Reproducible research

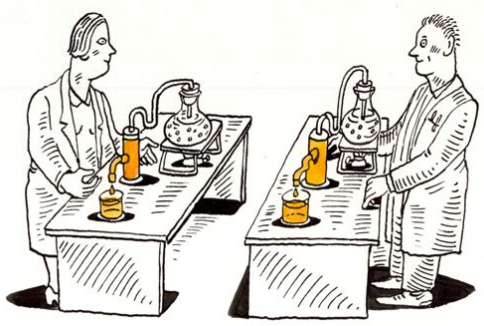
in data science

Mikhail Dozmorov

Overview

- What is reproducible research?
- Why do we care?
- Why reproducibility questions arise?
- The cost of reproducibility
- Reproducibility and statistics
- Current status of reproducibility
- What can we do?

What is reproducible research?

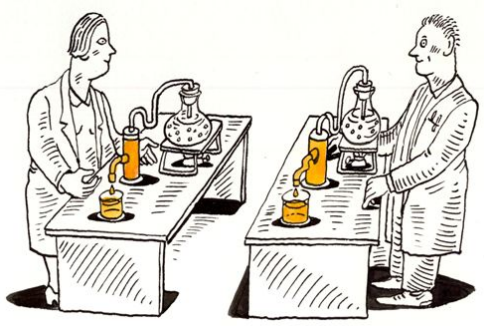


`http://blog.f1000research.com/2014/04/04/reproducibility-tweetchat-recap/`

Overview

- What is reproducible research?
- Why do we care?
- Why reproducibility questions arise?
- The cost of reproducibility
- Reproducibility and statistics
- Current status of reproducibility
- What can we do?

What is reproducible research?

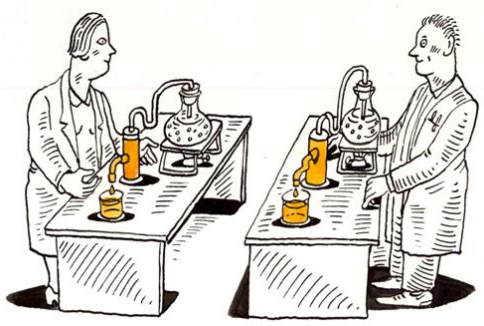


`http://blog.f1000research.com/2014/04/04/reproducibility-tweetchat-recap/`

Overview

- What is reproducible research?
- Why do we care?
- Why reproducibility questions arise?
- The cost of reproducibility
- Reproducibility and statistics
- Current status of reproducibility
- What can we do?

What is reproducible research?



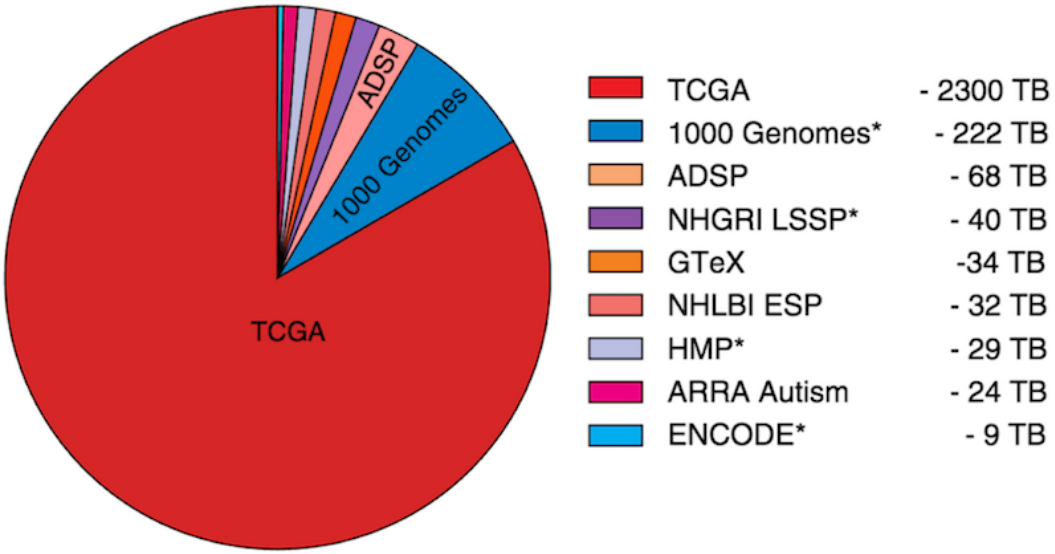
`http://blog.f1000research.com/2014/04/04/reproducibility-tweetchat-recap/`

Why do we care?

More data = more chance for errors

- High-throughput biology generates volumes of data
- Data-generating technologies are increasingly used to make clinical recommendations and treatment decisions
- A problem may be overlooked .. Published .. Get in clinical trials

More data = more chance for errors

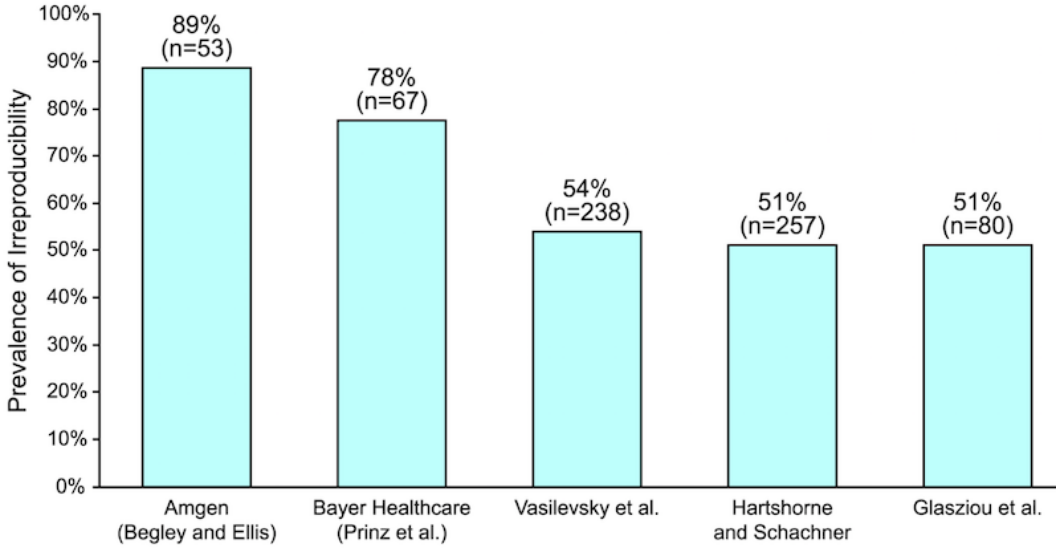


- Muir et al. "The Real Cost of Sequencing." Genome Biol. 2016

The cost of reproducibility

Figure 1:

Irreproducfibility ranges ~51% - 89%



Why reproducibility questions arise?

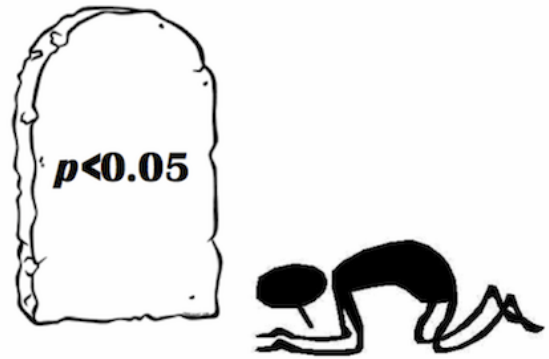


Figure 2:

Patterns in the noise

- Humans are good at recognizing patterns

Human beings do not have very many natural defenses. We are not all that fast, and we are not all that strong. We do not have claws or fangs or body armor. We cannot spit venom. We cannot camouflage ourselves. And we cannot fly. Instead, we survive by means of our wits. Our minds are quick. **We are wired to detect patterns** and respond to opportunities and threats without much hesitation.

Current status of reproducibility

Focus on preclinical research

Raise standards for preclinical cancer research

C. Glenn Begley and Lee M. Ellis propose how methods, publications and incentives must change if patients are to benefit.

Efforts over the past decade to characterize the genetic alterations in human cancers have led to a better

trials in oncology have the highest failure rate compared with other therapeutic areas. Given the high unmet need in oncology, it

investigators must reassess their approach to translating discovery research into greater clinical success and impact.

- Glenn Begley and Lee Ellis, "Drug Development. Raise standards for preclinical cancer research" Nature 2012

`http://www.nature.com/nature/journal/v483/n7391/full/483531a.html`

Focus on preclinical research

ing agencies to establish or enforce policies

What can we do | Tools to enhance reproducibility?

Flavors of reproducibility

- Empirical reproducibility
- Computational reproducibility
- Statistical reproducibility

Flavors of reproducibility

- Empirical reproducibility
- **Computational reproducibility**
- Statistical reproducibility

Steps in reproducible research

The most important is the mindset, when starting, that the

What can we do | Tools to enhance reproducibility?

Flavors of reproducibility

- Empirical reproducibility
- Computational reproducibility
- Statistical reproducibility

Flavors of reproducibility

- Empirical reproducibility
- **Computational reproducibility**
- Statistical reproducibility

Steps in reproducible research

The most important is the mindset, when starting, that the

What can we do | Tools to enhance reproducibility?

Flavors of reproducibility

- Empirical reproducibility
- Computational reproducibility
- Statistical reproducibility

Flavors of reproducibility

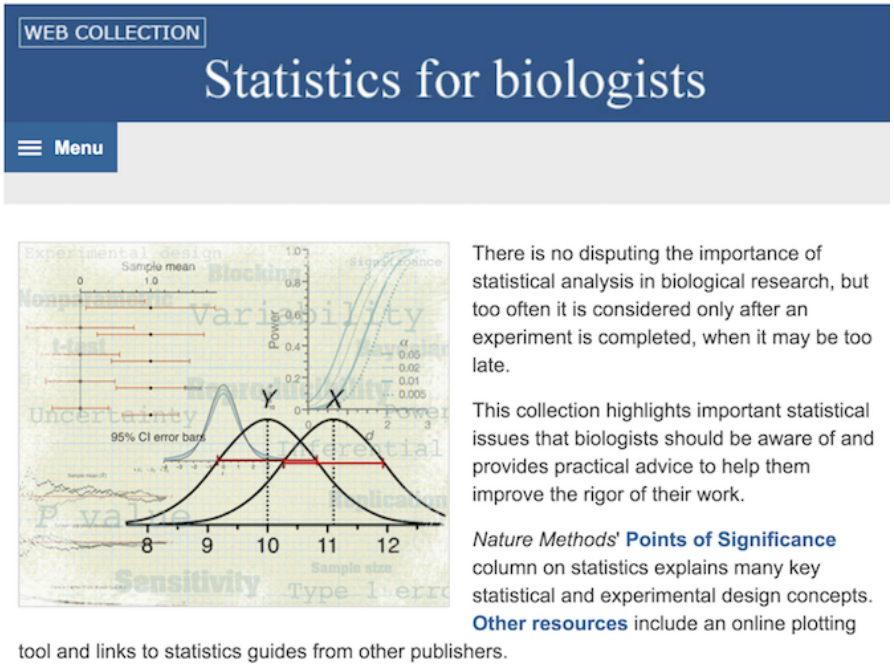
- Empirical reproducibility
- **Computational reproducibility**
- Statistical reproducibility

Steps in reproducible research

The most important is the mindset, when starting, that the

Learn more

Nature "Statistics for Biologists"



`http://www.nature.com/collections/qghhqm`

Reproducible research made simple

`https://www.ncbi.nlm.nih.gov/pubmed/24204232`