How Widespread Is The Problem?

“Why Most Published Research Findings Are False” (2005)

→ Uploaded paper on Moodle

→ Maybe take notes for discussion

Open access, freely available online

Essay

Why Most Published Research Findings Are False

John P. A. Ioannidis

Summary

There is increasing concern that most current published research findings are false. The probability that a research claim is true may depend on study power and bias, the number of other studies on the same question, and, importantly, the ratio of true to no relationships among the relationships probed in each scientific field. In this framework, a research finding is less likely to be true when the studies conducted in a field are smaller; when effect sizes are smaller; when there is a greater number and lesser preselection of tested relationships; where there is greater flexibility in designs, definitions, outcomes, and analytical modes; when there is greater financial and other interest and prejudice; and when more teams are involved in a scientific field in chase of statistical significance. Simulations show that for most study designs and settings, it is more likely for a research claim to be false than true. Moreover, for many current scientific fields, claimed research findings may often be simply accurate measures of the prevailing bias. In this essay, I discuss the implications of these problems for the conduct and interpretation of research.

Published research findings are sometimes refuted by subsequent evidence, with ensuing confusion and disappointment. Refutation and controversy is seen across the range of research designs, from clinical trials and traditional epidemiological studies [1–3] to the most modern molecular research [4,5]. There is increasing concern that in modern research, false findings may be the majority or even the vast majority of published research claims [6–8]. However, this should not be surprising. It can be proven that most claimed research findings are false. Here I will examine the key factors that influence this problem and some corollaries thereof.

The Essay section contains opinion pieces on topics of broad interest to a general medical audience.

Modeling the Framework for False Positive Findings

Several methodologists have pointed out [9–11] that the high rate of nonreplication (lack of confirmation) of research discoveries is a consequence of the convenient, yet ill-founded strategy of claiming conclusive research findings solely on the basis of a single study assessed by formal statistical significance, typically for a p-value less than 0.05. Research is not most appropriately represented and summarized by p-values, but, unfortunately, there is a widespread notion that medical research articles should be interpreted based only on p-values. Research findings are defined here as any relationship reaching formal statistical significance, e.g., effective interventions, informative predictors, risk factors, or associations. “Negative” research is also very useful. “Negative” is actually a misnomer, and the misinterpretation is widespread. However, here we will target relationships that investigators claim exist, rather than null findings.

It can be proven that most claimed research findings are false.

As has been shown previously, the probability that a research finding is indeed true depends on the prior probability of it being true (before doing the study), the statistical power of the study, and the level of statistical significance [10,11]. Consider a 2 × 2 table in which research findings are compared against the gold standard of true relationships in a scientific field. In a research field both true and false hypotheses can be made about the presence of relationships. Let R be the ratio of the number of “true relationships” to “no relationships” among those tested in the field. R is characteristic of the field and can vary a lot depending on whether the field targets highly likely relationships or searches for only one or a few true relationships among thousands and millions of hypotheses that may be postulated. Let us also consider, for computational simplicity, circumscribed fields where either there is only one true relationship (among many that can be hypothesized) or the power is similar to find any of the several existing true relationships. The pre-study probability of a relationship being true is $R/(R + 1)$. The probability of a study finding a true relationship reflects the power $1 - \beta$ (one minus the Type II error rate). The probability of claiming a relationship when none truly exists reflects the Type I error rate, α. Assuming that c relationships are being probed in the field, the expected values of the 2 × 2 table are given in Table 1. After a research finding has been claimed based on achieving formal statistical significance, the post-study probability that it is true is the positive predictive value, PPV. The PPV is also the complementary probability of what Wacholder et al. have called the false positive report probability [10]. According to the 2 × 2 table, one gets $PPV = (1 - \beta)R/(R - \beta R + \alpha)$. A research finding is thus

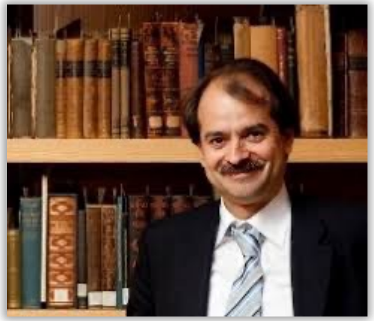
Citation: Ioannidis JPA (2005) Why most published research findings are false. PLoS Med 2(8): e124.

Abbreviation: PPV, positive predictive value

John P. A. Ioannidis is in the Department of Hygiene and Epidemiology, University of Ioannina School of Medicine, Ioannina, Greece, and Institute for Clinical Research and Health Policy Studies, Department of Medicine, Tufts-New England Medical Center, Tufts University School of Medicine, Boston, Massachusetts, United States of America. E-mail: jioannid@cc.uoi.gr

Competing Interests: The author has declared that no competing interests exist.

DOI: 10.1371/journal.pmed.0020124



John P. A. Ioannidis

How Widespread Is The Problem?

“Why Most Published Research Findings Are False” (2005)


Open access, freely available online

Essay

Why Most Published Research Findings Are False

John P. A. Ioannidis

It can be proven that most claimed research findings are false.

Analytic approach: “It can be *proven* that most claimed research findings are false.”

Ingredients:

- **Ratio of true to no true relationships**: $R = \dfrac{\text{True Effect } (E)}{\text{No Effect } (\neg E)}$
  - generally unknown and specific to a research field
  - $R > 1$: well-understood fields, but without much to learn
  - $R \leq 1$: more exploratory, speculative research fields (“speculativeness” increases with lower R)
  - Fraction of true relationships: $\frac{E}{\neg E + E} = \frac{R}{R+1}$
- $\boldsymbol{\alpha}$ **(type I error) rate**, the allowed probability of making a false positive; typically, $\alpha = 0.05$
- $\boldsymbol{\beta}$ **(type II error) rate**, the allowed probability of making a false negative; typically, $\beta = 0.20$
- **Power: $\mathbf{1-\boldsymbol{\beta}}$**, the probability of identifying an effect, *if* it exists
- $\boldsymbol{c}$: Number of studies in the literature

How Widespread Is The Problem?

“Why Most Published Research Findings Are False” (2005)

Research Finding	True Relationship		
	Yes	No	Total
Yes	$c(1-\beta)R/(R+1)$	$c\alpha/(R+1)$	$c(R+\alpha-\beta R)/(R+1)$
No	$c\beta R/(R+1)$	$c(1-\alpha)/(R+1)$	$c(1-\alpha+\beta R)/(R+1)$
Total	$cR/(R+1)$	$c/(R+1)$	c

How Widespread Is The Problem?

“Why Most Published Research Findings Are False” (2005)

True Positives

False Positives

All Positive Findings

Research Finding	True Relationship Yes	No	Total
Yes	$\boxed{c(1-\beta)R/(R+1)}$	$\boxed{c\alpha/(R+1)}$	$\boxed{c(R+\alpha-\beta R)/(R+1)}$
No	$c\beta R/(R+1)$	$c(1-\alpha)/(R+1)$	$c(1-\alpha+\beta R)/(R+1)$
Total	$cR/(R+1)$	$c/(R+1)$	c

How Widespread Is The Problem?

"Why Most Published Research Findings Are False" (2005)

True Positives | False Positives | All Positive Findings

Research Finding	True Relationship Yes	No	Total
Yes	$\boxed{c(1-\beta)R/(R+1)}$	$\boxed{c\alpha/(R+1)}$	$\boxed{c(R+\alpha-\beta R)/(R+1)}$
No	$c\beta R/(R+1)$	$c(1-\alpha)/(R+1)$	$c(1-\alpha+\beta R)/(R+1)$
Total	$cR/(R+1)$	$c/(R+1)$	c

Positive Predictive Value (PPV) $= P(E|\text{Study claims } E) = \dfrac{\text{True Positives}}{\text{All Positive Findings}}$

How Widespread Is The Problem?

"Why Most Published Research Findings Are False" (2005)

Research Finding	True Relationship: Yes	No	Total
Yes	$\boxed{c(1-\beta)R/(R+1)}$	$\boxed{c\alpha/(R+1)}$	$\boxed{c(R+\alpha-\beta R)/(R+1)}$
No	$c\beta R/(R+1)$	$c(1-\alpha)/(R+1)$	$c(1-\alpha+\beta R)/(R+1)$
Total	$cR/(R+1)$	$c/(R+1)$	c

$$\text{PPV} = \frac{c(1-\beta)R/(R+1)}{c(R+\alpha-\beta R)/(R+1)} = \frac{(1-\beta)R}{(1-\beta)R+\alpha}$$

How Widespread Is The Problem?

"Why Most Published Research Findings Are False" (2005)

Research Finding	True Relationship		
	Yes	No	Total
Yes	$c(1-\beta)R/(R+1)$	$c\alpha/(R+1)$	$c(R+\alpha-\beta R)/(R+1)$
No	$c\beta R/(R+1)$	$c(1-\alpha)/(R+1)$	$c(1-\alpha+\beta R)/(R+1)$
Total	$cR/(R+1)$	$c/(R+1)$	c

$$\text{PPV} = \frac{c(1-\beta)R/(R+1)}{c(R+\alpha-\beta R)/(R+1)} = \frac{(1-\beta)R}{(1-\beta)R+\alpha}$$

Example 1: Assuming an average power of 23% in psychotherapy trials (de Vries, 2022) and that 5 out of 10 new treatments are effective:

$$\text{PPV} = \frac{(1-\beta)R}{(1-\beta)R+\alpha} = \frac{0.23 \times \frac{5}{10-5}}{0.23 \times \frac{5}{10-5} + 0.05} = \mathbf{82\%}$$

How Widespread Is The Problem?

“Why Most Published Research Findings Are False” (2005)

Research Finding	True Relationship Yes	No	Total
Yes	$c(1-\beta)R/(R+1)$	$c\alpha/(R+1)$	$c(R+\alpha-\beta R)/(R+1)$
No	$c\beta R/(R+1)$	$c(1-\alpha)/(R+1)$	$c(1-\alpha+\beta R)/(R+1)$
Total	$cR/(R+1)$	$c/(R+1)$	c

$$\text{PPV} = \frac{c(1-\beta)R/(R+1)}{c(R+\alpha-\beta R)/(R+1)} = \frac{(1-\beta)R}{(1-\beta)R+\alpha}$$

Example 2: Assuming an average power of 23% in psychotherapy trials (de Vries, 2022) and that 5 out of **<u>100 new treatments</u>** are effective:

$$\text{PPV} = \frac{(1-\beta)R}{(1-\beta)R+\alpha} = \frac{0.23 \times \frac{5}{100-5}}{0.23 \times \frac{5}{100-5} + 0.05} = \mathbf{19\%}$$

How Widespread Is The Problem?

"Why Most Published Research Findings Are False" (2005)

Bias u: "factors that tend to produce research findings when they should not be produced."

- Can be caused by many things: file-drawer problem, questionable research practices, …
- Changes the number of findings vs. non-findings
- Changes the **ratio of false positives and true positives:**

Research Finding	True Relationship Yes	No	Total
Yes	$\boxed{(c[1-\beta]R + uc\beta R)/(R+1)}$	$\boxed{c\alpha + uc(1-\alpha)/(R+1)}$	$c(R+\alpha-\beta R+u-u\alpha+u\beta R)/(R+1)$
No	$(1-u)c\beta R/(R+1)$	$(1-u)c(1-\alpha)/(R+1)$	$c(1-u)(1-\alpha+\beta R)/(R+1)$
Total	$cR/(R+1)$	$c/(R+1)$	c

How Widespread Is The Problem?

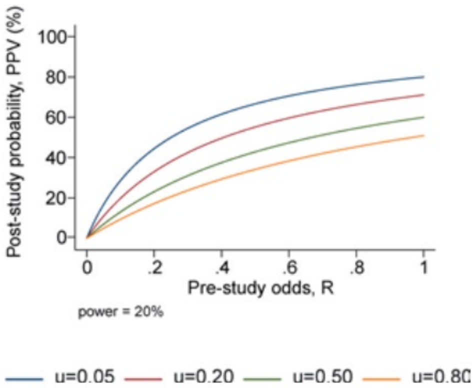
"Why Most Published Research Findings Are False" (2005)

Research Finding	True Relationship Yes	No	Total
Yes	$(c[1-\beta]R + uc\beta R)/(R+1)$	$c\alpha + uc(1-\alpha)/(R+1)$	$c(R+\alpha-\beta R+u-u\alpha+u\beta R)/(R+1)$
No	$(1-u)c\beta R/(R+1)$	$(1-u)c(1-\alpha)/(R+1)$	$c(1-u)(1-\alpha+\beta R)/(R+1)$
Total	$cR/(R+1)$	$c/(R+1)$	c

$$\text{PPV} = \frac{(1-\beta)R + u\beta R}{(1-\beta)R + \alpha + u(\beta R + 1 - \alpha)}$$

→ The stronger the bias u, the lower the PPV!

C



DOI: 10.1371/journal.pmed.0020124.g001

Figure 1. PPV (Probability That a Research Finding Is True) as a Function of the Pre-Study Odds for Various Levels of Bias, *u*

Panels correspond to power of 0.20, 0.50, and 0.80.

How Widespread Is The Problem?

“Why Most Published Research Findings Are False” (2005)

$1-\beta$	*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

Meta-Analysis

Almost certain that the research finding is false!

How Widespread Is The Problem?

“Why Most Published Research Findings Are False” (2005)

What parameters can we target to increase the PPV?

1. **True effect ratio $\boldsymbol{R}$:** Restrict exploratory analysis, massive uncorroborated testing
2. **Power $\mathbf{1}-\boldsymbol{\beta}$:** Increase sample sizes
3. **Bias $\boldsymbol{u}$:** “Researcher degrees of freedom”, systemic issues in research, financial interests

→ All things equal, meta-analysis is a great way to increase (2), and thus the PPV.

→ However, these attempts may become worthless in the presence of strong bias u

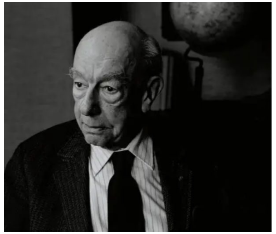
→ As meta-analysts, we cannot eradicate, but can try to **“adjust”** for the presence of u. This is what is typically attempted in **publication bias analyses**.

How Widespread Is The Problem?

“Why Most Published Research Findings Are False” (2005)

Limitations of Ioannidis’ Paper:

- Only a thought experiment: Ioannidis brings no data to prove his model
- Not a mathematical “proof”, although Ioannidis frames it this way (Goodman & Greenland, 2007)
- Based on an atomistic idea of truth: treats scientific fields as a collection of single independent hypotheses that are either strictly true or strictly false
    - Many philosophers of science would disagree with such a simple view (cf. the “Quine-Duhem thesis”, which states that all hypotheses are based on background assumptions and theories that can only be tested as a whole)



W. V. O. Quine