑΝΤΟΣ' ΕΛΑΥΝΟΜΕΝΟΙ, ΤΟ Δ'ΟΛΟΝ ΠΑΣ ΕΥΧΕΤΑΙ ΕΥΡΕΙΝ
- They were convinced only about whatever fragment each of them happened to hit upon, as they were overwhelmed on all sides by information, while certainly they would have wished to find the whole

# Advertisements and lost tragedies

- Current scientific papers are mostly summary advertisements of research
- Their relationship to research is of the same level as the relationship between a summary statement of the plot of Euripides and the unknown text of the tragedy per se
- Registration is also much like knowing that the tragedy existed, and bits and pieces about the plot
-

# A possible future goal

- All research data should be available in public by default
- This includes protocols, and analysis codes
- Open crowdsourcing for research may be feasible
- Consideration of live streaming of research-in-the-making
- Making the scientific record complete rather than fragmented and opportunistic

# Special thanks


- Shanthi Kappagoda, Stanford University
- Vish Nair, Stanford University
- Nazmus Saquib, Stanford University
- Juliann Saquib, Stanford University
- Despina Contopoulos-Ioannidis, Stanford University
- Jonathan Schoenfeld, Harvard University
- Thomas Pfeiffer, Harvard University
- Lars Bertram, Harvard University and Max Planck Institute
- David Chavalarias, Ecole Polytechnique, Paris
- Fainia Kavvoura, Oxford University
- Kostas Siontis, University of Ioannina
- George Siontis, University of Ioannina
- Vangelis Evangelou, University of Ioannina
- Muin Khoury, CDC and NCI
- Panagiotis Kyzas, University of Ioannina
- Orestis Panagiotou, University of Ioannina
- Jonathan Sterne, University of Bristol
- Alex Sutton, University of Leicester
- Daniele Fanelli, University of Edinburgh
- Julian Higgins, MRC Biostatistics Unit, Cambridge University
- Joseph Lau, ICRHPS, Tufts University
- Tiago Pereira, U Sao Paolo