Reproducibility: you can do data analysis without it, but should you?

Tiffany Timbers (UBC)

2022-08-06

bit.ly/timbers-jsm-2022

What is reproducibility?

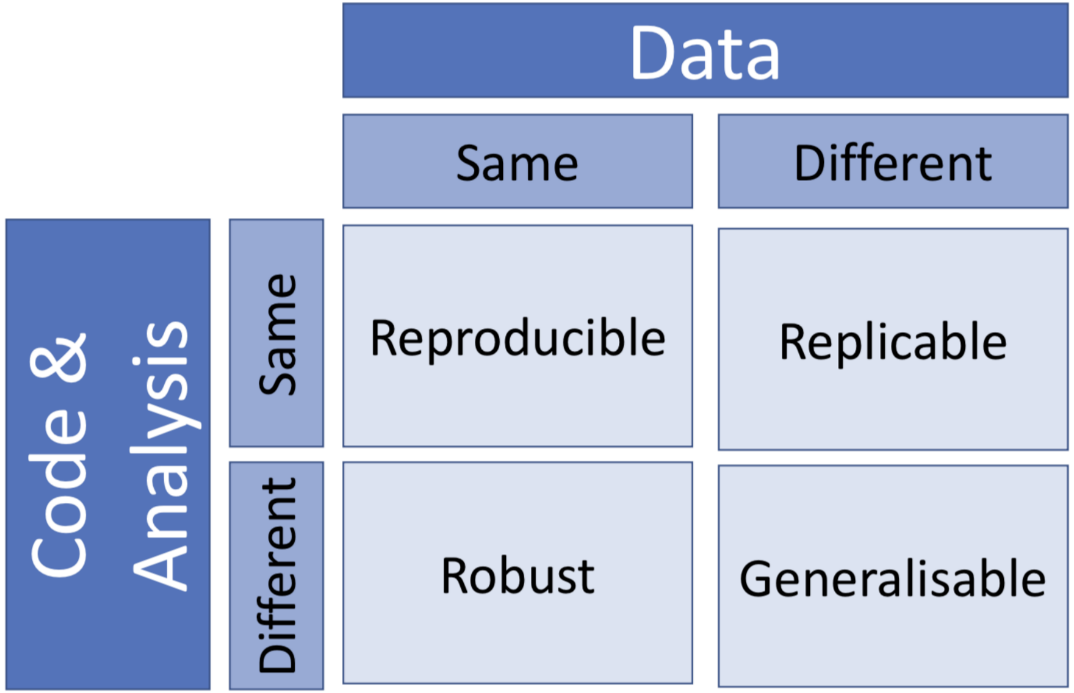
Reproducibility is reaching the same result given the same input, computational methods, and conditions[1]

Where:

input = data
computational methods = computer code
conditions = computational environment

[1] Committee on Reproducibility and Replicability in Science (2019), Reproducibility and Replicability in Science, Washington, D.C.: National Academies Press. https://doi.org/10.17226/25303

This is distinct from replicability, robust, and generalizable.



		Data	
		Same	Different
Code & Analysis	Same	Reproducible	Replicable
	Different	Robust	Generalisable

Source: Joelle Pineau et al., (2020). Improving Reproducibility in Machine Learning Research (A Report from the NeurIPS 2019 Reproducibility Program). arXiv:2003.12206

How does this fit in with a "Theory and Methods for Building Successful Data Analyses"?

Reproducible data analysis can "help" make an analysis successful by helping build "trust" in the analysis. This trust is important for stakeholders, of whom the application of the analysis impacts.

For example, it can:

1. provide evidence that the results or product could be regenerated given the same input computational methods, and conditions

2. is transparent - a human can audit all steps taken during analysis

3. if also version controlled, there is also a complete record of how and why analysis decisions were made

How does this fit in with a "Theory and Methods for Building Successful Data Analyses"?

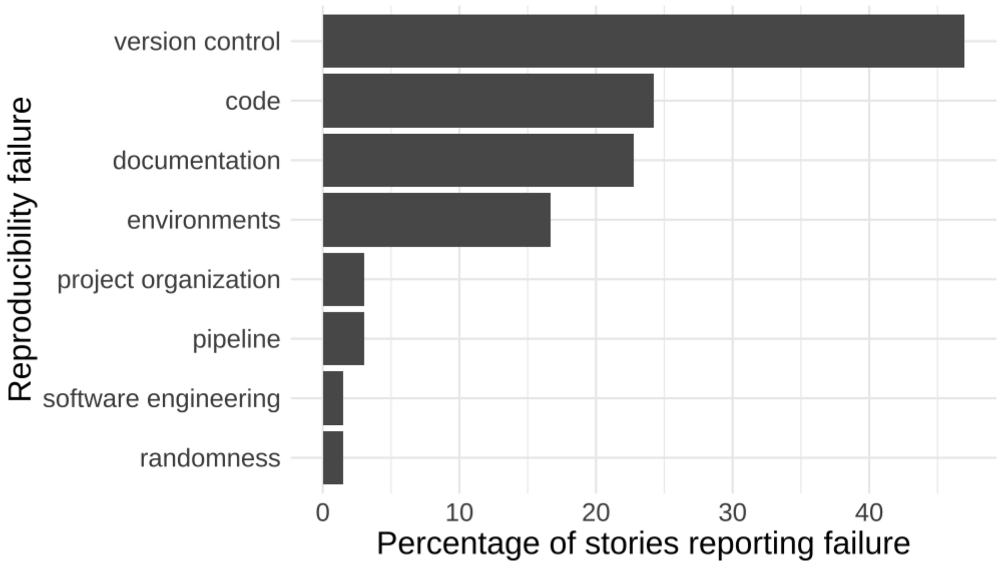
- Reproducible data analysis can "help" make an analysis successful by helping with efficiency of performing the analysis.
- This is less "important" from an ethical lense, but extremely important from a financial/business perspective.
- This appears to be the most obvious pain point for the data scientist and organization when reproducibility failures occur.

Reproducibility failures as reported by Data Scientists

We asked graduate and undergraduate Data Scientists (in training) for real life stories of data analysis workflow challenges and failures they had faced in the past (either at school or on the job).

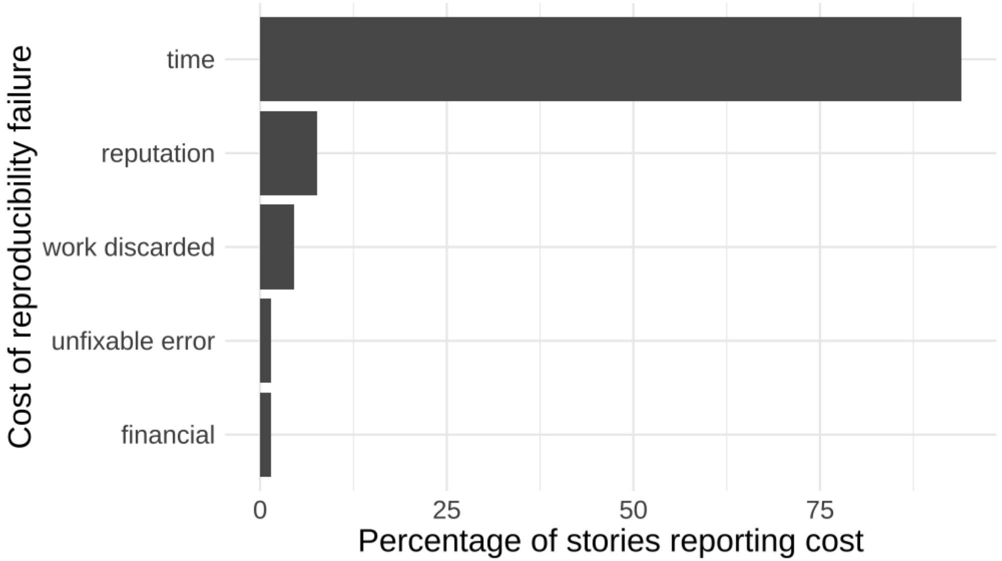
From the 66 stories we collected, four main themes related to failures in reproducibility emerged:

1. version control
2. code
3. documentation
4. environments



Impacts of reproducibility failures as reported by Data Scientists

These 66 stories also described the impacts as well as the reproducibility failures, with the most common by far being loss of time.



Some representative stories...

"When I was working with my last company, my team had a large dataset stored regarding some traffic data in the database and I wrote a script to retrieve the data and do data analysis in a single jupyter notebook, without putting proper documentation nor comments. The project was then tabled for a few months. But then a summer intern came and was assigned to pick up the project, so I had to explain everything in the notebook to her but I had totally no idea what I wrote since the logic was not quite straight-forward. It was also a hard time to set the environment up in her laptop since we had a bunch of libraries and dependencies used in the notebook. We finally took a week or two to set the environment up and transfer the knowledge to the intern, which was really a waste of time."

-- Master of Data Science student, UBC

- Reproducibility failure > insufficient documentation
- Impact of reproducibility failure > loss of time

"In my line of work, my colleagues and I heavily relied on SQL queries to withdraw data from database. She always forgot to save the queries while just sending me an excel file to read. Sometimes (Most of the time) when I cross checked the data, the result didn't match. It usually required hours of troubleshooting to figure out where the gap was between my query and her data. ¯_(ツ)_/¯"

-- Master of Data Science student, UBC

- Reproducibility failure > methods not recorded as code
- Impact of reproducibility failure > loss of time

“When I was working in my last job creating a database, I was working between two different computer systems. I had written the code on my local computer (a windows) but it was being deployed on a Linux command line system. To debug my code it was easier to do on my computer but certain lines of code were specific to the linux machine vs. the window machine. While I was using version control I had a really hard time keeping track of which script was which and it caused me to waste lots of time and energy trying to fix the issue”

-- Master of Data Science student, UBC

- Reproducibility failure > lack of a shippable and shareable computational environment
- Impact of reproducibility failure > loss of time

Levels that non-reproducible analysis can have an impact on

- Data Scientist
- Organization
- Beyond (e.g., Science & Society)

Potential impacts on the Data Scientist

- Time, which means decreased productivity!
- Reputation! Embarrassment! Loss of confidence!

Potential impacts on the Organization

- Increased salary costs paid out (due to analysis taking increased time)
- Reputation! Embarrassment!
- Work that must be discarded
- Liability!

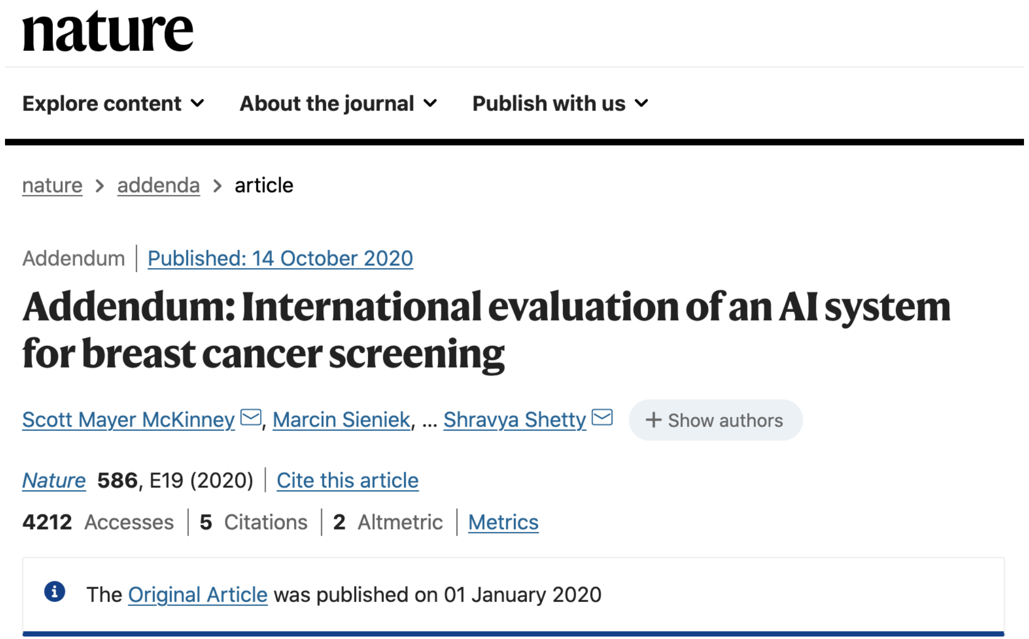
Potential impacts beyond

- Irreproducible (and thus, untrustworthy) analysis published in journals and taken as "truth"
- Retracted articles persisting in the literature from their citation previous to the retraction
- Delays in rollout of initiatives or products the analysis was used for

Two cases of analysis with reproducibility failures

1. McKinney et al. (2020a). International evaluation of an AI system for breast cancer screening. Nature. 577, 89–94 https://www.nature.com/articles/s41586-019-1799-6

2. Dabbous et al. (2021a). Safety and efficacy of favipiravir versus hydroxychloroquine in management of COVID-19: A randomised controlled trial. Scientific Reports. 11:7282 https://www.nature.com/articles/s41598-021-85227-0



nature

Explore content ⌄ About the journal ⌄ Publish with us ⌄

nature › addenda › article

Addendum | Published: 14 October 2020

Addendum: International evaluation of an AI system for breast cancer screening

Scott Mayer McKinney, Marcin Sieniek, ... Shravya Shetty + Show authors

Nature **586**, E19 (2020) | Cite this article

4212 Accesses | **5** Citations | **2** Altmetric | Metrics

The Original Article was published on 01 January 2020

- Original article
- Commentary
- Reply to commentary
- Addendum to article

A research paper was published in January 2020 that claimed they had created an artificial intelligence (AI) system that beat human experts in predicting breast cancer.

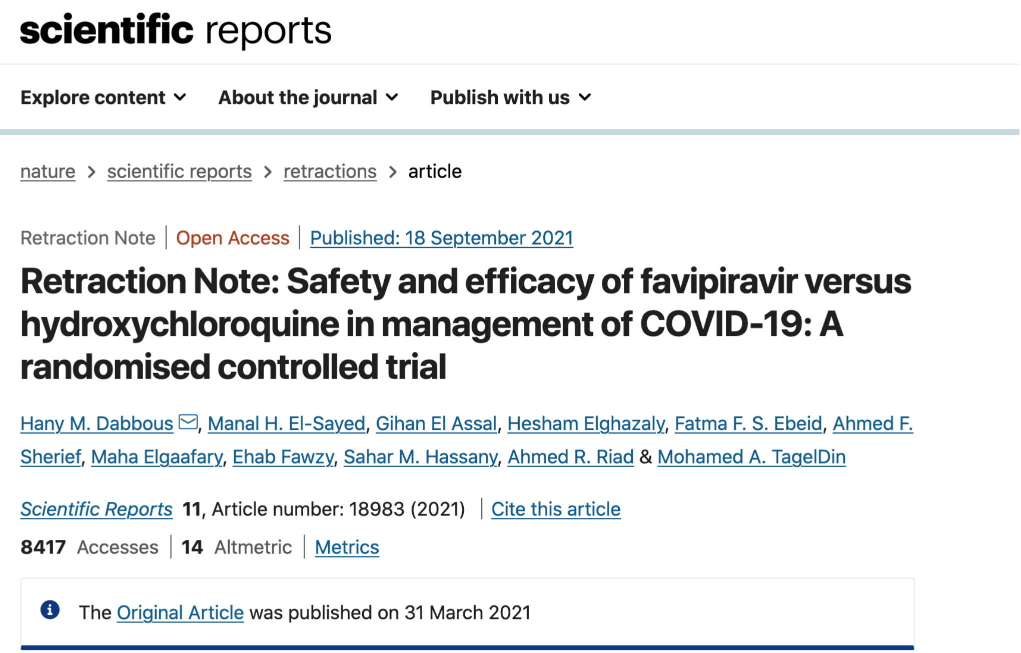
In October 2020, a matters arising commentary was published in response to the article, which in particular described many reproducibility issues with the original paper. The commentary also proposed recommendations to avoid such issues in the future.

In October 2020, the original authors published a reply to the commentary where they acknowledged some of the reproducibility issues and how they addressed them in an addendum also published that month, as well as attempted to justify why they were not able, or willing, to address other reproducibility issues pointed out in the commentary.

International evaluation of an AI system for breast cancer screening

- The reproducibility issues in this paper are not unique, and have been observed to occur in many machine learning (and other data analysis) papers [1].
- A positive result from this paper and commentary is the creation of a Machine Learning Reproducibility Checklist by Joelle Pineau
- The use of this checklist at NeurIPS 2019, a prominent machine learning conference, correlated with an increase in researchers including code with papers submitted to 74.4% (up from < 50% at NeurIPS 2018)[1].

[1] Pineau et al., (2020), Improving Reproducibility in Machine Learning Research (A Report from the NeurIPS 2019 Reproducibility Program) arXiv:2003.12206



Retraction Note | Open Access | Published: 18 September 2021

Retraction Note: Safety and efficacy of favipiravir versus hydroxychloroquine in management of COVID-19: A randomised controlled trial

Hany M. Dabbous, Manal H. El-Sayed, Gihan El Assal, Hesham Elghazaly, Fatma F. S. Ebeid, Ahmed F. Sherief, Maha Elgaafary, Ehab Fawzy, Sahar M. Hassany, Ahmed R. Riad & Mohamed A. TagelDin

Scientific Reports **11**, Article number: 18983 (2021) | Cite this article

8417 Accesses | **14** Altmetric | Metrics

The Original Article was published on 31 March 2021

A research paper was published in March 2021 that claimed that a drug, Favipiravir, was a safe and effective alternative to another drug, hydroxychloroquine (a medication commonly used to prevent or treat malaria), in mild or moderate COVID-19 infected patients.

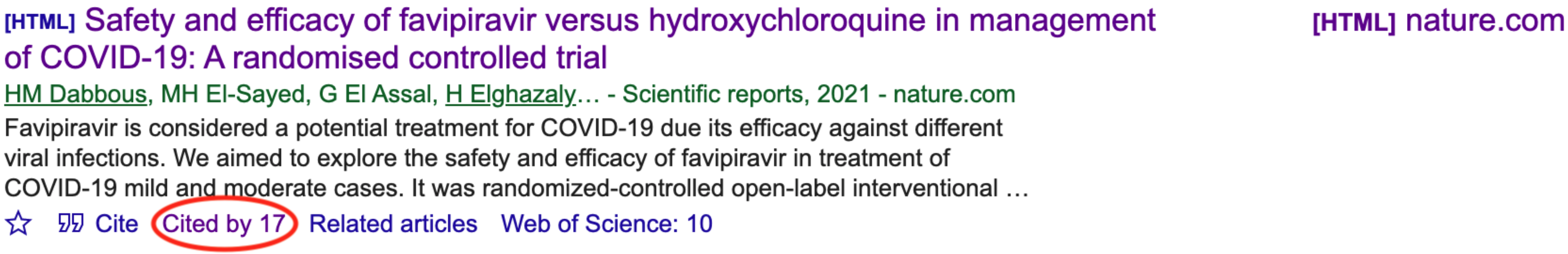
In September, 2021 the paper we retracted by the editors - in part due to reproducibility issues.

“After concerns were brought to the Editors’ attention after publication, the raw data underlying the study were requested. The authors provided several versions of their dataset. Post-publication peer review confirmed that none of these versions fully recapitulates the results presented in the cohort background comparisons, casting doubt on the reliability of the data. Additional concerns were raised about the randomisation procedure, as the equal distribution of male and female patients is unlikely unless sex is a parameter considered during randomisation. However, based on the clarification provided by the authors, sex was not considered during this process. The Editors therefore no longer have confidence in the results and conclusions presented.”

- Original article
- Retraction notice

Safety and efficacy of favipiravir versus hydroxychloroquine in management of COVID-19: A randomised controlled trial

The problem doesn't just stop once the article is retracted... Between the time the article was published and retracted, the article was cited 17 times!



[HTML] Safety and efficacy of favipiravir versus hydroxychloroquine in management of COVID-19: A randomised controlled trial

[HTML] nature.com

HM Dabbous, MH El-Sayed, G El Assal, H Elghazaly… - Scientific reports, 2021 - nature.com

Favipiravir is considered a potential treatment for COVID-19 due its efficacy against different viral infections. We aimed to explore the safety and efficacy of favipiravir in treatment of COVID-19 mild and moderate cases. It was randomized-controlled open-label interventional …

☆ ❞ Cite Cited by 17 Related articles Web of Science: 10

Summary

- Reproducibility means reaching the same result given the same data, code, and computational environment.

- Reproducibility should be included in a theory for successful data analysis because:

    1. It contributes to a trustworthy analysis
    2. It increases the efficiency of performing the analysis

- The most commonly reported reproducibility failures are version control, code, documentation and environments. The most commonly reported cost is time.

- Non-reproducible analysis can have an impact on:

    - Data Scientist
    - Organization
    - Beyond (e.g., Science & Society)

Acknowledgements

- UBC Master of Data Science students
- UBC DSCI 310 students
- Joel Ostblom, Assistant Professor of Teaching, Dept. of Statistics, UBC

References

1. Committee on Reproducibility and Replicability in Science (2019), Reproducibility and Replicability in Science, Washington, D.C.: National Academies Press. https://doi.org/10.17226/25303
2. Dabbous et al. (2021a). Safety and efficacy of favipiravir versus hydroxychloroquine in management of COVID-19: A randomised controlled trial. Scientific Reports. 11:7282 https://www.nature.com/articles/s41598-021-85227-0
3. Dabbous et al. (2021b). Retraction Note: Safety and efficacy of favipiravir versus hydroxychloroquine in management of COVID-19: A randomised controlled trial. Scientific Reports. 11:18983 https://www.nature.com/articles/s41598-021-98683-5
4. Haibe-Kains et al. (2020). Transparency and reproducibility in artificial intelligence. Nature. 586, E14–E16 https://www.nature.com/articles/s41586-020-2766-y
5. McKinney et al. (2020a). International evaluation of an AI system for breast cancer screening. Nature. 577, 89–94 https://www.nature.com/articles/s41586-019-1799-6
6. McKinney et al. (2020b). Reply to: Transparency and reproducibility in artificial intelligence. Nature. 586, E17–E18 https://www.nature.com/articles/s41586-020-2767-x
7. McKinney et al. (2020c). Addendum: International evaluation of an AI system for breast cancer screening. Nature. 586, E19 https://www.nature.com/articles/s41586-020-2679-9
8. Pineau, J. (2020). The Machine Learning Reproducibility Checklist https://www.cs.mcgill.ca/~jpineau/ReproducibilityChecklist.pdf
9. Pineau, J. et al., (2020). Improving Reproducibility in Machine Learning Research (A Report from the NeurIPS 2019 Reproducibility Program) arXiv:2003.12206

Reproducibility: you can do data analysis without it, but should you?

talk slides: *https://bit.ly/timbers-jsm-2022*