

Reproducible Analyses

What is it and why should I care?

Daniela Palleschi

4/12/23

Table of contents

Replication	2
An example from language research	2
Reproducibility	2
Replication vs. Reproducibility	3
Open Science: Why should I care?	3
What can I do?	3
How to do better science	4
What will we learn here?	4
R is for Reproducibility	5
Exercises	6
R	6
RStudio	6
Quarto	6
Quarto cont'd	7

```
# Create references.json file based on the citations in this script
# make sure you have 'bibliography: references.json' in the YAML
# rbbt::bibt_update_bib("_reproducibility.qmd")
```

Replication

- replication refers to the re-running a previous experiment with as few differences as possible
 - aim: determine whether the original results were *robust* and are *replicable*
 - if yes, great! the original findings are reliable
 - if no, hmm, maybe the original findings were false positives? or due to some other factor?
- (ioannidis__why__2005?): “There is increasing concern that in modern research, false findings may be the majority or even the vast majority of published research claims”
- in recent years, researchers have tried to *replicate* classic studies in their field
 - but in many cases, they did not get the same effects the original study reported (and were famous for)
- this began the *replication crisis*

An example from language research

- (nieuwland__large-scale__2018?): a *direct* replication
 - a multi-lab replication of (delong__probabilistic__2005?)’s impactful paper
 - * (delong__probabilistic__2005?): reported N400 effects elicited at unexpected nouns, but also on preceding determiners (English *a/an*) when it signalled an unexpected word,
 - e.g., *The day was breezy so the boy went outside to fly...a kite/*an airplane*
 - taken as evidence of pre-activation of phonological form, graded by cloze probability
 - * (nieuwland__large-scale__2018?): replicated N400 at noun, but not at adjective
- Laurinavichyute, Yadav, and Vasishth (2022); ; Peels (2019); Penders, Holbrook, and de Rijcke (2019); T. Roettger et al. (2022); T. B. Roettger (2021); Sönning and Werner (2021); Kathawalla, Silverstein, and Syed (2021)

Reproducibility

- reproducibility refers to the ability to *reproduce* somebody’s analyses with their
 - data
 - *and* code

Replication vs. Reproducibility

- **replication** of a study
 - repeating an **experiment**
 - getting *similar* results
- **reproducibility** of analyses
 - repeating **analyses** of the *same data*
 - getting the *same* results
- e.g., when you submit a paper to a journal, they make ask for your data and code so reviewers can *reproduce* your analyses
 - requires data and code
- if you have interesting findings, other researchers (or future you) may want to *replicate* your study to see if they can *replicate* your findings
 - (may require) stimuli, set-up and presentation information, participant demographics

Open Science: Why should I care?

1. It's good science
 - Science is cumulative. We should ensure we're building on reliable, robust findings
2. Because the field cares
 - replication/reproducibility are beginning to be foregrounded by e.g., journals/job advertisements
3. Helps future you
 - pre-registration, reproducible analyses, clean and sharable data all help *future you*

What can I do?

- there are aspects of open science that we can choose to implement
- some suggestions from Kathawalla, Silverstein, and Syed (2021):

Level: Easy

1. Journal Club
2. Project workflow

3. Pre-prints

Level: Medium

4. Reproducible code
5. Sharing data
6. Transparent manuscripts
7. Pre-registration

Level: Difficult

8. Registered reports

How to do better science

- don't be afraid of making mistakes
 - (most) researchers aren't statisticians or programmers
 - do the best you can, and ***be transparent***
- doing *some* of the steps is better than doing *none*

What will we learn here?

Design and Reporting

- Preregistration/Registered Reports
- Transparent writing

Analysis

- Reproducible code
 - with open source software (R, RStudio, packages)
 - dynamic reports with Quarto/Rmarkdown
- Project workflow
 - folder structure
 - * how to sensibly set up your folders
 - contained environments
 - * using RProjects and the **here** package

```
knitr::include_graphics(here::here("mats/day1/2-reproducibility/media/Kathawalla_research_
```

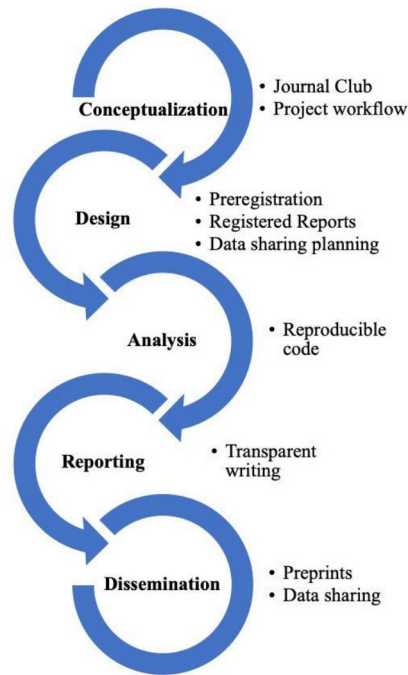


Figure 1. Open Science research practices across the research cycle

Figure 1: Image source: Kathawalla, Silverstein, and Syed (2021) (all rights reserved)

R is for Reproducibility

- we will be working with R, RStudio, Quarto, and RProjects
 - R: a programming language for statistical computing and graphics
 - RStudio: an integrated development environment (IDE)
 - * RStudio Desktop
 - * RStudio Server
 - Quarto (similar to Rmarkdown): dynamic reports
 - * combining text, code, and printed tables and figures
 - RProjects: a workflow tool
 - * contains all files necessary for a project
 - * works with *relative* file paths

Students: open R, then RStudio, then create an RProject Me: show them each on my computer
- R can run code and save a script - RStudio has so many more options (cheatsheet) - RProjects keep everything tidy and together

Exercises

R

1. Open R
 - write `3*32` and hit Enter
 - write `print("Hello World!")`
 - write `x <- 3*32`
 - write `y <- "Hello World!"`
 - write `x`
 - write `y`

RStudio

2. Open RStudio
 - locate the Environment, Files, and Console panes
 - File > New File > R script
 - write `3*32` and hit Enter
 - write `print("Hello World!")`
 - write `x <- 3*32`
 - write `y <- "Hello World!"`
 - write `x`
 - write `y`
 - add comments describing each step using `#`
 - File > Save As

Quarto¹

3. Create a new Quarto document
 - File > New File > Quarto Document
 - Read the instructions
 - Practice running the chunks individually
 - render the document

¹<https://r4ds.hadley.nz/quarto.html#workflow>

- verify that you can modify the code, re-run it, and see modified output
4. Create one new Quarto document for each of the three built-in formats: HTML, PDF and Word.
 - Render each of the three documents
 - How do the outputs differ?
 - How do the inputs differ?²

Quarto cont'd

- Choose a Quarto document:
 - give it a title, your name (author), and unclick ‘Use visual markdown editor’
- Render
- YAML:
 - title: "Eye-tracking during reading"
 - subtitle: "Lecture 2 notes"
 - author: [YOUR NAME]
 - date: ‘r Sys.Date()’
 - Render

References

- Kathawalla, Ummul-Kiram, Priya Silverstein, and Moin Syed. 2021. “Easing Into Open Science: A Guide for Graduate Students and Their Advisors.” *Collabra: Psychology* 7 (1): 18684. <https://doi.org/10.1525/collabra.18684>.
- Laurinavichyute, Anna, Himanshu Yadav, and Shravan Vasishth. 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: 12.
- Peels, Rik. 2019. “Replicability and Replication in the Humanities.” *Research Integrity and Peer Review* 4 (1): 2. <https://doi.org/10.1186/s41073-018-0060-4>.
- Penders, Holbrook, and de Rijcke. 2019. “Rinse and Repeat: Understanding the Value of Replication Across Different Ways of Knowing.” *Publications* 7 (3): 52. <https://doi.org/10.3390/publications7030052>.
- Roettger, Timo B. 2021. “Preregistration in Experimental Linguistics: Applications, Challenges, and Limitations.” *Linguistics* 59 (5): 1227–49. <https://doi.org/10.1515/ling-2019-0048>.

²You may need to install LaTeX in order to build the PDF output — RStudio will prompt you if this is necessary.

- Roettger, Timo, Agata Bochynska, Erin Buchanan, Joseph Casillas, Caitlin Halfacre, Liam Keeble, Melanie Röthlisberger, Irys-Amélie Champagne, and Kaidi Chen. 2022. “Reproducible Research Practices and Transparency in Linguistics,” August. <https://doi.org/10.17605/OSF.IO/J2Q5P>.
- Sönning, Lukas, and Valentin Werner. 2021. “The Replication Crisis, Scientific Revolutions, and Linguistics.” *Linguistics* 59 (5): 1179–1206. <https://doi.org/10.1515/ling-2019-0045>.