Published in final edited form as:

Bull Math Biol. 2018 December ; 80(12): 3095–3105. doi:10.1007/s11538-018-0500-9.

“Reproducible” research in mathematical sciences requires changes in our peer review culture and modernization of our current publication approach

Santiago Schnell^{1,2,*}

¹ Department of Molecular & Integrative Physiology, University of Michigan Medical School, Ann Arbor, MI 48109, USA

² Department of Computational Medicine & Bioinformatics, University of Michigan Medical School, Ann Arbor, MI 48109, USA

Abstract

The nature of scientific research in mathematical and computational biology allows editors and reviewers to evaluate the findings of a scientific paper. Replication of a research study should be the minimum standard for judging its scientific claims and considering it for publication. This requires changes in the current peer review practice, and a strict adoption of a replication policy similar to those adopted in experimental fields such as organic synthesis. In the future, the culture of replication can be easily adopted by publishing papers through dynamic computational notebooks combining formatted text, equations, computer algebra and computer code.

Scientists are awakening to the “reproducibility” crisis

Repeatability, replicability and reproducibility are held as the gold standards for scientific research. As a consequence, the legitimacy of any research paper depends on the ability of other scientists to repeat, replicate or reproduce the results published in an article since the appearance of the first scientific journals – *Journal des sçavans* and *Philosophical Transactions of the Royal Society* – in 1665 (Kronick 1976, Shapin and Schaffer 1985).

Over the last 15 years, there has been increasing concerns that many research findings published in influential peer-reviewed scientific journals cannot be replicated in the biomedical sciences (Hirschhorn et al. 2002, Ioannidis 2005, Prinz et al. 2011, Begley and Ellis 2012), psychology (Bakker and Wicherts 2011, John et al 2012, Pashler and Wagenmakers 2012, Van Bavel et al. 2016), sociology and economy (Bollen et al 2015, Ioannidis et al. 2017) and computer science (Donoho et al. 2009). There are even expository articles arguing that the failure to “reproduce” also occurs in the mathematical sciences (Redish et al. 2018). In a recent survey by the journal *Nature* (Baker 2016), respondent scientists agreed that there is a *reproducibility crisis*. Although the survey results may not be representative of the scientific community, individual scholars confirmed their skepticism by trying —and often failing—to repeat, replicate or reproduce results.

* schnells@umich.edu.

The “reproducibility” crisis has caused several academics, politicians and the public at large to call into question the reliability, value and trustworthiness of the scientific enterprise (Alberts et al. 2013, Macleod 2014, Van Bavel 2016, Wood and Randall 2016). The good news is that multiple studies show that there is no evidence that the failure to “reproduce” is the result of scientific misconduct (see Fanelli 2018 and reference therein). Less than 2% of scientists are fabricating or falsifying data (Shapiro and Charrow 1989, Steneck 2006, Fanelli 2009). Metaresearch studies suggest that the “reproducibility” crisis results from heterogeneous factors distributed across different areas of science that include bias in hypothesis testing (Fanelli 2010, Fanelli et al 2017), poor experimental design, statistical problems (John et al 2012, Button et al 2013) and lack of standardization to report and share experimental protocols, results and data (Wicherts et al. 2006, Nature Editorial 2014, Stark 2018).

Recommendations and policies adopted in response to the “reproducibility” crisis

As a response to the “reproducibility” crisis, the scientific community is encouraging journals to maintain high standards to increase the likelihood of reproducibility, and journals are responding accordingly (McNutt 2014). In the biomedical sciences, new guidelines have been adopted requiring detailed experimental protocols, establishing best-practices for image-based data, and statistical analysis (National Institutes of Health 2017). It has been recommended to modify the reward system in science by awarding scientists who aligned with best practices for reproducibility research (Ioannidis 2014). In the USA, federal funding agencies are now requiring training in rigor and reproducibility for scientists holding research grants. The National Institutes of Health Office of Extramural Research revised grant application instructions and review criteria to enhance reproducibility of research findings by adopting a rigor and reproducibility guidance (Collins and Tabak 2014, National Institutes of Health 2018). Similar approaches are being adopted for the psychological, social and economic sciences (Munafò 2017).

Most of the above reproducibility policies are geared towards the experimental biomedical sciences. However, mathematical and computational biology have not been ignored by the funding agencies and scientific community. In 2003, the US National Institutes of Health and National Science Foundation formed the Interagency Modeling and Analysis Group. This group provides an open forum of communication between funding agencies and scientists to discuss the development of new modeling and analysis methods in the biomedical sciences.¹ One of the working groups of Interagency Modeling and Analysis Group, the Model and Data Sharing Working Group, has been developing modeling standards, software, and repositories of models and experimental data to facilitate model reproducibility and model sharing.² Members of the Model and Data Sharing Working Group started to work on machine-readable description languages, such as Systems Biology Markup Language (SBML) (Hucka et al 2003) and CellML (Lloyd et al. 2004), to encode

¹<https://www.nibib.nih.gov/research-funding/interagency-modeling-and-analysis-group-imag>

²<https://www.imagwiki.nibib.nih.gov/content/model-and-data-sharing-working-group>

mathematical models, and facilitate their sharing and exchange. Simultaneously, model repositories, such as the CellML model repository (Lloyd et al., 2008), BioModels Database (Li et al. 2010, Chelliah et al 2015) and JWS online (Olivier et al. 2004), were established to store and distribute these models encoded with SBML and CellML. More recently, funding agencies have been funding the development of software tools to build reproducible dynamic models that are reusable, like Virtual Cell (Loew and Schaff 2001), CompuCell3d (Swate et al 2012) and Tellurium (Medley et al 2018).

In addition, scientific policy forums are recommending good practices for publishing scientific computing research (Stodden et al 2016). These include making the raw data analyzed computationally available in the paper (Peng 2011, Stodden et al 2013), adopting good lab notebook practices (Schnell 2015, Wilson et al 2017), and making computer code and scripts openly available (Yale RoundTable Participants 2010, Stodden et al 2012).

Have the “reproducibility” policies been effective in mathematical and computational biology?

The recommendations and policies available in the literature have been partially effective. Standardization of mathematical models and computer code, and code sharing have not been straight forward because they are highly variable in many disciplines and subdisciplines (Renear et al 2010). Despite of this, the scientific communication has been improved in the mathematical and computational sciences because it is becoming a more common practice to share models and computer code with other researchers to make findings “reproducible” after publication (Stodden et al 2013). Unfortunately, the effectiveness of model and code sharing has limited success in making scientific findings reproducible. An empirical study for articles published in *Science* shows that 44% of the authors respond to requests to share data and computer code after publication. From those, it was only possible to “reproduce” the computational results of 26% of the papers (Stodden et al 2018).

Simultaneously, there are also challenges with the curation of mathematical and computational models in databases. Model repositories like BioModels have nearly twice as many non-curated models as curated models (fully “reproducible” models).³ Although the non-curated models could be “reproducible”, the open science movement and rapid growth of the model repositories is exceeding the capacity of database curators to verify the reproducibility of models. There are still numerous models that are not reproducible in databases, because they are not translatable into SBML or CellML formats. Some models are not “reproducible”, because errors were introduced into the model during the publication process, such as authors mistyping parameter values or equations in the publication, or publishers inevitably committing typesetting errors during the article production (Lloyd et al 2008). In other instances, models are not “reproducible”, because they were not properly annotated following protocols, such as minimum information requested in the annotation of biochemical models (Novère et al 2005).

³For August 10th, 2018, the BioModels Database (<http://www.ebi.ac.uk/biomodels/>) has 641 manually curated (fully reproducible) models, and 1008 non-curated models

Given the limited success of the recommended policies by funding agency, scientific groups, and some publishers, is there a best-practice solution to the “reproducibility” crisis in the mathematical and computational sciences? Here I argue that most of the current discussions of the “reproducibility” crisis overlook the essential role that the editorial and peer review process has played in confirming “reproducibility” of experimental findings in scientific fields like organic syntheses for nearly a century. The current discussions also overlook that the current form of the scientific paper dating from the 1600s is outdated (Sommers 2018) and needs to change to keep up with current technologies available in the mathematical and computational sciences. However, before I proceed, it is important to settle what is meant by the conceptual framework of “reproducibility” in the research enterprise.

How can we define the R-words in mathematical and computational biology?

“Reproducibility” is an unsettled concept in most of the papers discussing it. The terms repeatability, replicability and reproducibility are used interchangeably in everyday language (Goodman et al. 2016). As a result, these R-terms are unfortunately used in nonstandard conceptual frameworks across the sciences. The definitions of the R-words, repeatability, replicability and reproducibility, vary between computer scientist subgroups (Association for Computing Machinery 2016, Rougier et al. 2017). They are also different between social, behavioral and economic sciences (Bollen et al 2015), computational neuroscience (Crook 2013, Plessner 2018) and biomedical sciences (Goodman et al. 2016).

The good news is that definition of the three R-words is well-established in the experimental quantitative sciences, such as engineering, physics and chemistry. The International Union of Pure and Applied Chemistry has clear definitions in the IUPAC Compendium of Chemical Terminology, informally known as the Gold Book (International Union of Pure and Applied Chemistry 1997). The International Organization for Standardization (ISO, 1994) describes the conditions necessary for repeatability, replicability and reproducibility under its International Standard ISO 5725–1, which were reiterated by the metrology community in its formal definition guides (Joint Committee for Guides in Metrology 2006). Inspired by the metrology community, the Association for Computing Machinery (2016) proposes a terminology in line with the experimental sciences. I am going to take the liberty of paraphrasing these definitions for the mathematical and computational biology community:

Repeatability (Same team, same methodological setup):

For experimental laboratory research, repeatability occurs when the measurement can be obtained with stated precision by the same team using the same measurement procedure, the same measuring system, under the same operating conditions, in the same location on multiple trials. *For mathematical and computational biology research, this means that a researcher can reliably repeat their own mathematical analysis and numerical computation.*

Replicability (Different team, same methodological setup):

For experimental laboratory research, replicability occurs when the measurement can be obtained with stated precision by a different team using the same measurement procedure, the same measuring system, under the same operating conditions, in the same or a different location on multiple trials. *For mathematical and computational biology research, this means that an independent group can obtain the same mathematical and computational results if they follow the methodology described in a published research paper, or reuse the same scripts and computer code.*

Reproducibility (Different team, different methodological setup):

For experimental laboratory research, reproducibility occurs when the measurement can be obtained with stated precision by a different team, a different measuring system, in a different location on multiple trials. *For mathematical and computational biology research, this means that an independent group can obtain the same mathematical and computational results using a methodology which they develop independently.*

Based on these definitions, failure to reproduce is a broader concept, which requires integrating potential conflicting observations, results and ideas into a coherent theory (Redish et al 2018). However, repeatability and replicability are clearly the two minimum gold standards of science. *Repeatability* is a necessary, but not sufficient condition, to determine if a novel research finding is ready to be submitted for publication. While *replicability* is the minimum standard by which any published scientific paper should be judged. If we establish a culture of replication, the scientific community will clear erroneous results and create a scientific community with a disciplined approach towards scientific discoveries (Peng 2011).

How to create a culture of replication in the mathematical and computational biology community?

The culture of replication comes from multiple directions, but it must start in the journals through peer review practice. Journals are starting to implement replication policies, though the fraction is small. An empirical analysis of journals considering mathematical and computational models shows that 22% of the journals had a computer code sharing policy, and 38% had a data sharing policy as of 2012 (Stodden et al 2013).

The challenge is that these data and code sharing policies are having a limited impact in creating a replicability culture. For example, *Biostatistics* implemented a “reproducible” research policy (Peng 2009), where papers in the journals are kite-marked R if both data and code are made available in the Supplementary Materials, and the *Biostatistics* Associate Editor for Reproducibility can use the data and code to replicate the results in the paper. Unfortunately, the “reproducibility” check is voluntarily, and only 4% of the paper received an “R” within the first two years of “reproducibility” policy implementation (Peng 2011). Since 2014, Nature journals require authors to provide a “Code Availability” statement indicating how code or algorithms could be accessed. However, peer reviewing code can be cumbersome. *Nature Methods*, *Nature Biotechnology* and *Nature Machine Intelligence* are

now running a trial in partnership with Code Ocean to enable authors to share computer code and facilitate its peer review while the manuscripts are under consideration for publication. Unfortunately, the trial is optional (Pastrana and Swaminathan, 2018). A similar voluntary replication review process is available in *ACM Transactions on Mathematical Software* (Heroux 2015). The *FEBS Journal*, *IET Systems Biology*, *Microbiology* and *Metabolomics* are collaborating with JWS Online to replicate models to be published in the journals, but the model replication is limited to ordinary differential equation kinetic models of biochemical pathways (Snoep 2005).

An empirical survey of 1,576 scientists suggests that 70% of scientists have tried and failed to replicate another scientist's findings (Baker 2016). I believe that most replication problems are bona fide results. These are scientific findings obtained by the original authors of a paper. However, replication is not possible for primarily two reasons. First, mathematical and computational biologists are not keeping a proper record how they obtained their results. Secondly, they are failing to report methodological information necessary for other researchers to replicate the results. The first problem can be addressed with a laboratory notebook policy in mathematical and computational biology research groups (Schnell 2015). The second requires mandatory replicability analysis of papers submitted for publication in mathematical and computational biology journals.

The latter suggestion seems radical, but it is not unprecedented in science. Since its inception 1921, *Organic Syntheses*⁴ publishes detailed and practical procedures for the synthesis of organic compounds. A unique feature of this journal is that every reaction and experimental procedure published is replicated in the laboratory of a member of the Board of Editors of *Organic Syntheses*. The Board is composed from the most distinguished members of the synthetic organic chemistry community. This ensures that the replicability experiments are supervised by the leading figures in the field. Submissions to *Organic Syntheses* need to provide a comprehensive list of the experimental protocols to allow replication. If any aspect of the procedures is unclear or ambiguous, authors are asked to provide additional details before the replication begins. The authors are also consulted for advice and assistance if problems are encountered during the replication experiments. Any article accepted for publication has been replicated at least twice in the checking editor's research group. In case that there is a slight difference in the replication results, the replication editor's results are reported in the final article (Danheiser 2011). Despite all these replicability measures, 7.5% of the articles submitted for publication to *Organic Syntheses* between 2010–2016 were rejected due to the inability of editor's labs to replicate the results shown in the papers (Bergman and Danheiser 2016).

Mathematical and computational biology papers can be the subject of identical replicability experiments since data, computer algebra calculations, computer code and software can be made available to the editors upon submission. Distinguished members of the Editorial Board should be responsible for maintaining high standards by leading replication efforts, while evaluating the adequacy of mathematical and computational methods to support the conclusions of papers. Simultaneously, expert reviewers should evaluate the significance,

⁴<http://www.orgsyn.org/>

quality, relevance, and clarity of the approach and findings of manuscripts under consideration.

The culture of replicability requires members of the editorial board to become responsible for the gold standard of science. The ultimate responsibility of editors and reviewers is not serving as gatekeepers of novelty and significance, but carefully examining and evaluating the validity of research findings and whether those findings support the conclusions presented in original research papers.

The modernization of the research paper can create a culture of replicability

Academic publishers also need to modernize the publication of the research paper to catalyze the replicability culture. Technological advances are making science more sophisticated and complex to communicate. Nowadays research papers are longer, contain more data and leave critical methods and research findings in supplementary materials, repositories and databases. Consequently, research papers are more difficult to review for publication. The publication process is to blame as the method for communicating scientific findings has not changed since the 1600s (Sommers 2018).

In the experimental methodological sciences, the publication model is evolving through Peer Reviewed Scientific Video Journal.⁵ Chemists are also starting to routinely record experiments with video cameras to make research more accessible and “reproducible” (Björnmalm et al 2016).

Can we redesign the journal article in the mathematical sciences? Yes, we can. We already have the right tools in our computers, but they are not currently used to describe procedures and algorithms, and share results in research papers. They are used as pedagogical tools to teach mathematical and computational modeling in the classroom. MapleSoft™ Maple Workshops, Wolfram Mathematica Notebook and MathWorks MATLAB® live scripts are interactive documents that combine scripts from those programs with formatted text, equations, and images to visually explore and analyze problems in the classroom. These tools can be used to write, execute and test code in an interactive environment to describe and explain scientific findings dynamically. A scientific paper submitted in the form of a “computational notebook” will be easy to replicate during the peer review processes, because a reader can execute the model, while they are reading the paper. This is something that we cannot achieve with the PDF file, because it provides a static picture of research findings by “modeling a piece of paper”. Unfortunately, the proprietary nature of Maple, Mathematica and MATLAB will make them hard to adopt broadly as a publication tool.

It is time to adopt open source web-based applications that allows us to create and share our research with live computer code, equations, text and visualizations. In the same way that TeX and LaTeX revolutionized typesetting and publishing for mathematical and computational scientists, the Jupyter Project⁶, through its Jupyter Notebook and JupyterLab,

⁵<https://www.jove.com/>

can be the modern solution to typeset and publish scientific papers in mathematical and computational sciences. The Jupyter Project supports the live execution of over 100 programming languages, including computer algebra algorithms and software, which are regularly used in mathematical and computational biology. Currently, Jupyter Notebooks are used as electronic notebooks, educational modeling laboratories, and supplemental material to scientific articles. It is time to extend the current uses of Jupyter Notebooks for academic publishing. Jupyter Notebooks allows us to typeset our research papers, share openly our computer codes, and present dynamically our research (Sommers 2018).

This proposal is highly ambitious, and maybe utopic on this point. It requires changing the nature of science communication, which has relied on publishing research using paper (or an electronic paper-like form) for over 400 years. Mathematical and computational scientists and academic publishers will need to be retrained to communicate research findings using web-based applications. The investment in the technology seems to be minimal as it is already available and accessible to many. However, the modernization of the publication technology will make mathematical and computational research more open, transparent and replicable by making the scientific paper coming alive through the “computational notebooks”.

Acknowledgements

I am very grateful for the helpful insights provided by Rick Danheiser (Editor-in-Chief of Organic Syntheses and Massachusetts Institute of Technology, USA), Edmund Crampin (University of Melbourne, Australia), and Wylie Stroberg (University of Michigan). This work was partially supported through the educational programs funded by NIGMS (T32 GM008322) and NIDDK (R25 DK088752).

References

- Alberts B, Stodden V, Young S, Choudhury S (2013) Testimony on scientific integrity & transparency. Available online at <https://science.house.gov/legislation/hearings/subcommittee-research-scientific-integrity-transparency>. Accessed August 10, 2018
- Association for Computing Machinery (2016). Artifact Review and Badging. Available online at: <https://www.acm.org/publications/policies/artifact-review-badging>. Accessed August 10, 2018
- Baker M (2016) 1,500 scientists lift the lid on reproducibility. *Nature* 533:452–454. [PubMed: 27225100]
- Bakker M, Wicherts JM (2011) The (mis)reporting of statistical results in psychology journals. *Behav Res Methods* 43:666–678. [PubMed: 21494917]
- Begley CG, Ellis LM (2012) Drug development: Raise standards for preclinical cancer research. *Nature* 483:531–533 [PubMed: 22460880]
- Bergman RG, Danheiser DL (2016) Reproducibility in chemical reaction. *Angew Chem Int Ed* 55:12548–12549
- Björnmalm M, Faria M, Caruso F (2016) Increasing the impact of materials in and beyond bio-nano science. *J Am Chem Soc* 138:13449–13456 [PubMed: 27672703]
- Bollen K, Cacioppo JT, Kaplan R, Krosnick J, Olds JL (2015) Social, behavioral, and economic sciences perspectives on robust and reliable science. National Science Foundation, Arlington, Virginia Available online at: https://www.nsf.gov/sbe/SBE_Spring_2015_AC_Meeting_Presentations/Bollen_Report_on_Replicability_SubcommitteeMay_2015.pdf. Accessed August 10, 2018

⁶<http://jupyter.org>

- Button KS, Ioannidis JP, Mokrysz C, Nosek BA, Flint J, Robinson ES, Munafò MR. (2013) Power failure: why small sample size undermines the reliability of neuroscience. *Nat. Rev. Neurosci* 14:365–376 [PubMed: 23571845]
- Chelliah V, Juty N, Ajmera I, Ali R, Dumousseau M, Glont M, Hucka M, Jalowicki G, Keating S, Knight-Schrijver V, Lloret-Villas A, Natarajan KN, Pettit JB, Rodriguez N, Schubert M, Wimalaratne SM, Zhao Y, Hermjakob H, Le Novère N, Laibe C (2015) BioModels: Ten-year anniversary. *Nucleic Acids Res* 43:D542–D548 [PubMed: 25414348]
- Crook S, Davison AP, Plesser HE (2013) Learning from the past: approaches for reproducibility in computational neuroscience In: Bower JM (ed) 20 Years in Computational Neuroscience. Springer, New York, pp. 73–102
- Collins FS, Tabak LA (2014) Policy: NIH plans to enhance reproducibility. *Nature* 505:612–613 [PubMed: 24482835]
- Danheiser DL (2011) Organic Syntheses: The “Gold Standard” in experimental synthetic organic chemistry. *Org Synth* 88:1–3
- Donoho DL, Maleki A, Rahman IU, Shahram M, and Stodden V (2009). 15 Years of reproducible research in computational harmonic analysis. *Comput. Sci. Eng* 11, 8–18.
- John LK, Loewenstein G, Prelec D (2012) Measuring the prevalence of questionable research practices with incentives for truth telling. *Psychol. Sci* 23:524–532 [PubMed: 22508865]
- Fanelli D (2009) How many scientists fabricate and falsify research? A systematic review and meta-analysis of survey data. *PLoS One* 4:e5738. [PubMed: 19478950]
- Fanelli D (2010) “Positive” results increase down the Hierarchy of the Sciences. *PLoS ONE* 5:e10068 [PubMed: 20383332]
- Fanelli D, Costas R, Ioannidis JPA (2017) Meta-assessment of bias in science. *Proc Natl Acad Sci USA* 114:3714–3719. [PubMed: 28320937]
- Fanelli D (2018) Opinion: Is science really facing a reproducibility crisis, and do we need it to? *Proc Natl Acad Sci USA* 115:2628–2631. [PubMed: 29531051]
- Goodman SN, Fanelli D, Ioannidis JPA (2016) What does research reproducibility mean? *Sci Transl Med* 8:341ps12.
- Heroux M (2015) Editorial: ACM TOMS Replicated Computational Results Initiative. *ACM Trans Math Softw* 41:art13
- Hucka M, Finney A, Sauro HM, Bolouri H, Doyle JC, Kitano H, Arkin AP, Bornstein BJ, Bray D, Cornish-Bowden A, Cuellar AA, Dronov S, Gilles ED, Ginkel M, Gor V, Goryanin II, Hedley WJ, Hodgman TC, Hofmeyr JH, Hunter PJ, Juty NS, Kasberger JL, Kremling A, Kummer U, Le Novère N, Loew LM, Lucio D, Mendes P, Minch E, Mjolsness ED, Nakayama Y, Nelson MR, Nielsen PF, Sakurada T, Schaff JC, Shapiro BE, Shimizu TS, Spence HD, Stelling J, Takahashi K, Tomita M, Wagner J, Wang J (2003) The systems biology markup language (SBML): A medium for representation and exchange of biochemical network models, *Bioinformatics* 19: 524–531 [PubMed: 12611808]
- Hirschhorn JN, Lohmueller K, Byrne E, Hirschhorn K (2002) A comprehensive review of genetic association studies. *Genet Med* 4:45–61. [PubMed: 11882781]
- International Organization for Standardization (1994) Applications of statistical methods. Technical Committee ISO/TC 69, Subcommittee SC 6, Measurement methods and results. Available online at: <https://www.iso.org/obp/ui/#iso:std:iso:5725:-1:ed-1:v1:en>. Accessed August 10, 2018
- International Union of Pure and Applied Chemistry (1997) IUPAC Compendium of Chemical Terminology 2nd ed. Compiled by AD McNaught and A Wilkinson. Blackwell Scientific Publications, Oxford Available online at: <https://goldbook.iupac.org/>. Accessed August 10, 2018
- Ioannidis JPA (2005) Why most published research findings are false. *PLoS Med* 2:e124 [PubMed: 16060722]
- Ioannidis JPA (2014) How to Make More Published Research True. *PLoS Med* 11: e1001747. [PubMed: 25334033]
- Ioannidis JPA, Stanley TD, Doucouliagos H (2017) The power bias in economics research. *Econ J* 127:F236–265.
- Joint Committee for Guides in Metrology (2006). International Vocabulary of Metrology: Basic and General Concepts and Associated Terms. 3rd Ed. Joint Committee for Guides in Metrology/

Working Group 2 Available online at: <https://www.nist.gov/sites/default/files/documents/pml/div688/grp40/International-Vocabulary-of-Metrology.pdf>. Accessed August 10, 2018

- John LK, Loewenstein G, Prelec D (2012) Measuring the prevalence of questionable research practices with incentives for truth telling. *Psychol Sci* 23:524–532. [PubMed: 22508865]
- Kronick DA (1976) *History of Scientific and Technical Periodicals*. 2nd ed. Scarecrow Press, Metuchen, New Jersey.
- Le Novère N, Finney A, Hucka M, Bhalla US, Campagne F, Collado-Vides J, Crampin EJ, Halstead M, Klipp E, Mendes P, Nielsen P, Sauro H, Shapiro B, Snoep JL, Spence HD, Wanner BL. (2005) Minimum information requested in the annotation of biochemical models (MIRIAM). *Nat Biotechnol* 23:1509–1515 [PubMed: 16333295]
- Li C, Donizelli M, Rodriguez N, Dharuri H, Endler L, Chelliah V, Li L, He E, Henry A, Stefan MI, Snoep JL, Hucka M, Le Novère N, Laibe C (2004) BioModels Database: An enhanced, curated and annotated resource for published quantitative kinetic models. *BMC Syst Biol* 4:92
- Lloyd CM, Halstead MD, Nielsen PF (2004) CellML: Its future, present and past, *Progr Biophys Mol Biol* 85: 433–450
- Lloyd CM, Lawson JR, Hunter PJ, Nielsen PF (2008) The CellML Model Repository. *Bioinformatics* 24:2122–2123. [PubMed: 18658182]
- Loew LM, Schaff JC (2001) The Virtual Cell: a software environment for computational cell biology. *Trends Biotechnol* 19:401–406 [PubMed: 11587765]
- Macleod MR, Michie S, Roberts I, Dirnagl U, Chalmers I, Ioannidis JPA, Salman RAH, Chan AW, Glasziou P (2014) Biomedical research: increasing value, reducing waste. *Lancet* 383:101–104. [PubMed: 24411643]
- McNutt M (2014) Journals unite for reproducibility. *Science* 346:6210
- Medley JK, Choi K, Konig M, Smith L, Gu S, Hellerstein J, Sealfon SC, Sauro HM (2018) Tellurium notebooks – An environment for reproducible dynamical modeling in systems biology. *PLoS Comput Biol* 14:e1006220 [PubMed: 29906293]
- Munafò MR, Nosek BA, Bishop DVM, Button KS, Chambers CD, Percie du Sert N, Simonsohn U, Wagenmakers EJ, Ware JJ, Ioannidis JPA (2017) A manifesto for reproducible science. *Nat Hum Behav* 1:0021.
- National Institutes of Health (2017) Principles and guidelines for reporting preclinical research. Available online at: <https://www.nih.gov/research-training/rigor-reproducibility/principles-guidelines-reporting-preclinical-research>. Accessed August 10, 2018
- National Institutes of Health (2018) Rigor and Reproducibility. Available online at: <https://grants.nih.gov/reproducibility/index.htm>. Accessed August 10, 2018
- Nature Editorial (2014) Journals unite for reproducibility. *Nature* 515:7
- Olivier BG, Snoep JL (2004) Web-based kinetic modelling using JWS Online. *Bioinformatics* 20:2143–2144 [PubMed: 15072998]
- Pashler H, Wagenmakers EJ (2012) Editors' introduction to the special section on replicability in psychological science: A crisis of confidence? *Perspect Psychol Sci* 7: 528–530. [PubMed: 26168108]
- Pastrana E, Swaminathan S (2018) Nature Research journals trial new tools to enhance code peer review and publication. of Scheme and Memes. A community blog from nature.com. 01 8 2018, 15:05 BST. Available online at: <http://blogs.nature.com/ofschemesandmemes/2018/08/01/nature-research-journals-trial-new-tools-to-enhance-code-peer-review-and-publication>. Accessed August 10, 2018
- Peng RD (2009) Reproducible research and Biostatistics. *Biostatistics* 10:405–408 [PubMed: 19535325]
- Peng RD (2011) Reproducible research in computational science. *Science* 334:1226–1227 [PubMed: 22144613]
- Plesser HS (2018) Reproducibility vs. Replicability: A brief history of a confused terminology. *Front. Neuroinform.* 11:76 [PubMed: 29403370]
- Prinz F, Schlange T, Asadullah K (2011) Believe it or not: How much can we rely on published data on potential drug targets? *Nat Rev Drug Discov* 10:712. [PubMed: 21892149]

- Redish AD, Kummerfeld E, Morris RL, Love AC (2018) Reproducibility failures are essential to scientific inquiry. *Proc Natl Acad Sci USA* 115:5042–5046. [PubMed: 29765001]
- Renear AH, Sacchi S, Wickett KM (2010) Definitions of dataset in the scientific and technical literature. *Proc Am Soc Inf Sci Technol* 47:1–4.
- Rougier NP, Hinsin K, Alexandre F, Arildsen T, Barba LA, Benureau FCY, et al. (2017). Sustainable Computational Science: The ReScience Initiative. Available online at: <https://arxiv.org/abs/1707.04393>. Accessed August 10, 2018
- Schnell S (2015) Ten simple rules for a computational biologist's laboratory notebook. *PLoS Comput Biol* 11: e1004385. [PubMed: 26356732]
- Shapin S, Schaffer S (1985) *Leviathan and the air-pump: Hobbes, Boyle, and the experimental life*. Princeton University Press, Princeton.
- Shapiro MF, Charrow RP (1989) The role of data audits in detecting scientific misconduct. Results of the FDA program. *JAMA* 261:2505–2511 [PubMed: 2704109]
- Snoep JL (2005) The Silicon Cell initiative: working towards a detailed kinetic description at the cellular level. *Curr Opin Biotechnol* 16:336–343 [PubMed: 15922580]
- Sommers J (2018) The scientific paper is obsolete. *The Atlantic*, 4 5, 2018 Available online at: <https://www.theatlantic.com/science/archive/2018/04/the-scientific-paper-is-obsolete/556676/>. Accessed August 10, 2018.
- Stark PB (2018) Before reproducibility must come preproducibility. *Nature* 557:613 [PubMed: 29795524]
- Steneck NH (2006) Fostering integrity in research: Definitions, current knowledge, and future directions. *Sci Eng Ethics* 12:53–74. [PubMed: 16501647]
- Stodden V, Mitchell I, LeVeque R (2012) Reproducible research for scientific computing: Tools and strategies for changing the culture. *Computing in Science and Engineering* 14:13–17.
- Stodden V, Guo P, Ma Z (2013) Toward reproducible computational research: An empirical analysis of data and code policy adoption by journals. *PLoS One* 8:e67111. [PubMed: 23805293]
- Stodden V, Seiler J, Ma Z (2018) An empirical analysis of journal policy effectiveness for computational reproducibility. *Proc Natl Acad Sci USA* 115:2584–2589 [PubMed: 29531050]
- Swat M, Thomas GL, Belmonte JM, Shirinifard A, Hmeljak D, Glazier JA (2012) Computational Methods in Cell Biology. *Methods Cell Biol* 110:325–366 [PubMed: 22482955]
- Bavel Van (2016) Why do so many studies fail to replicate? *New York Times*, May 27, 2016, page SR10 Available online at: <https://www.nytimes.com/2016/05/29/opinion/sunday/why-do-so-many-studies-fail-to-replicate.html>. Accessed August 10, 2018.
- Van Bavel JJ, Mende-Siedlecki P, Brady WJ, Reinero DA (2016) Contextual sensitivity in scientific reproducibility. *Proc Natl Acad Sci USA* 113:6454–6459 [PubMed: 27217556]
- Wicherts JM, Borsboom D, Kats J, Molenaar D (2006) The poor availability of psychological research data for reanalysis. *Am. Psychol* 61: 726–728 [PubMed: 17032082]
- Wilson G, Bryan J, Cranston K, Kitzes J, Nederbragt L, Teal TK (2017) Good enough practices in scientific computing. *PLoS Comput Biol* 13(6): e1005510. [PubMed: 28640806]
- Wood P, Randall D (2016) How bad is the government's science? *Wall Street Journal*, 4 16, 2018 5:56 ET Available online at <https://www.wsj.com/articles/how-bad-is-the-governments-science-1523915765>. Accessed August 10, 2018
- Yale RoundTable Participants (2010) Reproducible research: Addressing the need for data and code sharing in computational science. *Computing in Science and Engineering* 12:5