Reproducibility in the social sciences

Rationale, tools & best practice

Thomas de Graaff

June 24, 2021

Department of Spatial Economics, Vrije Universiteit Amsterdam

Replication crisis (?)—roadmap

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.

factors that influence this problem and some corollaries thereof.

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

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

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.

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 = $(1-\beta)R/(R$

- Why reproducibility?
- Which tools?
- How to write things down?
- Script everything. No, really!
- Versioning

Reproducibility in the social sciences

- In the social sciences few attention to what workflow to use (and why)
- more emphasis on transparency
    - scripts, data & additional analysis are openly shared
- increasing use of (large) datasets in the social sciences
    - More positive approach
- Early work Healy (2011) and Arribas-Bel and de Graaff (2015)

Why? Keeping sanity

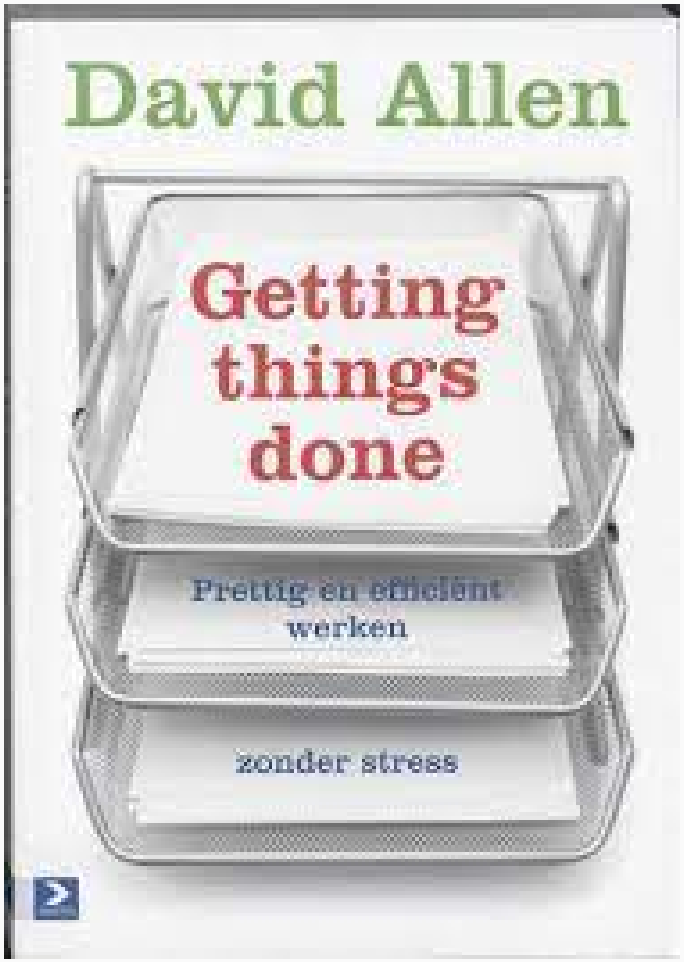
Because projects involve:

- whilst supervisor/referee not satisfied
  - whilst you are not satisfied
    1. read papers
    2. collect data;
    3. transform data;
    4. analyse data;
    5. write up results;
    6. present results;

- Circular? No, see wonderful time-lapse video

Why II? Do not loose your thoughts!

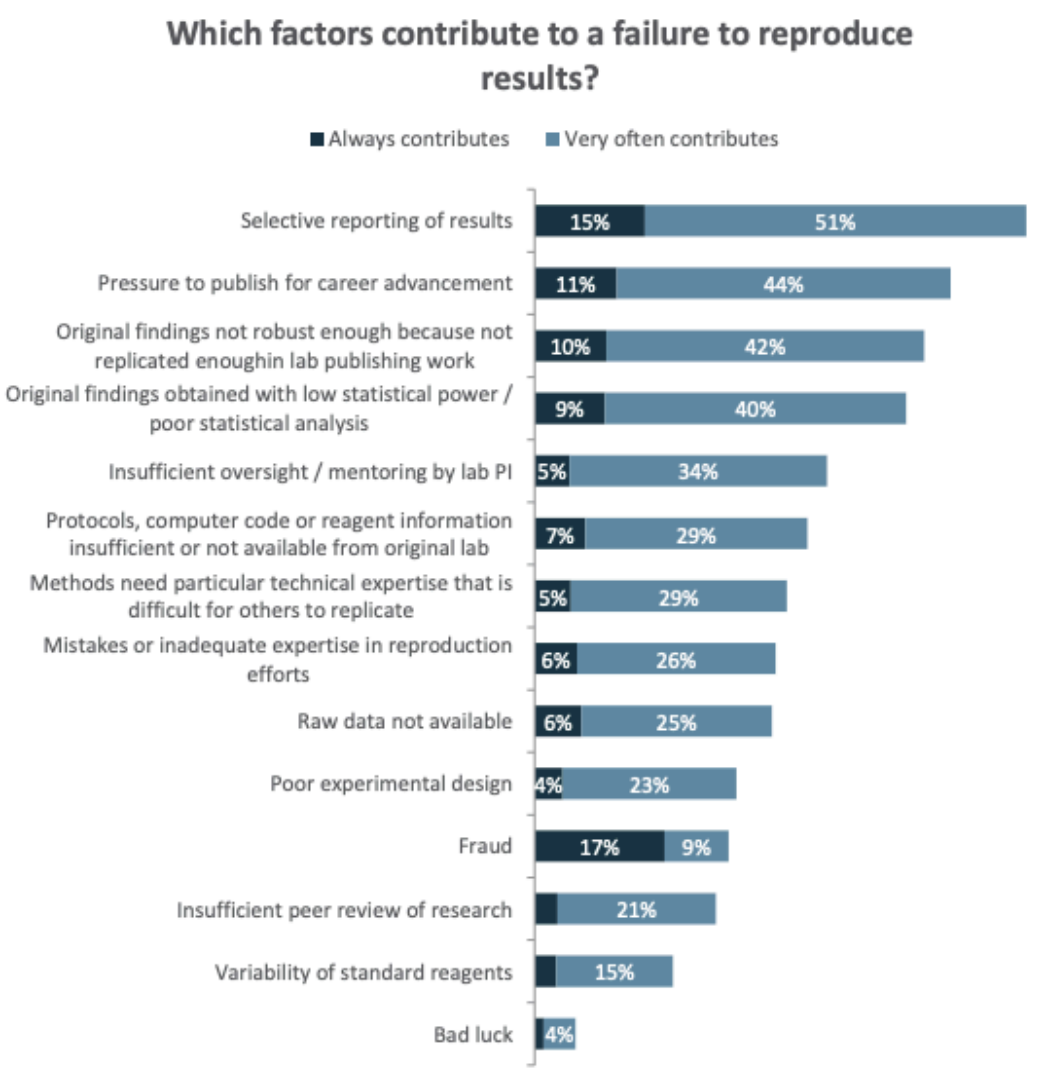
Notes aren't a record of my thinking process. They are my thinking process—Richard Feynman



More efficiency & creativity

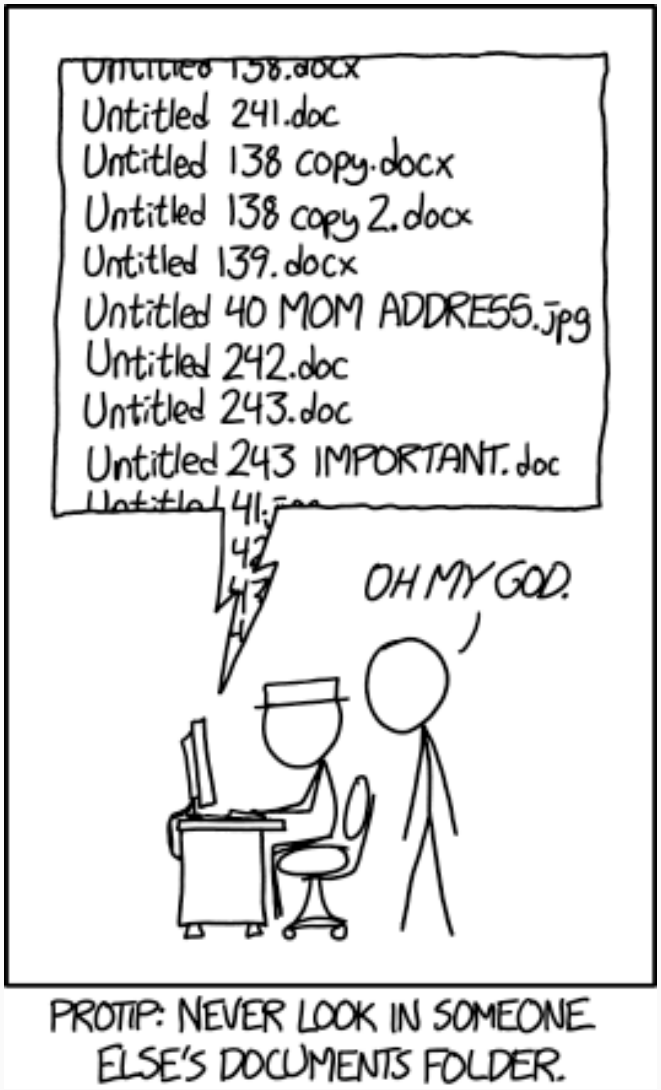
- "Never Have The Same Thought Twice. Unless You Like That Thought" (Allen, 2001)

The greater good (Nature, 2017)



Organisation, R and RStudio

Organise your stuff!




```
/
└── project_X
    ├── data
    │   ├── source
    │   └── derived
    ├── docs
    │   ├── notes
    │   └── paper
    ├── literature
    └── code
        └── r
            ├── data
            └── figures
```

Use projects

Multiple code files by functionality

Documentation

README files

Coding style

Creating documentation (advanced)

Script everything!

Reading data and data wrangling

Figures

Output

Functions (advanced)

Git and Github (advanced)

In conclusion

When should I adopt an open reproducable workflow?

- The sooner the better (now you have time)
- But think twice about tools to invest time in
    - choose well-maintained tools with large communities (R, Python)
    - invest some time in markup languages (RMarkdown, LaTeX)
    - really, really think about versioning (Git & Github)
- Start one step at a time

Questions/comments?

Get the source of this presentation from

`https://github.com/Thdegraaff/reproducibility_nscr`

References i

- Allen, D. (2001). *Getting Things Done: The Art of Stress-Free Productivity*. Penguin. 354 pp. Google Books: 7PoYBAAAQBAJ.
- Arribas-Bel, D. and T. de Graaff (2015). "WooW-II: Workshop on Open Workflows". In: *REGION* 2.2 (2), R1–R2.
- Healy, K. (2011). "Choosing Your Work Flow Applications". In: *The Political Methodologist* 18.2, pp. 9–18.