



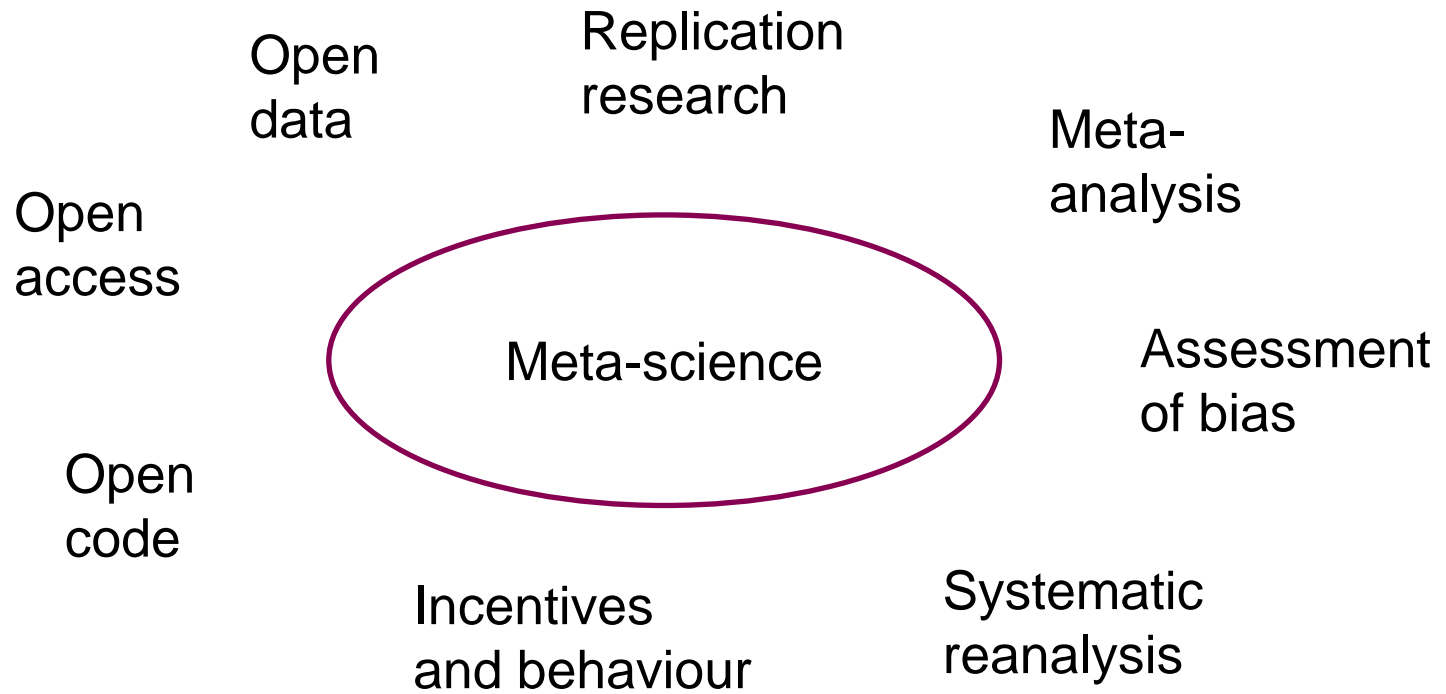
**Karolinska
Institutet**

Open data in research on humans – promises and challenges

Gustav Nilsson

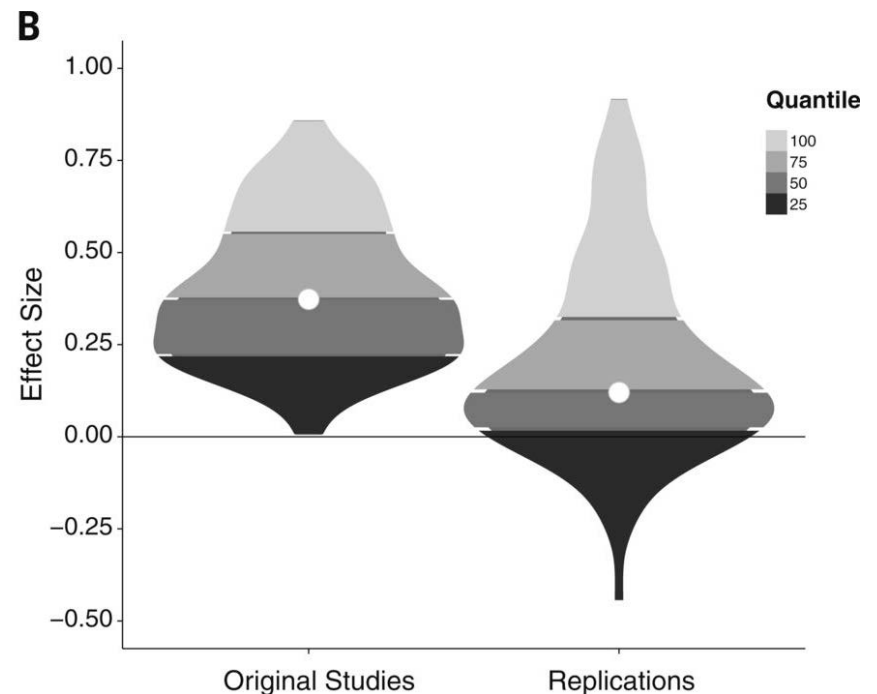
2019-08-22



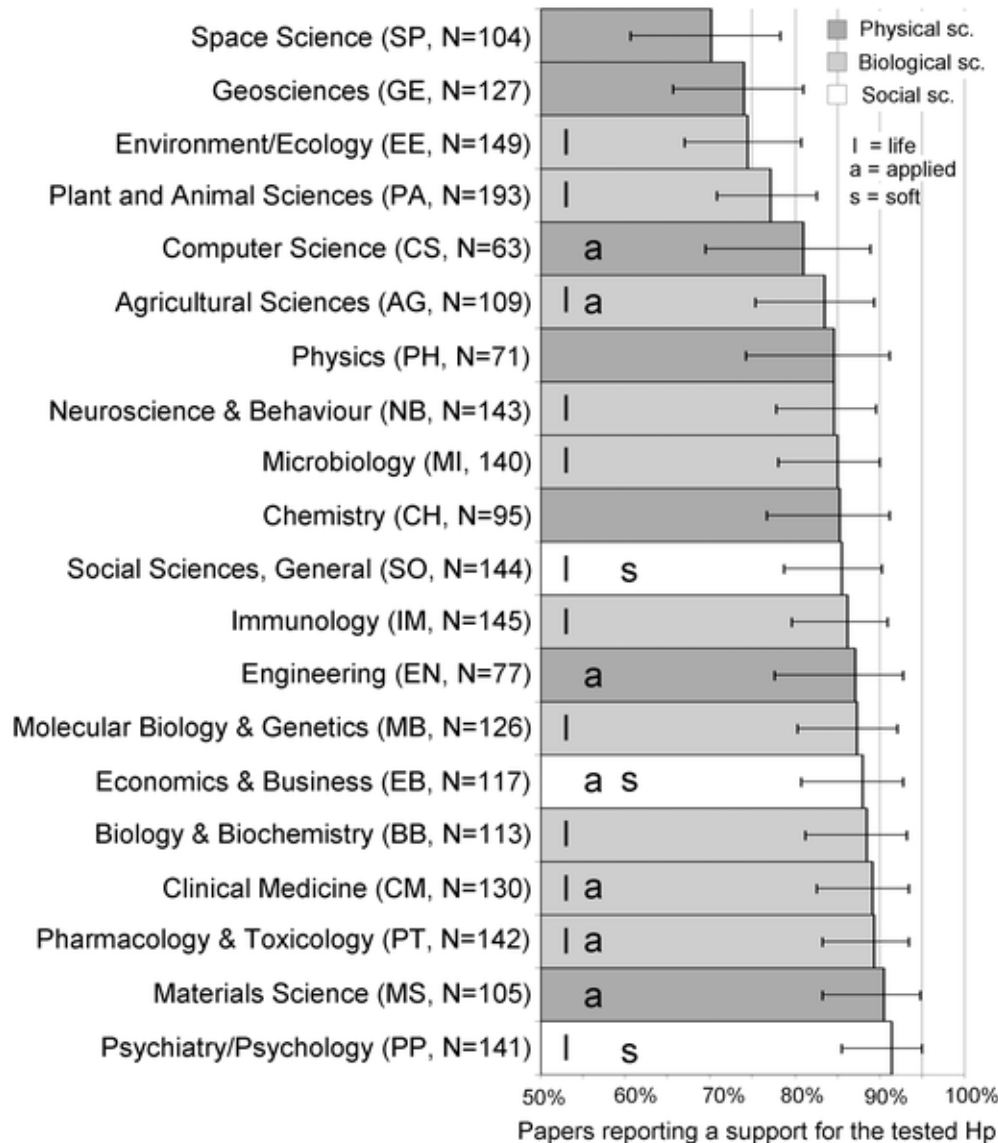


Reproducibility Project: Psychology

- Aim: to obtain an initial estimate of reproducibility in experimental psychology
- Attempted replication of 100 studies
- 36/97 studies showed a statistically significant effect in the same direction as the original study
- Effect size halved on average



Share of positive findings in different fields



ROYAL SOCIETY
OPEN SCIENCE

rsos.royalsocietypublishing.org

Research



Cite this article: Ingre M, Nilsson G. 2018
Estimating statistical power, posterior probability
and publication bias of psychological research
using the observed replication rate. *R. Soc. open
sci.* **5**: 181190.

<http://dx.doi.org/10.1098/rsos.181190>

Estimating statistical power, posterior probability and publication bias of psychological research using the observed replication rate

Michael Ingre^{1,2} and Gustav Nilsson^{1,3,4}

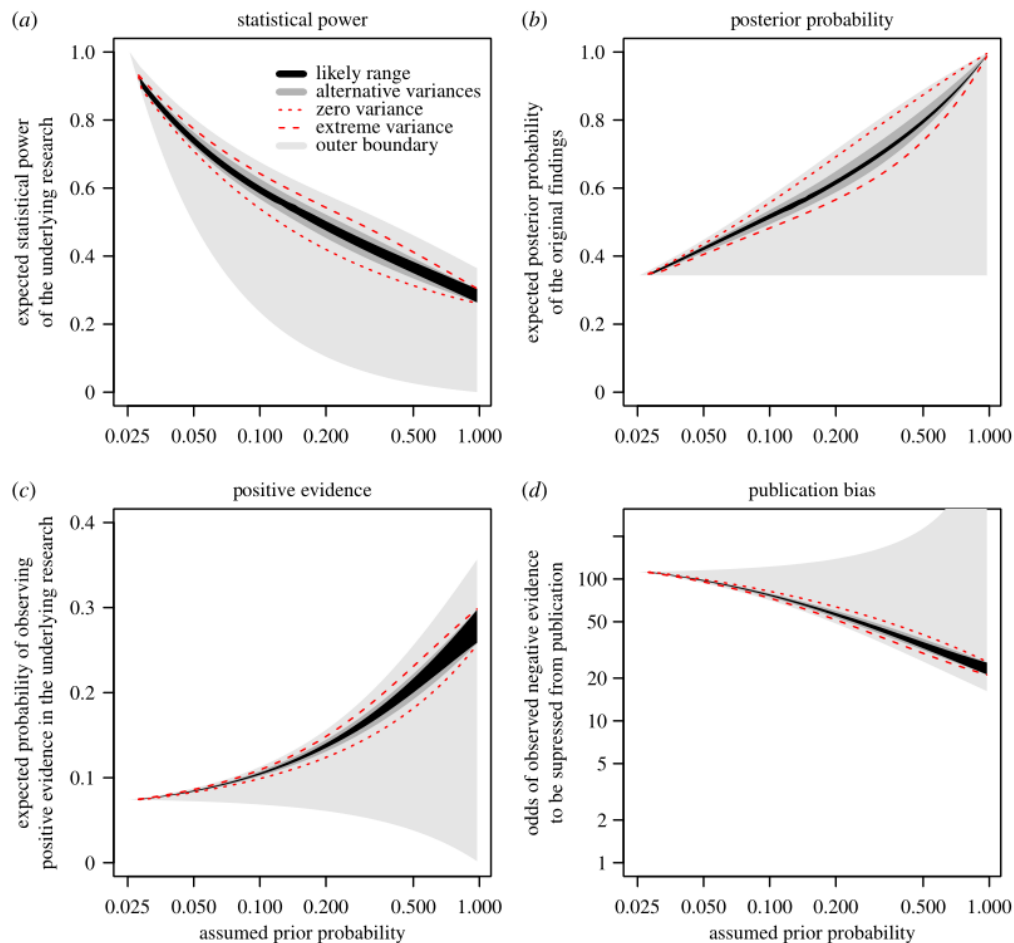
¹Department of Clinical Neuroscience, Karolinska Institutet, Solna, Sweden

²Institute for Globally Distributed Open Research and Education (IGDORE), Sweden

³Stress Research Institute, Stockholm University, Stockholm, Sweden

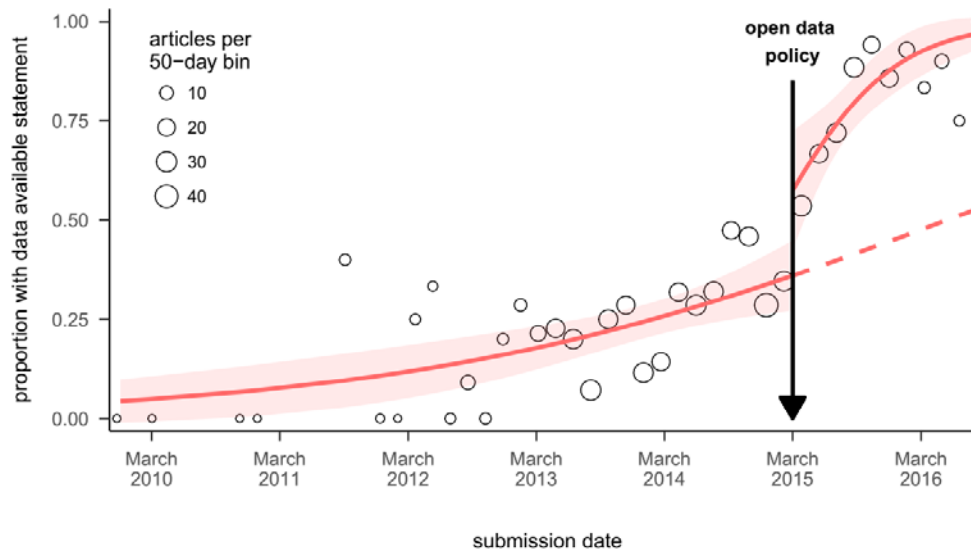
⁴Department of Psychology, Stanford University, Stanford, CA 94305, USA

Properties of a literature with 90% positive findings and 36% reproducibility rate



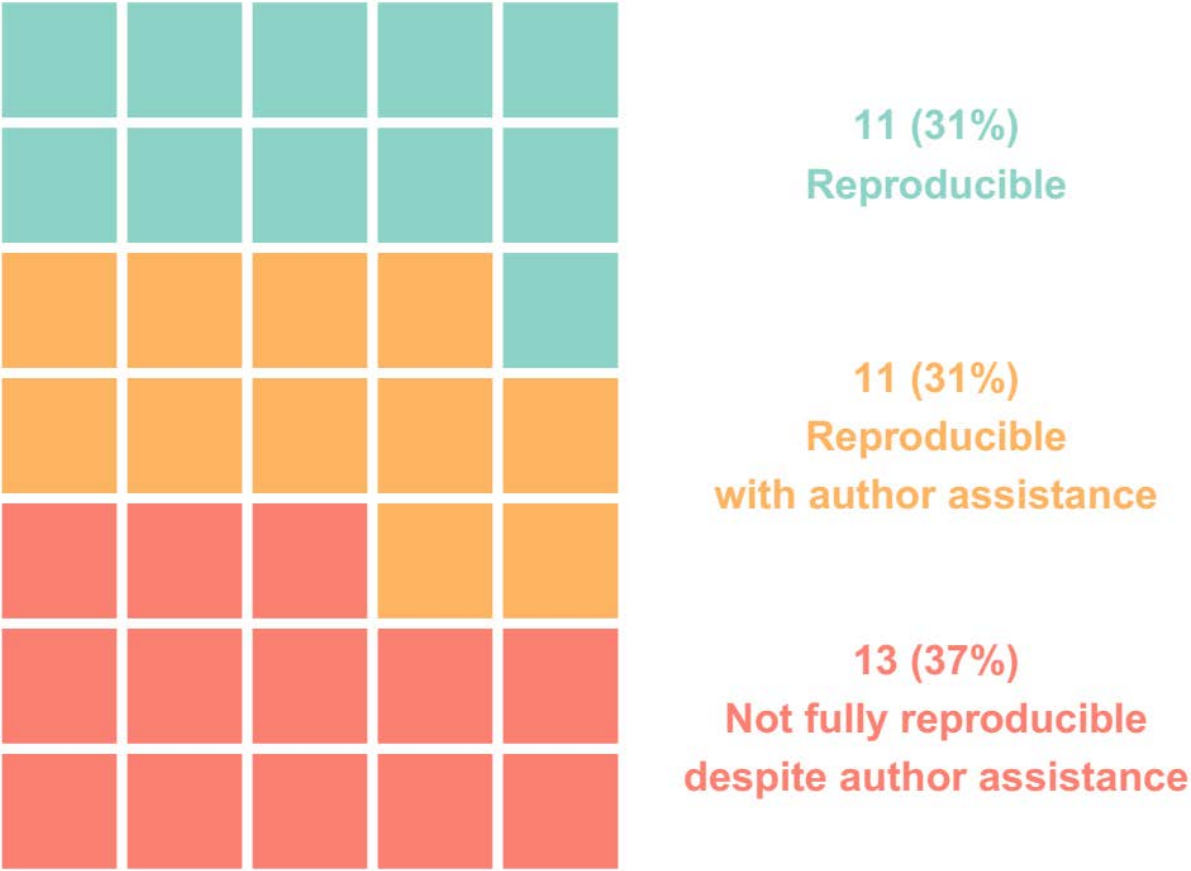
Computational reproducibility: a case study

- The journal Cognition introduced an open data policy in 2015
- We retrieved data from 35 papers and attempted to reproduce the main reported results

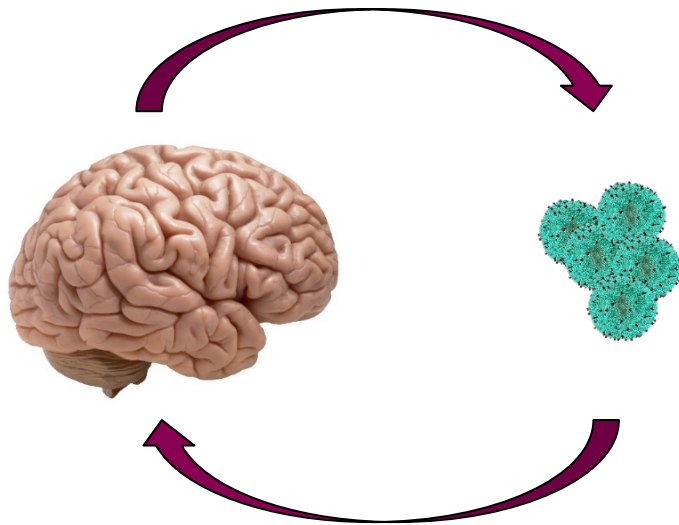


[Hardwicke et al, Royal Society Open Science 2018](#)

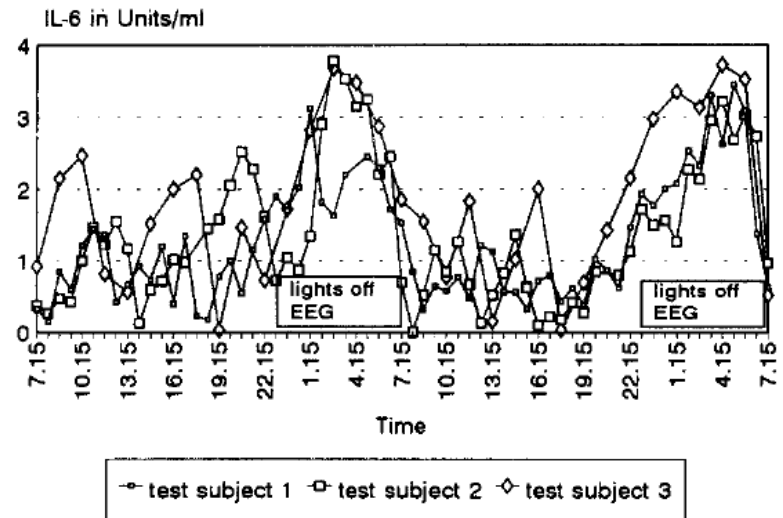
One third fully reproducible from data and published methods description



Does the immune system make us sleepy at night? Lessons from a meta-analysis

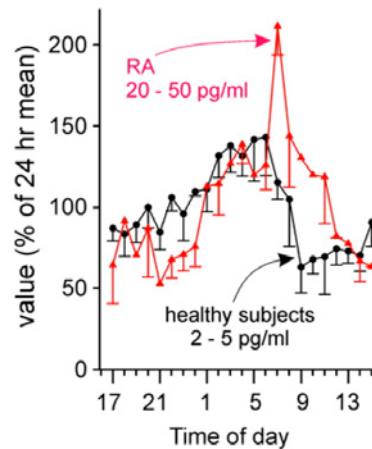
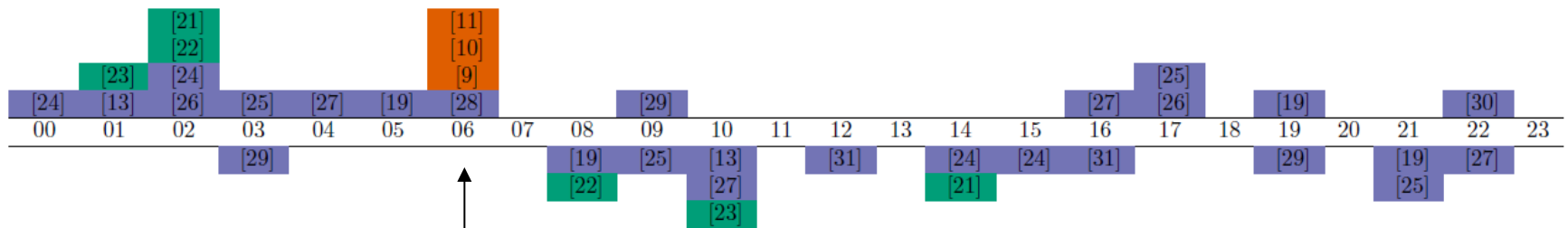


Interleukin-6



Bauer et al. Clin Investig. 1994 72(4):315

Interleukin-6 over the day: reported peaks and troughs

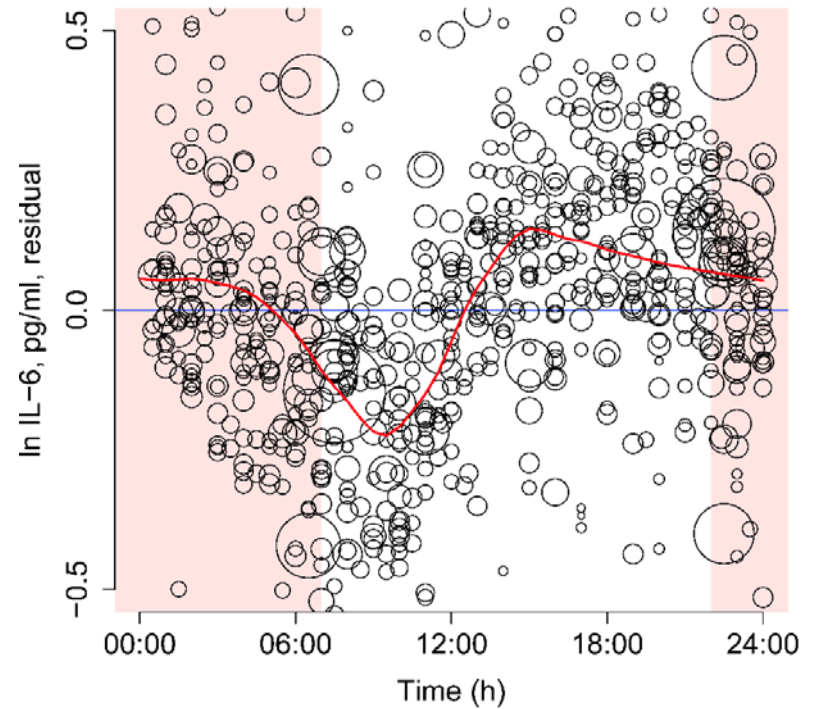
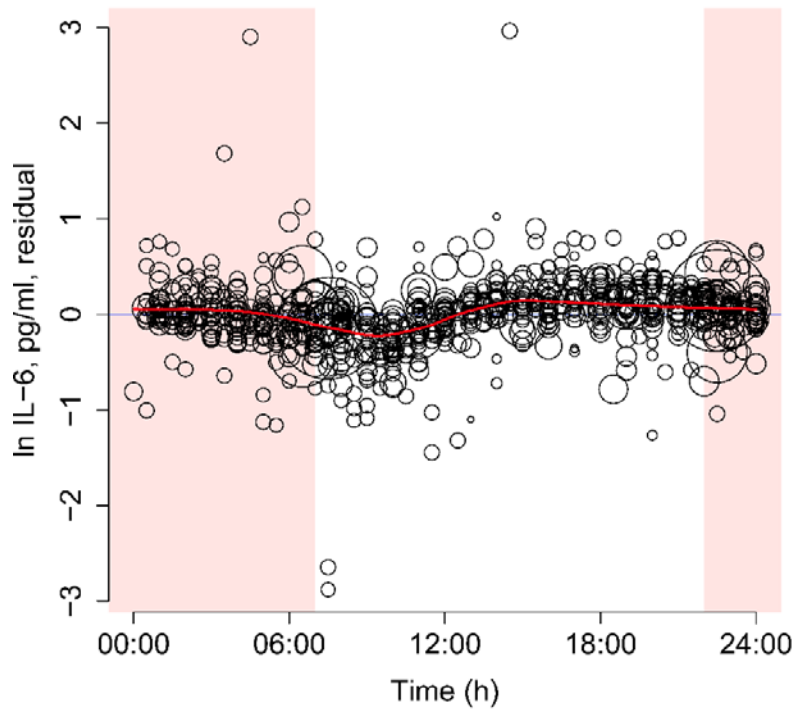


Cutolo & Straub. Autoimmunity Reviews 7 (2008) 223–228

Data for meta-analysis

- 43 studies with 56 datasets from a total of 1100 persons could be excluded in a meta-analysis
- Individual participant data were available from 3 datasets
- Another 36 studies fulfilled inclusion criteria, but could not be included because data were unavailable ($k = 25$), because data had been previously reported ($k = 7$) or because they reported data far from a physiological range ($k = 4$)

Meta-analysis

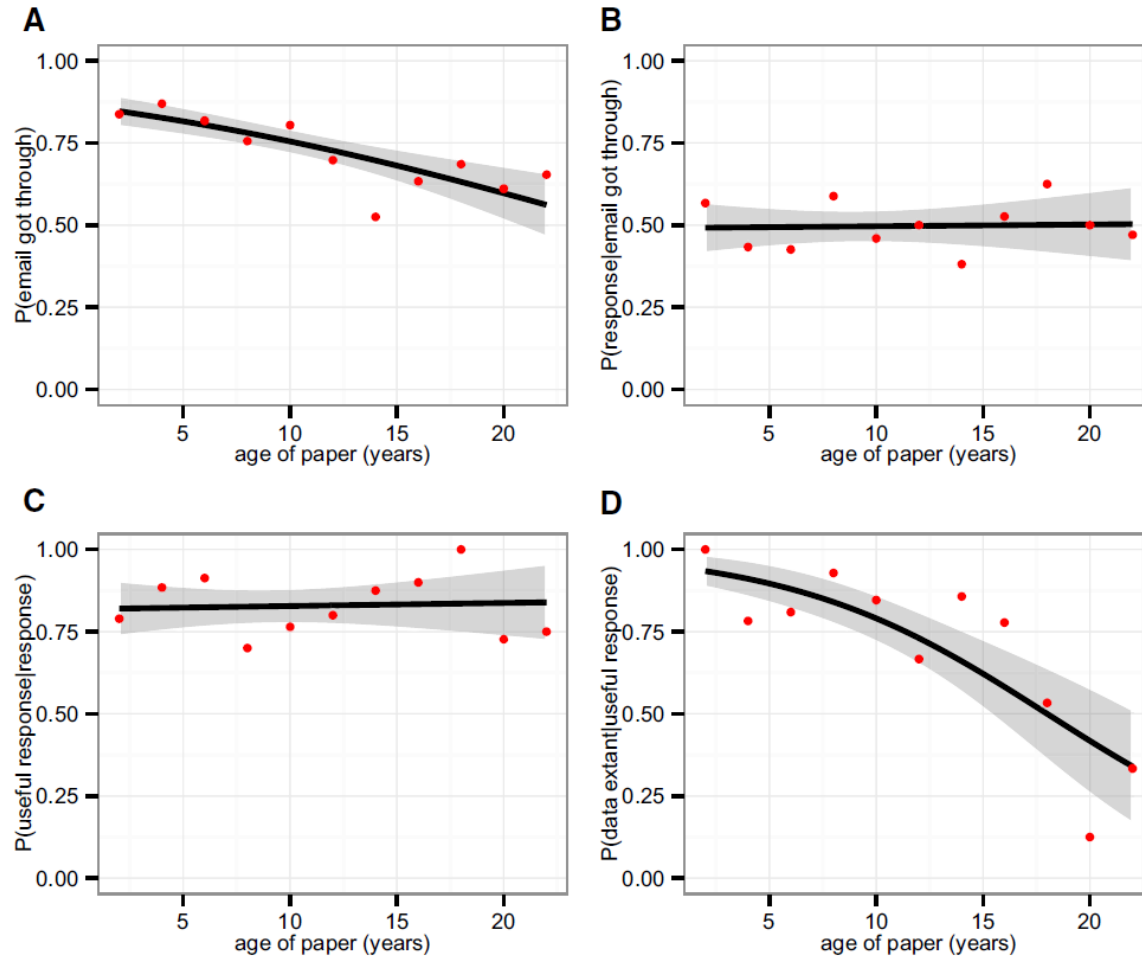


Nilsson et al. Plos One 2016. <https://doi.org/10.1371/journal.pone.0165799>

Open, closed, and hidden data

- The meta-analysis showed that interleukin-6 does not increase at night
- Data could often be estimated from published plots
- Data that could not be included increase the risk of bias and reduce precision of estimates of the diurnal curve
- Lack of individual participant data limits opportunities to model residual heterogeneity
- Closed and hidden data are a waste of resources and may be questioned on ethical grounds

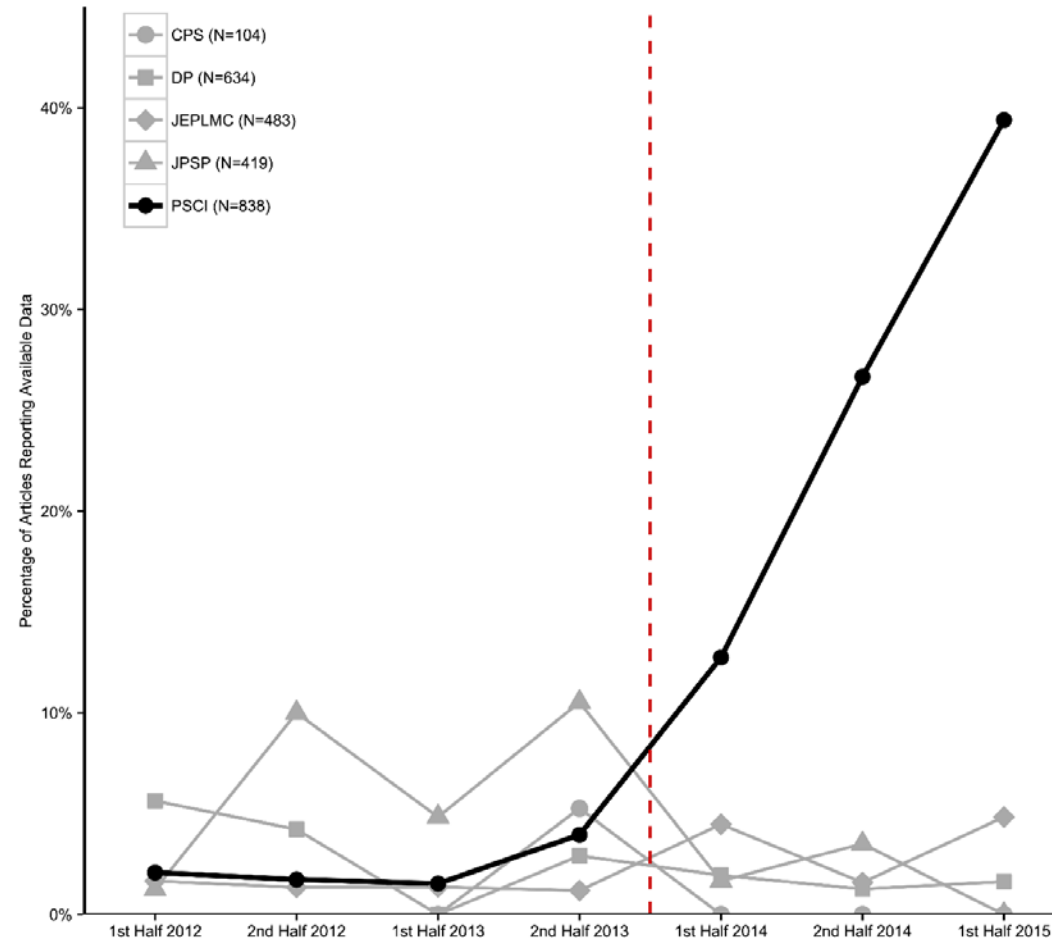
Survival of research data in ecology



Badges for open practices: an effective incentive?



Publication of open data in one journal after the introduction of badges



Data life cycle: new concepts

- **Data erosion:** Functional loss of data reusability. May be due to e.g:
 - Metadata not documented
 - Loss of support for old file formats
 - Loss of records
 - Intentional destruction of records
- **Loss of rawness** is erosion where incomplete aggregate results are preserved but raw data are lost
- **Salvage window:** The time during which data can be saved with an additional effort

Scientific inference: new concepts

- A scientific field that generates data will generate increasing knowledge if the growth in available data outmatches the growth in bias risk due to unavailable data
- A field that generates increasing knowledge can be called a **cumulating field**
- A field that generates new data but not increasing knowledge can be called a **dissipating field**
- Research organizations and funders should make efforts to counteract dissipation by supporting:
 - Publication of new data in appropriate repositories with structured formats and metadata
 - Salvage of old data by curation and publication

Some quick thoughts on incentives

- The scientific paper is canonized as bearer of merit
 - Prestige matters
 - Contributions hard to disentangle
 - Quality control operates mostly in a closed system and cannot be assessed
- Can we imagine a different system?

Conclusions

- Meta-science studies science itself
- Limited reproducibility suggests opportunities for improvement
 - Open data
 - Open code
 - More rigorous scientific practices, e.g. preregistration, larger samples
- Research data need to be salvaged and preserved in a systematic manner
- Changes in the ecosystem of research give cause for optimism