

# Workflows und Optionen bei Präregistrierungen

Dr Xenia Schmalz

Ludwig-Maximilians-Universität München

*xenia.schmalz@med.uni-muenchen.de*

February 14, 2022

- 1 Warum brauchen wir Präregistrierungen?
  - Reproduzierbarkeit und Replizierbarkeit
  - Apokalyptische Reiter der Replikationskrise
- 2 Was genau ist eine Präregistrierung?
  - Präregistrierung versus Registered Reports
  - Vor- und Nachteile: Registered Reports (RR) vs. Präregistrierung
- 3 Wie schreibe ich eine Präregistrierung?
  - Ein paar praktische Tipps
- 4 Aufgabe: Präregistrierung erstellen

[https://etherpad.wikimedia.org/p/Tue\\_winter-school](https://etherpad.wikimedia.org/p/Tue_winter-school)

# Warum brauchen wir Präregistrierungen?

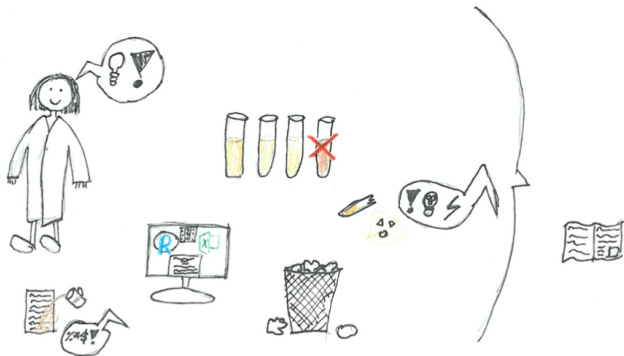
- Gute Forschung:
  - Reproduzierbar: Der Forschungsprozess muss nachvollziehbar sein
  - Replizierbar: Wenn wir das Prozedere wiederholen, finden wir dasselbe Ergebnis ( $\pm$ )



Image: [https://commons.wikimedia.org/wiki/File:Greek\\_drug\\_discovery.jpg](https://commons.wikimedia.org/wiki/File:Greek_drug_discovery.jpg)

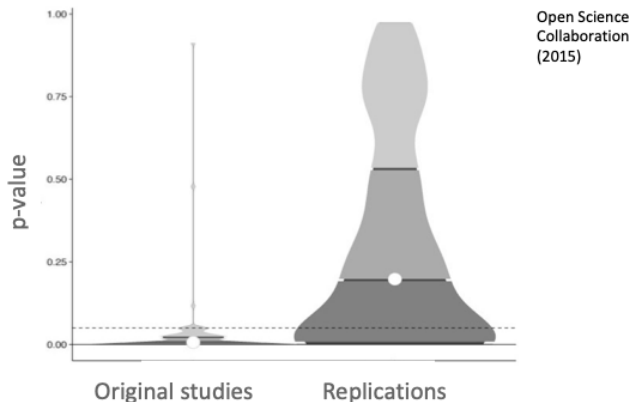
# Ist die Forschung reproduzierbar?

- Publizierung = Artikel (oft hinter Paywall)
- Materialien: Oft nur beschrieben, bei nicht publizierten Artikeln nicht zugänglich
- Cherry-Picking: Nur Experimente und Analysen, die “geklappt” haben
- Analysen in point-and-click-Software: oft fehlerhaft (s. Statcheck; Nujten et al., 2017)



# Ist die Forschung replizierbar?

- Psychologie: bei ca. 2/3 der Studien bekommt man bei einer Wiederholung der Experimente andere Ergebnisse



# Wieso???

# Wieso???

- Hypothesing After Results are Known (HARKING)
- Geringe statistische Power
- $p$ -hacking: Kreative Auslassung von Ausreißern, Datentransformationen (s. <http://flexiblemeasures.com/>), ...
- Publikationsbias: von Seiten der Forscher oder Journals

→ All das erhöht die Wahrscheinlichkeit, dass ein signifikantes Ergebnis ein falsch-positives statt einem echt-positivem Ergebnis ist

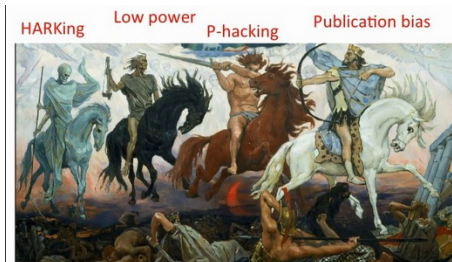


Image taken from Dorothy Bishop



Hypothesising after results are known: z.B. man schreibt die Einleitung nachdem man die Daten analysiert hat

- Problem: Explorative Analysen werden als konfirmativ dargestellt → zufällige Ergebnisse werden überbewertet
- Lösungen:
  - Open Lab Books: Transparenz zur Entwicklung der Hypothese
  - Pre-registration & Registered Reports: Festhalten der Hypothese, sowie der experimentellen Methode und Analysen bevor die Daten erhoben werden
  - Replikationsstudien: Zufallsbefunde, die Typ-I-Fehler (falsch-positiv) sind, werden mit hoher Wahrscheinlichkeit nicht repliziert

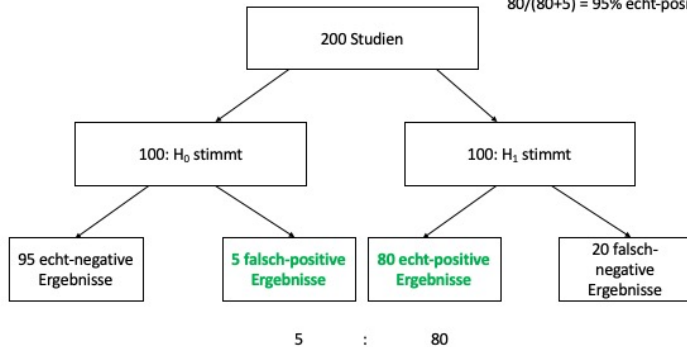
*Geringe statistische Power*: Die Anzahl der getesteten Versuchsteilnehmer\*innen reicht nicht aus, um mit hoher Wahrscheinlichkeit einen Effekt finden zu können

*Publikationsbias*: Studien mit negativen Ergebnissen werden nicht veröffentlicht

- Problem:
  - Keine Korrektur falsch-positiver Ergebnisse
  - Bei geringer statistischen Power plus Publikationsbias, werden viele falsch-negative Ergebnisse nicht publiziert → erhöhtes Verhältnis zwischen falsch-positiven und echt-positiven Ergebnissen
  - Überschätzung von Effektgrößen
  - Ressourcenverschwendung

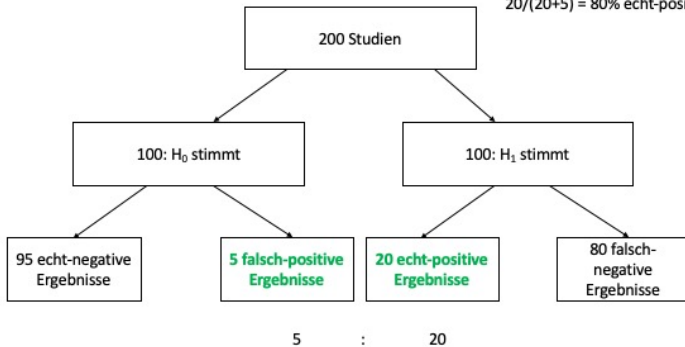
Bei 80% Power

Bei einem positiven Ergebnis:  
 $80/(80+5) = 95\%$  echt-positive



Bei geringer Power (20%)

Bei einem positiven Ergebnis:  
 $20/(20+5) = 80\%$  echt-positive



## • Lösungen

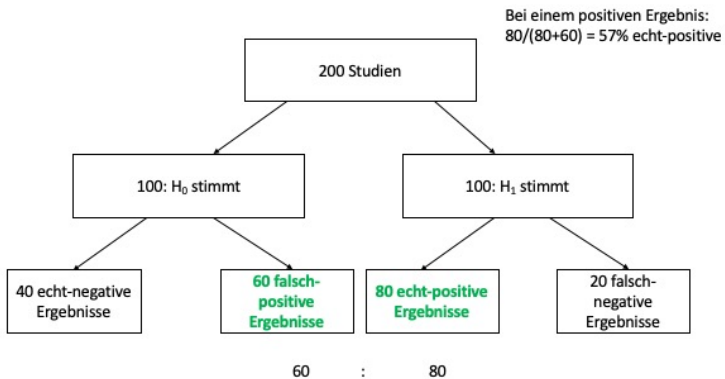
- Keine unterpowereten Studien durchführen → braucht strukturelle Veränderungen
- Preprints (Publikationsbias): Veröffentlichung von (nicht publizierten) Artikeln, zum Beispiel auf <https://edarxiv.org/>
- Pre-registration & Registered Reports (Publikationsbias)
- Meta-Analysen (geringe Power; hilft nicht bei Publikationsbias!)

p-Hacking: Post-hoc Praktiken, die die Wahrscheinlichkeit eines signifikanten Ergebnisses erhöhen, und somit auch die Wahrscheinlichkeit, ein falsch-positives Ergebnis zu erhalten

(<http://shinyapps.org/apps/p-hacker/>)

- Problem:
  - Erhöhte falsch-positiv-Rate (bis zu 60%!)
- Lösungen:
  - Transparenz in den Datenanalysen (Reproduzierbarkeit)
  - Offene Daten (wenn möglich)
  - Pre-registration & Registered Reports

## Bei p-Hacking



# Erste Aufgabe: p-hacke die Studie!

We tested 20 children with developmental dyslexia (12 females) and 20 children with average reading ability (10 females) on their visual attention skills, to test the hypothesis that visual attention skills are worse in children with developmental dyslexia (DD). We also recorded children's age and handedness. Children were classified as having DD if they had a previous diagnosis from a practitioner. We also measured their reading ability using a speeded word reading task, as well as their comprehension of written texts.

Each child performed a visual search task: they saw cluttered arrays of 8, 12, or 16 objects. Their task was to indicate if a target object was present in each display (yes/no decision). The target object (a cat, a green circle, or a teacup) was presented before each trial. In half of the trials, the target was absent. We recorded both the accuracy and the reaction time associated with each response.

Indeed, we found support for our hypothesis,  $p < 0.05!$



# Checkliste

- *t*-test: Accuracy in the visual search task in children with vs. without DD
- Including or excluding foil visual search trials (where target was absent)
- 2x3 ANOVA: DD vs. control, 3 different array sizes
  - Main effect of group (DD vs. control)
  - Interaction between group and array size
  - Post-hoc differences between groups for each array size
- Covariate analysis: age, gender, both
- Outlier exclusion: based on handedness, reading ability, or both
- Regression analysis with reading as a continuous variable (speeded word reading, comprehension)
- All of this with reaction times instead of accuracy as the measure of visual attention
  - Transform reaction time: z-score, logarithm, inverse transformation ...

# Was sind Registered Reports?

- Einleitung, Methoden und Analysenplan
- Registered Reports
  - Enthält eine komplette Einleitung, Methoden und Analysenplan
  - Wird bei einem Journal eingereicht, welches das Format annimmt
  - Gutachter geben Kommentare zu der Plausibilität der Hypothesen, Qualität und Angemessenheit der Methoden
  - Wenn die Gutachter mit dem Design einverstanden sind, bekommt man “in-principle acceptance”: Das Paper wird publiziert, unabhängig von den Ergebnissen
  - **NACHDEM** man die in-principle acceptance hat, beginnt man mit der Datenerhebung

# Was ist Präregistrierung?

- Einleitung, Methoden und Analysenplan
- Präregistrierung:
  - Nicht modifizierbares Dokument mit Zeitstempel
  - Zeitstempel muss von **VOR** Beginn der Datenerhebung stammen!
  - Flexibles Format (z.B. Stichpunkte)
  - Publikation in jedem beliebigen Journal, mit Link zu der Präregistrierung

# Vor- und Nachteile: Registered Reports (RR) vs. Präregistrierung

	RR	Präreg.
Hilft gegen $p$ -hacking	ja	ja
Hilft gegen Publikationsbias	ja	vielleicht
Hilft gegen HARKing	ja	ja
Garantierte Publikation	ja	nein
Qualitätscheck vor Datenerhebung	ja	vielleicht
Präregistrierungs-Badge	ja	ja
Flexibles Timing	nein	ja
Beseitigt Betrug	nein	nein

# Wie schreibe ich den Report?

- Im Prinzip: Ganz einfach Einleitung, Methoden und Analysenplan
- Worauf man achten sollte:
  - Ist die Studie geeignet? Ausnahmen: Explorative Studien, Methodenentwicklung
  - Soviel Detail wie möglich schreiben
  - Pilotieren: Funktioniert das Design so, wie ich es mir vorstelle?
  - Theoretische Überlegungen: Was kann ich schlussfolgern, wenn das Ergebnis nicht signifikant ist?
  - Bei Präregistrierung: Ist das Design gut? → Soviel Feedback wie möglich

# Wo kann ich Registered Reports einreichen?

<https://cos.io/rr/#journals>

- > 300 Journals (alle Fachdisziplinen)
- Wenn euer Lieblingsjournal nicht dabei ist, Editoren anschreiben:
  - Für ein Paper: Template for specific proposal for RR
  - Wenn ihr generell möchtet, dass das Journal RRs akzeptiert: Template for generic RR Outreach

<https://osf.io/3wct2/wiki/Journal%20Requests/>

# Fehler, die man vermeiden sollte

- Studien, die nicht informativ sind wenn as Ergebnis nicht signifikant ist
  - “Shower ideas”: geringe *a priori* Wahrscheinlichkeit
  - Geringe statistische Power
  - Schwache Manipulation, keine Manipulationschecks
- Versprechen, die man nicht halten kann
  - Aufgaben, die zu schwer für Probanden sind (z.B. Lernstudie)
  - Ausgefallene Analysen → An Pilotendaten oder simulierten Daten üben
  - Zu große Stichproben
- Mit höchster Wahrscheinlichkeit wird das Paper von der Registrierung abweichen → kein Problem, muss aber klar und begründet sein

# Timing

- Registered Reports: Kann durch mehrere Runden von Gutachten gehen!
- Durchschnittliche Zeit, die der RR im Editorial-System verbringt: 2 Monate, aber erfahrungsgemäß: rechnet mit 6-12 Monaten
- 'Insider'-Info von Cortex: Nachdem ein Registered Report durch den Editorial Triage geht, ist die Wahrscheinlichkeit von in-principle-acceptance  $\approx 90\%$ .

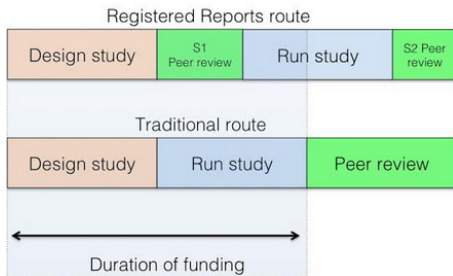


Image from Chris Chambers



- Power analyse: z.B. 90% power bei Cortex
- Alternative: Sequential sampling
  - Bei frequentistischen Analysen, s. Tutorial von Lakens (2014): <https://pdfs.semanticscholar.org/c871/0947cbf325f874e09b46d25b6ab170efecba.pdf>
  - Bei Bayesianischen Analysen, s. Tutorial von Schönbrodt & Wagenmakers (2018): <https://osf.io/w3s3s/>
- “Practical considerations”
  - Besonders bei Demographien, die schwer zu rekrutieren sind
  - Relevante Forschungsfrage + Stichprobe ist größer als bei vorherigen Studien zu dem Thema

In kleinen Gruppen oder alleine,  
denkt euch eine Studie aus  
und schreibt einen Registered  
Report in Stichpunkten.

# Tübingen Winter School

## Workshop: Präregistrierungen

### FRAGEN

1. Würden Sie bei einer explorativen Fragestellung eine Präregistrierung empfehlen, obwohl es keine direkten Hypothesen gibt?
2. Was genau unterscheiden RR und PCI RR Peer-Community-In Verfahren?
3. Würden Sie pre-regs mit oder ohne Template empfehlen?
4. Gibt es ein Beispiel für ein Open Lab Book? Ich kann mir nicht ganz vorstellen, wie das aussehen soll. Und gibt es hierfür entsprechende templates?  
→ Beispiel: <https://doi.org/10.5281/zenodo.5450517>  
→ Template/How-To: <https://openlabnotebooks.org/> (arbeitet zusammen mit Zenodo)
5. Inwieweit sind Präregistrierungen auch bei qualitativen Studien üblich/möglich?  
→ Publikationen zur Präregistrierung in **qualitativer** Forschung:
  - a. <https://doi.org/10.1177/1609406920976417>
  - b. <https://doi.org/10.1111/j.1365-2648.2006.03681.x>
  - c. <https://doi.org/10.1046/j.1365-2648.2003.02537.x>

### TIPPS

1. Hilfreiche Shiny App zur **Sample Size Justification**:  
[https://shiny.ieis.tue.nl/sample\\_size\\_justification/](https://shiny.ieis.tue.nl/sample_size_justification/)
2. **p-hack** a study!  
→ Tool von Felix Schönbrodt zum ausprobieren, wie einfach es ist Studien zu hacken: <http://shinyapps.org/apps/p-hacker/>
3. **Tool** für die Lehre, um zu zeigen, wie die **selektive Auswahl von Variablen** die stat. Signifikanz beeinflusst: <https://projects.fivethirtyeight.com/p-hacking/>
4. Journals in denen **Registered Reports** möglich sind:  
<https://www.cos.io/initiatives/registered-reports> (Tab "Participating Journals")

### ZUSAMMENFASSUNG:

**Reproduzierbarkeit** (gleiche Ergebnisse auf Basis der Datenanalysen generieren) vs.  
**Replizierbarkeit** (Versuch nochmals an einer anderen Stichprobe durchzuführen)

## "Threats to replicability/reproducibility"

- Forschende haben keinen Zugang zum Artikel (Paywall)
- Materialien nicht verfügbar
- "Cherry picking": Daten/Variablen werden ausgewählt, die "interessant" sind
- Open Science Collaboration (2015): Replikation von Studien problematischer als gedacht

## Questionable Research Practices:

- HARKING
- geringe Power
- p-hacking
- Publicationbias

## AUFGABE 1:

### Room 1:

- Outlier, deskriptive Analysen und Unterschiede vorher
- Auswertungsmethode, Analysen:
  - Objects: 8, 12, 16
  - Kovariate
- Outcomevariablen: Accuracy/reaction time → Operationalisierung: falsch/NA bei Nicht-Antwort
- Veränderung des sig. Niveaus (Alpha)

### Room 2:

- Im Nachhinein (begründet) nur Teile der Ausgangsgruppe betrachten, nur bestimmte Aufgaben, Akkuratessse anders definieren...

Data: No Data has been collected.

Hypothesis: Teilnehmende an der Winter School, präregistrieren mehr Studien als Personen die nicht an der Winter School teilgenommen haben.

Key Variables:

Conditions:

### Room 4:

#### Gruppen:

- Geschlecht
- Alter
- Händigkeit

- Einteilung in hoch/niedrig dyslexia bzw. wie haben sie in Test abgeschnitten (unterschiedliche Tests)
- Untergruppen von dyslexia definieren (z.B. Geschwindigkeit vs. Verständnisschwierigkeiten)
- Unterteilung in vorherige Diagnose oder nicht

Analysen:

- Ausreißer
- Ausschluss von Bedingungen
- t-Tests für alle Bedingungen

DVs (flexible measurements):

- Analyse auf 0/1-Daten
- Analyse Prozent-Korrekt oder Summenscore
- Reaktionszeit: Transformation, z.B. log, oder  $1/RT$ , ...