Teaching about Open and Reproducible Science

An idea for a transferable skills course

Stefanie Muff

January 18th 2023

Background

- In November 2022, Bob and I were involved in teaching a PhD course on *Open and Reproducible Science*.
- 2.5 credit points.
- Taking place in the Alpine center in Finse.
- Roughly 25 participants.

We received very positive feedback from the students.

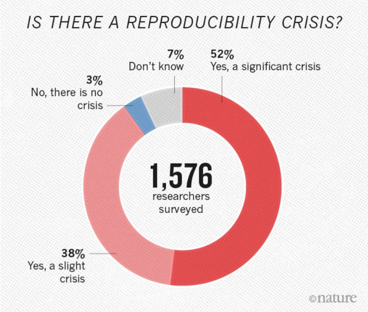
> *"I think this course should be continued."*
>
> *"There are many colleagues who love the idea of open science but they feel isolated nad do not have a direction to move ahead. I will recommend this course to them."*
>
> *"I would recommend the course to all PhDs and post-docs at my department/collaborators"*
>
> *"I hope that more young researchers can join the course in the future. I think the information and skills after such course can influence significantly someone's career or even change it."*
>
> *Our students: "This course should be on offer as a transferable skills course at NTNU!"*

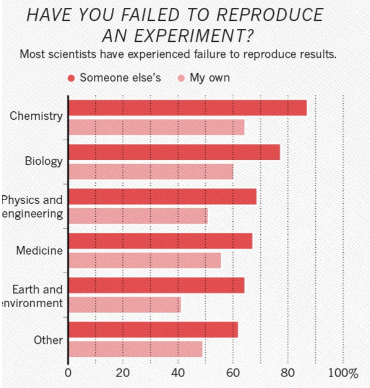
Course content

- Introduction on terminology (open science, reproducibility, transparancy)
- Data Management Plan (DMP)
- Data collection and data handling
- Data repositories
- FAIR data principles
- Using version control (git/GitHub)
- Reproducible work flows / neat coding strategies
- Best practice in data analysis and reporting results

Why is this relevant?

A survey carried out by *Nature* in 2016, sheds light on researcher's experiences and thoughts.





SCIENTIFIC INTEGRITY

What does research reproducibility mean?

Steven N. Goodman,* Daniele Fanelli, John P. A. Ioannidis

The language and conceptual framework of "research reproducibility" are nonstandard and unsettled across the sciences. In this Perspective, we review an array of explicit and implicit definitions of reproducibility and related terminology, and discuss how to avoid potential misunderstandings when these terms are used as a surrogate for "truth."

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.

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

STATISTICAL ERRORS

P values, the 'gold standard' of statistical validity, are not as reliable as many scientists assume.

BY REGINA NUZZO

A Dirty Dozen: Twelve *P*-Value Misconceptions

Steven Goodman

The *P* value is a measure of statistical evidence that appears in virtually all medical research papers. Its interpretation is made extraordinarily difficult because it is not part of any formal system of statistical inference. As a result, the *P* value's inferential meaning is widely and often wildly misconstrued, a fact that has been pointed out in innumerable papers and books appearing since at least the 1940s. This commentary reviews a dozen of these common misinterpretations and explains why each is wrong. It also reviews the possible consequences of these improper understandings or representations of its meaning. Finally, it contrasts the *P* value with its Bayesian counterpart, the Bayes' factor, which has virtually all of the desirable properties of an evidential measure that the *P* value lacks, most notably interpretability. The most serious consequence of this array of *P*-value misconceptions is the false belief that the probability of a conclusion being in error can be calculated from the data in a single experiment without reference to external evidence or the plausibility of the underlying mechanism.

Semin Hematol 45:135-140

COMMENT · 20 MARCH 2019

Scientists rise up against statistical significance

Valentin Amrhein, Sander Greenland, Blake McShane and more than 800 signatories call for an end to hyped claims and the dismissal of possibly crucial effects.

Valentin Amrhein, Sander Greenland & Blake McShane

Open and reproducible research...

- ...benefits the researcher (easy to track what you did, modify analyses, builds trust in your work, increases citation rates etc).
- ...benefits the research community (others can build on your data, code and results).
- ...benefits society (more insight for money, more trustworthy results, more scientific progress).
- ...are thus *basic skills any researcher in the future will HAVE to master*!

Other institutions are a step ahead

For example, the Center for Reproducible Science, University of Zurich:

https://www.crs.uzh.ch/en.html

Our idea

Given the positive feedback on the mentioned course (and the negative feedback on other TS courses) and the importance of the topic, we wondered:

- Is it possible to set up a transferable skills course at the department/faculty/entire NTNU?
- If yes, what would be the strategy forward?