

Anne Gärtner  
Faculty of Psychology

# Workshop Open Science Practices

## Part 1

Data Collection

# Overview

















## Time

13:00 – 13:10  
13:10 – 13:40  
13:40 – 13:55  
**13:55 – 14:40**  
14:40 – 14:55  
14:55 – 15:40  
15:40 – 16:10  
16:10 – 16:25  
16:25 – 16:40  
16:40 – 17:25  
17:25 – 17:40  
17:40 – 18:00

## Topic

Welcome  
Power (Alex)  
Discussion  
**Data Collection (Anne)**  
Break  
*P*-hacking (Anne)  
Publication Bias (Alex)  
Discussion  
Break  
Preregistration (Anne)  
Discussion  
Wrap Up, Evaluation

## Workshop material

- ◀  **MGK Open Science Module**
-  Registration
-  Introduction
- ▶  W1 - Good Scientific Practice
- ▶  W2 - Research Data Management
- ◀  W3 - Research Transparency
-  General Information
-  0. Introduction
-  1. Open Science
-  2. Open Access
-  3. Open Data, Materials, and Co
-  4. Reproducible Analyses
-  5. Preregistration
-  Opt.: Replication Research
-  **Workshop Slides**
-  Literaturverzeichnis

# Introduction



Bad data is like cracks in a foundation; building on it is beyond risky.

<https://3qdigital.com/blog/implications-poor-incomplete-data-collection/>

# Outline

## Issues to deal with

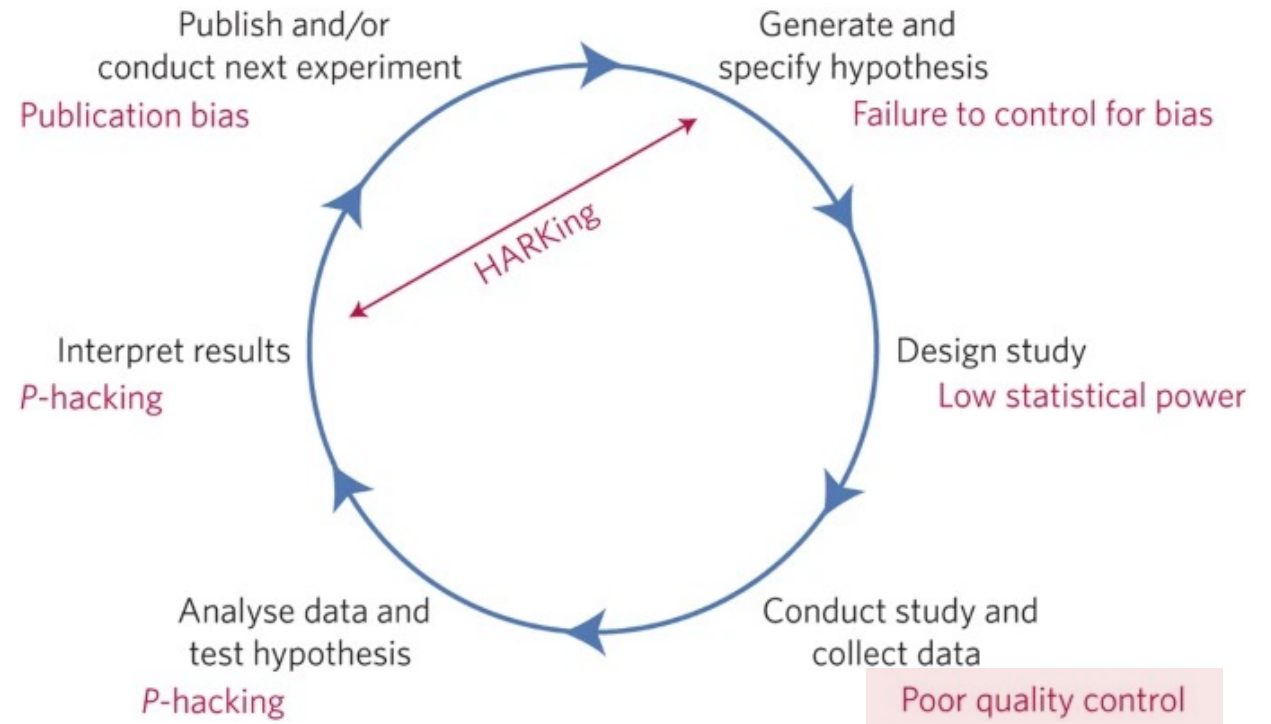
- Quality Control
- Data collection with REDCap

## Discussion

- What are your experiences? How do you ensure quality control?  
What leads to bad data quality and how can you avoid it?

## Further recommendations, possible best practices

- Lab manual
- Project manual



# Issues to deal with

# Issues to deal with

## Quality Control

### Errors producing “bad data”

*Bad data* are acquired through erroneous or sufficiently low-quality collection methods, study designs, or sampling techniques

One of the most annoying themes of errors because:

- no correct way to analyze bad data
- no scientifically justifiable conclusions can be reached about the original questions of interest
- may depend on information like the context in which the data are being used

### Real example



Self-reported energy intake has been used to estimate actual energy intake.

This method involves asking people to recall their dietary intake in one or more ways, and then deriving an estimate of metabolizable energy intake from these reports.



# Issues to deal with

## Quality Control

### Errors of data management

Are the result not of repeating others' errors, but of constructing customized methods of handling, storing, or otherwise managing data.

- like statistical analysis errors, these errors *might* be corrected by reanalysis of the data

### Possible solutions:

- Adhering to checklists
- Advances in data management (e.g., REDCap), data repositories (e.g., OSF, OPARA, ZPID), and data transparency (open data, well-documented metadata)

### Real examples

Accidentally use reverse-coded variables and make conclusions the opposite of what the data support

Incomplete dataset because entire categories of data are missing, when corrected, quantitative conclusions changed by a factor of >7

# Issues to deal with

## Data collection with REDCap

### REDCap

#### Research **E**lectronic **D**ata **C**apture

- Fantastic tool to collect data for your projects!
- Browser based web application using two-factor authentication
- Servers are hosted directly at each university that uses REDCap, so the data is safe and each account and project is managed by the respective university
- Allows you to have basically your entire project's data all in one place, specifically:
  - **Questionnaire data** collected directly in REDCap
  - Other data files (e.g. **physiological or behavioural**) uploaded to each subject's entry
  - **Meta data** like timestamps, and even quick analyses of basic sample characteristics like gender ratios



# Issues to deal with

## Data collection with REDCap

### Benefits of REDCap

Further, REDCap takes several burdens off your shoulders:

- Fields cannot be left empty unless you allow it, thereby **eliminating missings**
- Fields can have only certain ranges, thereby **eliminating impossible values**
- **Personalized invitations** to online questionnaires are sent at time points of your choice and automatically re-sent if the subject hasn't answered within a set time window
- Export the **automatically generated codebook** with the click of a button
- **Project calender** to schedule lab sessions with collaborators
- Export all data fields into different data formats like CSV, SPSS, XLS, etc.
- Working with the subject codes from the PaMS
- ...

# Infrastructure Support by INF

Research Data Management  
(last Workshop in 2021)

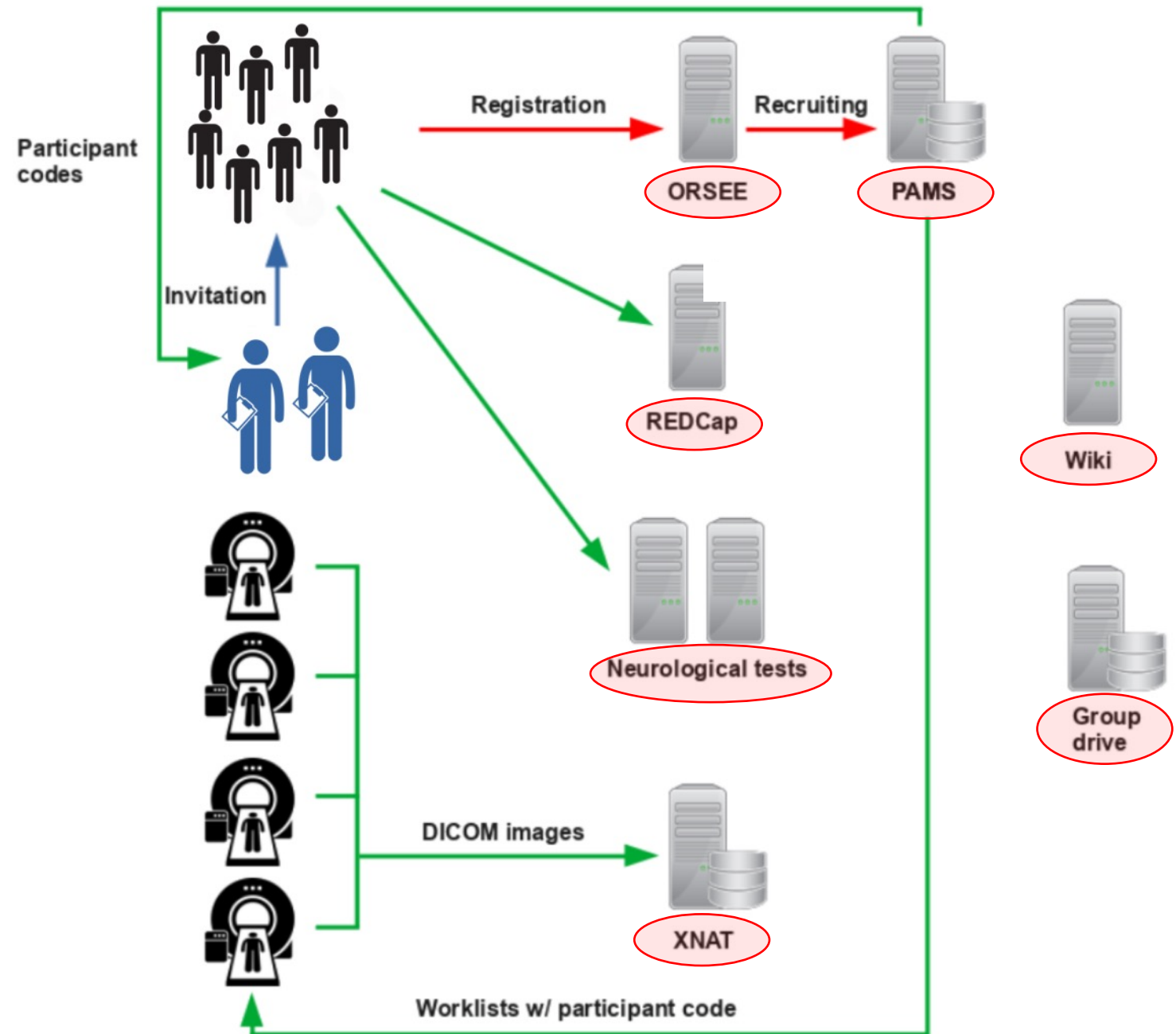
Contact for questions or problems:

Patricia Schimm (Student Assistant INF)  
[patricia.schimm@mailbox.tu-dresden.de](mailto:patricia.schimm@mailbox.tu-dresden.de)

Jens Strehle (ZIH)  
[jens.strehle@tu-dresden.de](mailto:jens.strehle@tu-dresden.de)

Alvaro Aguilera (ZIH)  
[alvaro.aguilera@tu-dresden.de](mailto:alvaro.aguilera@tu-dresden.de)

Further workshop in 2022 useful?  
Workshop on specific tools?



# MGK Open Science Module

## OPAL Course



TU Dresden | semesterübergreifend

### MGK Open Science Module

Verantwortliche/r: [Anne Gärtner](#) | [Alexander Strobel](#) | [Patricia Schimm](#) | [Josephine Zerna](#) | [Stefan Scherbaum](#)

*Doing research responsibly, efficiently and transparently*

With this e-learning module, we would like to support you in conducting your research efficiently and transparently according to the rules of good scientific practice. To this end, this module combines three workshops:

- [Workshop 1: Good Scientific Practice](#)
- [Workshop 2: Research Data Management](#)
- [Workshop 3: Research Transparency](#)

← Further material here

For each workshop, online content is provided, but there will also be virtual or presence meetings of one to two days to follow up on essential issues of each workshop. You are free to go through the online content of each workshop as your times allows, but you are required to attend the respective meetings to follow up on the online content of each workshop during the initial period of your doctoral studies. Be assured that you will find the topics dealt with most helpful for your scientific work.

# Discussion

# Discussion

## Key Questions

- What are your experiences?
- How do you ensure high quality control?
- What leads to bad data / low quality and how can you avoid it?



Save your ideas in Padlet:

<https://padlet.com/gaertneranne202/znhju8f84dzyl9u>

# Further recommendations, best practices

# Further recommendations

## Lab/project manual

README.md



## DPP-LabManual

---

Lab Manual of the Differential and Personality Psychology group (AG.DPP) at TU Dresden

So far, this manual is still under development and its contents have not yet been finally discussed and agreed upon by all team members. Still, it may already now be used for instructing new team members (e.g., master students) on how to use certain tools (see the [Wiki](#)) or perform certain routines (e.g., [creating a reproducible environment for collaborative work on research projects](#)). One major feature of this manual also will be to elaborate on what we as a group agree upon when it comes to [organizing a research project](#).



# Further recommendations

## Lab/project manual

### Benefits of a lab or project manual

- Enhances standardization
- Supports instruction for new team members (e.g., research assistants)
- Facilitates information management (where to find study documents, whom to contact for which purpose, which survey software to use)
- Provides recommended workflows (e.g., how to handle missings or outliers, how to name variables, how to analyze data)
- ... Design it according to your own (project) needs!

# Further recommendations

## Best practices from our discussion

- To be added during the workshop...

# Thank you!



**Break!**  
**(15 min)**