Reproducibility in cancer research

https://github.com/mdozmorov/presentations

Mikhail Dozmorov, Ph.D.
Department of Biostatistics, VCU
mikhail.dozmorov@vcuhealth.org
04/27/2016

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?

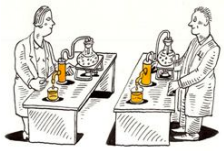


Image credit: http://blog.f1000research.com/2014/04/04/reproducibility-tweetchat-recap/

Reproducible research in science

- Science is the systematic enterprise of gathering knowledge about the universe and organizing and condensing that knowledge into testable laws and theories
- The success and credibility of science are anchored in the willingness of scientists to **expose their ideas and results to independent testing and replication by other scientists**.

http://www.aps.org/policy/statements/99_6.cfm

What is reproducible research?

- Reproducibility
- Replicability
- Repeatability
- Reliability
- Robustness
- Generalizability

Steve Goodman, Stanford, March 18, 2015

What is reproducible research?

Reproducible research is the ultimate standard for strengthening scientific evidence by independent:

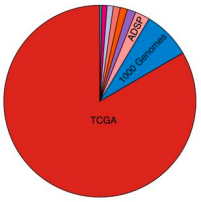
- Investigators
- Data
- Analytical methods
- Laboratories
- Instruments

Image source: Glenn B et al. "Institutions must do their part for reproducibility" Nature 2015

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



TCGA	- 2300 TB
1000 Genomes*	- 222 TB
ADSP	- 68 TB
NHGRI LSSP*	- 40 TB
GTeX	-34 TB
NHLBI ESP	- 32 TB
HMP*	- 29 TB
ARRA Autism	- 24 TB
ENCODE*	- 9 TB

Image credit: Muir et al., "The Real Cost of Sequencing," Genome Biol. 2016

Poor medical tests getting to patients

OvaSure diagnostic test for ovarian cancer

MISSING THE MARK

Why is it so hard to find a test to predict cancer?

BY LIZZIE BUCHEN

On 3 March, two studies appeared online that offered 19 pages of gloomy reading for anyone interested in cancer. They focused on biological molecules, or biomarkers, the presence of which in the blood might be used to detect the earliest glimmers of women — to ask whether these seemingly breakthrough biomarkers were better at identifying women with early-ovarian cancer than the one flawed biomarker that had been in use for almost 30 years, CA-125. None of them was[1,2]. "CA-125 remains the 'best of a bad contrast, detected 63%.) Mor's panel already had a tortured history. A primary research paper behind it had been criticized by other scientists for allegedly using inappropriate statistical calculations and for optimistically concluding that the test would help women.

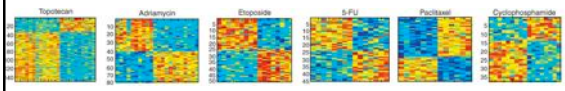
Buchen, "Cancer: Missing the mark" *Nature* 2011

Clinical trials based on flawed and fraudulent data

Genomic signatures to guide the use of chemotherapeutics

Anil Potti[1,2], Holly K Dressman[1,3], Andrea Bild[1,3], Richard F Riedel[1,2], Gina Chan[4], Robyn Sayer[4], Janiel Cragun[4], Hope Cottrill[4], Michael J Kelley[2], Rebecca Petersen[5], David Harpole[5], Jeffrey Marks[5], Andrew Berchuck[1,6], Geoffrey S Ginsburg[1,2], Phillip Febbo[1–3], Johnathan Lancaster[4] & Joseph R Nevins[1–3]

- Described drug response "gene signatures" in NCI60 cell lines
- Demonstrated these "signatures" correspond to patient-specific signatures and can be used to predict patient response to the drugs



Retraction Watch "The Importance of Being Reproducible: Keith Baggerly tells the Anil Potti story" 2011

Bioinformatics statisticians spot errors

How Bright Promise in Cancer Testing Fell Apart

Keith Baggerly, left, and Kevin Coombes, statisticians at M. D. Anderson Cancer Center, found flaws in research on tumors.

When Juliet Jacobs found out she had lung cancer, she was terrified, but realized that her hope lay in getting the best treatment medicine could offer. So she got a second opinion, then a third. In February of 2010, she ended up at Duke University, where she entered a research study whose promise seemed stunning.

New York Times "How Bright Promise in Cancer Testing Fell Apart" 2011

"Off-by-one" error

	Published	Replicated
...		
[3,]	1881_at	1882_g_at
[4,]	31321_at	31322_at
[5,]	31725_s_at	31726_at
[6,]	32307_r_at	32308_r_at
...		

New signatures continue to be published

Pharmacogenomic Strategies Provide a Rational Approach to the Treatment of Cisplatin-Resistant Patients With Advanced Cancer

David S. Hsu, Bala S. Balakumaran, Chaitanya R. Acharya, Vanja Vlahovic, Kelli S. Walters, Katherine Garman, Carey Anders, Richard F. Riedel, Johnathan Lancaster, David Harpole, Holly K. Dressman, Joseph R. Nevins, Phillip G. Febbo, and Anil Potti

Validation of gene signatures that predict the response of breast cancer to neoadjuvant chemotherapy: a substudy of the EORTC 10994/BIG 00-01 clinical trial

Hervé Bonnefoi, Anil Potti, Mauro Delorenzi, Louis Mauriac, Maria Campone, Michèle Tubiana-Hulin, Thierry Petit, Philippe Rouanet, Jacek Jassem, Emmanuel Blot, Véronique Becette, Pierre Farmer, Sylvie André, Chaitanya R Acharya, Sayan Mukherjee, David Cameron, Jonas Bergh, Joseph R Nevins, Richard D Iggo

More data added

	Sample ID	Response			
1	GSM44303	RES	11	GSM9694	RES
2	GSM44304	RES	12	GSM9695	RES
3	GSM9653	RES	13	GSM9696	RES
4	GSM9653	RES	14	GSM9698	RES
5	GSM9654	RES	15	GSM9699	SEN
6	GSM9655	RES	16	GSM9701	RES
7	GSM9656	RES	17	GSM9708	RES
8	GSM9657	RES	18	GSM9708	SEN
9	GSM9658	SEN	19	GSM9709	RES
10	GSM9658	SEN	20	GSM9711	RES

RES/SEN – resistant/sensitive

Summary of the Duke case

- A total of 162 co-authors
- 40 papers
- Two-thirds are partially or completely retracted

THE CANCER LETTER

Inside information on cancer research and drug development

publication date: nov 19, 2010

Duke University issued the following press release Nov. 19:

Duke Accepts Potti Resignation; Retraction Process Initiated with Nature Medicine

DURHAM, NC -- Anil Potti, MD, has voluntarily resigned from his positions as associate professor of medicine at Duke University School of Medicine and at the university's Institute for Genome Science & Policy. Dr. Potti's resignation is effective immediately.

In addition, Dr. Potti's collaborator, Joseph Nevins, Ph.D., has initiated a process intended to lead to a retraction request regarding a paper previously published in Nature Medicine. This process has been initiated due to concerns about the reproducibility of reported predictors, and their possible effect on the overall conclusions in this paper. Other papers published based on this science are currently being reviewed for any concerns.

The three clinical trials based on this science for which new enrollment was suspended in mid-July, have been closed.

http://retractionwatch.com/2011/05/04/the-importance-of-being-reproducible-keith-baggerly-tells-the-anil-potti-story/

IOM guidelines on translational omics

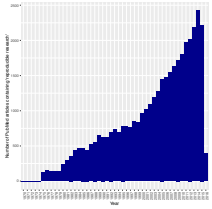
Kaiser J "Clinical medicine. Biomarker tests need closer scrutiny, IOM concludes." *Science* 2012

McShane LM et.al. "Criteria for the use of omics-based predictors in clinical trials." *Nature* 2013

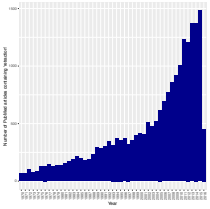
IOM report "Evolution of Translational Omics: Lessons Learned and the Path Forward" 2012

PubMed stats on "Reproducible research" vs. "Retraction"

"Reproducible research"



"Retraction"



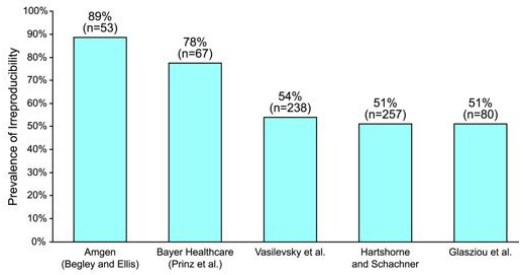
Number of publications per year, from 1970 through April 2016

Retraction Watch http://retractionwatch.com/ Tracking retractions as a window into the scientific process

THE COST OF REPRODUCIBILITY

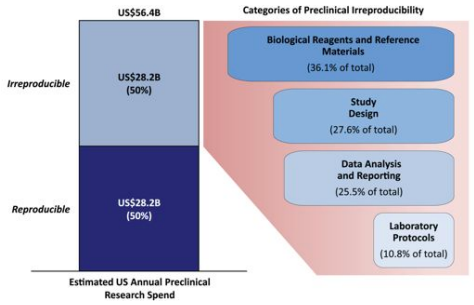
Image credit: The Cost of Living

Irreproducibility ranges from 51% to 89%



Leonard Freedman, Iain Cockburn, and Timothy Simcoe, "The Economics of Reproducibility in Preclinical Research," *PLOS Biol* 2015

Cost of irreproducibility



Leonard Freedman, Iain Cockburn, and Timothy Simcoe, "The Economics of Reproducibility in Preclinical Research," *PLOS Biol* 2015

WHY REPRODUCIBILITY QUESTIONS ARISE?

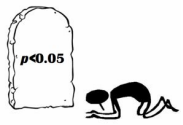


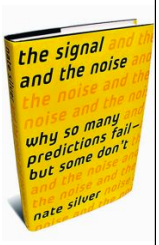
Image credit: Demystifying the p-value

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.

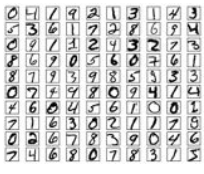
- Nate Silver



Nate Silver "The Signal and the Noise: Why So Many Predictions Fail--but Some Don't" 2015

Patterns in the noise

- Humans are good at recognizing patterns



Using neural nets to recognize handwritten digits
Spammers Use The Human Touch To Avoid CAPTCHA

Irreproducibility in high-throughput biology

- **Our intuition about patterns in high dimensional data quickly drops** with the increased dimensionality of the data
- We rely on computation to uncover patterns
- P values, the 'gold standard' of statistical validity, are not as reliable as many researchers assume.

The chance of being wrong

The probability of irreproducibility of studies with p-values in the range 0.005 to 0.05 **is roughly 0.33**

Regina Nuzzo, "Scientific Method: Statistical errors" *Nature* 2014
Dennis Boos and Leonard Stefanski "P-value precision and reproducibility" *Am Statistician*, 2001

The chance of being wrong

Why Most Published Research Findings Are False

John P. A. Ioannidis

Summary

There is increasing concern that most current published research findings are false. The probability that a research claim is true may depend on study power and

factors that influence this problem and some corollaries thereof.

Modeling the Framework for False Positive Findings

Several methodologists have

is characteristic of the field and can vary a lot depending on whether the field targets highly likely relationships or searches for only one or a few true relationships among thousands and millions of hypotheses that may

1. In evaluating any study try to take into account the amount of background noise. That is, remember that the more hypotheses which are tested and the less selection which goes into choosing hypotheses the more likely it is that you are looking at noise.
2. Bigger samples are better. (But note that even big samples won't help to solve the problems of observational studies which is a whole other problem).
3. Small effects are to be distrusted.
4. Multiple sources and types of evidence are desirable.
5. Evaluate literatures not individual papers.
6. Trust empirical papers which test other people's theories more than empirical papers which test the author's theory.
7. As an editor or referee, don't reject papers that fail to reject the null.

John Ioannidis "Why Most Published Research Findings Are False" *PLOS Medicine* 2005
Alex Tabarrok blog "Why Most Published Research Findings are False" 2005

Understanding the p-value

The ASA's statement on p-values: context, process, and purpose

DOI: 10.1080/00031305.2016.1154108
Ronald L. Wasserstein & Nicole A. Lazar
Publishing models and article dates explained
Received: 4 Feb 2016
Accepted: 9 Feb 2016

1. P-values can indicate how incompatible the data are with a specified statistical model.
2. P-values do not measure the probability that the studied hypothesis is true, or the probability that the data were produced by random chance alone.
3. Scientific conclusions and business or policy decisions should not be based only on whether a p-value passes a specific threshold.
4. Proper inference requires full reporting and transparency.
5. A p-value, or statistical significance, does not measure the size of an effect or the importance of a result.
6. By itself, a p-value does not provide a good measure of evidence regarding a model or hypothesis.

Ronald Wasserstein and Nicole Lazar "The ASA's statement on p-values: context, process, and purpose" *Am Stat* 2016
Monya Baker "Statisticians issue warning over misuse of P value" *Nature* 2016

P-value warning: consult a statistician before the experiment

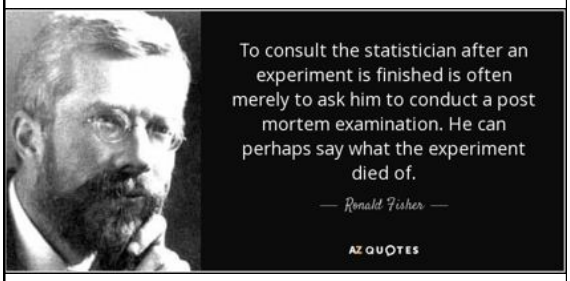


Image credit: azquotes.com

CURRENT STATUS OF REPRODUCIBILITY

Image credit: Badges to Acknowledge Open Practices

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.

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

shorter term, however, the checks and balances that once ensured scientific fidelity have been hobbled. This has compromised

ing agencies to establish or enforce policies that insist on data access.

PRECLINICAL PROBLEMS

Reproducibility is a problem in all scientific disciplines. However, human clinical trials seem to be less at risk because they are already governed by various regulations that stipulate rigorous design and independent oversight — including randomization, blinding, power estimates, pre-registration of outcome measures in standardized, public databases such as ClinicalTrials.gov and oversight by institutional review boards and data safety monitoring boards. Furthermore, the clinical trials community has taken important steps towards adopting standard reporting elements'.

Glenn Begley and Lee Ellis, "Drug Development." *Nature* 2012
Francis Collins and Lawrence Tabak, "Policy: NIH Plans to Enhance Reproducibility." *Nature* 2014

NIH focus on openness

- Public Access Policy
- Data Sharing Policies
- Genomics Data Sharing (GDS)
- Model Organism Sharing Policy
- Research Tools Policy

https://grants.nih.gov/reproducibility/index.htm

NSF stance on openness

- Responsible Conduct of Research (RCR)
- NSF Scientific Integrity Policy

https://www.nsf.gov/bfa/dias/policy/dmp.jsp

Reproducibility initiatives

"Enhancing reproducibility". *Nat Methods* 2013

Victoria Stodden et.al. "Reproducible Research" Yale Law School

https://cos.io/ https://osf.io/

Reproducibility guidelines

- **ARRIVE** – Animal Research Reporting of In Vivo Experiments
- **CONSORT** – Consolidated Standards of Reporting Trials
- **SPIRIT** – Standard Protocol Items: Recommendations for Interventional Trials
- **STROBE** – The Strengthening the Reporting of Observational Studies in Epidemiology
- **(STARD) TRIPOD** – Transparent Reporting of a multivariable prediction model for Individual PROgnosis of Diagnosis
- **REMARK** – REporting recommendations for tumour MARKer prognostic studies

http://www.equator-network.org/ - over 300 reporting guidelines

WHAT CAN WE DO TO ENHANCE REPRODUCIBILITY?

Image credit: Bill & Melinda Gates Foundation

Flavors of reproducibility

- Empirical reproducibility
- Computational reproducibility
- Statistical reproducibility

Steps in reproducible research

The most important is the mindset, when starting, that the end product will be reproducible. – Keith Baggerly

- Experimental design
- Data generation
- Data analysis
- Results interpretation
- Dissemination of results

Poor experimental design

Related Commentary, page 26 Research article

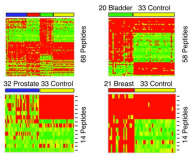
Differential exoprotease activities confer tumor-specific serum peptidome patterns

Josep Villanueva, David R. Shaffer, John Philip, Carlos A. Chaparro, Hediye Erdjument-Bromage, Adam B. Olshen, Martin Fleisher, Hans Lilja, Edi Brogi, Jeff Boyd, Marta Sanchez-Carbayo, Eric C. Holland, Carlos Cordon-Cardo, Howard I. Scher, and Paul Tempst

Memorial Sloan-Kettering Cancer Center, New York, New York, USA.

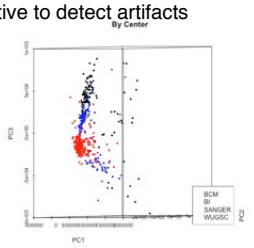
Villanueva J et.al. "Differential exoprotease activities confer tumor-specific serum peptidome patterns." *J Clin Invest.* 2006

- 100% sensitive and specific for **prostate cancer**
- Patient characteristics
  - Cancer patients: Age: mean 67 years, Sex: 100% male
  - Healthy controls: Age: mean 35 years, Sex: 58% women



Considerations for data generations

- Assay artifacts and batch effects
  - 'Omics' technologies are extremely well suited to detect biological features
  - They are exquisitly sensitive to detect artifacts
- Note everything:
  - changes in procedures
  - reagents
  - equipment
  - technician
  - date of experiment



Basics of reproducible research

- Lab notebook
  - Complete record of procedures, reagents, data, and thoughts to share with other researchers
  - Explanation of why experiments were initiated, how they were performed, what are the results
  - Legal document to prove your experiments against irreproducibility

Common approach: write report around results

Point and click approach

- Use MS Excel for data entry/cleaning/preparation, and possibly statistical analysis

Problems

- With point-and-click, there's no way to record/save the steps that generated the (copy/pasted) results
- Data files are kept separately from the analysis code, and from reports
- After modifications of one of the files, it becomes unclear which version corresponds exactly to the reported results
- Every time something changes, you have to regenerate the figures/results/reports by hand – very time consuming

Zeeberg BR et al. "Mistaken identifiers: gene name errors can be introduced inadvertently when using Excel in bioinformatics" *BMC Bioinformatics* 2004

Better approach: write report that generates results

- The report is automated via code
- Data is attached to the well-documented code
- History of any changes should be preserved

The final report should be self-sufficient and reproducible with a single command

Automate everything

- R – free/open source programming language
- Runs on Windows, Mac, and Linux
- Extensible with a very large collection of actively developing packages
- Excellent graphics & report-creating capabilities

Self-documenting code

- A report containing a stream of text and code chunks
- Each code chunk loads data, computes results, shows figures
- Each text chunk explains how the code chunks work
- The resulting report is human- and machine readable

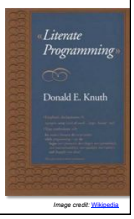
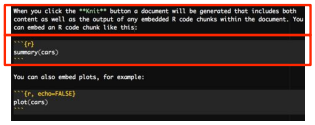


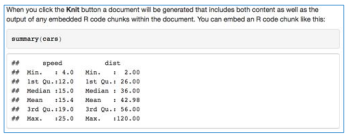
Image credit: Wikipedia

Self-documenting code

- Mix documentation with code



- Generate report with one command



Keeping history of changes

Image credit: PHD Comics

Version control – what and when did you do

- **Git** and GitHub – version control system
  - Each project stored in its own repository
  - History of changes – track what you did
  - Ability to go back if something breaks
  - Branch out, go creative, then merge or revert the changes
  - Collaborate through merging changes from multiple people

https://github.com/

The aims of computational reproducibility

- Fully automated report
- Including code
- And data

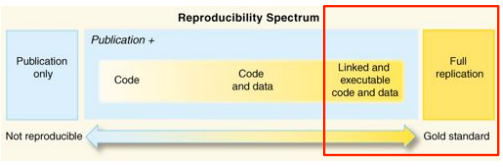


Image credit: Roger Peng "Reproducible Research in Computational Science" Science 2011

Publishing with Git

B-Cell and Monocyte Contribution to Systemic Lupus Erythematosus Identified by Cell-Type-Specific Differential Expression Analysis in RNA-Seq Data

Libertas Academica

Mikhail G. Dozmorov[1], Nicolas Dominguez[2], Krista Bean[2], Susan R. Macwana[2], Virginia Roberts[2], Edmund Glass[1], Judith A. James[2] and Joel M. Guthridge[2]

Implementation and availability. All RNA-seq data processing steps were performed in CentOS 6.6 high-performance cluster computing environment. All analyses were conducted in R/Bioconductor environment v 3.2.0.[36,37] All analytical scripts are available at https://github.com/mdozmorov/deconvolution.

https://www.ncbi.nlm.nih.gov/pubmed/26512198

Reproducibility 101

- **Reproducibility made easy**
  - Begin with the final product in mind
  - Use literate programming (self-documenting code)
  - History of changes via code versioning and sharing
- **Replication made easy**
  - Get basic statistics right
  - Set stringent cutoffs, correct p-values for multiple testing
  - Be critical, consider batch effects, visualize, do sanity checks, use random controls, cross-validation
  - Follow reporting guidelines

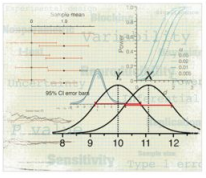
LEARN MORE

Nature "Statistics for Biologists"

WEB COLLECTION

Statistics for biologists

Menu



There is no disputing the importance of statistical analysis in biological research, but too often it is considered only after an experiment is completed, when it may be too late.

This collection highlights important statistical issues that biologists should be aware of and provides practical advice to help them improve the rigor of their work.

Nature Methods' **Points of Significance** column on statistics explains many key statistical and experimental design concepts. **Other resources** include an online plotting tool and links to statistics guides from other publishers.

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

Reproducible research made simple

OPEN ACCESS Freely available online

PLOS | COMPUTATIONAL BIOLOGY

Editorial

Ten Simple Rules for Reproducible Computational Research

Geir Kjetil Sandve[1,2]*, Anton Nekrutenko[3], James Taylor[4], Eivind Hovig[1,5,6]

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

Best Practices for Scientific Computing

Greg Wilson *, D.A. Aruliah †, C. Titus Brown ‡, Neil P. Chue Hong §, Matt Davis ¶, Richard T. Guy ||, Steven H.D. Haddock **, Katy Huff ††, Ian M. Mitchell ‡‡, Mark D. Plumbley §§, Ben Waugh ¶¶, Ethan P. White ***, Paul Wilson †††

https://www.ncbi.nlm.nih.gov/pubmed/24415924

Practical reproducibility

- BOIS 692 "Reproducible Research Tools"
- June 13-16, 2016
- https://mdozmorov.github.io/BIOS692/

Mikhail Dozmorov, Ph.D.
Department of Biostatistics, VCU
mikhail.dozmorov@vcuhealth.org