RUTGERS

BIOMEDICAL AND
HEALTH SCIENCES

Rigor and Reproducibility

(*the new R&R for bioscientists*)

Topics In Advanced Biotechnology

September 30, 2016

Ann Stock

Why Are We Discussing Rigor and Reproducibility?

Why Are We Discussing Rigor and Reproducibility?

- NIH began focusing on rigor and reproducibility in 2014.
- New requirements have been introduced since then.
- Additional requirements will continue to be rolled out, including requirements for trainees, anticipated in 2017.

Why Is the NIH Focused on Rigor and Reproducibility?

2011 Scientists at Bayer publish a Correspondence in *Nature Reviews Drug Discovery* reporting inconsistencies between published data and in-house data related to company projects.

2012 Scientists at Amgen publish a Comment in *Nature* reporting problems with reproducibility of preclinical research findings.

2013 An article about unreliable research is published in the *Economist*.

2014 NIH Directors publish a Comment in *Nature* announcing NIH plans to enhance reproducibility.

Scientists at Bayer Report Inconsistencies between Published Data and In-House Data (2011)

CORRESPONDENCE

NATURE REVIEWS | DRUG DISCOVERY
10, 712 (September 2011) | doi:10.1038/nrd3439-c1

Believe it or not: how much can we rely on published data on potential drug targets?

Florian Prinz, Thomas Schlange and Khusru Asadullah

Prinz, F., Schlange, T. & Asadullah, K. (2011) *Nat. Rev. Drug Discov.* 10: 712.

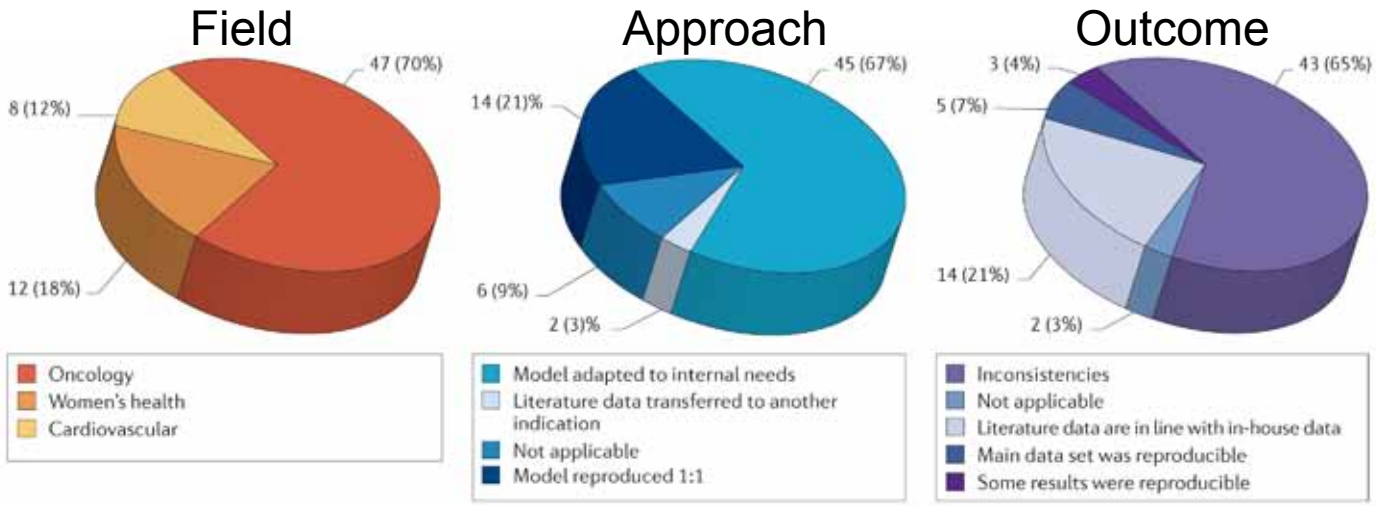
Scientists at Bayer Report Inconsistencies between Published Data and In-House Data (2011)

Methods:

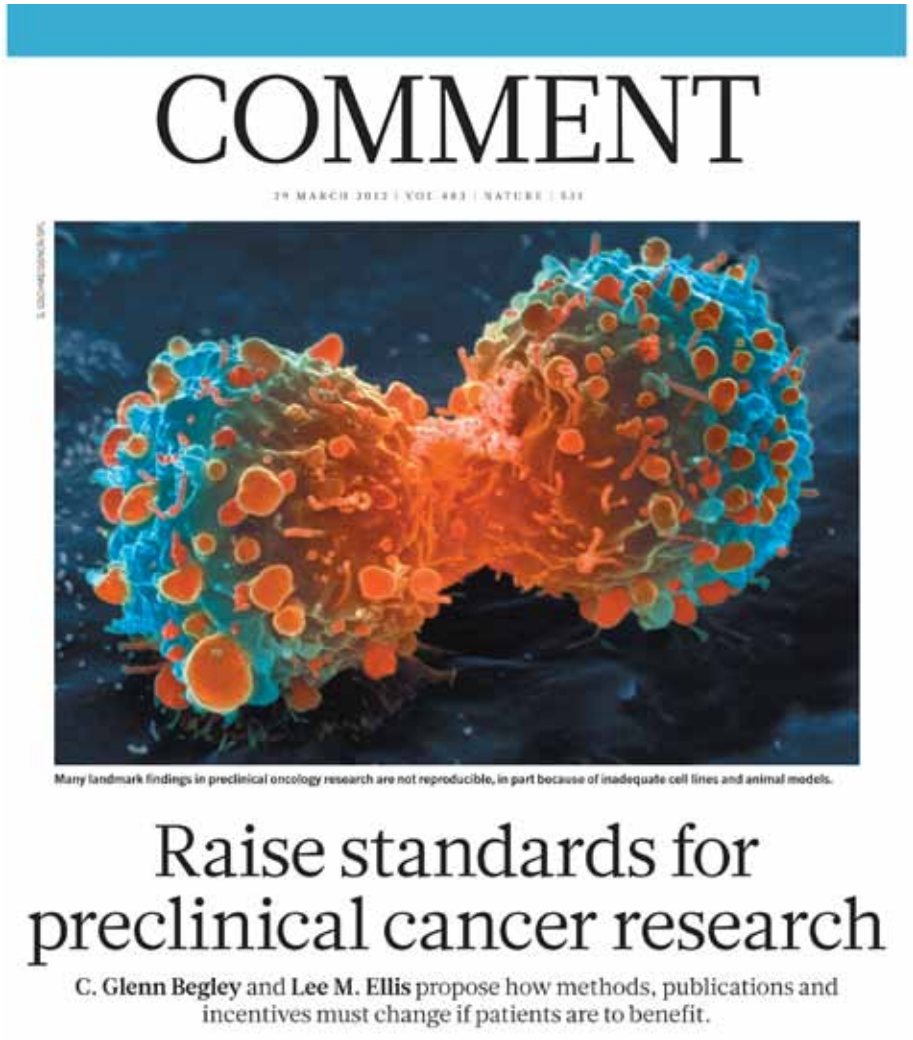
- Analysis of early in-house projects by questionnaire
- Comparison of in-house to published data and project outcome
- Oncology, women's health, cardiovascular disease in recent 4 years
- Responses from 23 scientists, representing 67 projects

Results:

- 20-25% of projects had published data in agreement with in-house findings
- ~2/3 had inconsistencies that prolonged target validation or resulted in project termination for lack of sufficient evidence to validate the therapeutic hypothesis



Scientists at Amgen Report Irreproducibility of Preclinical Research Findings (2012)



COMMENT

29 MARCH 2012 | VOL 483 | NATURE | 531

Many landmark findings in preclinical oncology research are not reproducible, in part because of inadequate cell lines and animal models.

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.

Begley, C.G. & Ellis, L.M. (2012) *Nature* 483: 531-533.

Scientists at Amgen Report Irreproducibility of Preclinical Research Findings (2012)

Methods:

- Before pursuing a line of research, scientists tried to confirm published findings
- Haematology and oncology department
- 53 papers were deemed landmark studies
- Acknowledgment that papers were selected for describing something completely new

Results:

- Scientific findings were confirmed in only 6 of 53 cases (11%)

REPRODUCIBILITY OF RESEARCH FINDINGS

Preclinical research generates many secondary publications, even when results cannot be reproduced.

Journal impact factor	Number of articles	Mean number of citations of non-reproduced articles*	Mean number of citations of reproduced articles
>20	21	248 (range 3–800)	231 (range 82–519)
5–19	32	169 (range 6–1,909)	13 (range 3–24)

Results from ten-year retrospective analysis of experiments performed prospectively. The term 'non-reproduced' was assigned on the basis of findings not being sufficiently robust to drive a drug-development programme.
*Source of citations: Google Scholar, May 2011.

Unreliability of Research Reaches the Popular Press (2013)

The Economist

Unreliable research

Trouble at the lab

Scientists like to think of science as self-correcting. To an alarming degree, it is not

Oct 19th 2013 | From the print edition

"I SEE a train wreck looming," warned Daniel Kahneman, an eminent psychologist, in an open letter last year. The premonition concerned research on a phenomenon known as "priming". Priming studies suggest that decisions can be influenced by apparently irrelevant actions or events that took place just before the cusp of choice. They have been a boom area in psychology over the past decade, and some of their insights have already made it out of the lab and into the toolkits of policy wonks keen on "nudging" the populace.

Unreliability of Research Reaches the Popular Press (2013)

- Cites irreproducibility in psychology (priming) and biomedical (Bayer, Amgen) research
- Governments spent $59 billion on biomedical research in 2012 as a presumed basis for private drug-development work
- The assumption that the system is self-correcting is not supported by data

Many factors contribute to the problem:

- Statistical mistakes
- Research is poorly designed and/or executed
- Peer review is poor at detecting errors
  - A concocted, highly flawed, pseudonymous paper was accepted at 157 of 304 peer-reviewed journals
  - 1998, the *BMJ* editor sent an article with 8 deliberate mistakes to 200 reviewers – none detected all mistakes
- Replication is difficult and thankless
  - A study of 238 articles in 84 journals found less than half identified all reagents
  - Access to data, proprietary software hinder replication

Unreliability of Research Reaches the Popular Press (2013)

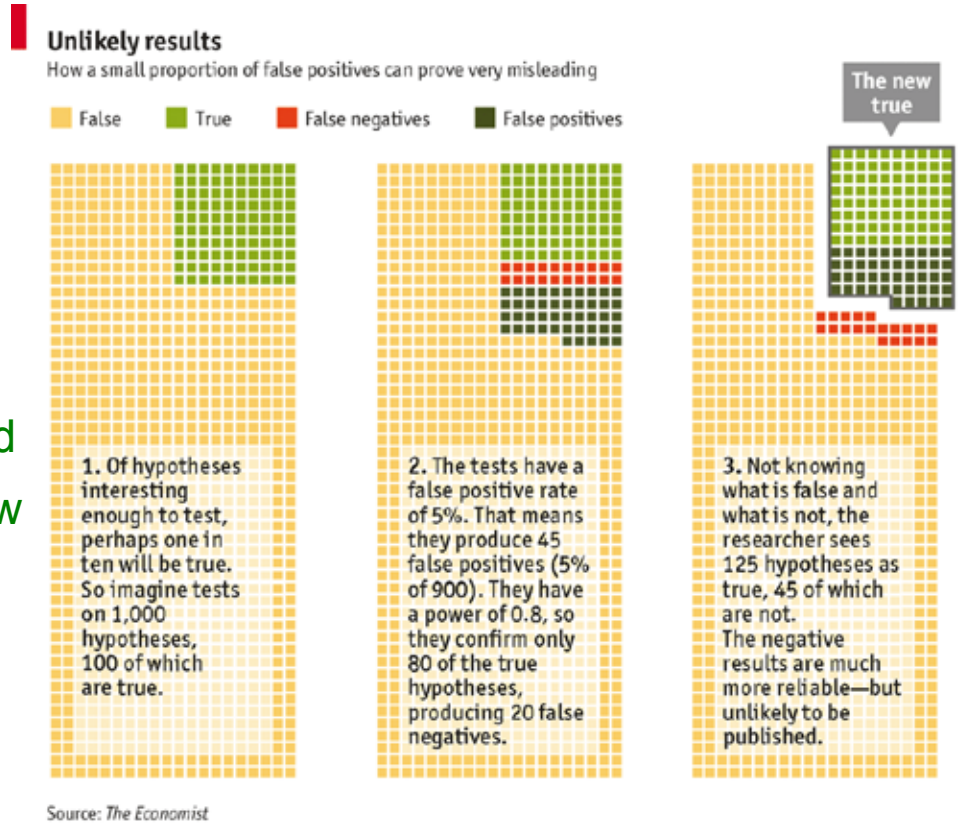
Type 1 Error:	Thinking something is true when it is not	False Positive
Type 2 Error:	Thinking something is not true when it is	False Negative

General Assumption: If likelihood of a false positive is <5% results are "statistically significant"

John Ioannidis (2005) "*most published research findings are probably false*"

Assumption is invalid because it ignores:

- statistical power
  (ability to detect false negatives –
  accepted value 0.8, estimated ~0.35)
- unlikeliness of the hypothesis being tested
- bias favoring publication of something new
  (reporting of positive, not negative results)



http://www.economist.com/blogs/graphicdetail/2013/10/daily-chart-2

Ioannidis, J. (2005) Why most published research findings are false. *PLoS*. 2: 696-701.

Unreliability of Research Reaches the Popular Press (2013)

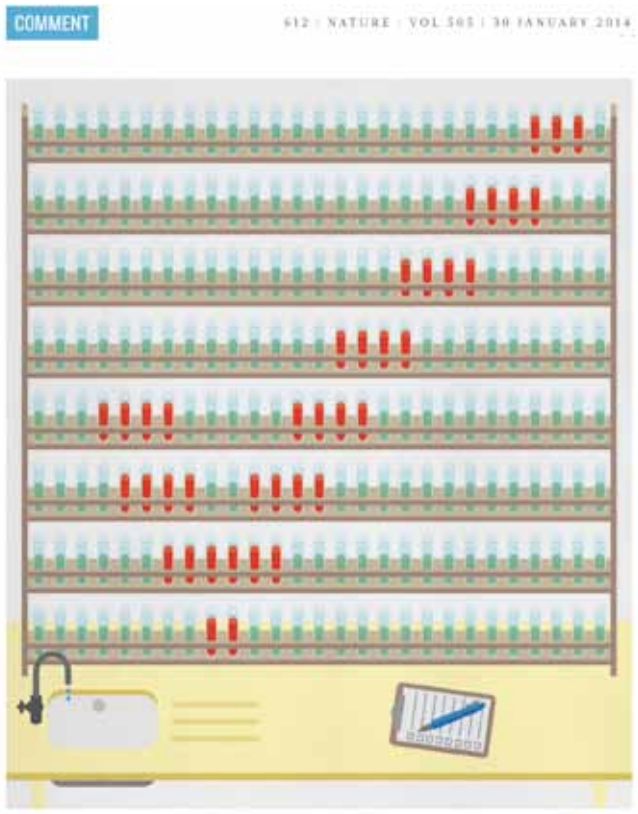
Bruce Alberts – Testimony to Congress, March 2013

- Journals should enforce standards
- Trainees should be taught technical skills and statistics
- Researchers should be judged on quality not quantity of publications
- Funding agencies should encourage replications (and lower barriers to reporting inconsistencies)
- Failures to reproduce results should be attached to the original publication

Culture Change:

"need to develop a value system where simply moving on from one's mistakes without publicly acknowledging them severely damages, rather than protects, a scientific reputation."

NIH Leadership Responds (2014)



Collins, F.S. & Tabak, L.A. (2014) Policy: NIH plans to enhance reproducibility. *Nature* 505: 612-613.

NIH Leadership Responds (2014)

- Acknowledged problem
  - The issue is greater for preclinical than clinical research (highly regulated)
- Proposed steps that NIH will take immediately
  - Develop required training module incorporated into ethical conduct (intramural)
  - Pilot checklist for systematic evaluation of grant proposals
  - Develop Data Discovery Index to access unpublished primary data
- Emphasized need for community engagement
  - Scientific publishers (methods, primary data, statistical analyses during review)
  - University tenure and promotion committees should emphasize quality over quantity
  - NIH Biosketch reformatted to emphasize scientific contributions, not # of papers

NIH Website: Resources for Rigor and Reproducibility

U.S. Department of Health & Human Services

NIH National Institutes of Health
Turning Discovery Into Health

Search NIH

NIH Employee Intranet | Staff Directory | En Español

Health Information | Grants & Funding | News & Events | Research & Training | Institutes at NIH | About NIH

Home » Research & Training

RIGOR AND REPRODUCIBILITY

Rigor and Reproducibility

- Principles and Guidelines
- Expanded Guidelines
- Application Instructions
- Training
- Funding Opportunities
- Meetings and Workshops
- Publications

Updated Application Instructions to Enhance Rigor and Reproducibility

Two of the cornerstones of science advancement are rigor in designing and performing scientific research and the ability to reproduce biomedical research findings. The application of rigor ensures robust and unbiased experimental design, methodology, analysis, interpretation, and reporting of results. When a result can be reproduced by multiple scientists, it validates the original results and readiness to progress to the next phase of research. This is especially important for clinical trials in humans, which are built on studies that have demonstrated a particular effect or outcome.

Johns Hopkins University students in a laboratory. *Johns Hopkins University*

In recent years, however, there has been a growing awareness of the need for rigorously designed published preclinical studies, to ensure that such studies can be reproduced. This webpage provides information about the efforts underway by NIH to enhance rigor and reproducibility in scientific research.

Email Updates

Sign up to receive email updates about rigor and reproducibility.

Sign up for updates

Related Links

Letter from Dr. Stephen I. Katz: An Update on the NIH Initiative to Enhance Research Rigor and Reproducibility

Contact Us

Please send email to NIHReprodEfforts@od.nih.gov.

Scientific Societies Are Developing Standards and Training for Their Own Disciplines

Society for Neuroscience
- Training Module Videos (on NIH website)

American Society for Cell Biology
- Report on Reproducibility

FASEB
- Overarching Recommendations
- Recommendations Specific to Research Using Antibodies
- Recommendations Specific to Research Using Mouse and Other Animal Models

NIH Introduces New Requirements for Grant Applications (2016)

"NIH's Rigor and Transparency efforts are intended to clarify expectations and highlight attention to four areas that may need more explicit attention by applicants and reviewers:"

- Scientific premise
- Scientific rigor
- Consideration of relevant biological variables, such as sex
- Authentication of key biological and/or chemical resources

Review Criteria for Rigor and Transparency of Research (R01 Grants)

	Applies to which applications?	Where is it included in the application?	Addition to review criteria	Affect overall impact score?
Scientific Premise	All	Research Strategy (Significance)	Is there a strong scientific premise for the project?	Yes (Significance)
Scientific Rigor	All	Research Strategy (Approach)	Are there strategies to ensure a robust and unbiased approach?	Yes (Approach)
Consideration of Relevant Biological Variables, Such as Sex	Projects with vertebrate animals and/or human subjects	Research Strategy (Approach)	Are adequate plans to address relevant biological variables, such as sex, included for studies in vertebrate animals or human subjects?	Yes (Approach)
Authentication of Key Biological and/ or Chemical Resources	Project involving key biological and/or chemical resources	New Attachment	Comment on plans for identifying and ensuring validity of resources.	No

Scientific Premise

GOAL: Ensure that the underlying **scientific foundation** of the project (concepts, previous work, and data, when relevant) is sound.

- Pertains to the **underlying evidence/data** for the project
- Address under Significance (R applications)
- Addition to the review criteria: "*Is there a strong scientific premise*?"
- Specifically, has the applicant:
  - Provided sufficient justification for the proposed work?
  - Cited appropriate work and/or preliminary data?
  - Appropriately identified strengths and weaknesses in prior work in the field?
  - Proposed to fill a significant gap in the field?
  - OR has the applicant explained why this is not possible?

Scientific Rigor

GOAL: Ensure a strict application of scientific method that supports robust and unbiased design, analysis, interpretation, and reporting of results, and sufficient information for the study to be assessed and reproduced. Give careful consideration to the methods and issues that matter in your field.

- Pertains to the **proposed research**
- Address under **Approach** (R applications)
- Addition to review criteria: Are there *"strategies to ensure a robust and unbiased approach, as appropriate for the work proposed?"*
- Possible considerations, if appropriate for the scientific field and research question, include plans for:
  - determining group sizes
  - analyzing anticipated results
  - reducing bias
  - ensuring independent and blinded measurements
  - improving precision and reducing variability
  - including or excluding research subjects
  - managing missing data

Relevant Biological Variables

GOAL: Ensure that the research accounts for sex and other relevant biological variables in developing research questions and study designs. The ways in which sex and other biological variables need to be accounted for will differ across research questions and fields of study.

- Pertains to the **proposed research (**vertebrate animals, human subjects)
- Address in **Approach** (R applications)
- Addition to review criteria: Are there "*adequate plans to address relevant biological variables for studies in vertebrate animals or human subjects?*"
- Consideration of sex is required in all studies involving human subjects or vertebrate animals.
- Specific considerations to assess include:
  - Applies broadly to all biological variables relevant to the research such as sex, age, source, weight, or genetic strain.
  - Has the applicant considered biological variables, such as sex, that are relevant to the experimental design?
  - Will relevant biological variables be controlled or factored into the study design appropriately?

Plan for Resource Authentication

GOAL: Ensure processes are in place to identify and regularly validate key resources used in their research and avoid unreliable research as a result of misidentified or contaminated resources.

- Researchers are expected to authenticate key biological and/or chemical resources used in their research, to ensure that the resources are genuine.
- New additional review consideration: "*Authentication of Key Biological and/or Chemical Resources: For projects involving key biological and/or chemical resources, reviewers will comment on the brief plans proposed for identifying and ensuring the validity of those resources.*"
- Does not affect criterion scores or overall impact score (rated as acceptable or unacceptable)

Will these Steps Enhance Reproducibility?

Will these Steps Enhance Reproducibility?

Awareness of the problem is an important first step.

- Reproducibility is being studied, and research published

New initiatives raise awareness

- Training
- Journal standards
- NIH grant application criteria

However, a major cultural change may be required for sustained impact.

Pressure to Publish Selects for Irreproducibility

ROYAL SOCIETY OPEN SCIENCE

rsos.royalsocietypublishing.org

Research

The natural selection of bad science

Paul E. Smaldino[1] and Richard McElreath[2]

[1]Cognitive and Information Sciences, University of California, Merced, CA 95343, USA

[2]Department of Human Behavior, Ecology, and Culture, Max Planck Institute for Evolutionary Anthropology, Leipzig, Germany

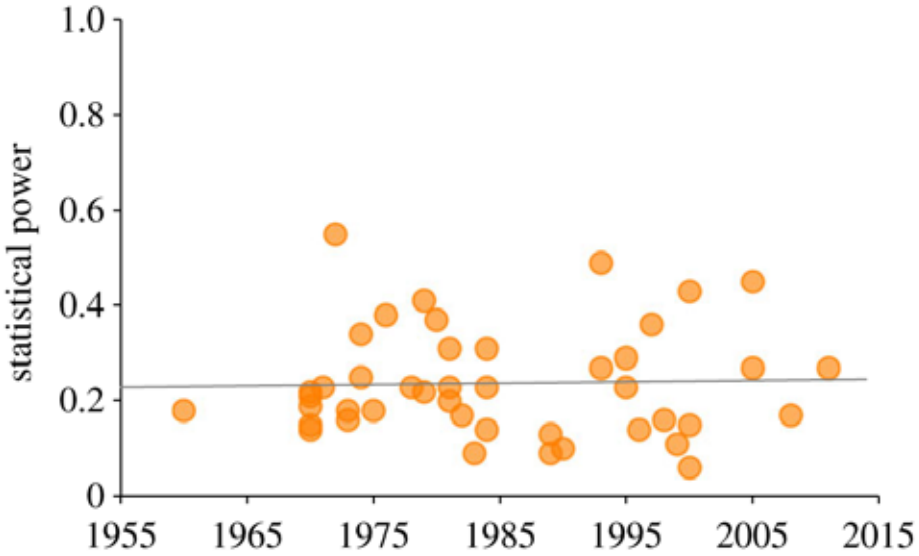
"Poor research design and data analysis encourage false-positive findings. Such poor methods persist despite perennial calls for improvement, suggesting that they result from something more than just misunderstanding.

The persistence of poor methods results partly from incentives that favour them, leading to the natural selection of bad science."

Smaldino, P.E. & McElreath, R. (2016) *R. Soc. Open Sci.* 2016 3: 160384
Published 21 September 2016

Poor Methods Persist Despite Calls for Improvement

Statistical power is low and has not improved for 50 years.



Average statistical power from 44 reviews of papers published in social and behavioral science journals

Smaldino, P.E. & McElreath, R. (2016) *R. Soc. Open Sci.* 2016 3: 160384
Published 21 September 2016

Institutional Incentives for Scientific Researchers

- Increases in publication rate
  - average pubs of newly hired biologist: 22 in 2013, 12.5 in 2005
- Pressure to portray work as groundbreaking
  - 25-fold increase in “innovative”, “groundbreaking” and “novel” in PubMed abstracts
- Overuse of h-index

- Researchers are rewarded for publications
- Positive results are easier to publish and more prestigious than negative result

Researchers who can obtain more positive results (whatever their truth) will have an advantage

Smaldino, P.E. & McElreath, R. (2016) *R. Soc. Open Sci.* 2016 3: 160384
Published 21 September 2016

Less Rigorous Methods Produce More Positive Results

- Methods that generate false positives:
  - generate output at higher rates (less replication)
  - are more likely to generate publishable results

- Low penalties for publication of false positives:
  - false discoveries are rarely detected
    (e.g., <1% of psychology research is replicated)
  - discredited research is frequently cited

Smaldino, P.E. & McElreath, R. (2016) *R. Soc. Open Sci.* 2016 3: 160384
Published 21 September 2016

An Evolutionary Model of Science

- Each lab has a characteristic *power*, the ability to positively identify a true association.
- Increasing power also increases the rate of false positives, unless *effort* is exerted.
- Increasing effort decreases the productivity of a lab, because it takes longer to perform rigorous research.
- Labs receive “pay-offs” for publishing their research (prestige, funding, etc.; can be positive or negative)
- Labs die randomly and those with higher pay-offs reproduce. Labs inherit attributes of their parent lab (but with mutation probabilities).

Smaldino, P.E. & McElreath, R. (2016) *R. Soc. Open Sci.* 2016 3: 160384
Published 21 September 2016

An Evolutionary Model of Science

Table 1. Global model parameters.

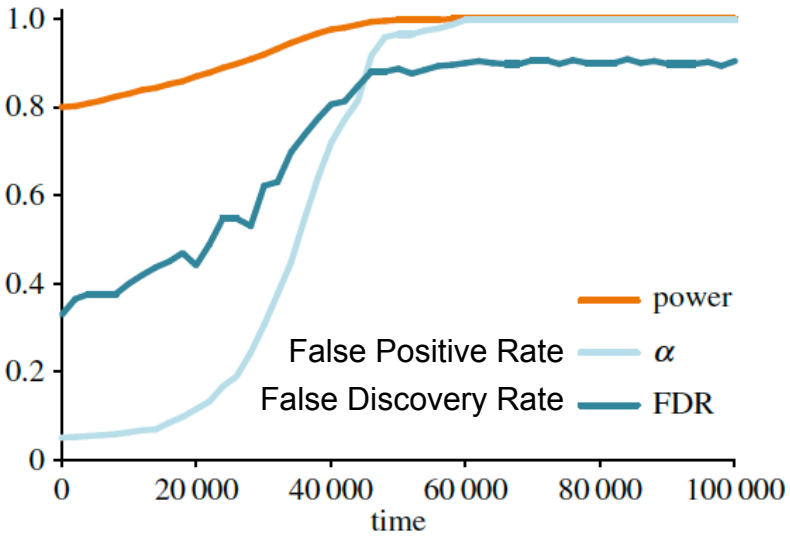
parameter	definition	values tested
N	number of labs	100
b	base rate of true hypotheses	0.1
r_0	initial replication rate for all labs	{0, 0.01, 0.2, 0.5}
e_0	initial effort for all labs	75
w_0	initial power for all labs	0.8
η	influence of effort on productivity	0.2
c_{R+}	probability of publishing positive replication	1
c_{R-}	probability of publishing negative replication	1
V_N	pay-off for publishing novel result	1
V_{R+}	pay-off for publishing positive replication	0.5
V_{R-}	pay-off for publishing negative replication	0.5
V_{O+}	pay-off for having novel result replicated	0.1
V_{O-}	pay-off for having novel result fail to replicate	−100
d	number of labs sampled for death and birth events	10
μ_r	probability of r mutation	{0, 0.01}
μ_e	probability of e mutation	{0, 0.01}
μ_w	probability of w mutation	{0, 0.01}
σ_r	standard deviation of r mutation magnitude	0.01
σ_e	standard deviation of e mutation magnitude	1
σ_w	standard deviation of w mutation magnitude	0.01

Smaldino, P.E. & McElreath, R. (2016) *R. Soc. Open Sci.* 2016 3: 160384
Published 21 September 2016

An Evolutionary Model of Science

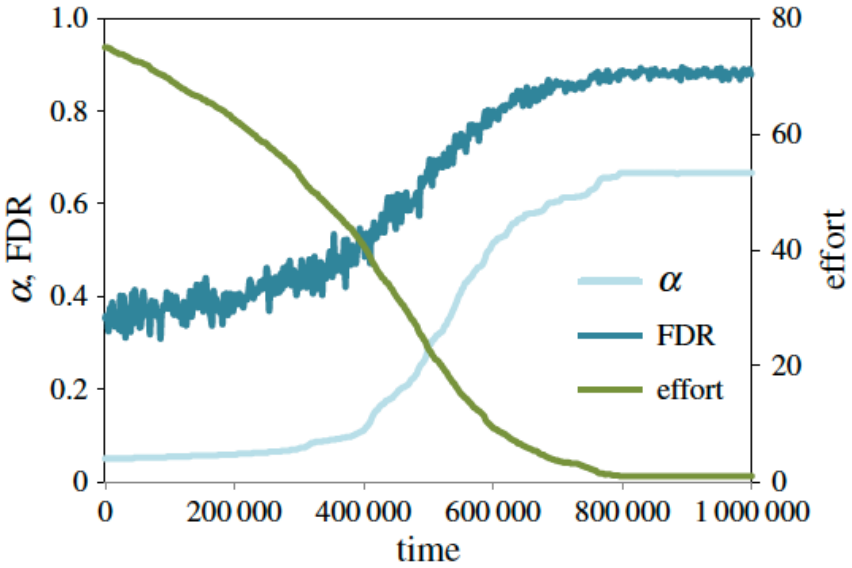
Simulation with Constant Effort

As power increases,
the rate of false positives increases.



Effort Evolves

As effort decreases,
the rate of false positives increases.



Smaldino, P.E. & McElreath, R. (2016) *R. Soc. Open Sci.* 2016 3: 160384
Published 21 September 2016

A Cultural Change is Required to Promote Reproducibility

Science is a cultural activity and can change through evolutionary processes.

Incentives drive cultural evolution.

"Some of the most powerful incentives in contemporary science actively encourage, reward and propagate poor research methods and abuse of statistical procedures."

Smaldino, P.E. & McElreath, R. (2016) *R. Soc. Open Sci.* 2016 3: 160384
Published 21 September 2016