Unreliable research

Trouble at the lab

Reproducible Research

Scientists like to think of science as self-correcting. To an alarming degree, it is not

Oct 19th 2013 | From the print edition

The Economist

Prof. Naomi Altman
Dept. of Statistics
dePamphilis Lab
2014

The Atlantic

BRAVE THINKERS NOVEMBER 2010

Lies, Damned Lies, and Medical Science

Much of what medical researchers conclude in their studies is misleading, exaggerated, or

Altman's rule:

A result is scientifically interesting only if it reproducible.

Altman's rule:

A result is scientifically interesting only if it reproducible.

Altman's proviso:

Not everything that is reproducible is scientifically interesting.

Altman's rule:

A result is scientifically interesting only if it reproducible.

Altman's proviso:

Not everything that is reproducible is scientifically interesting.

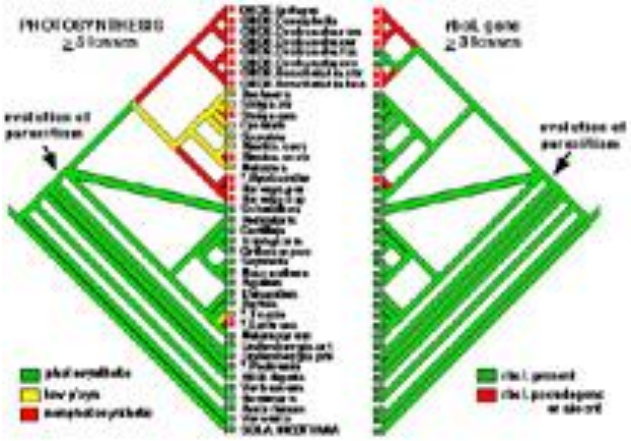
Mistakes and biases are often very reproducible (at least by the same lab).

2 types of reproducibility

Reproducibility in the lab

Paulson lab

Reproducibility of the bioinformatics analysis

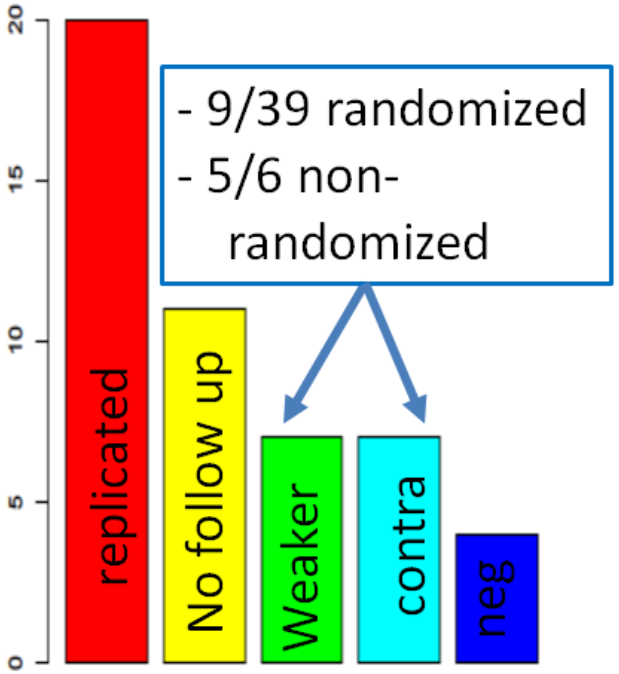


From dePamphilis lab

Altman, 9/14

Reproducibility in the Lab

- Ioannidis 2005 *JAMA*: "Contradicted and initially stronger effects in highly cited clinical research."
- Ioannidis 2005 *PLoS Medicine* "Why Most Published Research Findings are False"



Altman, 9/14

Reproducibility in the Lab

Ioannidis' arguments:

- Testing for rare events – most discoveries are false
- Publication bias: Only "interesting" events are published.
- Detection bias: Selective or distorted reporting, conflicts of interest, deliberate manipulation
- Lack of independent replication
- Selection of most significant events instead of proper meta-analysis when there is replication

Reproducibility in the Lab

High false positive rates should be expected when:

- Low power
- Small studies: if the false positive rate is controlled, the false negative rate is high
- Small effect sizes:
- Large number of relationships tested without preliminary findings: lower prior probability of effect
- High flexibility in designs, definitions, outcomes and analyzes: the search for significance
- **Rewards of research: winner takes all, so it pays to be first and to find something**
- Hot areas of research: research is rushed, lots of studies with low prior probability of effect

Altman, 9/14

Reproducibility in the Lab

Obvious (I think):

1. Follow lab procedures.
2. Document everything including failures.
3. Keep a good lab book.

Reproducibility in the Lab

Reproducible ≠ Correct

We can ALWAYS get reproducible results if we:

- Do biased experiments
- Always accept (reject) our hypothesis

Correct is more important!!

Reproducibility in the Lab

Statistical testing:

We do a test (e.g. t-test) and reject if $P<\alpha$

(Reject means that we declare a "discovery")

Usually we pick $\alpha=0.05$ or 0.01

What is the probability that we have **actually** discovered something if we reject?

Reproducibility in the Lab

Statistical testing:

Prob(correct|P<α)=

$$\frac{Prob(\text{P}<\alpha|true\ discovery)P(true\ discovery)}{P(\text{P}<\alpha|true\ discovery)P(true\ discovery)+P(\text{P}<\alpha|true\ null)P(true\ null)}$$

↑ Power

Prob Null is true	α	Power	Prob(correct|P<α)	Prob(reproducible|Correct reject)
95%	0.05	0.80	46%	0.64
95%	0.01	0.47	71%	0.22
50%	0.05	0.80	94%	0.64
50%	0.01	0.47	98%	0.22

Reproducibility in the Lab

Statistical testing:

Prob(correct|P<α)=

$$\frac{Prob(\text{P}<\alpha|true\ discovery)P(true\ discovery)}{P(\text{P}<\alpha|true\ discovery)P(true\ discovery)+P(\text{P}<\alpha|true\ null)P(true\ null)}$$

Power

Lessons: Improving P(true discovery) is critical.

How?

Reproducibility in the Lab

Statistical testing:

Prob(correct|P<α)=

$$\frac{Prob(\text{P}<\alpha|true\ discovery)P(true\ discovery)}{P(\text{P}<\alpha|true\ discovery)P(true\ discovery)+P(\text{P}<\alpha|true\ null)P(true\ null)}$$

Power

Lessons: Improving P(true discovery) is critical.

How?

Do the science	
Control variability	
Adequate sample size	

Reproducibility in the Lab

Statistical testing:

Prob(correct|P<α)=

$$\frac{Prob(\text{P}<\alpha|true\ discovery)P(true\ discovery)}{P(\text{P}<\alpha|true\ discovery)P(true\ discovery)+P(\text{P}<\alpha|true\ null)P(true\ null)}$$

Power

Lessons: Improving P(true discovery) is critical.

How?

Do the science	**Preliminary studies, literature search**
Control variability	**Good experimental design**
Adequate sample size	**Sample size or power computation**

Reproducible Analysis

(2009) Repeatability of published microarray gene expression analyses.

Ioannidis JP[1], Allison DB, Ball CA, Coulibaly I, Cui X, Culhane AC, Falchi M, Furlanello C, Game L, Jurman G, Mangion J, Mehta T, Nitzberg M, Page GP, Petretto E, van Noort V.

Evaluated 18 articles for which the data were available.

- One table or figure from each article was independently evaluated by two teams of analysts.
- Reproduced:
  - two analyses in principle
  - six partially or with some discrepancies
  - ten could not be reproduced.

Reproducible Analysis

Why can't we reproduce the analysis?

My experience (similar to the paper):

1. Data not *actually* published.
2. Data available but poorly annotated.
3. Data preprocessing steps not documented.
4. Data analysis not adequately described.

Reproducible Analysis

Lessons:

1. Make sure that the data are properly uploaded to repository.
2. Check uploaded annotation.
3. Also upload preprocessed data.
4. Keep history of analysis using software – e.g. Galaxy or R-Sweave.
5. Upload analysis scripts with data.

Ireproducible is unethical

1. **Incorrect results lead to wrong policies to solve pressing problems.**
2. **Research resources are finite and should not be wasted chasing wrong results.**

Reproducible Research

1. **Do the background research before launching into resource-intense work.**
2. **Document everything (even if not publishable because "no discovery").**
3. **Be honest about the failures as well as the successes.**
4. **Make the data and the analyses available.**