Research report – why?

Rob Davies

2023-11-14

## 1 PSYC401: Classes weeks 6-10

* My name is Dr Rob Davies, I am an expert in communication, individual differences, and methods

|  |
| --- |
| Tip |
| **Ask me anything**:   * questions during class on slido or just ask; * all other questions on discussion forum |

## 2 PSYC401: Classes weeks 6-10

We are working together to develop skills in a series of classes

1. *Week 6* — Research report
2. *Week 7* — Associations, hypotheses
3. *Week 8* — Visualizing data
4. *Week 9* — Linear models: predictions
5. *Week 10* — Linear models: development

## 3 This week: we focus on the research report, especially why

* **Why**: what is the motivation for the assignment?
* Explain this, to help to make sense of what we are doing, and what you can learn

## 4 How to work with the materials

* **Watch** these lectures for an outline
* **Read** the notes to get a detailed explanation
* **Follow** references to sources to build expertise

What to do and how to do it: see notes, and discuss in class

## 5 Why this? Key ideas

1. The difference between science “…being done, science in the making, and science already done, a finished product …” [Bourdieu (2004); p.2]
2. Published reported analyses are not *necessary* or *sufficient* to the data or the question

## 6 Wider context: Credibility revolution

Science (including psychological science) has undergone a rolling series of crises:

* Replicability or replication crisis (Pashler & Wagenmakers, 2012)
* Statistical crisis (Gelman & Loken, 2014)
* Generalizability crisis (Yarkoni, 2022)

## 7 Wider context: Credibility revolution

|  |
| --- |
| Nosek et al. (2022): replication outcomes for three systematic replication studies |

## 8 The triggers for crisis

* Failures to replicate influential claims (Nosek et al., 2022)
* Questionable research practices (John et al., 2012)
* Questionable measurement practices (Flake & Fried, 2020)
* Limited samples (Button et al., 2013; Henrich et al., 2010; Wild et al., 2022)

## 9 The credibility revolution: responses

* Pre-registration (Nosek et al., 2018, 2019) and registered reports (Nosek & Lakens, 2014)
* Replication studies (e.g. Aarts et al., 2015)
* Identification of open science principles (Munafò et al., 2017)

## 10 What is the motivation?

Opportunity to do original research work

* Important elements of the hard work in trying to make science work better is led by PhD students and junior researchers (e.g., Herndon et al., 2014)
* Identification of problems in the literature because of independent **post-publication review work**

|  |
| --- |
| Student Herndon discovered that a famous claim – countries with more debt have low growth – based on error |

## 11 Let’s take a break

* End of part 1

## 12 Multiverses: a fresh perspective

* We introduce the idea that your analysis work will flow through the stages of a *pipeline*
* From getting the data |> to presenting your findings
* Because, next, we examine how pipelines can *multiply*

## 13 Data analysis: we get from raw data to presenting results in stages

|  |
| --- |
| Figure 1: The data analysis pipeline or workflow |

## 14 The data analysis pipeline (workflow)

* Get some data
* Process or tidy the data
* Explore, visualize, and analyze the data
* Present or report your findings

|  |
| --- |
| Tip |
| * Identify the elements and the order in *your* work as the parts of a pipeline or the stages in a *workflow* |

## 15 The garden of forking paths

|  |
| --- |
| Figure 2: Forking paths in data analysis |

## 16 The garden of forking paths: pathways multiply at each choice

* It is possible to have **multiple potential different paths** from the data to the results
* The results may vary, depending on the path we take

## 17 Credibility and multiplicity

* Certainty over data processing or the uncertainty over analysis methods revealed in **multiverse analyses** (next)
* Plus limitations of data and code sharing, and incompleteness of reports
* Drives problems replicating or reproducing claims

## 18 Multiple pathways and secret lives

|  |
| --- |
| Carp (2012): Proportions of studies that reported using each of 21 procedures for data collection and analysis. |

## 19 Data processing is always necessary

* When you collect or access data for a research study
* The complete raw dataset you receive is *almost never* the complete dataset you analyze or whose analysis you report

## 20 Data multiverse analyses

* The impact of dataset construction choices on analysis results
* Same data constructed in different ways, given reasonable alternate choices
* p-values of tests vary for same tests across data versions

|  |
| --- |
| Steegen et al. (2016); figure 1 |

## 21 Data multiverse: lessons

* We approach our study with the same research question, and prediction
* We begin with the exact same data
* We could construct different datasets depending on *equally reasonable* choices
* As a result, we may see different results for the analyses of the different datasets

## 22 Analysis multiverse

Different researchers may adopt different approaches (Silberzahn & Uhlmann, 2015)

|  |
| --- |
| Silberzahn and Uhlmann (2015): Twenty-nine research teams reached a wide variety of conclusions using different methods on the same data set to answer the same question (about football players’ skin colour and red cards). |

## 23 Analysis multiverse

In multiverse analyses, we typically see variation in:

* how psychological constructs are operationalized;
* how data are processed or datasets constructed;
* *what* statistical techniques are used;
* *how* those techniques are used;
* with associated variation in results

## 24 Conclusions

If we see one analysis reported in a paper that does not mean only one analysis can reasonably be applied

|  |
| --- |
| Tip |
| If you read the methods or results section of a paper, you should reflect:   * What other analysis methods could be used here? * How could variation in analysis method — in what or how you do the analysis — influence the results? |

## 25 Let’s take a break

* End of Part 2

## 26 Kinds of reproducibility

(Gilmore et al., 2017; following Goodman et al., 2016) present three kinds of reproducibility:

* Methods reproducibility
* Results reproducibility
* Inferential reproducibility

We are mostly focused — in thinking about the **research report** — on methods and inferential reproducibility

## 27 Methods reproducibility

* Other researchers should be able to get the same results if they use the analysis methods with the same data

|  |
| --- |
| Hardwicke et al. (2021): Frequency of reproducibility outcomes by value type. |

## 28 Inferential reproducibility

If researchers repeat a study (results reproducibility) or re-analyze original data (methods reproducibility) then they *should* come to similar conclusions as original authors

But …

* reproducibility attempt could reveal problems, uncertainty over choices
* different researchers could apply different prior expectations over the probability of possible effects

## 29 The match between open science ideas and practices

The lessons learned from crises mean we now hope to see researchers:

* Share data and code;
* Publish research reports in ways that enable others to check or query analyses.
* What do we see, when we look at current practices?

## 30 Data and code sharing

* Analyses by Kidwell et al. (2016) and analyses reviewed by Nosek et al. (2022): study data increasingly available

|  |
| --- |
| Nosek et al. (2022): Yearly counts of users, sharing of files (research data, materials, code), and registration of studies on OSF and AsPredicted. |

## 31 Data and code sharing

It is great that data are shared but analyses show they are not always readily usable (Towse et al., 2021) but *should be*

* **completeness**: are all the data and the data descriptors supporting a study’s findings publicly available?
* **reusability**: how readily can the data be accessed and understood by others?

## 32 Enabling others to check or query analyses

Overall, the *good news* is

* for many published reports, *if* data are shared and *if* the shared data are accessible and reusable *then*
* most of the time, the researchers **can reproduce** the results presented by the original study authors (Hardwicke et al., n.d.; Hardwicke et al., 2018; Laurinavichyute et al., 2022; Minocher et al., n.d.; Obels et al., 2020; but see Crüwell et al., n.d.)

Yet studies of reproducibility identify challenges

## 33 Data challenges

1. Data Availability Statements or open science badges indicate data are shared but data may not be directly accessible
2. The data are shared and accessible but there may be missing or incorrect information *about* the data
3. Original study authors may share raw and processed data or just processed or just raw data
4. There may be mismatches between the variables referred to in the report and the variables named in the data file

## 34 Analysis challenges

1. The description of the analysis procedure may be incomplete or ambiguous
2. It may be difficult to identify the key analysis
3. It is easier to reproduce results if both data and code are shared but code is not always shared
4. Sometimes, analysis code is shared but it is difficult to use
5. Sometimes, there are errors in the analysis

## 35 Summary: why this?

1. We are in the middle of a credibility revolution
2. Focusing on data analysis, it is useful to think about the whole *data pipeline*
3. At every stage of the data pipeline, there are choices: forking paths
4. We can share data and enable others to check, query analysis choices

|  |
| --- |
| Tip |
| * Learning to do data analysis requires learning good workflow practices |

## 36 Opportunities

* The conventions of normal scientific practice can work so that if we encounter an *anomaly* we blame ourselves (Kuhn, 1970)
* Sometimes, if you think you have found an error or a problem in a published analysis or a shared dataset, remember: **you may be right** (Herndon et al., 2014)

## 37 What we learn: critical reading skills

* Learn to critically read research reports and data documentation
* Develop proficiency in analysis workflow
* Develop critical skills in analysis

## 38 What we learn: good practice sense

* The credibility revolution requires us to think about and to teach good open science practices that safeguard the value of evidence in psychology

## 39 What we learn: practical sense

* People make mistakes
* Different choices are often reasonable
* We *always* need to check the evidence

## 40 What we learn: do better

* Share data and code
* Publish research reports in ways that *enable* others to check or query analyses

## 41 References

Aarts, E., Dolan, C. V., Verhage, M., & Van der Sluis, S. (2015). Multilevel analysis quantifies variation in the experimental effect while optimizing power and preventing false positives. *BMC Neuroscience*, *16*(1), 1–15. <https://doi.org/10.1186/s12868-015-0228-5>

Bourdieu, P. (2004). *Science of Science and Reflexivity*. Polity.

Button, K. S., Ioannidis, J. P. A., Mokrysz, C., Nosek, B. A., Flint, J., Robinson, E. S. J., & Munafò, M. R. (2013). Power failure: Why small sample size undermines the reliability of neuroscience. *Nature Reviews Neuroscience*, *14*(5), 365–376.

Crüwell, S., Apthorp, D., Baker, B. J., Colling, L., Elson, M., Geiger, S. J., Lobentanzer, S., Monéger, J., Patterson, A., Schwarzkopf, D. S., Zaneva, M., & Brown, N. J. L. (n.d.). *What’s in a badge? A computational reproducibility investigation of the open data badge policy in one issue of psychological science*. <https://doi.org/10.31234/osf.io/729qt>

Flake, J. K., & Fried, E. I. (2020). Measurement Schmeasurement: Questionable Measurement Practices and How to Avoid Them. *Advances in Methods and Practices in Psychological Science*, *3*(4), 456–465. <https://doi.org/10.1177/2515245920952393>

Gelman, A., & Loken, E. (2014). The statistical crisis in science. *American Scientist*, *102*(6), 460–465. <https://doi.org/10.1511/2014.111.460>

Gilmore, R. O., Diaz, M. T., Wyble, B. A., & Yarkoni, T. (2017). Progress toward openness, transparency, and reproducibility in cognitive neuroscience. *Annals of the New York Academy of Sciences*, *1396*, 5–18. <https://doi.org/10.1111/nyas.13325>

Goodman, S. N., Fanelli, D., & Ioannidis, J. P. A. (2016). What does research reproducibility mean? *Science Translational Medicine*, *8*(341).

Hardwicke, T. E., Bohn, M., MacDonald, K., Hembacher, E., Nuijten, M. B., Peloquin, B. N., deMayo, B. E., Long, B., Yoon, E. J., & Frank, M. C. (n.d.). Analytic reproducibility in articles receiving open data badges at the journal psychological science: An observational study. *Royal Society Open Science*, *8*(1), 201494. <https://doi.org/10.1098/rsos.201494>

Hardwicke, T. E., Mathur, M. B., MacDonald, K., Nilsonne, G., Banks, G. C., Kidwell, M. C., Hofelich Mohr, A., Clayton, E., Yoon, E. J., Henry Tessler, M., Lenne, R. L., Altman, S., Long, B., & Frank, M. C. (2018). Data availability, reusability, and analytic reproducibility: evaluating the impact of a mandatory open data policy at the journal Cognition. *Royal Society Open Science*, *5*(8), 180448. <https://doi.org/10.1098/rsos.180448>

Henrich, J., Heine, S. J., & Norenzayan, A. (2010). The weirdest people in the world? *The Behavioral and Brain Sciences*, *33*(2-3). <https://doi.org/10.1017/S0140525X0999152X>

Herndon, T., Ash, M., & Pollin, R. (2014). Does high public debt consistently stifle economic growth? A critique of Reinhart and Rogoff. *Cambridge Journal of Economics*, *38*(2), 257–279. <https://doi.org/10.1093/cje/bet075>

John, L. K., Loewenstein, G., & Prelec, D. (2012). Measuring the prevalence of questionable research practices with incentives for truth telling. *Psychological Science*, *23*(5), 524–532. <https://doi.org/10.1177/0956797611430953>

Kidwell, M. C., Lazarević, L. B., Baranski, E., Hardwicke, T. E., Piechowski, S., Falkenberg, L. S., Kennett, C., Slowik, A., Sonnleitner, C., Hess-Holden, C., Errington, T. M., Fiedler, S., & Nosek, B. A. (2016). Badges to acknowledge open practices: A simple, low-cost, effective method for increasing transparency. *PLoS Biology*, *14*(5), 1–15. <https://doi.org/10.1371/journal.pbio.1002456>

Kuhn, T. S. (1970). *The structure of scientific revolutions* ([2d ed., enl). University of Chicago Press.

Laurinavichyute, A., Yadav, H., & Vasishth, S. (2022). Share the code, not just the data: A case study of the reproducibility of articles published in the Journal of Memory and Language under the open data policy. *Journal of Memory and Language*, *125*, 104332. <https://doi.org/10.1016/j.jml.2022.104332>

Minocher, R., Atmaca, S., Bavero, C., McElreath, R., & Beheim, B. (n.d.). Estimating the reproducibility of social learning research published between 1955 and 2018. *Royal Society Open Science*, *8*(9), 210450. <https://doi.org/10.1098/rsos.210450>

Munafò, M. R., Nosek, B. A., Bishop, D. V. M., Button, K. S., Chambers, C. D., Percie Du Sert, N., Simonsohn, U., Wagenmakers, E. J., Ware, J. J., & Ioannidis, J. P. A. (2017). A manifesto for reproducible science. *Nature Human Behaviour*, *1*(1), 1–9. <https://doi.org/10.1038/s41562-016-0021>

Nosek, B. A., Beck, E. D., Campbell, L., Flake, J. K., Hardwicke, T. E., Mellor, D. T., van?t Veer, A. E., & Vazire, S. (2019). Preregistration is hard, and worthwhile. *Trends in Cognitive Sciences*, *23*(10), 815–818.

Nosek, B. A., Ebersole, C. R., DeHaven, A. C., & Mellor, D. T. (2018). The preregistration revolution. *Proceedings of the National Academy of Sciences*, *115*(11), 2600–2606.

Nosek, B. A., Hardwicke, T. E., Moshontz, H., Allard, A., Corker, K. S., Dreber, A., Fidler, F., Hilgard, J., Kline Struhl, M., Nuijten, M. B., Rohrer, J. M., Romero, F., Scheel, A. M., Scherer, L. D., Schönbrodt, F. D., & Vazire, S. (2022). Replicability, Robustness, and Reproducibility in Psychological Science. *Annual Review of Psychology*, *73*, 719–748. <https://doi.org/10.1146/annurev-psych-020821-114157>

Nosek, B. A., & Lakens, D. (2014). Registered reports: A method to increase the credibility of published results. *Social Psychology*, *45*(3), 137–141. <https://doi.org/10.1027/1864-9335/a000192>

Obels, P., Lakens, D., Coles, N. A., Gottfried, J., & Green, S. A. (2020). Analysis of open data and computational reproducibility in registered reports in psychology. *Advances in Methods and Practices in Psychological Science*, *3*(2), 229–237. <https://doi.org/10.1177/2515245920918872>

Pashler, H., & Wagenmakers, E. J. (2012). Editors’ introduction to the special section on replicability in psychological science: A crisis of confidence? *Perspectives on Psychological Science*, *7*(6), 528–530. <https://doi.org/10.1177/1745691612465253>

Silberzahn, R., & Uhlmann, E. L. (2015). Crowdsourced research: Many hands make tight work. *Nature*, *526*(7572), 189–191. <https://doi.org/10.1038/526189a>

Towse, J. N., Ellis, D. A., & Towse, A. S. (2021). Opening Pandora’s Box: Peeking inside Psychology’s data sharing practices, and seven recommendations for change. *Behavior Research Methods*, *53*(4), 1455–1468. <https://doi.org/10.3758/s13428-020-01486-1>

Wild, H., Kyröläinen, A.-J., & Kuperman, V. (2022). How representative are student convenience samples? A study of literacy and numeracy skills in 32 countries. *PLOS ONE*, *17*(7), e0271191. <https://doi.org/10.1371/journal.pone.0271191>

Yarkoni, T. (2022). The generalizability crisis. *Behavioral and Brain Sciences*, *45*, e1. <https://doi.org/10.1017/S0140525X20001685>