Fragwürdige Forschungspraktiken in empirischen Projekten im Rahmen des Psychologie-Studiums

PsyFaKo Open Science AG

08.09.2019

# Zusammenfassung

# Einleitung

Testzitat (Simmons, Nelson, & Simonsohn, 2018)

# Methode

# Ergebnisse

## Deskriptive Ergebnisse

### Verbreitung fragwürdiger Forschungspraktiken

Den Ergebnissen in diesem Abschnitt liegen die Daten der EF (Empirische Erfahrung)-Teilstichprobe (N = 1413) zugrunde.

Aus den Angaben der Teilnehmenden ergibt sich ein mittlerer Einsatz von 1,26 (SD = 1,39) fragwürdigen Forschungspraktiken pro Projekt. Wie in Abbildung @ref(fig:plot.nqrps) zu sehen ist, zeigen sich dabei Unterschiede zwischen den abgefragten Projekten. Besonders viele fragwürdige Praktiken scheinen demnach mit durchschnittlich 1,42 (SD = 1,63) Praktiken pro Projekt im Expra verwendet zu werden, während Masterarbeiten mit durchschnittlich 0,52 (SD = 0,84) verwendeten fragwürdigen Praktiken den niedrigsten Wert aufweisen. Bachelor- und Projektarbeiten sowie andere Projekte (durchschnittlich jeweils 1,02 (SD = 1,33), 1,30 (SD = 1,51) und 0,85 (SD = 1,37) fragwürdige Praktiken pro Projekt) liegen dazwischen. Betrachtet man die Projekte, die sich eindeutig nach Studienfortschritt ordnen lassen, also Expra, Bachelorarbeit und Masterarbeit, so zeigt sich ein deutlicher Trend: Mit fortschreitendem Studium wurden immer weniger fragwürdige Forschungspraktiken von den Teilnehmenden eingesetzt.

In Abbildung @ref(fig:plot.rp\_details) ist die Einsatzhäufigkeit der einzelnen abgefragten Forschungspraktiken dargestellt. Als am häufigsten eingesetzte fragwürdige Forschungsmethoden zeigten sich fehlende Stichprobenplanung und selektives Berichten abhängiger Variablen mit einem Anteil von 34,30 %, bzw. 23,10 % betroffenen Projekten. Die von den Teilnehmenden am seltensten eingesetzten Methoden waren das Abrunden von p-Werten und das Nacherheben von Versuchspersonen zum Zweck der Herbeiführung statistischer Signifikanz mit 1,90 %, bzw. 6,40 % betroffenen Projekten. Die beiden offenen, bzw. positiven Forschungspraktiken, die ebenfalls Teil der Befragung waren, rangieren am oberen Ende der Einsatzhäufigkeit: Laut Angaben der Teilnehmenden wurde für 33,70 % der erfassten Projekte eine Power-Analyse zur Stichprobenplanung durchgeführt, und 22,40 % der erfassten Projekte wurden präregistriert. Die Präregistrierung fand am häufigsten bei einer Lehrperson statt (282 mal gewählt). Es folgen in nach Häufigkeit absteigender Reihenfolge als Plattformen für Präregistrierungen das Open Science Framework (OSF, 101 mal gewählt), aspredicted.org (87 mal gewählt), “andere” (65 mal gewählt) und ein Journal (11 mal gewählt). Von den Teilnehmenden gaben 704 Personen an, nie eines ihrer empirischen Projekte präregistriert zu haben.

Anzumerken ist, dass hier *nicht* wie bei John, Loewenstein, & Prelec (2012) der Anteil der Teilnehmenden gezeigt wird, die die jeweilige Forschungspraktik *mindestens einmal* verwendet haben, sondern der Anteil der pro Forschungspraktik betroffenen Projekte. Diese Berechnung erlaubt eine detailliertere Betrachtung, da Informationen über alle 2681 im Datensatz enthaltenen Projekte dargestellt werden, anstelle sich auf einen (ungenaueren) Datenpunkt pro Teilnehmer\*in zu beschränken. Aus diesem Grund weichen wir hier von unserer Präregistrierung [(https://osf.io/k85r4/)](https://osf.io/k85r4/) ab, die eine Darstellung analog zu John et al. (2012) vorsah (“Praktikaggregation”). Diese Darstellung kann in den [ergänzenden Online-Materialien](https://jobrachem.github.io/Students-Open-Science-Survey/supplementary_data-edited.pdf)(LINK) dennoch abgerufen werden und zeigt generell ein ähnliches Bild, wobei die Prävalenzschätzungen durch den Informationsverlust überhöht sind. Wir verzichten zugunsten einer prägnanteren Darstellung hier außerdem auf die in der Registrierung vorgesehene “Detailaggregation”, eine Aufschlüsselung der Praktikaggregation nach Forschungsprojekt. Sie kann ebenfalls in den Online-Materialien abgerufen werden.

### Zusätzliche Ergebnisse

Den Ergebnissen in diesem Abschnitt liegen die Daten der vollen Stichprobe (N = 1413) zugrunde.

#### Behandlung der Replikationskrise in der Lehre

Zu den hier dargestellten Ergebnissen finden sich ausführliche Tabellen mit allen Daten in den [ergänzenden Online-Materialien](https://jobrachem.github.io/Students-Open-Science-Survey/supplementary_data-edited.pdf) (LINK).

**Thematisierung der Replikationskrise**. Insgesamt gaben 75,4 % der Teilnehmenden an, dass die Replikationskrise an ihrer jeweils aktuellen Universität in Lehrveranstaltungen thematisiert wurde; 20,6 % gaben an, dass sie nicht thematisiert wurde. Die restlichen 4,1 % der Teilnehmenden antworteten mit “Weiß ich nicht”. Erfreulich ist dabei, dass selbst unter den Studienanfänger\*innen (Studierende im ersten oder zweiten Bachelor-Semester) 60,2 % der Teilnehmenden angaben, in Lehrveranstaltungen von der Replikationskrise gehört zu haben. Für alle Studierenden im Bachelor und im Master zeigte sich im Allgemeinen ein ähnliches Bild mit 75,3 %, bzw. 77,4 % positiven Antworten.

**Betrachtung nach Lehrveranstaltungen**. Die Behandlung der Replikationskrise fand größtenteils in der Methodenlehre (inkl. Expra) statt: 54,5 % der Teilnehmenden gaben an, dass hier die Replikationskrise thematisiert wurde. In der Lehre der Sozialpsychologie, der am stärksten im Fokus der Kritik stehenden Teildisziplin, wurde am zweithäufigsten über die Replikationskrise berichtet (27,4 %). In den 13 übrigen Fachbereichen (inkl. “Andere”) bewegt sich die Abdeckung zwischen 19,0 % (Allgemeine Psychologie) und 0,8 % (Forensische Psychologie). Dabei sollte beachtet werden, dass wohl nur ein Bruchteil der Befragten Module in Forensischer Psychologie belegt haben dürfte.

**Behandlung fragwürdiger Praktiken**. Studierende erleben im Studium durchaus eine kritische thematisierung einzelner fragwürdiger Forschungspraktiken. So gaben 74,0 % der Teilnehmenden an, eine Problematisierung von HARKing (Hypothesizing After the Results are Known, das Aufstellen von Hypothesen nach Kenntnis der Ergebnisse) erfahren zu haben. HARKing stellte sich als die am häufigsten in der Lehre kritisierte fragwürdige Forschungspraktik heraus. Auf den folgenden Plätzen finden sich fehlende Stichprobenplanung (65,3 %) und selektives Berichten von Variablen (64,1 %). Am seltensten wurden Flexibilität in der Datenanalyse (50,7 %) und das Abrunden von p-Werten (39,9 %) kritisch thematisiert.

#### Einstellung von Studierenden

Abbildung @ref(fig:plot.attitudes) zeigt die Einschätzung der Wichtigkeit des Themenkomplexes “Open Science und Replikationskrise” durch die Teilnehmenden, ihr Interesse daran und ihre Einschätzung dazu, wie gut sie über das Thema informiert sind. Deutlich wird, dass das Thema von der überwältigenden Mehrheit der Teilnehmenden für wichtig gehalten wird, dass eine Mehrheit der Teilnehmenden Interesse an dem Thema hat, und dass sich ein substanzieller Teil der Teilnehmenden als nicht gut informiert einschätzt.

#### Korrelationen

Die einfachen Korrelationen zwischen den numerischen Variablen finden sich in Tabelle 1.

## Explorative Analysen

Zur näheren Untersuchung der Daten führten wir eine Reihe von explorativen Analysen durch, von denen wir die wichtigsten hier berichten. Alle weiteren oder nur kurz erwähnten Analysen sind in den Online-Materialien (LINK) abrufbar. Die Analysen wurden in R (Zitat) mit dem Paket “lme4” (Zitat) ausgeführt. Für alle Analysen wurden Bonferroni-korrigierte Alpha-Niveaus zugrunde gelegt.

### Einsatz fragwürdiger Forschungsmethoden

In einem generalisierten linearen gemischten Modell für dichotome abhängige Variablen mit logit-Linkfunktion nahmen wir alle verfügbaren, potentiell informativen Kovariaten auf, um die Odds für den Einsatz fragwürdiger Forschungspraktiken zu modellieren. Abhängige Variable war ein Indikator dafür, ob eine spezifische Forschungspraktik in einem spezifischen Projekt eingesetzt wurde (0 = “Nein”, 1 = “Ja”). Das Modell enthält zufällige Achsenabschnitte für den/die jeweilige/n Teilnehmer\*in (SD = 1,24, (1) = 1055,68, p <.001), für die Universität, an der das Projekt durchgeführt wurde (SD = 0,51, (1) = 40,89, p <.001), und für die spezifische Forschungspraktik (SD = 1,03, (1) = 1545,68, p <.001). Die numerischen Prädiktoren wurden z-standardisiert. Die Schätzungen der fixen Koeffizienten und die jeweiligen Odds Ratios sind in Tabelle 5 dargestellt. Zugrunde lagen die Daten aus der EE-Teilstichprobe.

Einzelne Beobachtungen wurden wegen inkonsistenter Antworten oder unklarer Zuordnung des jeweiligen Projektes zu einer Universität ausgeschlossen. Ersteres geschah in Fällen, in denen Teilnehmende bei der Frage nach der Anwendung einer spezifischen Forschungspraktik sowohl die Antwortmöglichkeit “Keines” (Praktik wurde in keinem Projekt verwendet), als auch mindestens ein oder mehrere Projekte auswählten (bei QRPs n = 95 Beobachtungen). Zweiteres geschah bei den Projekten “Projektarbeit” und “Anderes”, wenn die Teilnehmenden ihre Universität zwischen Bachelor und Master gewechselt hatten, oder bei Alumni/Alumnae, da deren Bachelor-Universität nicht erfasst wurde (n = 539 Projekte). Zwischen beiden Kriterien gab es Überschneidungen. Insgesamt wurden so 4934 Datenpunkte zu 539 Projekten von -251 Teilnehmenden in der Analyse nicht berücksichtigt.

Das Effektstärkemaß Odds Ratio gibt die multiplikative Änderung im Verhältnis der *Odds* (dt.: Chancen) von “Forschungspraktik angewandt” und “Forschungspraktik nicht angewandt” bei Änderung des Prädiktors um eine Einheit, bzw. bei kategorialen Prädiktoren im Vergleich zur Referenzgruppe an. Ein Odds Ratio größer als 1 bedeutet, dass die Anwendung der Forschungspraktik wahrscheinlicher wird, ein Odds Ratio unter 1, dass sie weniger wahrscheinlich wird. Zunächst lässt sich beobachten, dass QRPs umso weniger eingesetzt wurden, je eher Teilnehmende den Themenkomplex “Replikationskrise und Open Science” als wichtig ansahen und je eher sie sich als gut informiert einschätzten (Odds Ratios von 0,82 und 0,73).

Darüber hinaus lassen sich Unterschiede zwischen den einzelnen Projekttypen erkennen, wobei alle Projekttypen im Vergleich zur Referenzkategorie “Expra” eine geringere Wahrscheinlichkeit des Einsatzes von QRPs aufweisen. Die Effektstärken variieren dabei von OR = 0,27 (Masterarbeiten) bis OR = 0,77 (Projektarbeit). Bei einer Masterarbeit ist die Chance, dass eine fragwürdige Forschungspraktik eingesetzt wird, also weniger als ein Drittel der Chance, dass sie in einem Expra eingesetzt wird. Die übrigen untersuchten Kovariaten wiesen bei Bonferroni-adjustiertem Alpha-Niveau von = 0.00333 keinen signifikanten Zusammenhang mit dem Einsatz von QRPs auf (siehe Tabelle 5).

In einem nächsten Schritt untersuchten wir im Detail, ob sich Hinweise für eine systematische Veränderung der Einsatzwahrscheinlichkeit von QRPs im Verlauf des Studiums finden. Dazu betrachteten wir ausschließlich Expras, Bachelorarbeiten und Masterarbeiten, da diese Projekte eindeutig verschiedenen Phasen des Studiums zugeordnet werden können. Die Analyse wurde auch hier mit einem generalisierten linearen gemischten Modell für dichotome abhängige Variablen mit logit-Linkfunktion mit der gleichen abhängigen Variable und, abgesehen vom Projekttyp, den gleichen Kovariaten wie im vorherigen Modell durchgeführt. Der Projekttyp wurde nun als geordneter Faktor in das Modell mit aufgenommen, so dass ein linearer Trend über die drei eingeschlossenen Projekttypen untersucht werden konnte. Tatsächlich zeigte sich ein linearer Trend, OR = 0.4 [0.32, 0.50], p <.001, der darauf hindeutet, dass mit fortschreitendem Studium, bzw. in anspruchsvolleren Projekten mit geringerer Wahrscheinlichkeit QRPs angewandt werden.

### Einsatz positiver Forschungspraktiken

Die Ergebnisse bezüglich positiver Forschungspraktiken fassen wir hier nur kurz zusammen. Die vollständigen Tabellen zu allen Modellen können in den Online-Materialien (LINK) abgerufen werden. Wir untersuchten hier ebenfalls in generalisierten gemischten linearen Modellen die gleichen Prädiktoren und zufälligen Achsenabschnitte wie bei der Exploration zu QRPs (Fußnote: Mit Ausnahme von “Praktik gelehrt”, da diese Frage nur für QRPs gestellt wurde). Auf Bonferroni-adjustiertem Alpha-Niveau zeigten sich signifikante Unterschiede zwischen Master- und Bachelorstudierenden (B = -0,38, SE = 0,13, OR = 0,68, 95 % KI = [0.53, 0.87], p = .002). Master-Studierende wiesen demnach eine geringere Wahrscheinlichkeit für den Einsatz positiver Forschungspraktiken auf als Bachelor-Studierende. Diese Beobachtung zeigte sich noch deutlicher, wenn nur Expras und Bachelorarbeiten untersucht wurden, so dass ein etwas direkterer Vergleich zwischen länger zurückliegenden Projekten (Expras und Bachelorarbeiten von aktuellen Master-Studierenden) und kürzer zurückliegenden Projekten (von aktuellen Bachelor-Studierenden) gezogen werden kann (siehe Online-Materialien, LINK).

Auch in diesem Modell zeigte sich ein Einfluss der subjektiv eingeschätzten Informiertheit der Teilnehmenden, je informierter sie sich sahen, desto wahrscheinlicher war der Einsatz positiver Forschungspraktiken (B = 0,34, SE = 0,05, OR = 1,41, 95 % KI = [1.27, 1.56], p <.001). In Projektarbeiten (B = -0,72, SE = 0,13, OR = 0,49, 95 % KI = [0.38, 0.62], p <.001) und “Anderen” Projekten (B = -0,87, SE = 0,16, OR = 0,42, 95 % KI = [0.31, 0.57], p <.001) wurden jeweils mit auf Bonferroni-adjustiertem Alpha-Niveau signifikant geringerer Wahrscheinlichkeit positive Praktiken eingesetzt als im Expra (Details siehe Online-Materialien, LINK).

Ein linearer Trend über den Studienverlauf wie bei QRPs zeigte sich in einer weiteren Analyse nicht, wohl aber ein quadratischer Trend (B = 0,34, SE = 0,10, OR = 1,40, 95 % KI = [1.16, 1.69], p <.001\*\* ). Die Unterschiede zwischen Bachelorarbeiten und Expras (B = -0,24, SE = 0,10, OR = 0,79, 95 % KI = [0.65, 0.96], p = .017) sowie zwischen Masterarbeiten und Expras (B = 0,21, SE = 0,18, OR = 1,23, 95 % KI = [0.87, 1.74], p = .249) waren in der vorgehenden Analyse statistisch nicht signifikant, die Odds Ratios deuten aber tendenziell darauf hin, dass verglichen mit Expras in Bachelorarbeiten mit geringerer und in Masterarbeiten mit höherer Wahrscheinlichkeit positive Forschungspraktiken angewendet werden. Vor diesem Hintergrund erscheint der quadratische Trend für den Studienverlauf plausibel.

### Eindruck von Informiertheit

Welche Faktoren hängen damit zusammen, ob Studierende sich über den Themenkomplex gut informiert fühlen?

In einer weiteren explorativen Analyse untersuchten wir den Zusammenhang zwischen den verfügbaren, potentiell informativen Kovariaten und der subjektiven Informiertheit unserer Teilnehmenden über den Themenkomplex “Replikationskrise und Open Science”. Für die Analyse nutzten wir ein lineares gemischtes Modell mit einem zufälligen Achsenabschnitt für die aktuelle (bei Alumni/Alumnae: letzte) Universität (SD = 0,38, (1) = 135,66, p <.001). Die Schätzungen der fixen Koeffizienten sind in Tabelle 6 dargestellt.

Die Analyse deutet auf signifikante positive Zusammenhänge der subjektiven Informiertheit mit der Anzahl von durchgeführten oder begonnenen Projekten und der Wichtigkeit, die dem Themenkomplex “Replikationskrise und Open Science” beigemessen wird. Ob es de facto eine im Netzwerk der Open Science Initiativen gelistete Open Science Initiative an der jeweiligen Universität gibt scheint nicht mit der subjektiven Informiertheit der Teilnehmenden zusammenzuhängen, wohl aber das Wissen der Teilnehmenden um eine Open Science Initiative an ihrer Universität: Sowohl Teilnehmende, die sich über diese Frage unsicher sind, als auch solche, die eine Intiative an ihrer Universität kennen (etwas größerer Zusammenhang als bei unsicheren Teilnehmenden), schätzen sich als besser informiert ein, als Teilnehmende, die keine Initiative an ihrer Universität kennen. Schließlich zeigt sich auch ein deutlicher Zusammenhang zwischen einer berichteten Behandlung der Replikationskrise in der Lehre und der subjektiven Informiertheit der Teilnehmenden: Mit einem Beta-Gewicht von 0,63 ist dies sogar der stärkste in dieser Analyse beobachtete Zusammenhang. Demnach schätzten sich Studierende, die über eine Behandlung der Replikationskrise in der Lehre an ihrer Universität berichteten im Mittel um 0,63 Standardabweichungen besser informiert ein (Referenzkategorie: Keine Behandlung in der Lehre).

Weitere explorative Analysen führten wir zur Behandlung der Replikationskrise in der Lehre und zur subjektiv eingeschätzten Wichtigkeit des Themenkomplexes “Replikationskrise und Open Science” durch unsere Teilnehmenden durch. Die Ergebnisse können in den zusätzlichen Online-Materialien (LINK) abgerufen werden.

# Diskussion

# Abbildungen

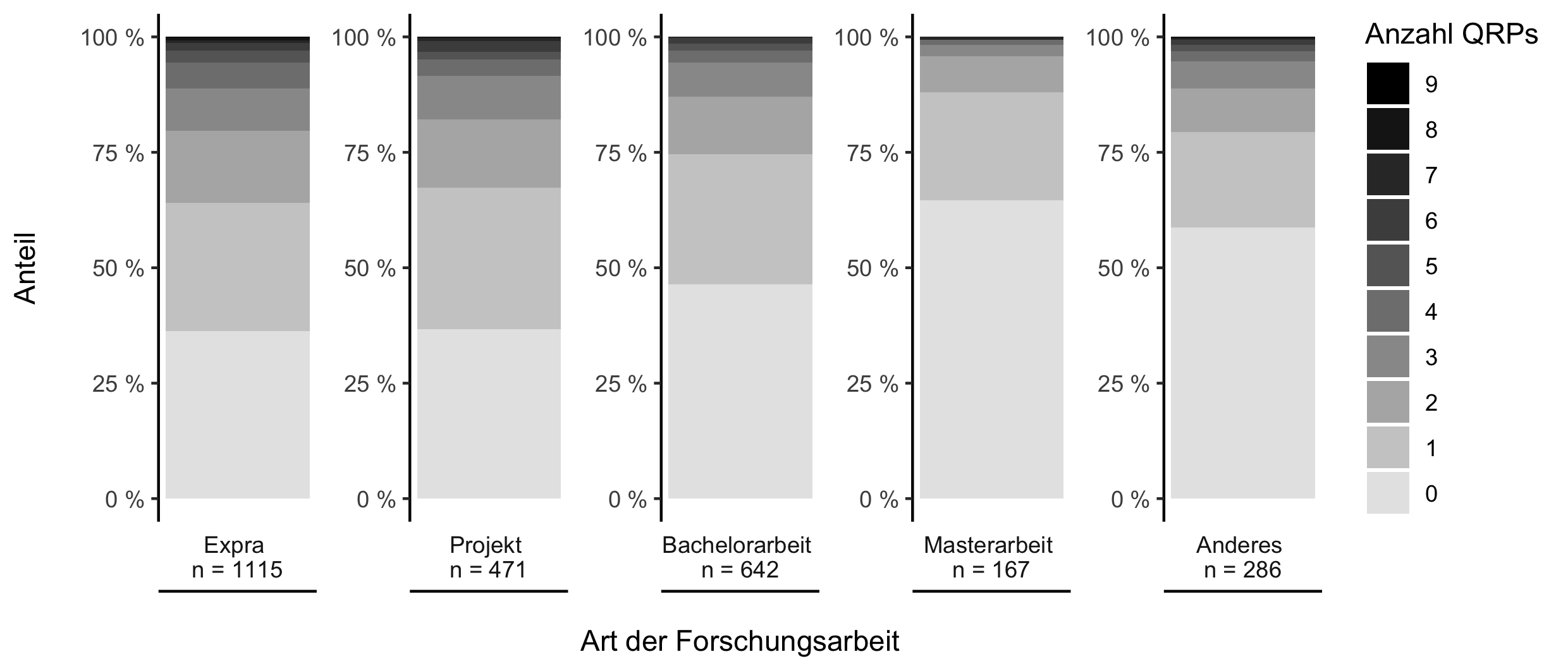


Figure 1: Anzahl eingesetzter fragwürdiger Forschungspraktiken aufgeschlüsselt nach Art des Projekts.

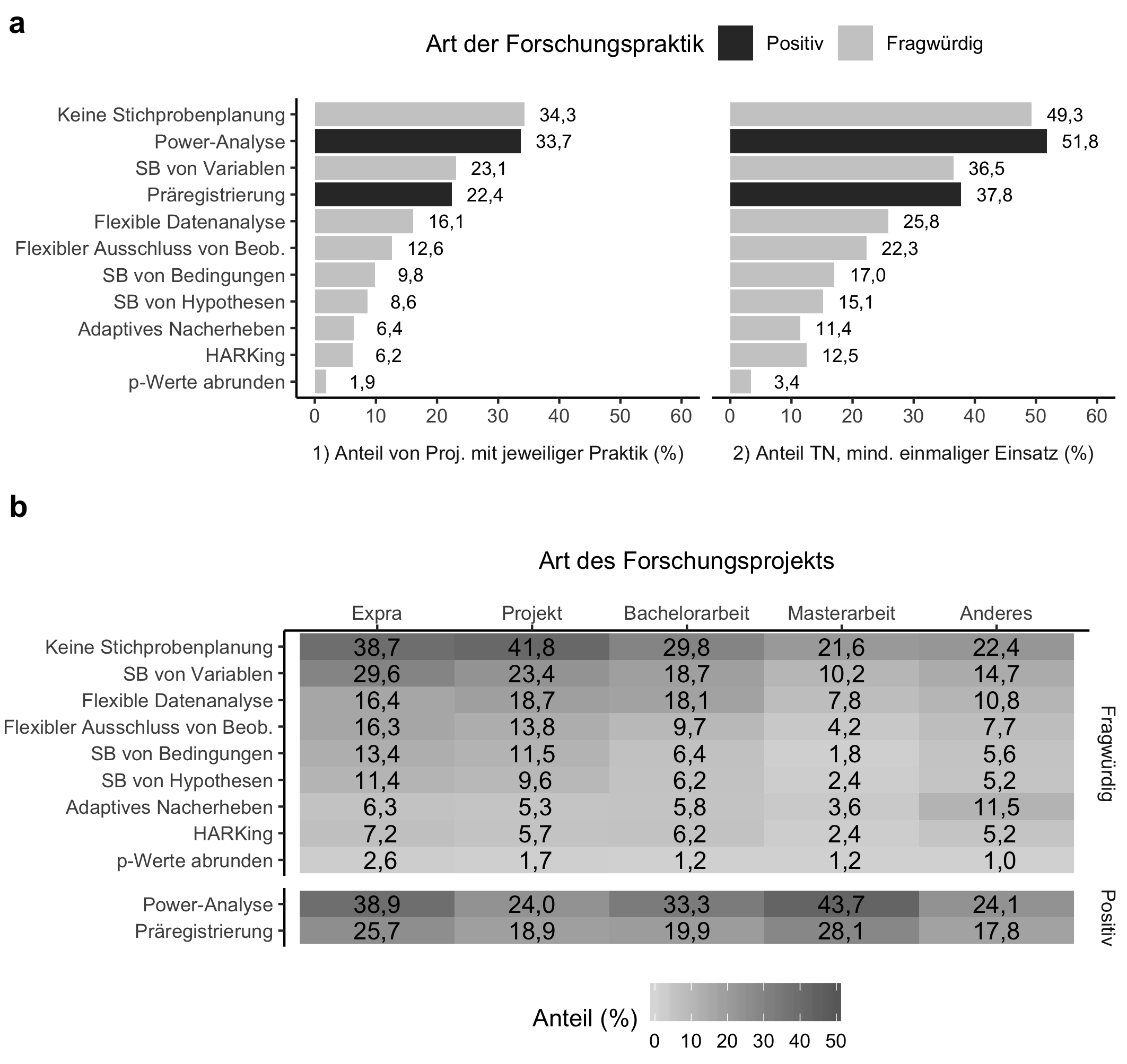


Figure 2: Anteil von Projekten, in denen die jeweilige Forschungspraktik eingesetzt wurde. S.B. = Selektives Berichten.

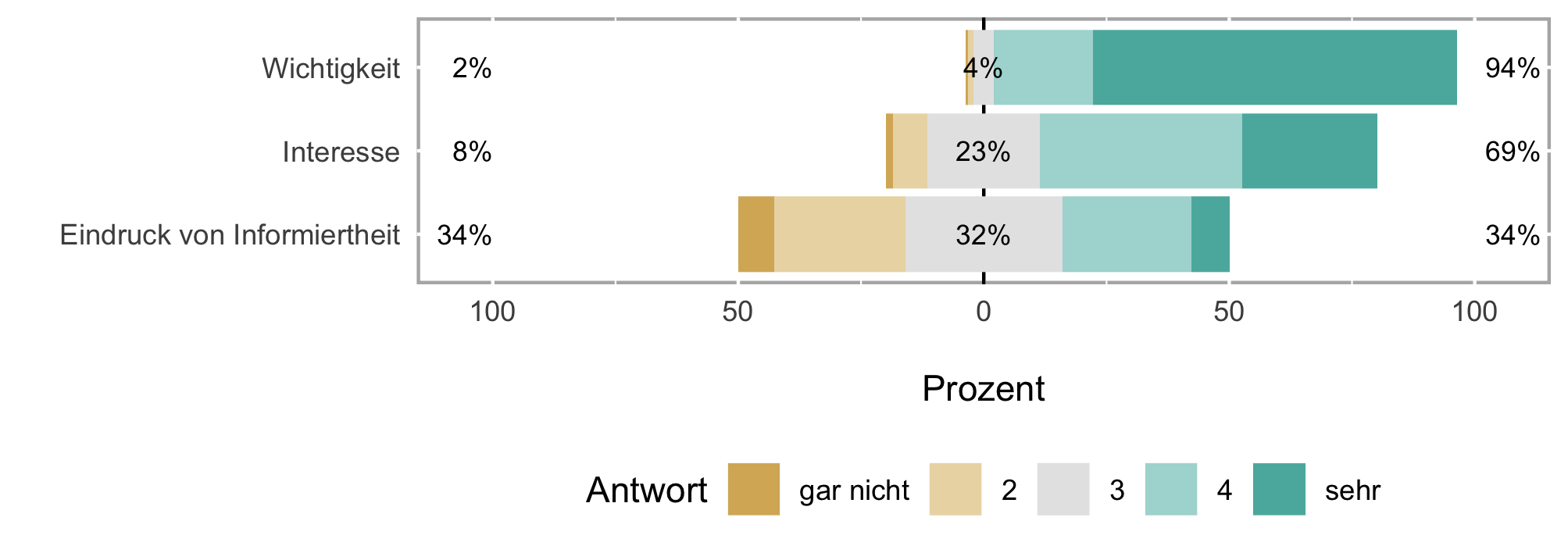


Figure 3: Einstellungen der Teilnehmenden zum Themenkomplex “Open Science und Replikationskrise” auf einer 5-Punkte Likert-Skala von “gar nicht” bis “sehr”. Bedeutung der Prozentzahlen: “gar nicht” & “2” (links), “3” (mittig), “4” & “sehr” (rechts).

2

# Tabellen

Tabelle 1. Korrelationen zwischen numerischen Variablen.

|  | Alter | Semester | Wichtigkeit | Subj. Inform. | Interesse | Anz. Projekte |
| --- | --- | --- | --- | --- | --- | --- |
| Alter | 1.00 |  |  |  |  |  |
| Semester | 0.15\*\*\* | 1.00 |  |  |  |  |
| Wichtigkeit | 0.10\*\*\* | 0.07\* | 1.00 |  |  |  |
| Subj. Inform. | 0.06 | 0.10\*\*\* | 0.27\*\*\* | 1.00 |  |  |
| Interesse | 0.15\*\*\* | 0.04 | 0.49\*\*\* | 0.19\*\*\* | 1.00 |  |
| Anz. Projektea | 0.32\*\*\* | 0.23\*\*\* | 0.13\*\*\* | 0.18\*\*\* | 0.13\*\*\* | 1.00 |
| Markierungen: \* p < 0.0083, \*\* p < 0.0017, \*\*\* p < 0.0002 (adjustierte Alpha-Niveaus nach Bonferroni) | | | | | | |
| aAnzahl der von der/dem jeweiligen Teilnehmenden begonnenen oder durchgeführten Projekte. | | | | | | |

Tabelle 2. Abgefragte Forschungspraktiken und Begleittexte.

| Bezeichnung (kurz) | Begleittext |
| --- | --- |
| **Fragwürdige Forschungspraktiken** | |
| Datenerhebung ohne konkrete Stichprobenplanung beginnen (Keine Stichprobenplanung) | Es wird eine willkürliche Zahl von Versuchspersonen erhoben, oder es werden Daumenregeln angewandt (z.B. mindestens 20 VP pro Bedingung). Im Vorfeld eines Projekts findet keine Power-Analyse statt, um die Stichprobengröße zu bestimmen. |
| Über zusätzliche Erhebung entscheiden, nachdem die Ergebnisse bekannt sind (Adaptives Nacherheben) | Beispiel: Bei einem knapp nicht signifikanten p-Wert 10 zusätzliche Versuchspersonen rekrutieren (z.B. in der Hoffnung, doch noch ein signifikantes Ergebnis zu erzielen) |
| Flexible Datenanalyse ohne Kennzeichnung (Flexible Datenanalyse) | Beispiel: Verschiedene, im Vorfeld nicht genau festgelegte statistische Tests für ein und dieselbe Hypothese durchführen oder den Einfluss von Kovariaten (z.B. Geschlecht) untersuchen, über die es ursprünglich keine spezifischen Hypothesen gab. Die Analyse wird nicht als explorativ gekennzeichnet. |
| Flexibler Ausschluss von Ausreißern (Flexibler Ausschluss von Beob.) | Beispiel: Der ursprüngliche Test einer Hypothese ergab einen p-Wert von p = .13. Nach dem Ausschluss von vier Ausreißern (±3 Standardabweichungen über dem Mittelwert) wird ein p-Wert von p = .043 erreicht. Daraufhin wird entschieden, die Ausreißer auszuschließen, um den p-Wert signifikant zu machen. Im Bericht wird nur die zweite Analyse berichtet. |
| Hypothesen an Ergebnisse anpassen (HARKing) | Beispiel: Vor der Analyse der Daten gab es die Hypothese, dass eine untersuchte Therapie-Methode wirksam ist. Der Effekt wird nicht signifikant, aber es zeigt sich eine Interaktion mit dem Geschlecht der Patient\*innen. Deshalb wird die Hypothese dieser Studie geändert, die Rolle des Geschlechts wird in die Hypothese mit aufgenommen. |
| Selektives Berichten von Hypothesen (SB von Hypothesen) | Beispiel: Von den ursprünglichen fünf Hypothesen lieferten zwei signifikante Ergebnisse. Nur diese beiden Hypothesen werden berichtet. |
| Selektives Berichten von experimentellen Bedingungen (SB von Bedingungen) | Beispiel: Von drei Bedingungen (Kontrollgruppe, Experimentalgruppe 1 und Experimentalgruppe 2) nur die ersten beiden im Artikel/Bericht beschreiben, zwischen denen es einen signifikanten Unterschied gab. |
| Selektives Berichten von Variablen | Beispiel: Von zehn erhobenen abhängigen Variablen nur die drei Variablen im Bericht/Artikel beschreiben, die statistisch signifikant mit der unabhängigen Variable zusammenhingen. |
| p-Werte abrunden | Beispiel: Der Test für die zentrale Hypothese ergibt einen p-Wert von p = .058, im Bericht wird p = .05 geschrieben. |
| **Positive Forschungspraktiken** | |
| Poweranalyse durchführen (Power-Analyse) | Beispiel: Vor der Datenerhebung wurde ausgerechnet, dass die Stichprobe aus mindestens 60 Personen bestehen muss, um den zu untersuchenden Effekt mit einer Wahrscheinlichkeit von z.B. 80% finden zu können. |
| Ein Experiment oder Teile davon präregistrieren (Präregistrierung) | Beispiel: Vor der Datenerhebung wurde schriftlich festgehalten, welche Daten erhoben, welche Hypothesen untersucht und welche statistischen Tests verwendet werden. |

Tablle 3. Anteil von Teilnehmenden, die in der jeweiligen Lehrveranstaltung an ihrer aktuellen (Alumni: letzten) Universität eine Thematisierung der Replikationskrise erfahren haben (N = 1413).

| Lehrveranstaltung | n | Anteil (%) |
| --- | --- | --- |
| Methodenlehre, Statistik, Expra | 770 | 54,50 |
| Sozialpsychologie | 387 | 27,40 |
| Allgemeine Psychologie | 268 | 19,00 |
| Einführungsveranstaltung | 263 | 18,60 |
| Diagnostik | 244 | 17,30 |
| Differentielle Psychologie | 197 | 13,90 |
| Andere | 164 | 11,60 |
| Klinische Psychologie | 121 | 8,60 |
| Biologische / Neuropsychologie | 106 | 7,50 |
| Arbeits-, Betriebs- und Organisationspsychologie | 92 | 6,50 |
| Pädagogische Psychologie | 71 | 5,00 |
| Entwicklungspsychologie | 62 | 4,40 |
| Gesundheitspsychologie | 16 | 1,10 |
| Medien- / Kommunikationspsychologie | 13 | 0,90 |
| Forensische Psychologie | 11 | 0,80 |

Tablle 4. Anteil von Teilnehmenden, die an ihrer aktuellen (bei Alumni: letzten) Universität in mindestens einer Lehrveranstaltung eine kritische Betrachtung der jeweiligen Forschungspraktik erfahren haben (N = 1413).

| Fragwürdige Forschungspraktik | n | Anteil (%) |
| --- | --- | --- |
| HARKing | 1046 | 74.0 |
| Keine Stichprobenplanung | 923 | 65.3 |
| S.B. von Variablen | 906 | 64.1 |
| S.B. von Hypothesen | 864 | 61.1 |
| Flexible Stichprobengröße | 854 | 60.4 |
| Flexibler Ausschluss von Beob. | 847 | 59.9 |
| S.B. von Bedingungen | 832 | 58.9 |
| Flexible Datenanalyse | 717 | 50.7 |
| Runden von p-Werten | 564 | 39.9 |
| Nicht sicher | 130 | 9.2 |
| Keine | 57 | 4.0 |

Tabelle 5. Modell zur Untersuchung der Zusammenhänge von Kovariaten mit dem Einsatz fragwürdiger Forschungspraktiken. Schätzungen der fixen Effekte. Abhängige Variable: Forschungspraktik angewendet (0 – Nein, 1 – Ja).

| Prädiktor | B | SE | OR | KI (95.00 %) | z | p |
| --- | --- | --- | --- | --- | --- | --- |
| (Intercept) | -2,42 | 0,39 | 0,09 | [0.04, 0.19] | -6,26 | <.001\*\*\* |
| Alter | 0,05 | 0,05 | 1,06 | [0.96, 1.16] | 1,08 | .281 |
| Geschlecht - Männlich | 0,05 | 0,11 | 1,05 | [0.84, 1.31] | 0,45 | .650 |
| Geschlecht - Divers | -0,11 | 1,00 | 0,89 | [0.13, 6.30] | -0,11 | .910 |
| Geschlecht - Keine Antwort | -0,39 | 0,49 | 0,68 | [0.26, 1.75] | -0,80 | .421 |
| SF - Studiere nicht (mehr) | 0,71 | 0,29 | 2,03 | [1.15, 3.56] | 2,46 | .014 |
| SF - Master | 0,18 | 0,13 | 1,20 | [0.93, 1.54] | 1,41 | .159 |
| SF - Andere | -0,87 | 1,15 | 0,42 | [0.04, 3.97] | -0,76 | .450 |
| Anzahl emp. Projekte | -0,09 | 0,06 | 0,92 | [0.81, 1.04] | -1,33 | .183 |
| Interesse | 0,03 | 0,05 | 1,03 | [0.93, 1.14] | 0,51 | .612 |
| Wichtigkeit | -0,19 | 0,05 | 0,82 | [0.75, 0.91] | -3,92 | <.001\*\* |
| Eindruck von Informiertheit | -0,32 | 0,05 | 0,73 | [0.66, 0.81] | -5,94 | <.001\*\*\* |
| OSI an der Uni - Ja | 0,15 | 0,16 | 1,16 | [0.86, 1.58] | 0,97 | .331 |
| OSI bekannt - Unsicher | 0,03 | 0,12 | 1,03 | [0.81, 1.31] | 0,26 | .798 |
| OSI bekannt - Ja | -0,27 | 0,12 | 0,76 | [0.60, 0.97] | -2,22 | .027 |
| RK gelehrt - Unsicher | 0,52 | 0,26 | 1,69 | [1.01, 2.81] | 2,00 | .046 |
| RK gelehrt - Ja | -0,06 | 0,13 | 0,94 | [0.73, 1.21] | -0,50 | .618 |
| Praktik gelehrt - Ja | -0,08 | 0,06 | 0,92 | [0.82, 1.04] | -1,34 | .180 |
| Projekt – Anderes | -0,79 | 0,11 | 0,45 | [0.37, 0.56] | -7,29 | <.001\*\*\* |
| Projekt – Projektarbeit | -0,27 | 0,08 | 0,77 | [0.66, 0.89] | -3,38 | .001\* |
| Projekt – Bachelorarbeit | -0,46 | 0,07 | 0,63 | [0.55, 0.72] | -6,86 | <.001\*\*\* |
| Projekt – Masterarbeit | -1,31 | 0,17 | 0,27 | [0.19, 0.37] | -7,89 | <.001\*\*\* |
| Anmerkungen. Datenpunkte insg. = 19 195, Projekte = 2 142, Teilnehmende = 1 078 | | | | | | |
| Referenzkategorien: Weiblich (Geschlecht), Bachelor (Studienfortschritt), Nein (OSI an der Uni), Nein (OSI bekannt), Nein (RK gelehrt), Nein (Praktik gelehrt), Expra (Projekt). | | | | | | |
| Abkürzungen: B – Beta-Gewicht, SE – Standard Error [Standardfehler], OR – Odds Ratio [Chancenverhältnis], KI - Konfidenzintervall, RK - Replikationskrise, OSI - Open Science Initiative, QRP – Questionable Research practice [Fragwürdige Forschungspraktik], SB – Selektives Berichten, SF - Studienfortschritt | | | | | | |
| Markierungen: \* p < 0.00238, \*\* p < 0.00048, \*\*\* p < 0.00005 (adjustierte Alpha-Niveaus nach Bonferroni) | | | | | | |

Tabelle 6. Modell zur Untersuchung von Zusammenhängen der Kovariaten mit dem Eindruck der Teilnehmenden, wie gut sie über den Themenkomplex ‘Replikationskrise und Open Science’ informiert sind. Schätzungen der fixen Effekte. Abhängige Variable: Eindruck von Informiertheit (numerisch, min: 1 (gar nicht), max: 5 (sehr), z-standardisiert).

|  | B | SE | t | p |
| --- | --- | --- | --- | --- |
| (Intercept) | -0,88 | 0,08 | -11,62 | <.001\*\*\* |
| Alter | -0,02 | 0,03 | -0,78 | .436 |
| Geschlecht - Männlich | 0,14 | 0,06 | 2,38 | .017 |
| Geschlecht - Divers | -0,27 | 0,49 | -0,56 | .574 |
| Geschlecht - Keine Antwort | 0,07 | 0,22 | 0,34 | .733 |
| SF - Studiere nicht (mehr) | -0,07 | 0,13 | -0,49 | .623 |
| SF - Master | -0,01 | 0,07 | -0,10 | .918 |
| SF - Andere | 0,08 | 0,29 | 0,28 | .776 |
| Anzahl emp. Projekte | 0,10 | 0,02 | 4,34 | <.001\*\*\* |
| Interesse | 0,07 | 0,03 | 2,74 | .006 |
| Wichtigkeit | 0,14 | 0,03 | 5,24 | <.001\*\*\* |
| OSI an der Uni - Ja | 0,24 | 0,10 | 2,34 | .028 |
| OSI bekannt - Unsicher | 0,27 | 0,06 | 4,74 | <.001\*\*\* |
| OSI bekannt - Ja | 0,42 | 0,06 | 6,97 | <.001\*\*\* |
| RK gelehrt - Unsicher | 0,09 | 0,12 | 0,76 | .450 |
| RK gelehrt - Ja | 0,63 | 0,06 | 10,54 | <.001\*\*\* |
| Anmerkungen. N = 1 413 | | | | |
| Referenzkategorien: Weiblich (Geschlecht), Bachelor (Studienfortschritt), Nein (OSI an der Uni), Nein (OSI bekannt), Nein (RK gelehrt). | | | | |
| Abkürzungen: B – Beta-Gewicht, SE – Standard Error [Standardfehler], OR – Odds Ratio [Chancenverhältnis], KI - Konfidenzintervall, SF - Studienfortschritt, OSI - Open Science Initiative, RK - Replikationskrise | | | | |
| Markierungen: \* p < 0.00333, \*\* p < 0.00067, \*\*\* p < 0.00007 (adjustierte Alpha-Niveaus nach Bonferroni) | | | | |

# Literatur

John, L. K., Loewenstein, G., & Prelec, D. (2012). Measuring the Prevalence of Questionable Research Practices With Incentives for Truth Telling. *Psychological Science*, *23*(5), 524–532. <https://doi.org/10.1177/0956797611430953>

Simmons, J. P., Nelson, L. D., & Simonsohn, U. (2018). False-Positive Citations. *Perspectives on Psychological Science*, *13*(2), 255–259. <https://doi.org/10.1177/1745691617698146>