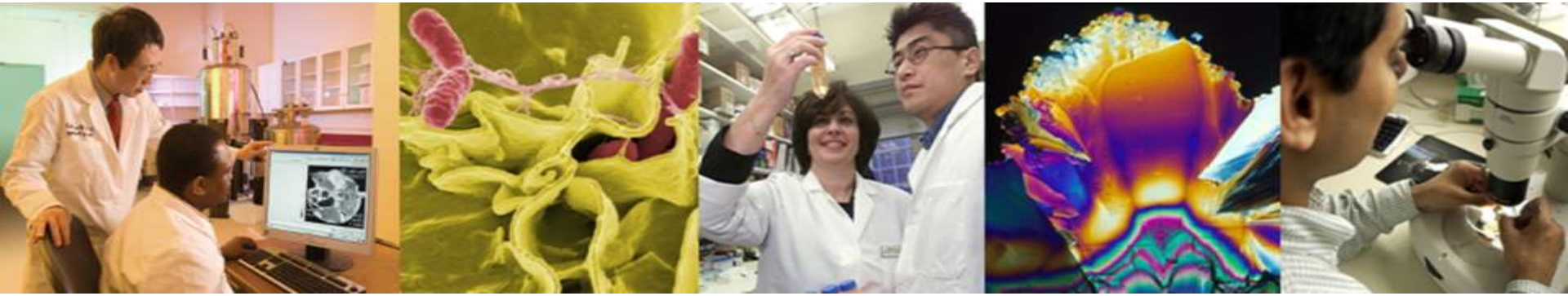
On the Reproducibility of Biomedical Research

Sixth Annual Lecture on Biomedical Ethics

April 15, 2016



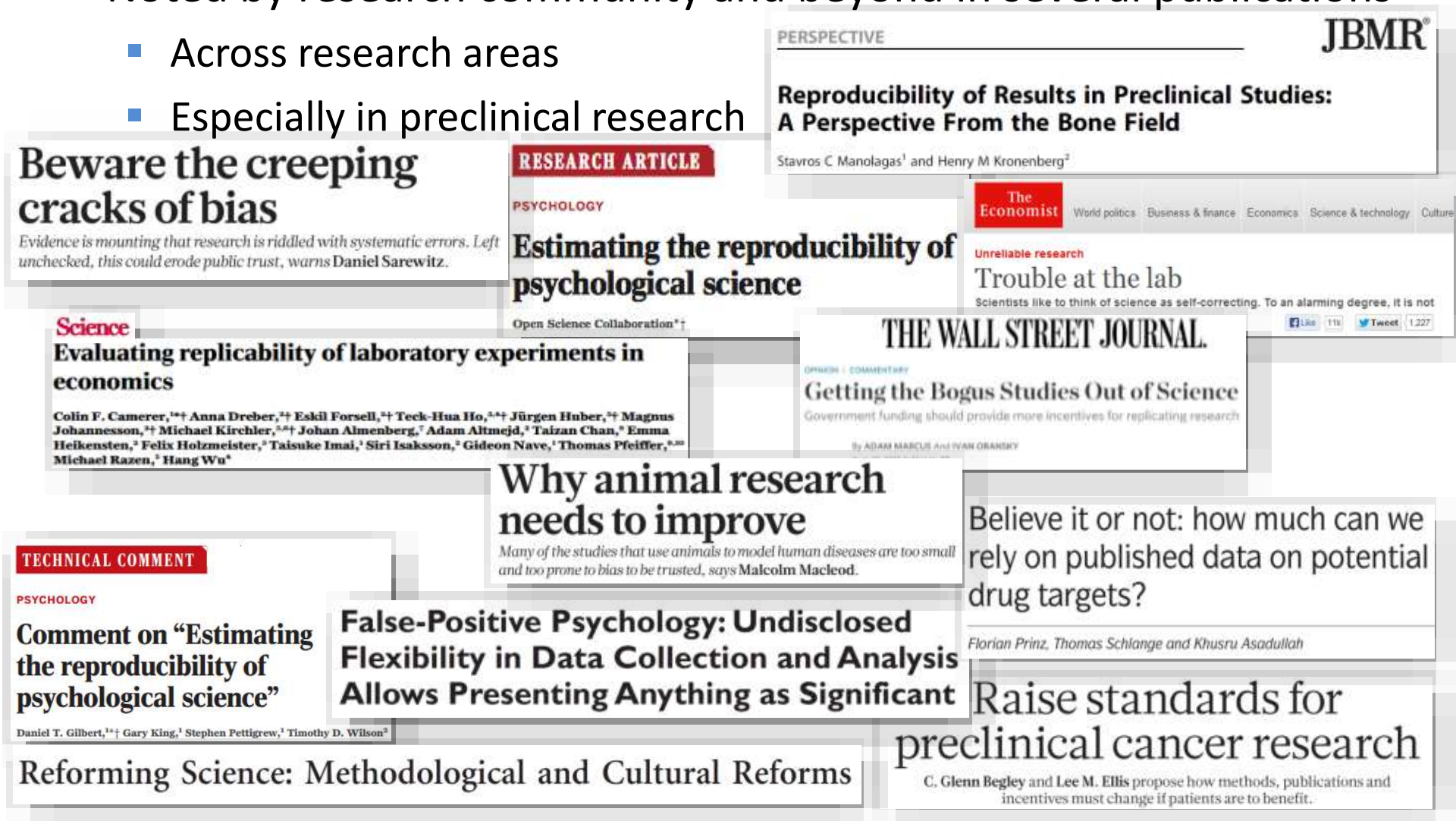
Lawrence A. Tabak, DDS, PhD
Principal Deputy Director, NIH
Department of Health and Human Services

Reproducibility

No, not that kind of reproducibility…

The Growing Challenge

- Noted by research community and beyond in several publications
  - Across research areas
  - Especially in preclinical research



Science is “self-correcting”

“In experimental philosophy we are to look upon propositions inferred by general induction from phenomena as accurately or very nearly true…till such time as other phenomena occur, by which they may either be made more accurate, or liable to exception.”

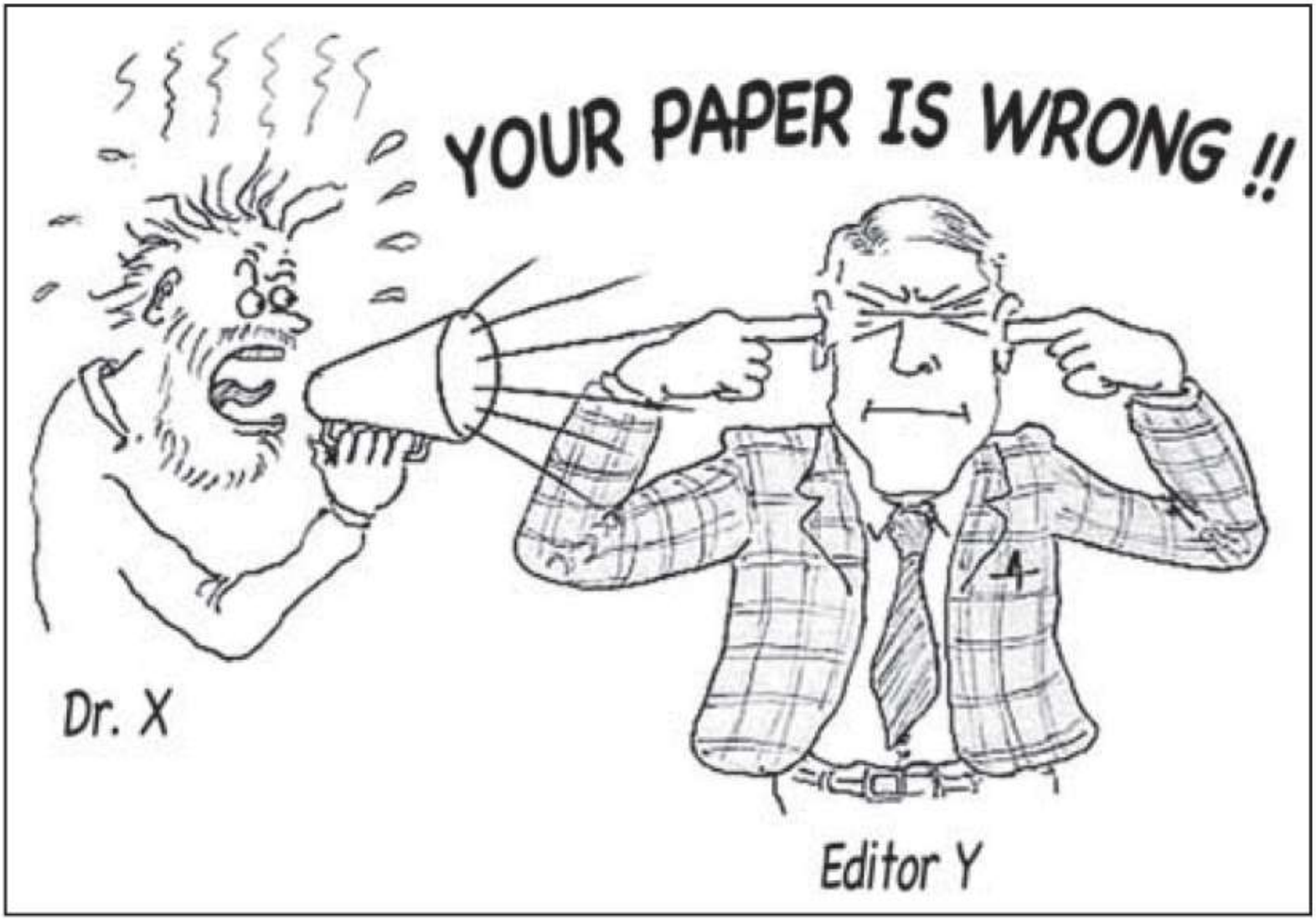
Isaac Newton, *Mathematical Principles of Natural Philosophy*

Science is “self-correcting”

“the really valuable part of the Fourth Rule is that which implies that a *constant verification*, and, if necessary, rectification, of truths discovered by induction, should go on in the scientific world. Even when the law is, or appears to be, most certainly exact and universal, it should be constantly exhibited to us afresh in the form of experience and observation.”

William Whewell, *On the Philosophy of Discovery*

Science is “self-correcting”



So what has gone awry?

Challenges to Ensuring Rigor and Transparency in Reporting Science: Underlying Issues

Publish or perish!

Need for grant support

Misuse of impact factors

Your work must be "novel"

Misaligned Incentives

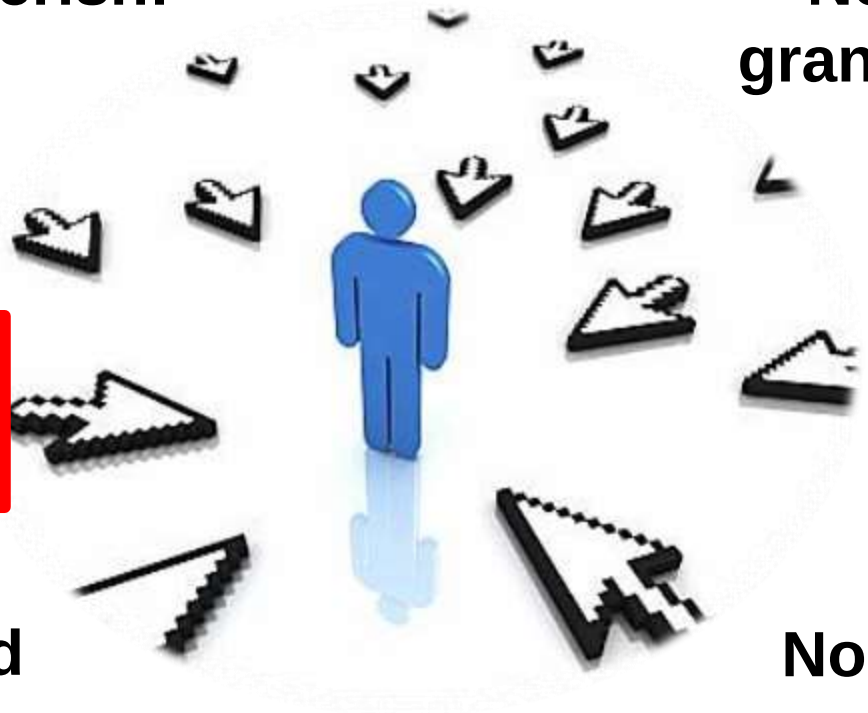
No negative data

Poor training

Challenges to Ensuring Rigor and Transparency in Reporting Science: Underlying Issues

Publish or perish!

Need for grant support



Misuse of impact factors

Your work must be “novel”

Misaligned Incentives

No negative data

Poor training

Challenges to Ensuring Rigor and Transparency in Reporting Science: Underlying Issues

Publish or perish!

Need for grant support

Misuse of impact factors

Your work must be "novel"

No negative data

Misaligned Incentives

Poor training

Challenges to Ensuring Rigor and Transparency in Reporting Science: Underlying Issues

Publish or perish!

Need for grant support

Misuse of impact factors

Your work must be "novel"

Misaligned Incentives

No negative data

Poor training

Challenges to Ensuring Rigor and Transparency in Reporting Science: Underlying Issues

Publish or perish!

Need for grant support

Misuse of impact factors

Your work must be "novel"

Misaligned Incentives

No negative data

Poor training

Deficiencies in Experimental Procedures

- Insufficient Reporting in publications – blinding, replication & randomization, sample size outliers and exclusion criteria

Insufficient Reporting of Methodological Approaches is Evident for Pre-Clinical Studies

Table 3. Prevelence of selected quality characteristics in other experimental models

	Number of publications	Randomisation (%)	Blinded assessment of outcome (%)	Sample-size calculation (%)
Transgenic stroke studies	157	n/a	3	0
Stroke pathophysiology studies	166	5	18	0
Parkinson's disease	118	12	15	0
Multiple sclerosis	183	2	11	0

Trends Neurosci 2007; 30: 433-439

Adapted from Dr. S. Silberberg, NINDS

Amyotrophic Lateral Sclerosis

ISSN: 1748-2968 (Print) 1471-180X (Online) Journal homepage: http://www.tandfonline.com/loi/iafd19

Design, power, and interpretation of studies in the standard murine model of ALS

Sean Scott, Janice E. Kranz, Jeff Cole, John M. Lincecum, Kenneth Thompson, Nancy Kelly, Alan Bostrom, Jill Theodoss, Bashar M. Al-Nakhala, Fernando G. Vieira, Jeyanthi Ramasubbu & James A. Heywood

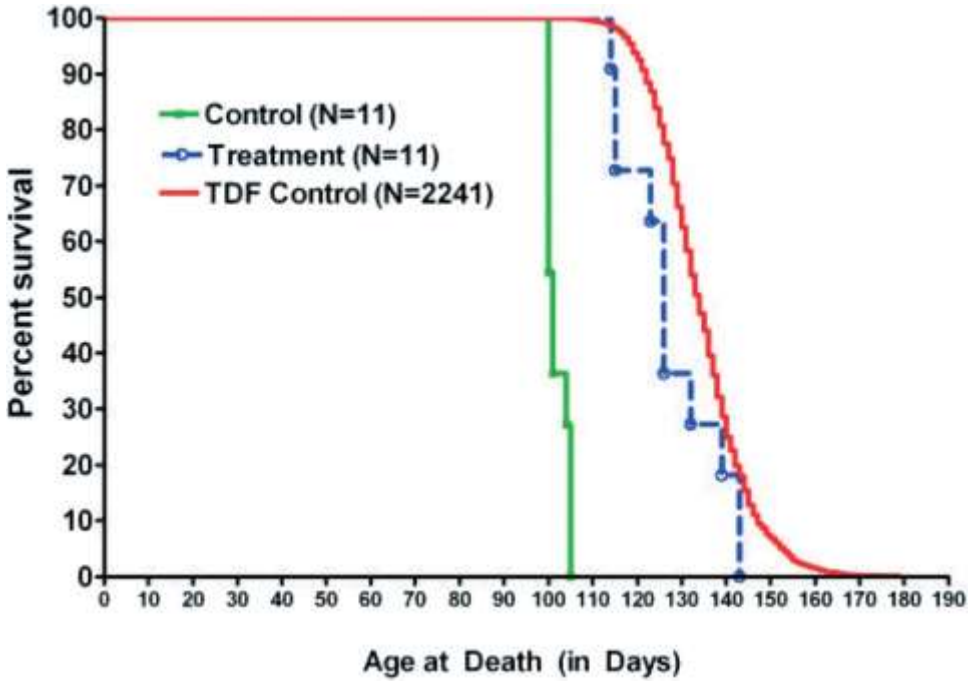


Figure 4. Survival analysis. Control and treated SOD1^{G93A} mice from one publication compared to all of our 2241 control animals (acquired over four years – data from Table S2) that died of ALS.

Amyotroph Lateral Scler 2008; 9: 4-15

Deficiencies in Experimental Procedures (cont.)

- Insufficient Reporting in publications – blinding, replication & randomization, sample size outliers and exclusion criteria
- "P-Hacking"

1521-0103/351/1/200–205$25.00
THE JOURNAL OF PHARMACOLOGY AND EXPERIMENTAL THERAPEUTICS
http://dx.doi.org/10.1124/jpet.114.219170
J Pharmacol Exp Ther 351:200–205, October 2014

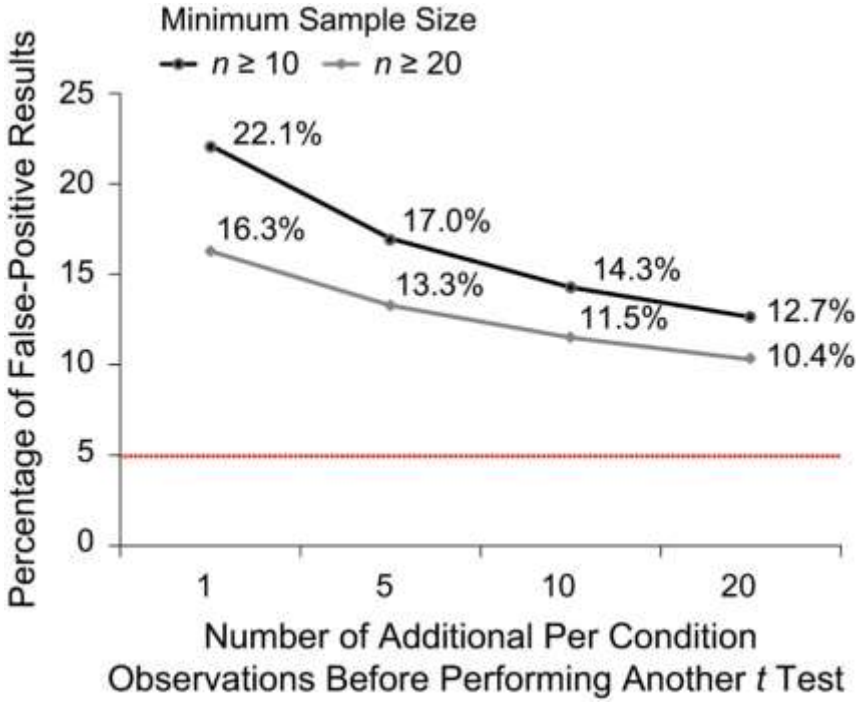
Commentary

Common Misconceptions about Data Analysis and Statistics

Harvey J. Motulsky
GraphPad Software Inc., La Jolla, California
Received August 8, 2014; accepted August 8, 2014

1) P-hacking, which is when you reanalyze a data set in many different ways, or perhaps reanalyze with additional replicates, until you get the results you want; 2) overemphasis on *P* values rather than on the actual size of the observed effect; 3) overuse of statistical hypothesis testing, and being seduced by the word "significant"; and 4) over-reliance on standard errors, which are often misunderstood.

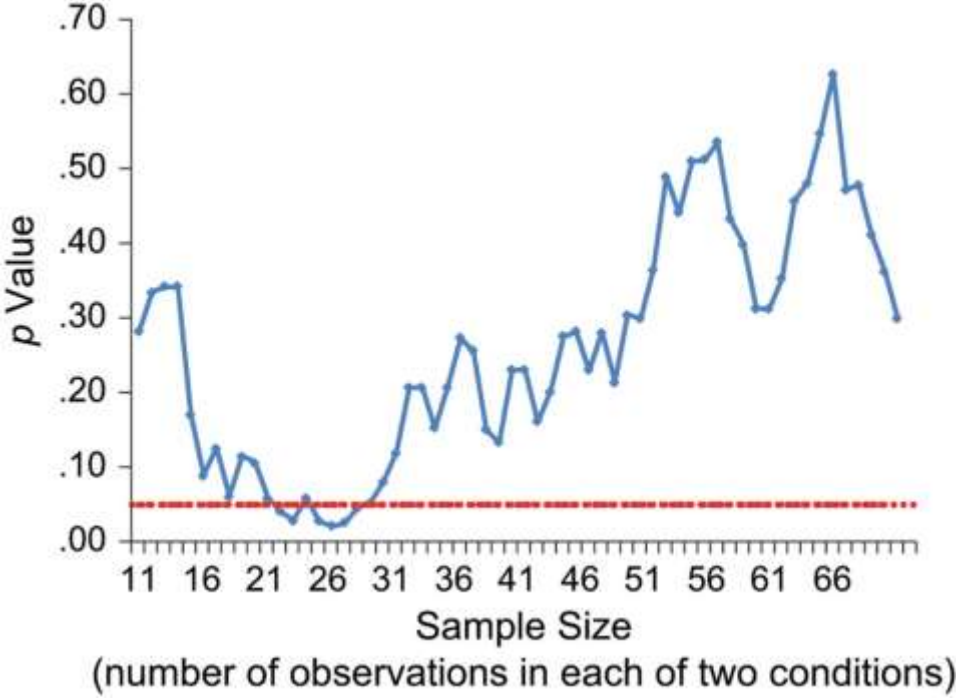
Deficiencies in Experimental Procedures (cont.)
Researcher's "Degrees of Freedom"



Likelihood of obtaining a false-positive result when data collection ends upon obtaining significance ($p \leq .05$, highlighted by the dotted line).

Simmons et al. *Psychological Science* 2011;22:1359-1366

Deficiencies in Experimental Procedures (cont.)
Researcher's "Degrees of Freedom"



Simulation of p values obtained by a researcher who continuously adds an observation to each of two conditions, conducting a t test after each addition

More isn't always better!

Simmons et al. *Psychological Science* 2011;22:1359-1366

Deficiencies in Experimental Procedures (cont.)
Researcher's "Degrees of Freedom"

Table 2. Simple Solution to the Problem of False-Positive Publications

Requirements for authors

1. Authors must decide the rule for terminating data collection before data collection begins and report this rule in the article.
2. Authors must collect at least 20 observations per cell or else provide a compelling cost-of-data-collection justification.
3. Authors must list all variables collected in a study.
4. Authors must report all experimental conditions, including failed manipulations.
5. If observations are eliminated, authors must also report what the statistical results are if those observations are included.
6. If an analysis includes a covariate, authors must report the statistical results of the analysis without the covariate.

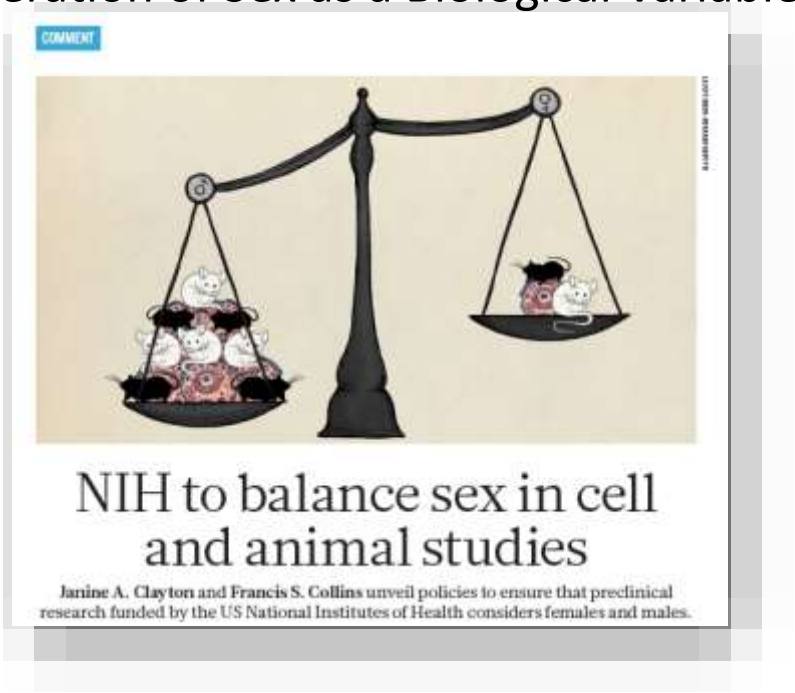
Guidelines for reviewers

1. Reviewers should ensure that authors follow the requirements.
2. Reviewers should be more tolerant of imperfections in results.
3. Reviewers should require authors to demonstrate that their results do not hinge on arbitrary analytic decisions.
4. If justifications of data collection or analysis are not compelling, reviewers should require the authors to conduct an exact replication.

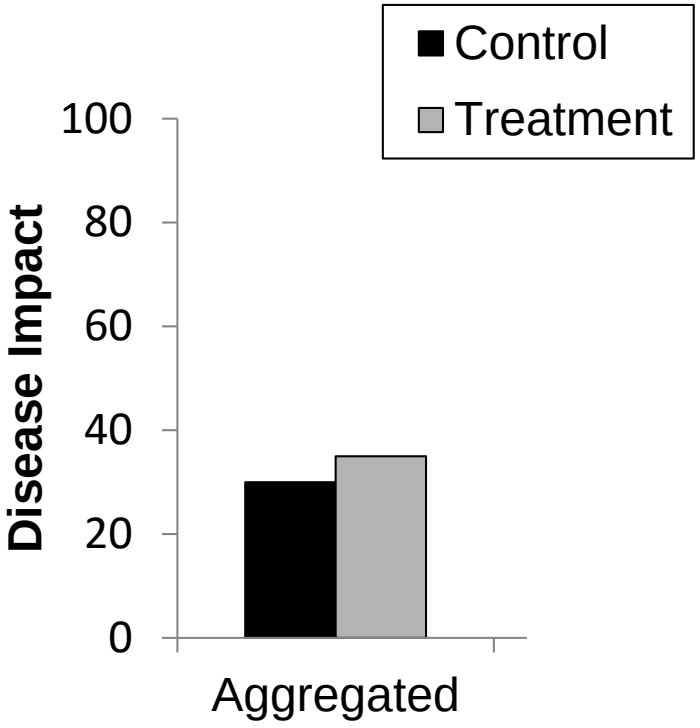
Simmons et al. *Psychological Science* 2011;22:1359-1366

Deficiencies in Experimental Procedures (cont.)

- Insufficient Reporting in publications – blinding, replication & randomization, sample size outliers and exclusion criteria
- "P-Hacking"
- Researcher's "Degrees of Freedom"
- Lack of Consideration of Sex as a Biological Variable



Biological/Disease Impact of Experimental Design



Real Life

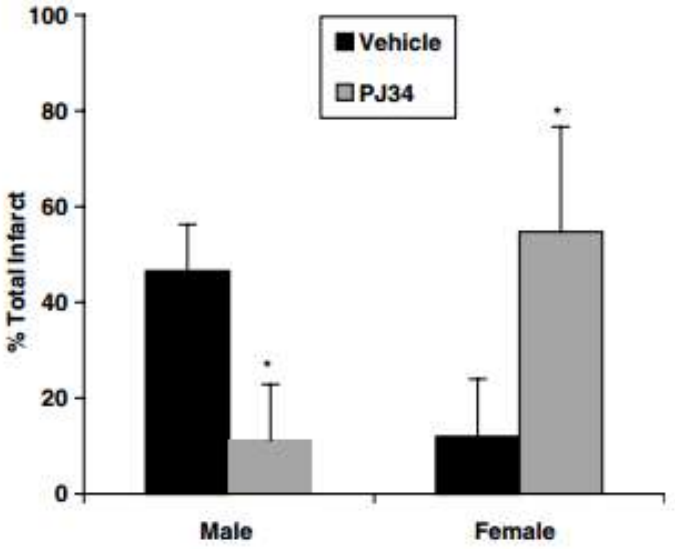
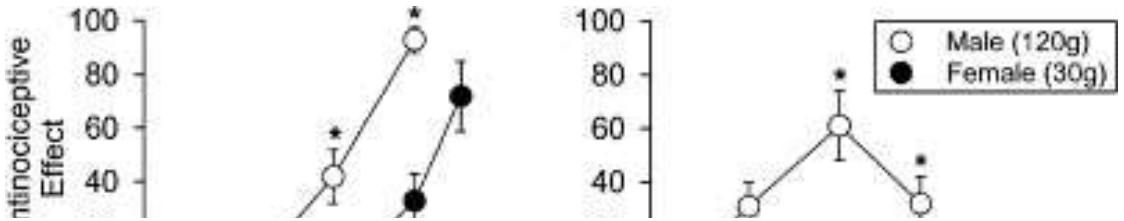


Figure 9 The effects of the selective poly-ADP ribose polymerase (PARP-1) inhibitor PJ-34 in wild-type (WT) mice of both genders. Treatment with PJ-34 at ischemic onset reduced total infarction in male mice compared with saline-treated controls (*$P<0.001$). A significant increase in ischemic damage was seen in PJ-34-treated females compared with control (*$P<0.001$).

McCullough LD, et al. *Journal of Cerebral Blood Flow & Metabolism* (2005) **25,** 502–512.

Importance of Sex as a Biological Variable



1521-0081/68/2/242–263$25.00
PHARMACOLOGICAL REVIEWS

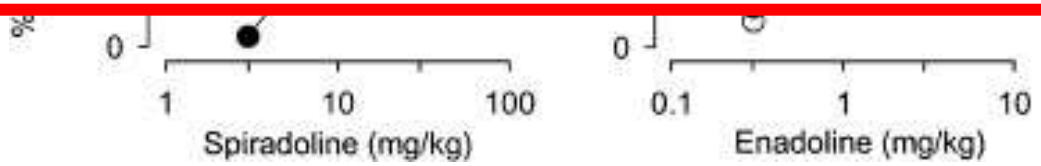
http://dx.doi.org/10.1124/pr.115.011163
Pharmacol Rev 68:242–263, April 2016

ASSOCIATE EDITOR: MICHAEL M. GOTTESMAN

Sex Differences in Animal Models: Focus on Addiction

Jill B. Becker[1] and George F. Koob[1]

Molecular & Behavioral Neuroscience Institute, Department of Psychiatry, Department of Psychology, University of Michigan, Ann Arbor, Michigan (J.B.B.); and Director, National Institute on Alcohol Abuse and Alcoholism, National Institutes of Health, Bethesda, Maryland (G.F.K.)

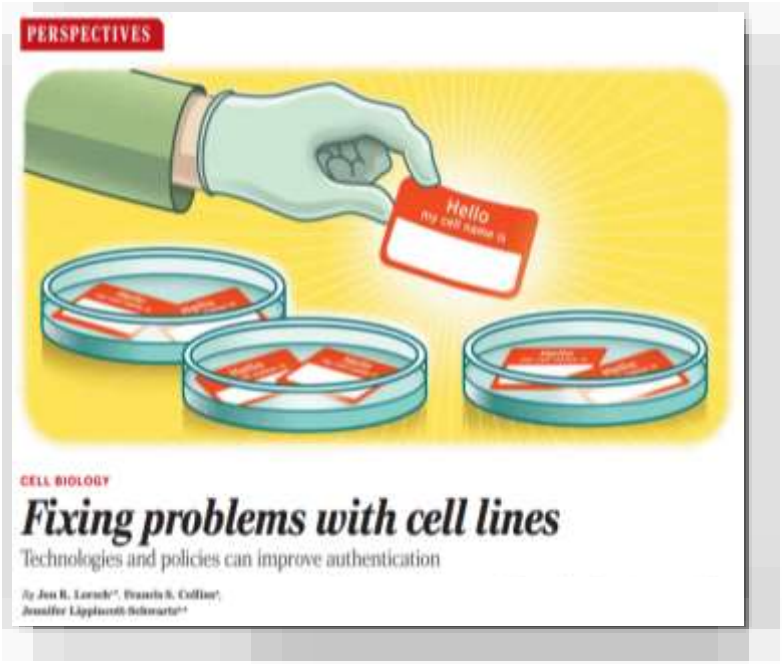


Morphine was 2.3-fold more potent in males and buprenorphine produced a 61% effect in males and only a 5% effect in females

Barrett AC, et al. *European Journal of Pharmacology* (2002) **452,** 163–173.

Deficiencies in Experimental Procedures (cont.)

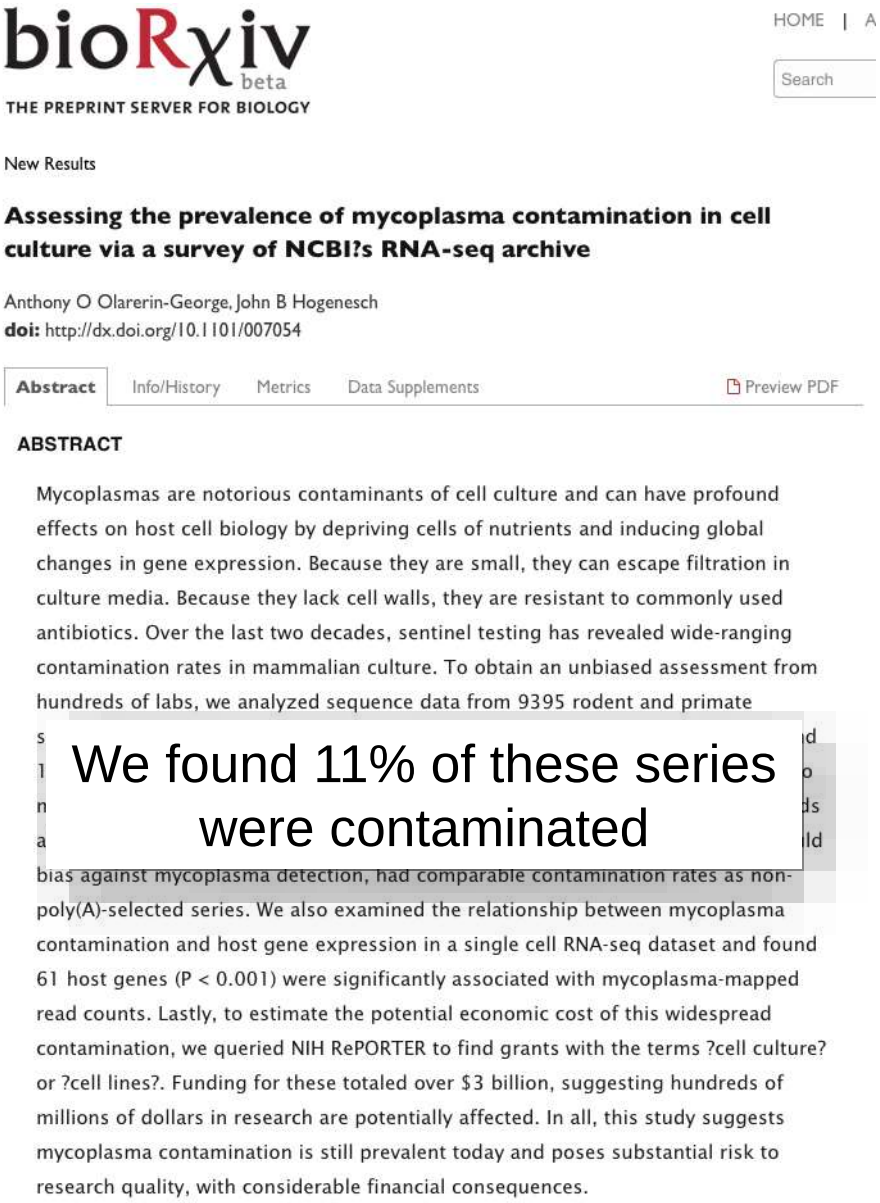
- Insufficient Reporting in publications – blinding, replication & randomization, sample size outliers and exclusion criteria
- "P-Hacking"
- Lack of Consideration of Sex as a Biological Variable
- Problems with Authentication of Cell Lines



Reproducibility in Cell Culture Studies

- >400 misidentified cell lines have been cataloged, dating back to the 1960s
- ~70% of researchers surveyed in 2004 had never checked the identity of their cell lines
- Major repositories report that 14-30% of cell lines submitted are contaminated
- In a 2013 survey <50% of cell lines had an unambiguous identifier and source in publications
- Standards for cell line authentication and affordable methods for cell authentication now available

Reproducibility in Cell Culture Studies



bioRxiv beta
THE PREPRINT SERVER FOR BIOLOGY

HOME | A

Search

New Results

Assessing the prevalence of mycoplasma contamination in cell culture via a survey of NCBI?s RNA-seq archive

Anthony O Olarerin-George, John B Hogenesch
doi: http://dx.doi.org/10.1101/007054

Abstract | Info/History | Metrics | Data Supplements | Preview PDF

ABSTRACT

Mycoplasmas are notorious contaminants of cell culture and can have profound effects on host cell biology by depriving cells of nutrients and inducing global changes in gene expression. Because they are small, they can escape filtration in culture media. Because they lack cell walls, they are resistant to commonly used antibiotics. Over the last two decades, sentinel testing has revealed wide-ranging contamination rates in mammalian culture. To obtain an unbiased assessment from hundreds of labs, we analyzed sequence data from 9395 rodent and primate

We found 11% of these series were contaminated

bias against mycoplasma detection, had comparable contamination rates as non-poly(A)-selected series. We also examined the relationship between mycoplasma contamination and host gene expression in a single cell RNA-seq dataset and found 61 host genes ($P < 0.001$) were significantly associated with mycoplasma-mapped read counts. Lastly, to estimate the potential economic cost of this widespread contamination, we queried NIH RePORTER to find grants with the terms ?cell culture? or ?cell lines?. Funding for these totaled over $3 billion, suggesting hundreds of millions of dollars in research are potentially affected. In all, this study suggests mycoplasma contamination is still prevalent today and poses substantial risk to research quality, with considerable financial consequences.

Importance of Cell Line Authentication

Primary tissue	Short tandem repeat locus								*TP53* mutation
	D21S11	TH01	D3S1358	FGA	TPOX	D8S1179	vWA	D5S818	
SK-GT-2	29, 32.2	8, 9	15, 17	25, 26	9, 12	13, 15	15, 18	10, 12	c.524G>A
SK-GT-5	28, 32.2	9	15, 17	21, 22	7, 8	10, 13	17	12	c.916C>T

Esophageal Adenocarcinoma cell line (EAC), SK-GT-5, is in fact the gastric fundus carcinoma cell line SK-GT-2!

- More than 100 scientific publications using SK-GT-5 or two other misidentified EAC cell lines have been identified
- Almost half of these reports were based solely on the use of cell lines not representative for EAC

Jurjen J. Boonstra et al. *JNCI* J Natl Cancer Inst 2010;102:271-274

Principles for Addressing Underlying Issues

- Raise community awareness
- Enhance formal training
- Protect the quality of funded and published research by adoption of more systematic review processes
- Share information/data
- Increase stability for investigators

Principles for Addressing Underlying Issues

- **Raise community awareness**
- Enhance formal training
- Protect the quality of funded and published research by adoption of more systematic review processes
- Share information/data
- Increase stability for investigators

Raise Community Awareness

NIH Rigor and Reproducibility Web-portal

http://www.nih.gov/science/reproducibility

Raise Community Awareness

- Workshop in Summer 2014 with PhRMA to identify areas of common interest with industry
- Workshop in Summer 2014 with Journal Editors to identify common opportunity areas
- **Over 135 journals** endorsed the principles, which were broadly shared in November 2014 through editorials and other notifications

JCB CellPress nature jbc THE JOURNAL OF BIOLOGICAL CHEMISTRY PLOS ONE Science Circulation The Journal of Neuroscience

http://www.nih.gov/science/reproducibility/principles-guidelines.htm

Raise Community Awareness

Efforts by Other Organizations: Recent Example

AMERICAN STATISTICAL ASSOCIATION RELEASES STATEMENT ON STATISTICAL SIGNIFICANCE AND *P*-VALUES

Provides Principles to Improve the Conduct and Interpretation of Quantitative Science

"We teach it because it's what we do; we do it because it's what we teach."

[http://amstat.tandfonline.com/doi/abs/10.1080/00031305.2016.1154108#.Vt2XlOaE2MN]. The ASA

"The *p*-value was never intended to be a substitute for scientific reasoning"

chosen techniques, properly conducted analyses, and correct interpretation.

http://amstat.tandfonline.com/doi/abs/10.1080/00031305.2016.1154108

Principles for Addressing Underlying Issues

- Raise community awareness
- **Enhance formal training**
- Protect the quality of funded and published research by adoption of more systematic review processes
- Share information/data
- Increase stability for investigators

Enhance Formal Training

- NINDS, IRP, and Office of the Director (OD) developed training modules in experimental design, which are being used within the IRP and are available publicly

- NIGMS (with 9 other ICs) is supporting the development of training modules to enhance reproducibility
  - Funded 6 awards, supported by 8 ICs
  - Planning to re-issue the RFA
- IRP workshops on data interpretation considerations for various experimental techniques – “potentials and pitfalls”

http://www.nih.gov/science/reproducibility/training.htm

Principles for Addressing Underlying Issues

- Raise community awareness
- Enhance formal training
- **Protect the quality of funded and published research by adoption of more systematic review processes**
- Share information/data
- Increase stability for investigators

Application and Review Processes

- The NIH Office of Extramural Research (OER) **clarified and revised application instructions and review criteria** to enhance reproducibility of research findings
- Enhancing reproducibility through rigor and transparency
  - Scientific premise of proposed research
  - Rigorous experimental design
  - Consideration of sex and other relevant biological variables
  - Authentication of key biological and/or chemical resources
- Considering sex as a biological variable in NIH-funded research
- Applies to application submitted **Jan. 25, 2016** and beyond

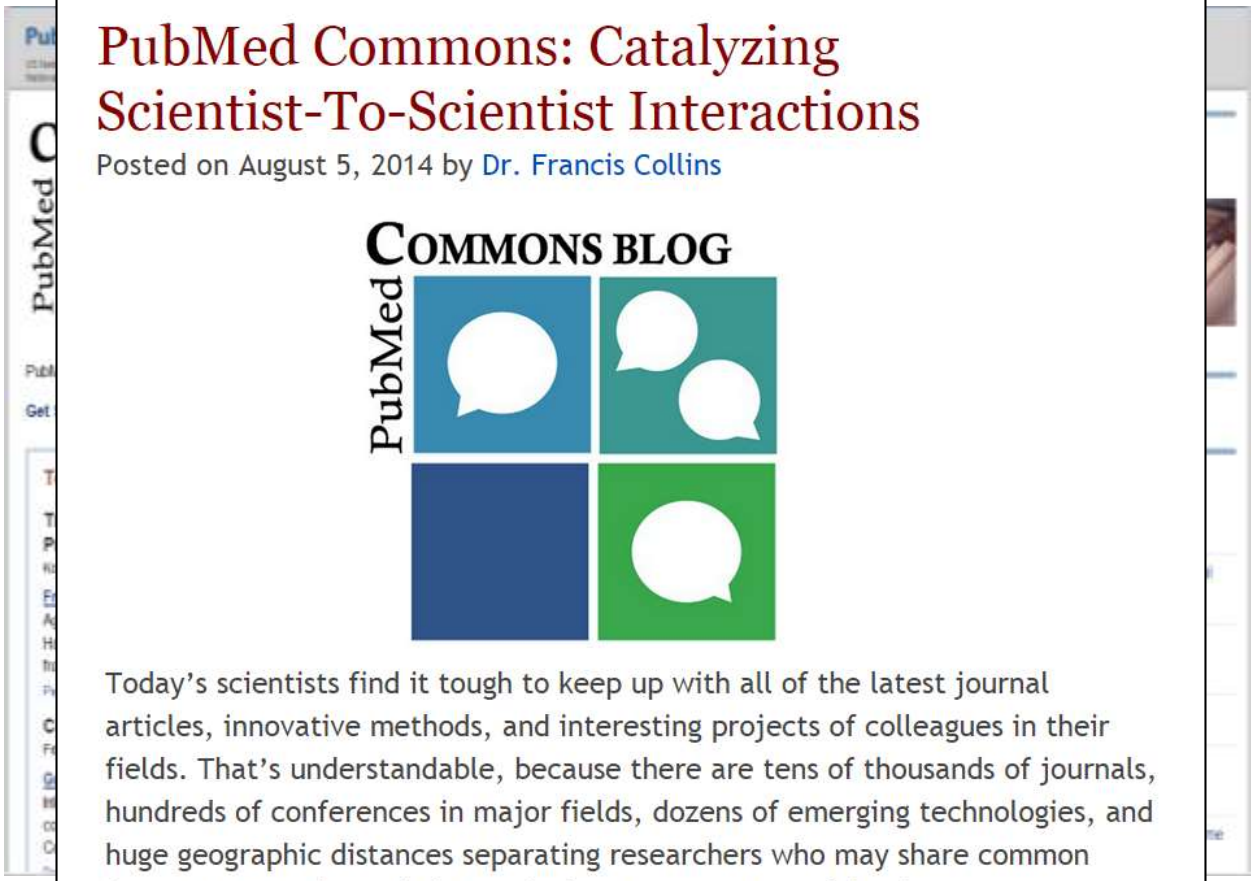
http://grants.nih.gov/reproducibility

Principles for Addressing Underlying Issues

- Raise community awareness
- Enhance formal training
- Protect the quality of funded and published research by adoption of more systematic review processes
- **Share information/data**
- Increase stability for investigators

Share Information/Data

PubM ers to
shar Med



NIH DIRECTOR'S BLOG

PubMed Commons: Catalyzing Scientist-To-Scientist Interactions

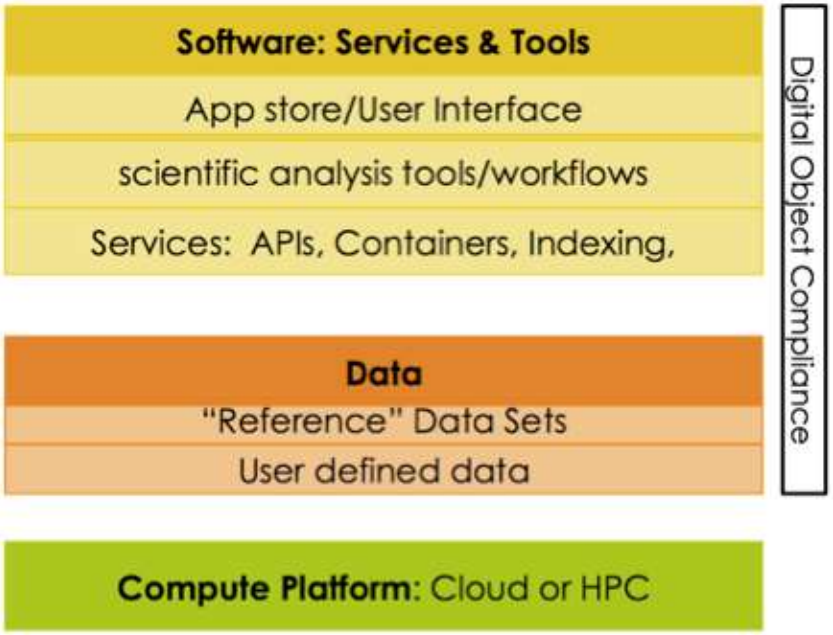
Posted on August 5, 2014 by Dr. Francis Collins

PubMed COMMONS BLOG

Today's scientists find it tough to keep up with all of the latest journal articles, innovative methods, and interesting projects of colleagues in their fields. That's understandable, because there are tens of thousands of journals, hundreds of conferences in major fields, dozens of emerging technologies, and huge geographic distances separating researchers who may share common interests. But science is increasingly a team sport—and it's important to provide scientists with as many avenues as possible through which to interact, including commenting on each other's work.

Share Information/Data

NIH Data Commons: **F**indable, **A**ccessible, **I**nteroperable and **R**eusable (FAIR)

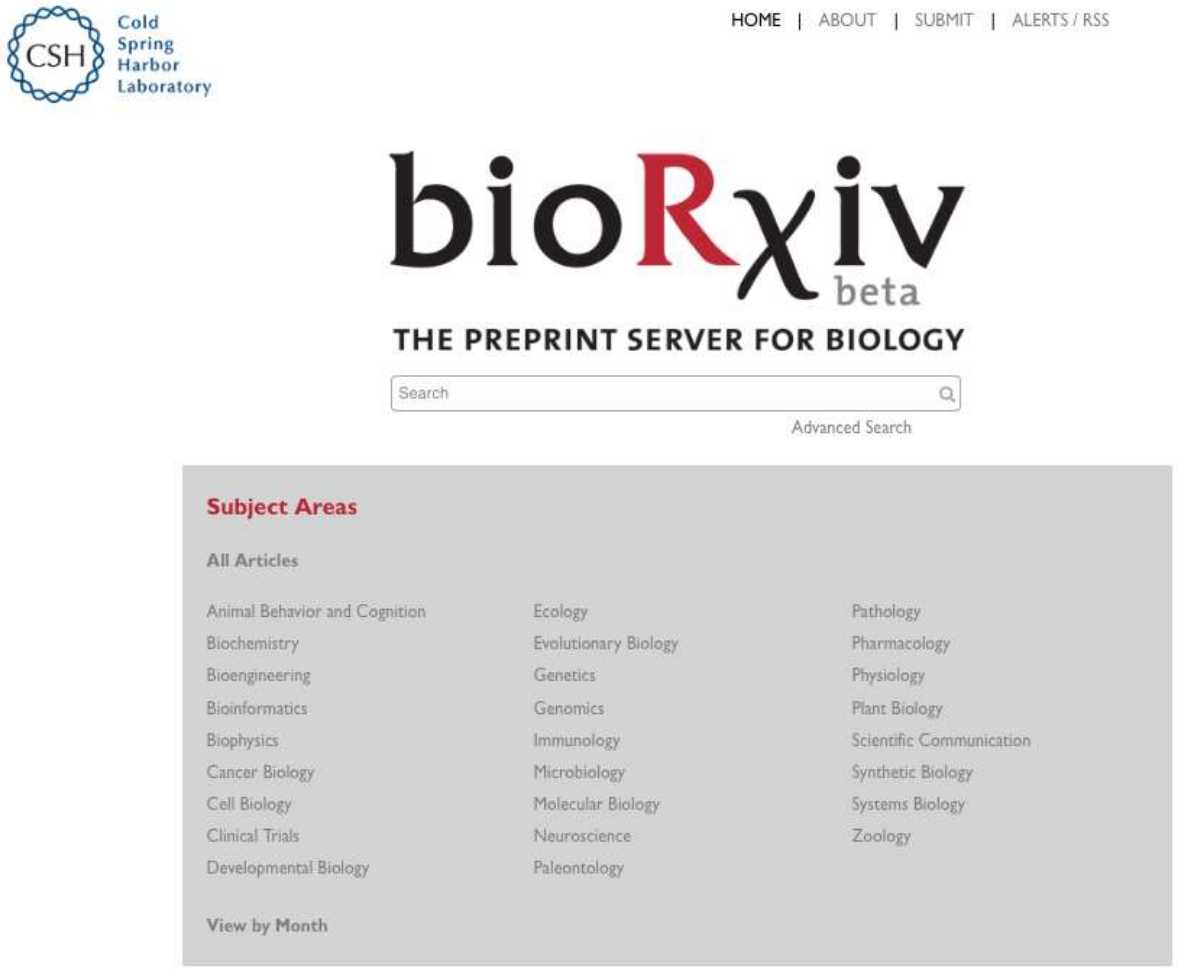


- A **computing environment**, such as the cloud or High Performance Computing (HPC) resources, which supports access, utilization, and storage of digital objects
- Publicly available **datasets** that adhere to a *Commons* digital object compliance model
- **Software** services and tools to facilitate access to and use of data, both the data in the Commons or elsewhere
- A **digital object compliance model** that describes the properties of digital objects that enable them to be findable, accessible, interoperable, and reproducible (FAIR)

http://datascience.nih.gov/commons

Share Information/Data

Efforts by Other Organizations: Recent Example



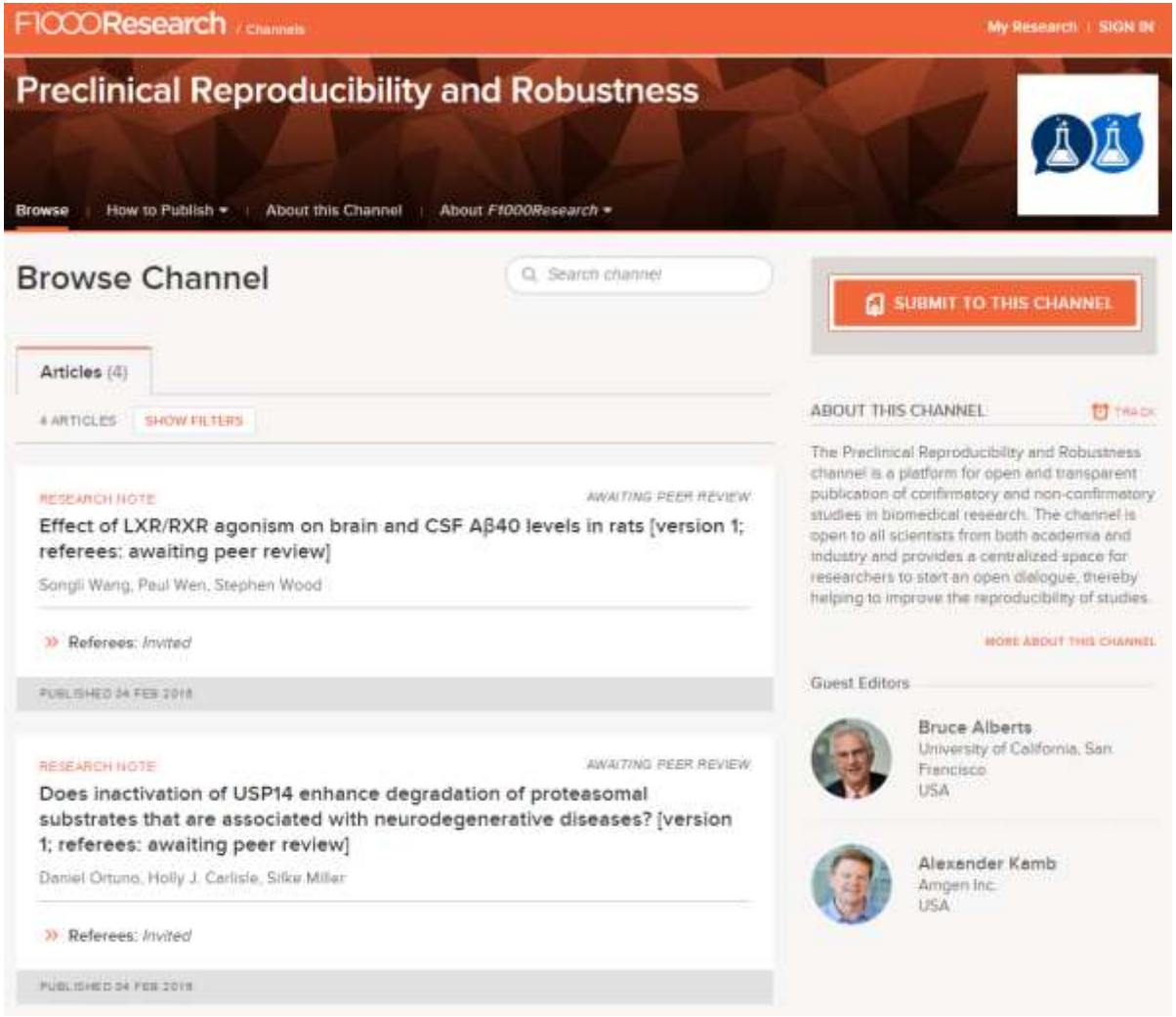
Share Information/Data

Team up with industry

Combining commercial and academic incentives and resources can improve science, argues **Aled Edwards**.

Share Information/Data

Efforts by Other Organizations: Recent Example



http://f1000research.com/channels/PRR

Principles for Addressing Underlying Issues

- Raise community awareness
- Enhance formal training
- Protect the quality of funded and published research by adoption of more systematic review processes
- Share information/data
- **Increase stability for investigators**

Investigator Stability

NIH DIRECTOR'S BLOG

Formula for Innovation: People + Ideas + Time

Posted on July 17, 2014 by Dr. Sally Rockey and Dr. Francis Collins

In these times of tight budgets and rapidly evolving science, we must consider new ways to invest biomedical research dollars to achieve maximum impact—to turn scientific discoveries into better health as swiftly as possible. We do this by thinking strategically about the areas of research that we support, as well as the process by which we fund that research.

COMMENT

NIH plans to enhance reproducibility

Francis S. Collins and **Lawrence A. Tabak** discuss initiatives that the US National Institutes of Health is exploring to restore the self-correcting nature of preclinical research.

A growing chorus of concern, from scientists and laypeople, contends that the complex system for ensuring the reproducibility of biomedical research is failing and is in need of restructuring[1,2]. As leaders of the US National Institutes of Health (NIH), we share this concern and here explore some of the significant interventions that we are planning.

Science has long been regarded as 'self-correcting', given that it is founded on the replication of prior work. Over the long term, that principle remains true. In the shorter term, h
balances that onc
have been hobble
the ability of today
others' findings.

Let's be clear:
have no evidenc
ducibility is abo
In 2011, the Offi
the US Departm
Services pursu
Even if this repr
the actual proble

"Efforts by the NIH alone will not be sufficient to effect real change in this unhealthy environment."

Role for Individual Scientists

What you can do:

- Stimulate discussion amongst societies/organizations
- Increase transparency
- Promote training in experimental design
- Encourage data and material sharing
- Consider publication of refutations




NIH…

Lawrence.Tabak@nih.gov

Turning Discovery Into Health