

# Was müssen Replikationsstudien leisten können?

Edgar Erdfelder  
Universität Mannheim

# Was müssen Replikationsstudien leisten können?

- A) Ziele von Replikationsstudien
- B) Förderkriterien für Replikationsstudien
- C) Priorisierung von Replikationsstudien
- D) Reichen bestehende Förderformate aus?

# A) Ziele von Replikationsstudien

- Übergeordnetes Ziel:
  - Empirische Prüfung von wissenschaftlichen Theorien oder (technologischen) Hypothesen.
  - Dies setzt ein theoriebezogenes Fundament an verlässlichen empirischen Basissätzen voraus.
- Verlässlichkeit impliziert
  - Objektivität und Replizierbarkeit (Stabilität).
- Replizierbarkeitsnachweis mittels Replikation.
- Direkte oder konzeptuelle Replikationen?
- Das o.g. Ziel verlangt direkte Replikationen.

# B) Förderkriterien für Replikationsstudien

1. Begründung der Wichtigkeit
2. Genauigkeit der Replikationsmethodik
3. Unabhängigkeit des Forscherteams
4. Kooperation mit Forschern der Originalstudie
5. Strategie der Replikationsüberprüfung
6. A-Priori-Definition der relevanten Effektstärke
7. Fundierte Stichprobenumfangsplanung
8. Publikationsverpflichtung
9. Transparenz

# 1) Begründung der Wichtigkeit

- Nicht jede beliebige Studie sollte repliziert werden.
- Nur wichtige Befunde rechtfertigen die i.d.R. hohen Kosten einer Replikationsstudie.
- Folglich ist der besondere theoretische und/oder technologische Stellenwert der zu replizierenden Studie zu begründen.
- Die theoretischen oder technologischen Implikationen einer erfolgreichen / nicht erfolgreichen Replikation sind abzuleiten.
- Der zu erwartende Erkenntnisgewinn durch beide denkbaren Ausgänge ist darzulegen.

## 2) Genauigkeit der Replikationsmethodik

- Die dokumentierte Methodik der Originalstudie ist so genau wie möglich zu replizieren.
- Dies betrifft in erster Linie:
  - die Methode der Vpn-Gewinnung bzw. -Ausschließung
  - Versuchsdesign, -material und –prozedur
  - Verwendete Maße für abhängige Variablen
- Unproblematische / unvermeidbare Abweichungen:
  - Versuchsleiter
  - Grundgesamtheit der Vpn sowie Stichprobenumfang
  - Raum bzw. Labor der Untersuchung
  - Sprache
  - Ort und Zeit

### 3) Unabhängigkeit des Forscherteams

- Replikationen durch die Autoren der Originalstudie bergen eine höhere Gefahr der Ergebnisverzerrung durch Versuchsleitereffekte.
- Statt Zerlegung einer Stichprobe in Analyse- und Replikationsstichprobe besser Zusammenfassung beider Stichproben zwecks Teststärke-Maximierung!
- Replikationsstudien sollten daher immer von (personell und institutionell) unabhängigen Forscherteams durchgeführt werden.
- Befangenheitskriterien sind zu beachten.
- Mögliche Befangenheitsgründe sind bei der Antragstellung transparent zu machen.

# 4) Kooperation mit Forschern der Originalstudie

- Soweit möglich, ist Kooperation mit Forschern der Originalstudie anzustreben.
- Kooperation
  - erleichtert exakte Replikation;
  - ermöglicht Zugriff auf nichtveröffentlichte Dokumente zur Versuchsmethodik (Programme, Materialien etc.);
  - Ermöglichkeit Abstimmung über Implementation der Versuchsprozedur.
- Es ist darzulegen, welche Kooperationsbemühungen unternommen wurden.
- Kommt Kooperation nicht zustande, sind die Gründe darzulegen.

# 5) Strategie der Replikationsüberprüfung

- Vermeidung von p-hacking durch Exploration verschiedener Auswertungsmethoden.
- Die (statistische) Überprüfungsmethodik zur Replikationshypothese ist a priori festzulegen.
- Zwei Möglichkeiten:
  - (1) Test auf Abweichung vom Originalergebnis
  - (2) Replikation des Inferenzverfahrens der Originalstudie.
- Im Regelfall macht der zu prüfende empirische Basissatz keine genaue Aussage über die Stärke des zu replizierenden Effekts, nur über dessen Vorzeichen.
- Ergo: Replikation des statistischen Inferenzverfahrens der Originaluntersuchung ist die Methodik der Wahl.

# 6) A-Priori-Definition der relevanten Effektstärke

- Unabhängig von der gewählten Strategie muss der für die Replikation kritische Effekt a priori definiert werden.
- Strategie (1): Standardisierte Abweichung vom Originalergebnis, ab der die Replikation als misslungen gilt.
- Strategie (2): Standardisierter Effekt, der erreicht werden muss, damit die Replikation als gelungen gilt.
- Mögliche Wege zur Festlegung von Effektgrößen:
  - (a) Effektgrößenkonventionen (z.B. Cohen, 1988)
  - (b) Effektgrößenschätzungen aus Originalstudien.
- Weg (b) ist problematisch aufgrund von
  - Überlagerung des wahren Effekts durch Stichprobenfehler;
  - systematischer Überschätzung durch Publication Bias.

# Überschätzung von Cohens $d$ aufgrund von Publication Bias beim 2-Gruppen- $t$ -Test (Quelle: Ulrich, Miller & Erdfelder, under review)



# Korrelationsüberschätzung aufgrund von Publication Bias beim Korrelations- $t$ -Test

(Quelle: Ulrich, Miller & Erdfelder, under review)



# 7) Fundierte Stichprobenumfangsplanung

- Stichprobenumfänge ( $N$ ) bedürfen grundsätzlich einer nachvollziehbaren Begründung.
- Übereinstimmung mit dem  $N$  der Originalstudie ist keine zulässige Begründung.
- Das  $N$  ist so zu wählen, dass fälschliche Bewährungs-/Nichtbewährungsurteile zur Replikationshypothese nachweislich unwahrscheinlich sind.
- I.d.R. balancierte Fehlerraten (z.B.  $\alpha = \beta = .05$ ).
- Basiert die Effektgrößenspezifikation auf publizierten Ergebnissen, so ist eine unverzerrte Schätzmethode anzuwenden, die Publication Bias berücksichtigt (Ulrich, Miller & Erdfelder, under review).

# Konsequenzen verzerrter Effektgrößenschätzungen

- Beispiel: Open Science Collaboration (Science, 2015).
- Insgesamt 16 *t*-Test Replikationen.
- Power für unkorrigierte Effektstärke: mindestens .90.
- Replikationsquote gesamt: 9 von 16 (= 56%).

# Konsequenzen verzerrter Effektgrößenschätzungen

- Beispiel: Open Science Collaboration (Science, 2015)
- Insgesamt 16 *t*-Test Replikationen.
- Power für unkorrigierte Effektstärke: mindestens .90.
- Replikationsquote gesamt: 9 von 16 (= 56%).
  
- In 9 Fällen: Korrigierte  $\approx$  unkorrigierte Effektstärke
- Replikationsquote hierfür: 6 von 9 (= 66%)

# Konsequenzen verzerrter Effektgrößenschätzungen

- Beispiel: Open Science Collaboration (Science, 2015)
- Insgesamt 16 *t*-Test Replikationen.
- Power für unkorrigierte Effektstärke: mindestens .90.
- Replikationsquote gesamt: 9 von 16 (= 56%)
  
- In 9 Fällen: Korrigierte  $\approx$  unkorrigierte Effektstärke
- Replikationsquote hierfür: 6 von 9 (= 66%)
  
- In 7 Fällen: Korrigiert  $\ll$  unkorrigierte Effektgröße
- Replikationsquote hierfür: 3 von 7 (= 43%)

# 8) Publikationsverpflichtung

- Beihilfeempfänger müssen sich zur Publikation des Replikationsergebnisses verpflichten.
- Attraktives Format: Registered Reports
- Grundprinzip:
  - Entscheidung über Publikationsannahme nachweislich vor der Datenerhebung (Greve, Bröder & Erdfelder, 2013).
- Alternativen:
  - Standardartikel in breit verfügbaren Fachzeitschriften mit Begutachtungssystem;
  - Verlässliche Repositorien (im Bereich der Psychologie z.B. ZPID).

# 9) Transparenz

- Anerkennung der Dokumentationspflicht für alle Details der Replikationsstudie.
- Archivierung in einem verlässlichen internet-basierten Repository (z.B. PsychData von ZPID).
- Bereitschaft, alle Materialien (Versuchbeschreibung, Versuchsmaterial, anonymisierte Urliste, Makros und Computerprogramme) an andere Forscher zwecks Reanalyse oder ggf. Durchführung einer weiteren Replikationsstudie herauszugeben.

# C) Priorisierung von Replikationsstudien

- Erkenntnisgewinn
  - Höhe des Erkenntnisgewinns durch die beiden denkbaren Ausgänge der Replikationsstudie.
- Dringlichkeit
  - in Anbetracht einer aktuellen theoretischen Debatte;
  - in Anbetracht einer aktuellen Debatte über eine technologische Hypothese.
- Kosten
  - eines fälschlich positiven Originalbefunds
  - der Replikationsstudie (value for money).

# D) Reichen bestehende Förderformate aus?

- Einführung eines neuen Förderformats würde die Fehlkonzeption von Replikationen als Selbstzweck forcieren.
- Bestehende Förderformate bieten ausreichend Raum für wissenschaftlich gebotene Replikationsstudien.
- *Wichtig:* Kommunikation von Förderkriterien für Replikationsstudien an
  - Antragstellerinnen und Antragsteller
  - Gutachterinnen und Gutachter.

Vielen Dank für Ihre  
Aufmerksamkeit!