

**Quantitative Research on Representation:  
How to Replicate Published Research, Improve Transparency and Why This  
All Matters  
2020-03-15**

NOTE: Because of the Covid-19 pandemic, some information in this outline is provisional and might need to be adapted.

instructor: Prof. Ingo Rohlfing, PhD

office hours: Tuesday, 10am-12; by appointment; open door policy

NOTE: Because of the consequences of Covid-19, I only offer online office hours for now. You can send me an email with small questions or arrange a Skype call by email.

room: Herbert-Lewin-Str. 2, 313.c (right next to the staircase at the South of the building)

phone: +4922147089973

NOTE: Because of the consequences of Covid-19, I am practicing home office for now.

email: i.rohlfing@uni-koeln.de

first session: 9.4.2020 (NOTE: to be confirmed)

last session: 16.7.2020

no session: 21.5.2020; 4.6.2020; 11.6.2020

room: Institutshörsaal, Gottfried-Keller-Str. 6

time: 12.00-13.30 (see course plan for changes of course schedule)

NOTE: As of today, it is very likely that courses will take place online. Further information will follow ahead of time.

Registration for exam in KLIPS2 (for Master students)

Please also regularly check the CCCP information on teaching on the internet:

<http://www.cccp.uni-koeln.de/en/public/teaching/>

---

An important element for the verification of descriptive and causal inferences is that one is able to understand *how* a study arrived at the inferences. If one cannot fully reconstruct how the empirical analysis was implemented, one cannot fully judge the credibility of the claims derived from the analysis. This idea is central to recent calls for making empirical research *reproducible* and *replicable* and to reassess the quality of published empirical research. A study is reproducible when the same data and procedures that were used in an original study produce the same results. A study is replicable when new data and the same procedures as in the original study produce the same results.

Using *quantitative* research on representation as an example, this course starts with introducing participants to the reasons why published research might not be reproducible or replicable. Participants are introduced to problems such as HARKing and p-hacking and tools devised to diminish them. In the second part of the course, participants will learn simple tools and hands-on

practices for the program R that help making quantitative research more accessible and reproducible.

The main goal and task for the participants is to reproduce a published quantitative study substantively dealing with representation. This gives participants the opportunity to make first-hand experiences in how reproducibility assessments work and why they can be more difficult to implement than one might think. More information about the details of the exercise will be added in due time. The type of exam in this course is a *portfolio exam*. Participants will have to perform multiple, smaller tasks during the course such as selecting a published article for the reproduction analysis, setting up R an RStudio properly and more.

It is absolutely *essential* that participants have taken quantitative methods courses using R/RStudio before taking this course. This course is not about learning R from scratch and it is *strongly discouraged* to take the course without any expertise in R and quantitative methods. I recommend to only take this course if you have taken a quantitative R course with Prof. Proksch or Dr. Castanho Silva before. If you have doubts whether you have the required skills, please send me an email or drop by my office to discuss this.

## Topics and readings

### Part 1: What could go wrong with empirical research and what could be done about it

09.04.20: What open science is and why it is important

- Open science (data access and material)
  - Lupia, Arthur and Colin Elman (2014): Openness in Political Science: Data Access and Research Transparency. *PS: Political Science & Politics* 47 (1): 19-42.
- Examples for importance of data access
  - Ziemann, Mark, Yotam Eren and Assam El-Osta (2016): Gene Name Errors Are Widespread in the Scientific Literature. *Genome Biology* 17 (1): 177.
  - Gabelica, Mirko, Jakica Cavar and Livia Puljak (2019): Authors of Trials from High-Ranking Anesthesiology Journals Were Not Willing to Share Raw Data. *Journal of Clinical Epidemiology* 109: 111-116.

16.04.20: Replication and reproduction of empirical research

- Clarification of terms
  - Freese, Jeremy and David Peterson (2017): Replication in Social Science. *Annual Review of Sociology* 43 (1): 147-165.
- Examples for relevance:
  - Rohlfing, Ingo, Lea Königshofen, Susanne Krenzer, Ayjeren Rozyjumayeva and Jan Schwalbach (2020): A reproduction analysis of QCA studies. *typescript*.
  - Open Science Collaboration (2015): Estimating the Reproducibility of Psychological Science. *Science* 349 (6251)

23.04.20: Questionable research practices (QRPs) and countermeasures I

- QRPs
  - Neuroskeptic (2012): The Nine Circles of Scientific Hell. *Perspectives on Psychological Science* 7 (6): 643-644.
- HARKing
  - Kerr, Norbert L. (1998): Harking: Hypothesizing after the Results Are Known. *Personality and Social Psychology Review* 2 (3): 196-217.

#### 30.04.19: Questionable research practices and countermeasures II

- p-hacking
  - Franco, Annie, Neil Malhotra and Gabor Simonovits (2015): Underreporting in Political Science Survey Experiments: Comparing Questionnaires to Published Results. *Political Analysis* 23 (2): 306-312.
  - Gerber, Alan and Neil Malhotra (2008): Do Statistical Reporting Standards Affect What Is Published? Publication Bias in Two Leading Political Science Journals. *Quarterly Journal of Political Science* 3 (3): 313-326.
- Preregistration and preanalysis plans
  - Monogan, James E. III (2015): Research Preregistration in Political Science: The Case, Counterarguments, and a Response to Critiques. *Political Science & Politics* 48 (3): 425-429.

#### 07.05.20: Questionable research practices and countermeasures III

- Multiverse/specification curve
  - Simonsohn, Uri, Joseph A. Simmons and Leif D. Nelson (2015): Specification Curve: Descriptive and Inferential Statistics on All Reasonable Specifications. <https://dx.doi.org/10.2139/ssrn.2694998>
- Registered reports
  - Wiseman, Richard, Caroline Watt and Diana Kornbrot (2019): Registered Reports: An Early Example and Analysis. *PeerJ* 7: e6232.

#### 14.05.20: Recap and presentation of selected papers

- As an exercise in presenting papers and getting an idea about the papers some participants are working on, we will have presentations and discussions of two to three selected papers. More information about how the presentation should look like and how it will be determined who will present will be shared at a later point in time.

#### **21.05.20: NO COURSE (public holiday)**

#### Part 2: Skills and tools for enhancing reproducibility and replicability

This part is about learning some skills and tools for making code more accessible and reproducible. It has a more applied component and requires less reading of material.

#### 28.05.20: Organization of projects and script annotation and style I

- Setting up a project in RStudio
  - Gandrud, Christopher (2018): *Reproducible Research with R and R Studio (2nd Ed)*. New York: Chapman and Hall/CRC: sections 4.1-4.3.
- Code style: <https://style.tidyverse.org/>

#### **04.06.20: NO COURSE (Pentecost)**

#### **11.06.20: NO COURSE (public holiday)**

#### 18.06.20: Organization of projects and script annotation and style II

#### 25.06.20: Reproducible reports and results I

- Reproducible reports/RMarkdown
  - <https://rmarkdown.rstudio.com/>
- knitr
  - <https://yihui.org/knitr/>

#### 02.07.20: Reproducible reports and results II

- Reproducible reports/RMarkdown: <https://rmarkdown.rstudio.com/>
- knitr: <https://yihui.org/knitr/>

#### 09.07.20: Data sharing

- Open Science Framework: <https://osf.io/>
- Example for journal requirements: <https://ajps.org/ajps-verification-policy/>