

Epidemiology of a scientific paper

Notes from lecture on Methodology of Science and Bioinformatics at 2nd Faculty of Medicine, Charles University in Prague

What did we discuss: How Science shares information? which problems can occur along the way

Example-Werner Bezwoda

an example was given on a trial for breast cancer patients. The patients in the trial had had a bone marrow transplant prior to their chemotherapy and were able, to receive a higher dosage of chemotherapeutics. He reported great results

-in 1999 other studies contradicted this.

They found out that there is no proof that the study actually happened, and the heavy side effects of that treatment were not reported.

-10 000 patients received this treatment

What motivates Scientists?

-Do good science/help mankind

- Finances/Grants
- Career → high social status of Science+ Competition
- Prestige
- And more

→It's all connected

How Science spread?

Official Channels:

- Journals
- Books
- Textbooks
- Conference proceedings
- Conferences

they differ in various fields, f.e. Medicine-mostly Journals; Computer science-mostly Conferences

Unofficial Channels:

- Preprints (paper is ready for proofreading and already published)
- Blogs
- Social media (Twitter, Mastodon...)
- Press release

Social media is very important, discussion on scientific publishing very early on

How do we detect/correct mistakes made in Science?

In Theory:

- Peer review barrier
- Letters to editor
- Papers in response (f.e. criticising the work that has been done)
- Expression of concern
- Retraction (when Fraud or mistake was proven it gets removed/corrected)

Peer review barrier

-Peer review is a poor barrier, results of peer review have a big element of randomness (e.g. the NIPS peer review experiment (<http://blog.mrtz.org/2014/12/15/the-nips-experiment.html>)). Who you get as peer reviewer determines the outcome (we can have sloppy reviewers)

-it was never designed to detect Fraud!

-Example: Brian Wansink: former american professor, which published questionable papers, where the Participants percentages did not match up. All of those papers passed peer review.

Letters to editor

-Can take months, Or not happen at all

-Limited words, Defending party gets usually more Space

Papers in Response

-Incumbent advantage: means there is an advantage of the paper that was published first, other papers have to prove you wrong

-Example: Didier Raoult and Hydroxychloroquine: Trial in Covid-19, Death of Patients were left out of trial → Fraud? every paper that came after this trial, had to justify this paper and proof them wrong

Expression of concern

-If authors disagree, can take years

-Example: Wakefield the Lancet MMR Autism Fraud: he claimed to have found a linkage between enterocolitis and autism. No other scientists were able to reproduce his findings. *It took 12 years to retract it.*

-Retracted papers still cited (even positively)

We dont really know how many Scientific papers are fraudulent!

However, there are a few persons, that tried to proof it:

Elisabeth Bik: found 3,8% that are copy paste (only found this specific type of fraud)

Carlisle: estimated after years of evaluation, 20% showed false data (f.e. same Patient twice etc..)

Consequences of Fraud

Problems:

Publishing more, also gets rewarded more →there is a drive to publish a lot.

Scientific Journals like surprising claims, and publish more controversial topics.

-What is the difference between big negligence and intent? in the end it does not change the situation for the Patients!

-Example: star surgeon Paolo Macchiarini: transplanted Trachea from Cadavers, populated them with stem cells for better outcome of the acceptance of the graft. he published successful operations, but some people died, which was not mentioned. He got a lot of money from funding. People working close with him got fired for mentioning the fraud.

Quantitative Metrics

by Citations:

-shows usefulness of previous studies, but they dont measure Quality!

H Index:

-measurement for number of Citations

-high h Index = high number of Citations

Impact Factor:

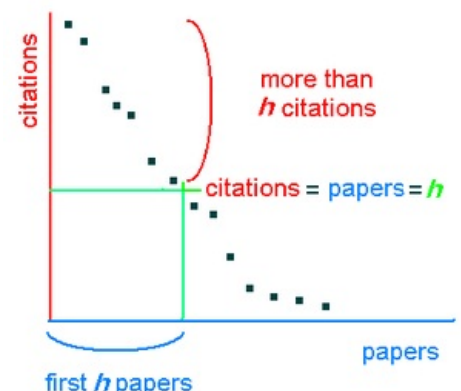
-used to measure frequency of with which the average article in a journal has been cited in a particular year.

-not really an average since Journals can ask to leave it out

Problems of Citations:

1) *Goodhart's / Strathern Law*: When a measure becomes a target, it ceases to be a good measure (turns into a competition)

2) *Matthew effect*: rich get richer (result with most citations get up to the top of Google/ Paper with the most Citations gets more Citation)



H-Index

3) *biases*:

- Mathilda effect: women excluded from work
- Rosalind Frankling: excluded from work on DNA

Grants:

- good scientific metrics
- Competetive (taking weeks of work for application, could maybe be wasted time)
- Result oriented (possible to return money)

Preregistration

- specifying your research plan in advance of your study and submitting it to a registry
- Researcher degrees of freedom
- Publication bias
- Registered reports