

# Reproducible Research (RR)

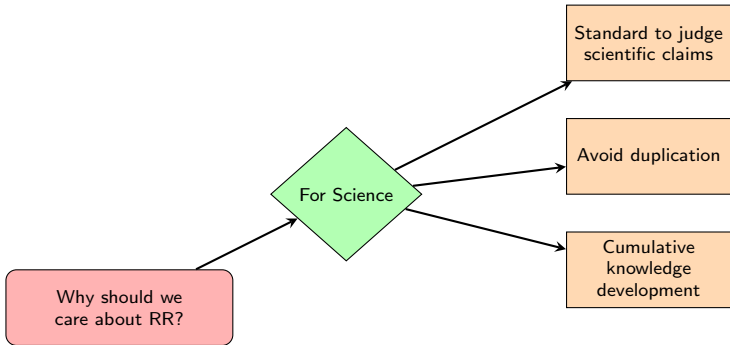
## Some Comments

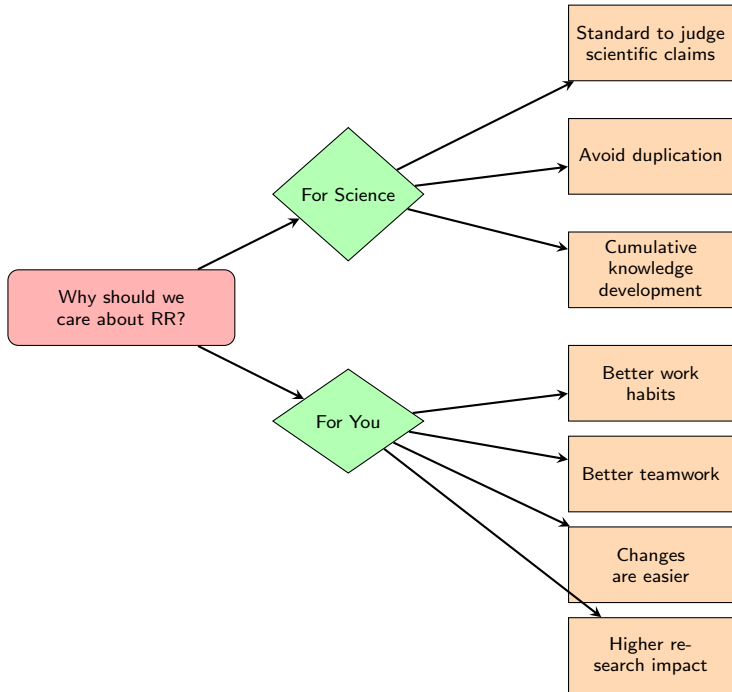
Greg Voisin and Sahir Bhatnagar<sup>1</sup>

June 18, 2015

---

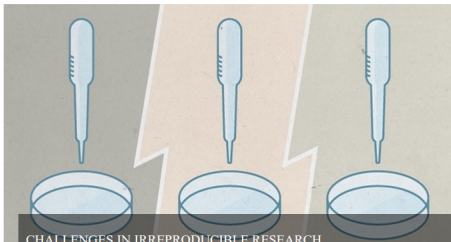
<sup>1</sup><http://admingreenwoodlab.github.io/tutorials/>





SPECIAL

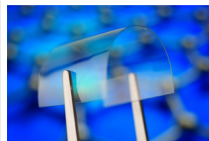
▶ See all specials



CHALLENGES IN IRREPRODUCIBLE RESEARCH

No research paper can ever be considered to be the final word, and the replication and corroboration of research results is key to the scientific process. In studying complex entities, especially animals and human beings, the complexity of the system and of the techniques can all too easily lead to results that seem robust in the lab, and valid to editors and referees of journals, but which do not stand the test of further studies. *Nature* has published a series of articles about the worrying extent to which research results have been found wanting in this respect. The editors of *Nature* and the *Nature* life sciences research journals have also taken substantive steps to put our own houses in order, in improving the transparency and robustness of what we publish. Journals, research laboratories and institutions and funders all have an interest in tackling issues of irreproducibility. We hope that the articles contained in this collection will help.

Recommended



Graphene booms in factories but lacks killer app

Although the wonder material is being made in record volume, commercial success is elusive.

Recent

1. **Why the Pope's letter on climate matters**  
*Nature* | 18 June 2015
2. **Podcast: Positive memories fight depression, a Mars-sized exoplanet and clever plants**  
*Nature* | 18 June 2015
3. **Activating happy memories cheers moody mice**  
*Nature* | 17 June 2015
4. **How gravity kills Schrödinger's cat**  
*Nature* | 17 June 2015
5. **Plant science: Rediscovering the bush telegraph**  
*Nature* | 17 June 2015

Figure 1 : <http://www.nature.com/news/reproducibility-1.17552>

**Annals of Internal Medicine**

## ACADEMIA AND CLINIC

**Reproducible Research: Moving toward Research the Public Can Really Trust**

Christine Laine, MD, MPH; Steven N. Goodman, MD, PhD, MHS; Michael E. Griswold, PhD; and Harold C. Sox, MD

A community of scientists arrives at the truth by independently verifying new observations. In this time-honored process, journals serve 2 principal functions: evaluative and editorial. In their evaluative function, they winnow out research that is unlikely to stand up to independent verification; this task is accomplished by peer review. In their editorial function, they try to ensure transparent (by which we mean clear, complete, and unambiguous) and objective descriptions of the research. Both the evaluative and editorial functions go largely unnoticed by the public—the former only draws

public attention when a journal publishes fraudulent research. However, both play a critical role in the progress of science. This paper is about both functions. We describe the evaluative processes we use and announce a new policy to help the scientific community evaluate, and build upon, the research findings that we publish.

*Ann Intern Med.* 2007;146:450–453.

For author affiliations, see end of text.

[www.annals.org](http://www.annals.org)

Figure 2 : Annals of Internal Medicine (*Liane et al. 2007*)

BIOCONDUCTOR PROJECT WORKING  
PAPERS

## Statistical Analyses and Reproducible Research

[Robert Gentleman](#), *Department of Biostatistics, Harvard University*

Follow

[Duncan Temple Lang](#), *Department of Statistics, University of California, Davis*

Follow

[Download](#)**Abstract**

For various reasons, it is important, if not essential, to integrate the computations and code used in data analyses, methodological descriptions, simulations, etc. with the documents that describe and rely on them. This integration allows readers to both verify and adapt the statements in the documents. Authors can easily reproduce them in the future, and they can present the document's contents in a different medium, e.g. with

Included in  
[Bioinformatics Commons](#),  
[Computational Biology Commons](#),  
[Numerical Analysis and Computation Commons](#)

Figure 3 : Bioconductor (*Gentleman and Lang 2004*)

## Reproducible research and *Biostatistics*

ROGER D. PENG

### 1. INTRODUCTION AND MOTIVATION

The replication of scientific findings using independent investigators, methods, data, equipment, and protocols has long been, and will continue to be, the standard by which scientific claims are evaluated. However, in many fields of study there are examples of scientific investigations that cannot be fully replicated because of a lack of time or resources. In such a situation, there is a need for a minimum standard that can fill the void between full replication and nothing. One candidate for this minimum standard is “reproducible research”, which requires that data sets and computer code be made available to others for verifying published results and conducting alternative analyses.

The need for publishing reproducible research is increasing for a number of reasons. Investigators are

Figure 4 : Biostatistics (*Peng 2009*)

# Biostatistics requirements for RR

Motivations

Acknowledgement  
of RR

Media

Challenges

Summary

References

- 1 data analysis script
- 2 other code
- 3 data
- 4 script for results used in paper
- 5 knitr file (`.Rnw`)
- 6 resulting `.tex` file from compiling with `knitr`
- 7 `bibTEX`file



# CRAN has a dedicated Task View for RR

Motivations

Acknowledgement  
of RR

Media

Challenges

Summary

References

<http://cran.r-project.org/web/views/>

Motivations

Acknowledgement  
of RR

Media

Challenges

Summary

References

The screenshot shows the Coursera website interface. At the top left is the Coursera logo. To its right is a navigation bar with a 'Catalog' button, a search bar containing the text 'Search catalog', and a search icon. Further right are links for 'Institutions', 'Log In', and 'Sign Up'. The main content area features the Johns Hopkins Bloomberg School of Public Health logo and the course title 'Reproducible Research'. Below the title is a link to 'Part of the Data Science Specialization'. A short description follows: 'Learn the concepts and tools behind reporting modern data analyses in a reproducible manner. This is the fifth course in the Johns Hopkins Data Science Specialization.' To the right of the text is a large graphic showing a hand holding a pen, with a play button overlay and the text 'Watch Intro Video'. The background of the graphic includes a stylized globe and a brain with neural connections.

Figure 5 : <https://www.coursera.org/course/repdata>

Motivations

Acknowledgement  
of RR

Media

Challenges

Summary

References



PERSPECTIVE

## The Economics of Reproducibility in Preclinical Research

Leonard P. Freedman<sup>1\*</sup>, Iain M. Cockburn<sup>2</sup>, Timothy S. Simcoe<sup>2,3</sup>

**1** Global Biological Standards Institute, Washington, D.C., United States of America, **2** Boston University School of Management, Boston, Massachusetts, United States of America, **3** Council of Economic Advisors, Washington, D.C., United States of America

\* [lfreedman@gbsi.org](mailto:lfreedman@gbsi.org)

### Abstract

Low reproducibility rates within life science research undermine cumulative knowledge production and contribute to both delays and costs of therapeutic drug development. An analysis of past studies indicates that the cumulative (total) prevalence of irreproducible preclinical research exceeds 50%, resulting in approximately US\$28,000,000,000 (US \$28B)/year spent on preclinical research that is not reproducible—in the United States alone. We outline a framework for solutions and a plan for long-term improvements in reproducibility rates that will help to accelerate the discovery of life-saving therapies and cures.

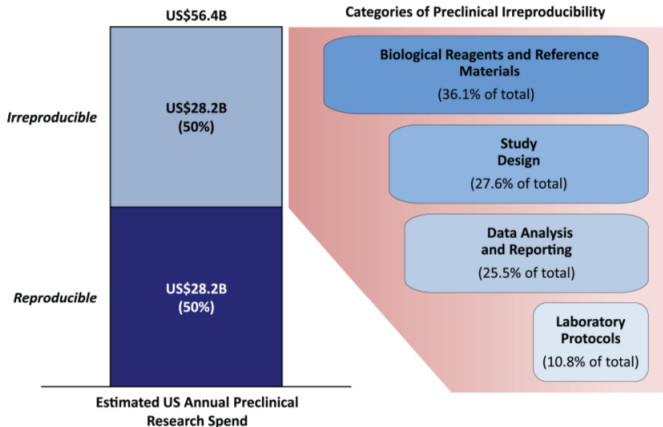


CrossMark  
[click for updates](#)

 OPEN ACCESS

# How did they come up with that number?

- Motivations
- Acknowledgement of RR
- Media
- Challenges
- Summary
- References



Science

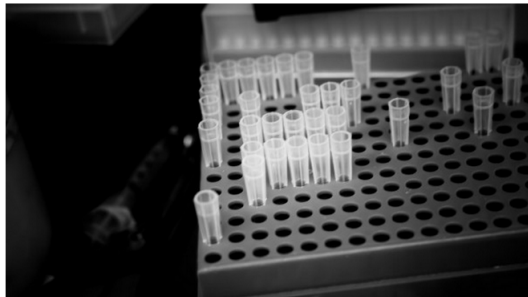
AAAS NEWS SCIENCE JOURNALS CAREERS MULTIMEDIA

News Home Hot Topics Categories From the Magazine ScienceInsider

News > Biology > Study claims \$28 billion a year spent on irreproducible biomedical research

## SCIENCEINSIDER

Breaking news and analysis from the world of science policy



BILL DICKINSON/Flickr (CC BY-NC-ND 2.0)

Study claims \$28 billion a year spent on  
irreproducible biomedical research

Motivations

Acknowledgement  
of RR

Media

Challenges

Summary

References



The image is a screenshot of a mobile phone displaying a tweet. At the top, the status bar shows the time as 11:01 AM, 92% battery, and various connectivity icons. The tweet header is blue with a back arrow, the Twitter logo, the word "Tweet", a search icon, and a three-dot menu icon. The tweet is from "Nature News&Comment" (@NatureNews), which has a verified account. The profile picture is the Nature logo, and there is a "+1" button next to it. The main content of the tweet is an illustration of two orange capsules, one above the other, with green dollar signs falling out of the opening of the top capsule. Below the illustration, the text reads: "Irreproducible biomedical research costs \$28 \*billion\* a year. And that's just in the US. [ow.ly/OerZ6](https://ow.ly/OerZ6)". The tweet is dated "11:10 AM · 13 Jun 15". Below the text, it shows "108 RETWEETS" and "52 FAVORITES". At the bottom of the tweet, there are four icons: a reply arrow, a retweet icon, a star icon, and a share icon. Below these icons is a text input field with the placeholder "Reply to Nature News&Comment".

# Challenges

Motivations

Acknowledgement  
of RR

Media

**Challenges**

Summary

References

# Challenges

Motivations

Acknowledgement  
of RR

Media

**Challenges**

Summary

References

- Large data/computations



# Challenges

Motivations

Acknowledgement  
of RR

Media

**Challenges**

Summary

References

- Large data/computations
- Complicated pipelines

# Challenges

Motivations

Acknowledgement  
of RR

Media

**Challenges**

Summary

References

- Large data/computations
- Complicated pipelines
- Privacy issues

# Challenges

Motivations

Acknowledgement  
of RR

Media

Challenges

Summary

References

- Large data/computations
- Complicated pipelines
- Privacy issues
- Getting PI's on board

# Caution

Motivations

Acknowledgement  
of RR

Media

Challenges

**Summary**

References

# Caution

Motivations

Acknowledgement  
of RR

Media

Challenges

**Summary**

References

- Reproducible doesn't make it right

# Caution

Motivations

Acknowledgement  
of RR

Media

Challenges

Summary

References

- Reproducible doesn't make it right
- Not Reproducible doesn't make it wrong

# If you can only take away one thing from today's discussion...

Motivations

Acknowledgement  
of RR

Media

Challenges

**Summary**

References

$$\text{Reproducibility} \propto \frac{1}{\text{copy paste}}$$

# References I

Motivations

Acknowledgement  
of RR

Media

Challenges

Summary

References



Christopher Gandrud, *Reproducible research with r and rstudio*, Chapman and Hall-CRC The R Series, 2013.



David Smith, *Did an excel error bring down the london whale?*, <http://blog.revolutionanalytics.com/2013/02/did-an-excel-error-bring-down-the-london-whale.html>.



C. Laine, S. N. Goodman, M. E. Griswold, and H. C. Sox, *Reproducible research: moving toward research the public can really trust*, *Ann. Intern. Med.* **146** (2007), no. 6, 450–453.



New York Times, *Reporters find science journals harder to trust, but not easy to verify*, [http://www.nytimes.com/2006/02/13/business/media/13journal.html?\\_r=0&adxnnl=1&pagewanted=all&adxnnlx=1390399611-aqm52MhkXkIFF7Azx7irCg](http://www.nytimes.com/2006/02/13/business/media/13journal.html?_r=0&adxnnl=1&pagewanted=all&adxnnlx=1390399611-aqm52MhkXkIFF7Azx7irCg).



R. D. Peng, *Reproducible research and Biostatistics*, *Biostatistics* **10** (2009), no. 3, 405–408.



# References II

Motivations

Acknowledgement  
of RR

Media

Challenges

Summary

References



Sergey Fomel and Jon F. Claerbout, *Guest editor's introduction: Reproducible research*, Computing in Science and Engineering (Jan/Feb 2009).



Yihui Xie, *Dynamic documents with r and knitr*, Chapman and Hall-CRC The R Series, 2013.