A biased mind: Significance as a publication booster

Julia Jerke, Antonia Velicu, Heiko Rauhut

University of Zurich

Publication bias

We take a different approach

- Publication bias is the results of decisions and behavior by individual actors
- Each of the involved actors has own motives, perceptions, etc.
- So, in our survey we go a step back and look at their beliefs about significance

We want to know

- What beliefs about the significance of results contribute to publication bias?
- Do researchers associate significance with quality?

University of Zurich[UZH]

Institute of Sociology

@SurveyAcademics

https://linktr.ee/ZSoA

The Zurich Survey of Academics

The *Zurich Survey of Academics*

Project team: Prof. Dr. Heiko Rauhut, Dr. David Johann, Dr. Julia Jerke, Justus Rathmann, Antonia Velicu

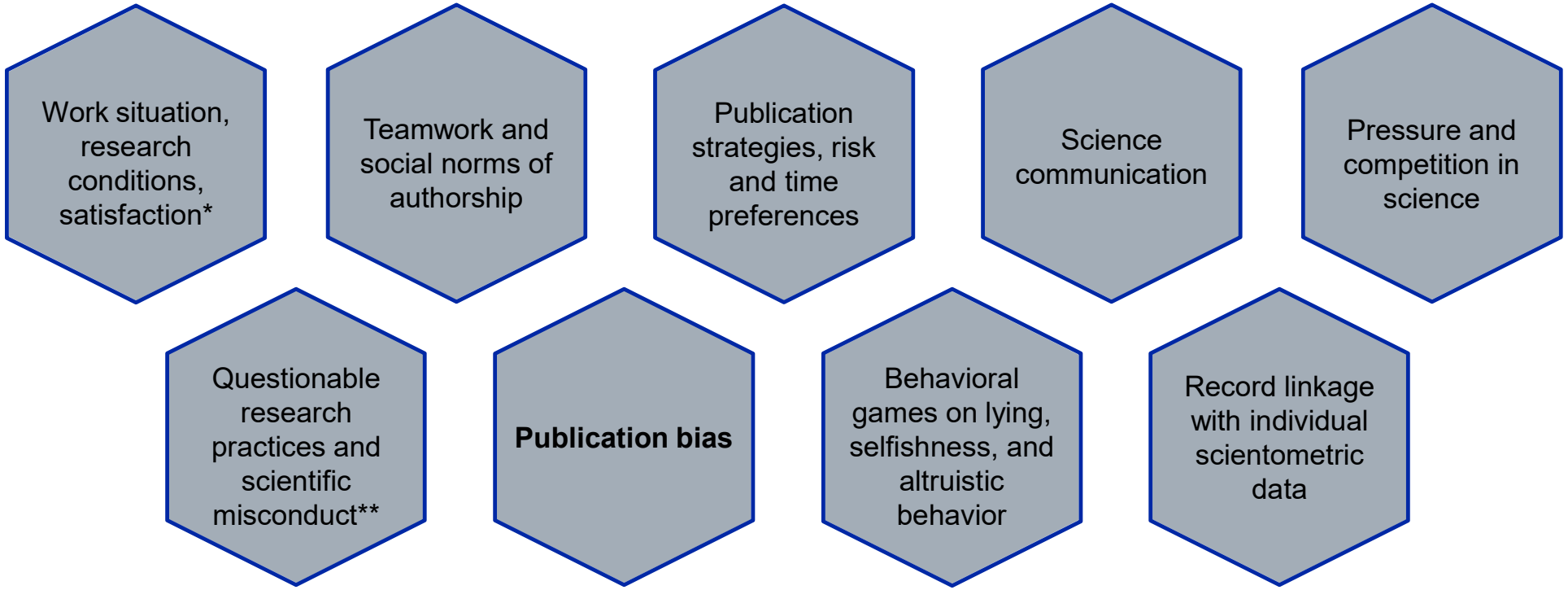
Overview

- Large-scale web survey among scientists at universities in Austria, Germany, and Switzerland (DACH region)
- Part of the SNF project ”Social Norms, Cooperation and Conflict in Scientific Collaborations” (CONCISE)
- Aim: in-depth insights into the everyday work of researchers in the DACH region and into how researchers deal with conflicts and increasing competition and pressure to publish
- Omnibus survey: topics cover among others working conditions in science, publication pressure, collaborations, authorship conflicts, scientific misconduct and **publication bias**

Implementation

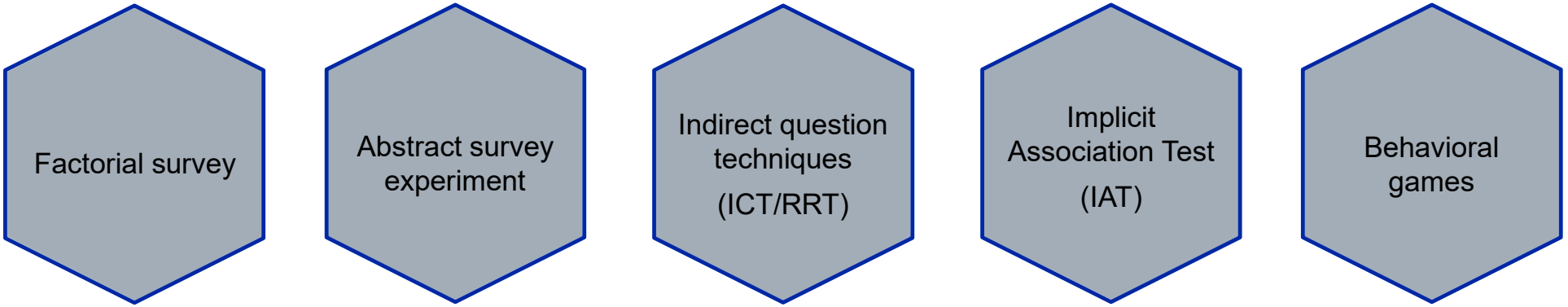
- Survey period: February to April 2020
- Approx. 150.000 scientists in the DACH region have been contacted
- Final sample: N = 15’778 scientists from 263 universities (response rate approx. 11.2%)

Topics covered in the survey



* Raabe, I. J., Ehlert, A., Johann, D., & Rauhut, H. (2020). Satisfaction of scientists during the COVID-19 pandemic lockdown. *Humanities and Social Sciences Communications*, *7*(1), 1-7.
**Jerke, J., Johann, D., Rauhut, H., Thomas, K., & Velicu, A. (2020). Handle with Care: Implementation of the List Experiment and Crosswise Model in a Large-Scale Survey on Academic Misconduct. *Field Methods*.

Implementation of a variety of methods



Survey report **"The Zurich Survey of Academics: Methods, Design, and Data"**

- Rauhut, H., Johann, D., Jerke, J., Rathmann, J., Velicu, A. (2021). The Zurich Survey of Academics: Methods, Design, and Data. Zurich: University of Zurich.
- https://www.zora.uzh.ch/id/eprint/204689/

Sample description

- Majority of the respondents are mid-level faculty
- But we also have a substantial number of professors in the sample
- Humanities and social sciences are the most strongly represented subjects in the sample
- More than half of the respondents are from Germany (not surprising)

	N	%
Gender		
Male	8'790	55.7
Female	6'882	43.6
Divers	91	0.6
Academic status		
Professor	3'275	20.8
Postdoc	6'014	38.1
Predoc	6'489	41.1
Academic field		
Humanities and social sciences	6'687	42.4
Life sciences	2'653	16.8
Natural sciences	2'762	17.5
Engineering	2'247	14.3
Other	1'421	9.0
Country		
Germany	8'182	51.8
Austria	2'771	17.6
Switzerland	4'825	30.6
	N = 15'778	

University of Zurich[UZH]

Institute of Sociology

@SurveyAcademics

https://linktr.ee/ZSoA

Publication Bias Project in the ZSoA

Publication bias

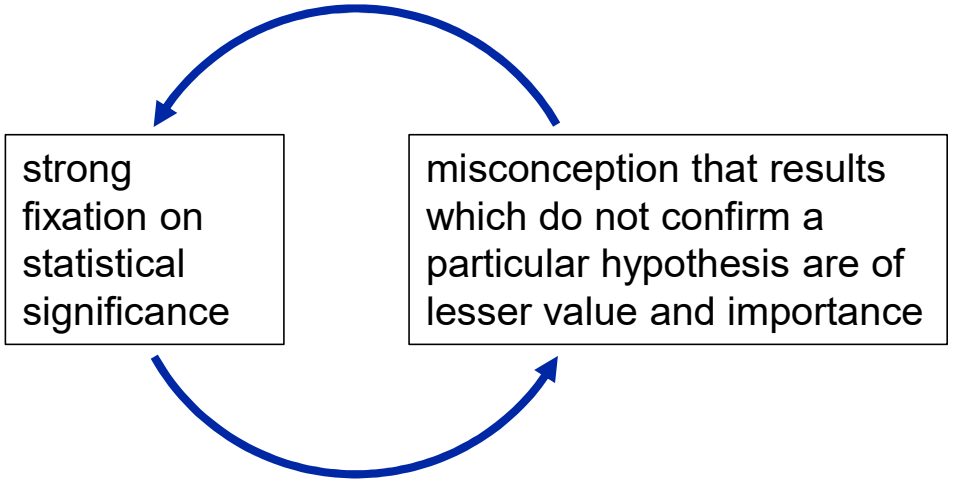
Framework

- PB describes the tendency to preferentially publish significant over insignificant results
- PB closely related to the widespread use of null-hypothesis testing in combination with judgments on the statistical significance of results as a criterion for evaluating scientific evidence

Overall, negative results seem to disappear from the scientific literature (e.g., Fanelli 2012; Sterling et al. 1995)

Studies with negative results are less likely to be published (e.g., Franco et al. 2014; Turner et al. 2008)

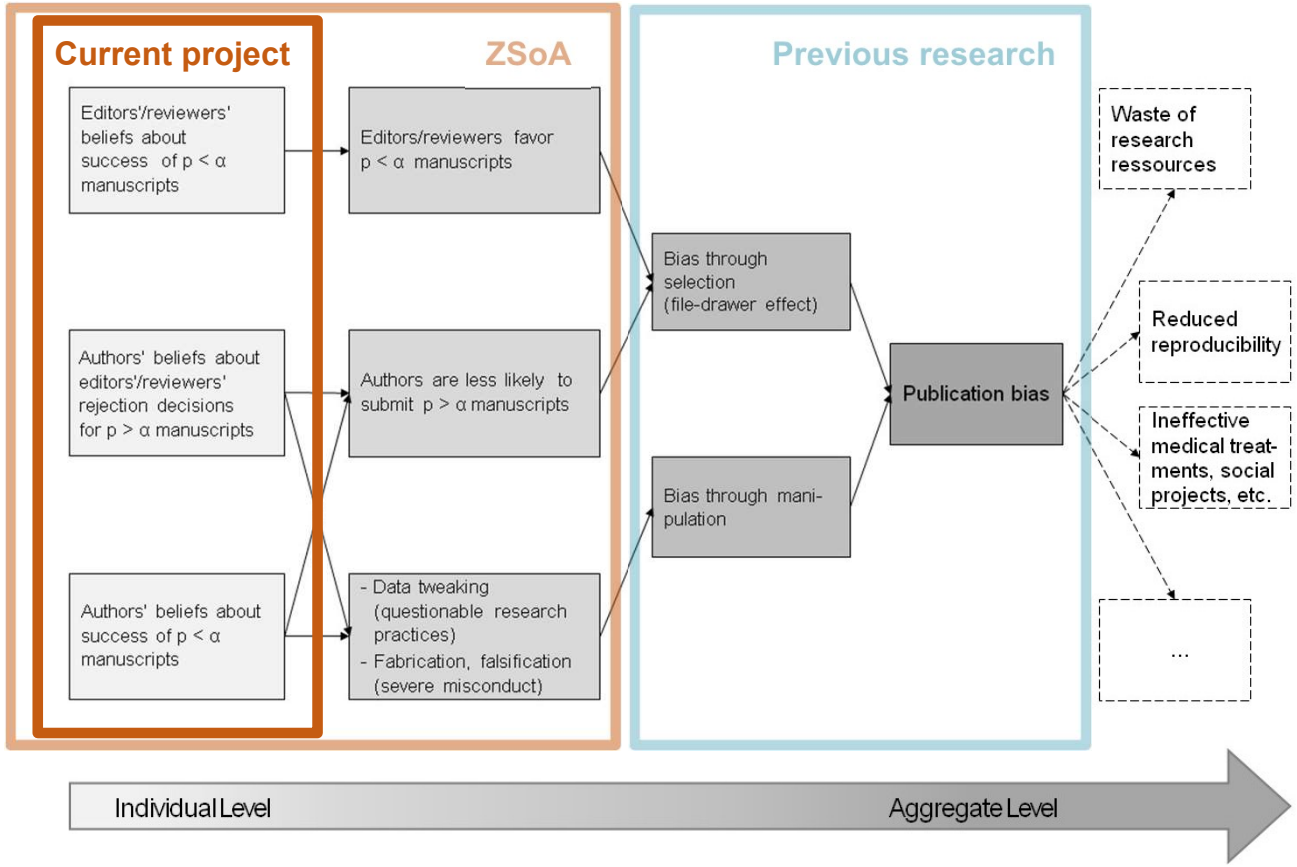
Publication bias has been documented in a variety of disciplines (e.g., Auspurg et al. 2014; Brodeur et al. 2016; Gerber & Malhotra 2008; Hedin et al. 2016)



In-depth analysis of scientists' perceptions and experiences with regard to significance and how these affect the scientists' behavior

Abstract survey experiment: evaluation of a hypothetical study for which we vary the strength of the statistical results

Publication bias and the *Zurich Survey of Academics*



In-depth analysis of scientists' perceptions and experiences with regard to significance and how these affect the scientists' behavior

Abstract survey experiment: evaluation of a hypothetical study for which we vary the strength of the statistical results

* Source of the scheme:
Jerke, J. (2020). Doctoral dissertation, University of Zurich.

Implementation of the survey experiment

Abstract vignette design

Basic idea

- Aim: understanding the cognitive concepts that lead to publication bias
- Approach:
  - Presenting respondents with a hypothetical study for which we vary the statistical significance of the results
  - Respondents are asked to evaluate that study
- Strategy: comparing respondents that receive a study with significant results with respondents that receive a study with insignificant results

Implementation of the survey experiment

Quantitative part: Abstract evaluation questions

Treatment 1

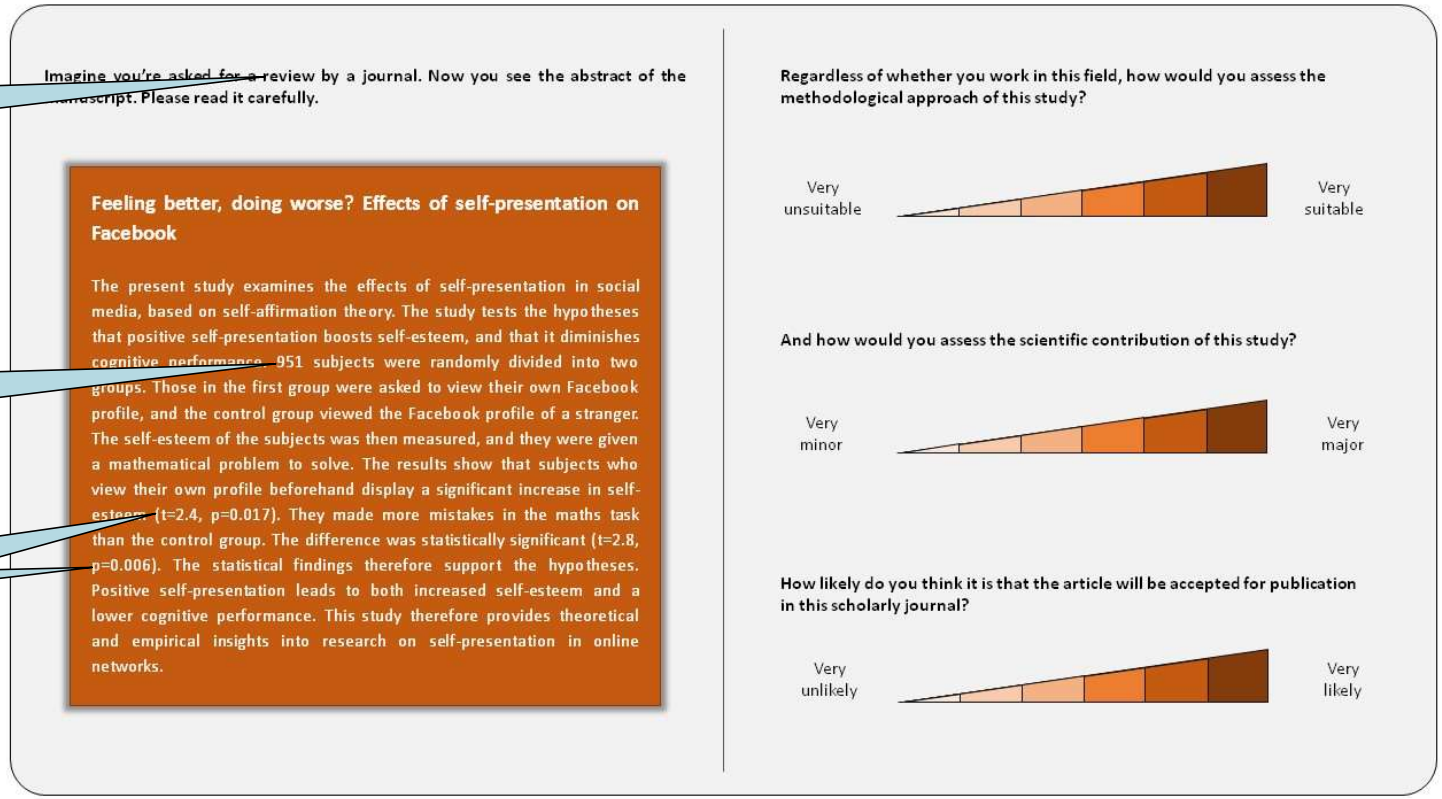
F_{expert} = review
$F_{neutral}$ = conference

Treatment 2

N_{small} = 159
N_{large} = 951

Treatment 3

R_{sig} = significant
R_{insig} = insignificant



Implementation of the survey experiment

Quantitative part: Abstract evaluation questions

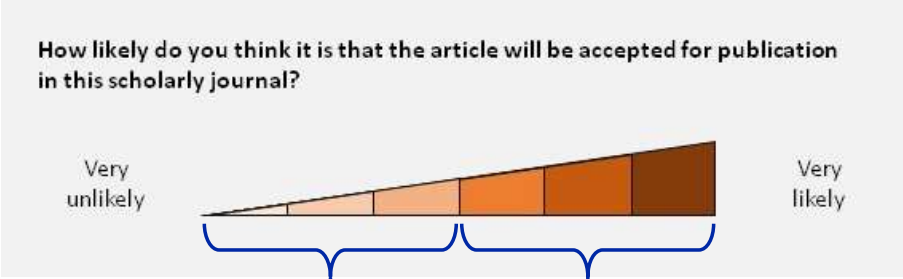
Implementation of the abstract vignettes (2 x 2 x 2 = 8):

- Treatment 1 - *Frame*: review vs. conference
- Treatment 2 - *Sample size*: small (159) vs. large (951)
- Treatment 3 - *Statistical significance of the results*: significant vs. not significant
- Respondents were randomly assigned to either one of the 8 abstract versions

➢ **This setup allows an investigation of causal relationships between the statistical significance of research results and judgements on the quality, relevance, and publication potential of an empirical study**

Implementation of the survey experiment

Qualitative part: Explanation of the assessment



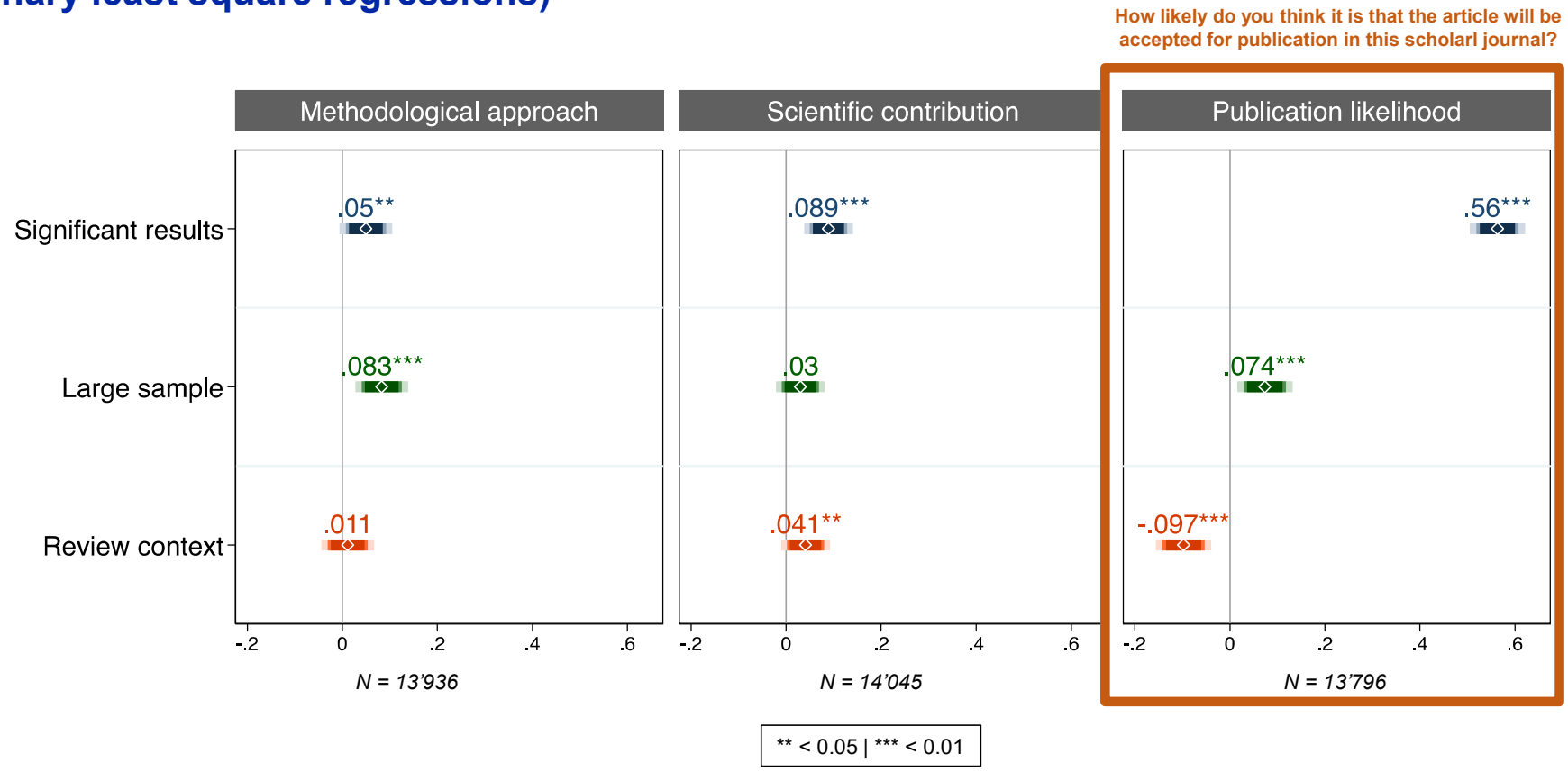
"You've stated that it is ***unlikely*** that the article will be accepted. On what do you base your assessment?"

"You've stated that it is ***likely*** that the article will be accepted. On what do you base your assessment?"

- **We received 11'522 detailed comments from the researchers in our survey**

Quantitative results

Main effects of the treatments
(Ordinary least square regressions)



Are there interaction effects?
(Ordinary least square regressions with inclusion of variable interactions)

- **None** of the interactions between *frame*, *sample size* and *significance* **were substantial**
- This indicates that there is no relationship between the different treatments

- Most surprisingly, there is **no interaction between *sample size* and *significance***
  - ➢ Researchers do not seem to take statistical power into account!
  - ➢ Insignificant results are “penalized” (or significant results “rewarded”) in any case, independent of whether the study may have had sufficient power or not

Do we see differences in academic status?

Status-specific variation

- Notably larger bias among junior researchers

  $b_{predoc} = 0.66,$

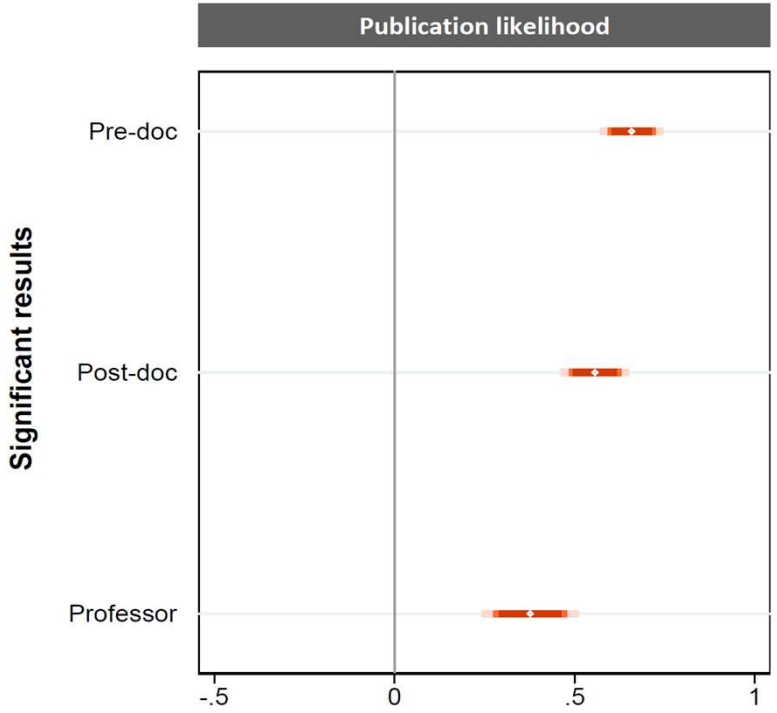
  $b_{postdoc} = 0.56,$

  $b_{prof} = 0.38$

- They are either…

  … more prone to contribute to such a bias, or

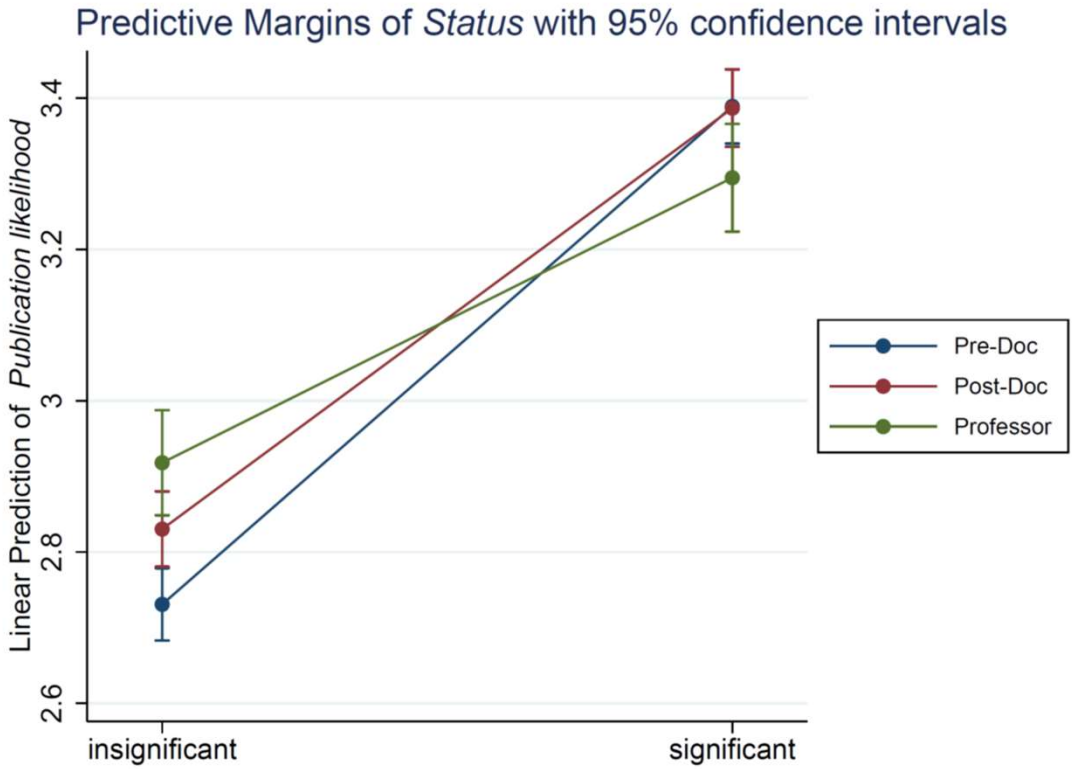
  … more sensitized to this bias



Do we see differences in academic status?

Looking at the predicted margins

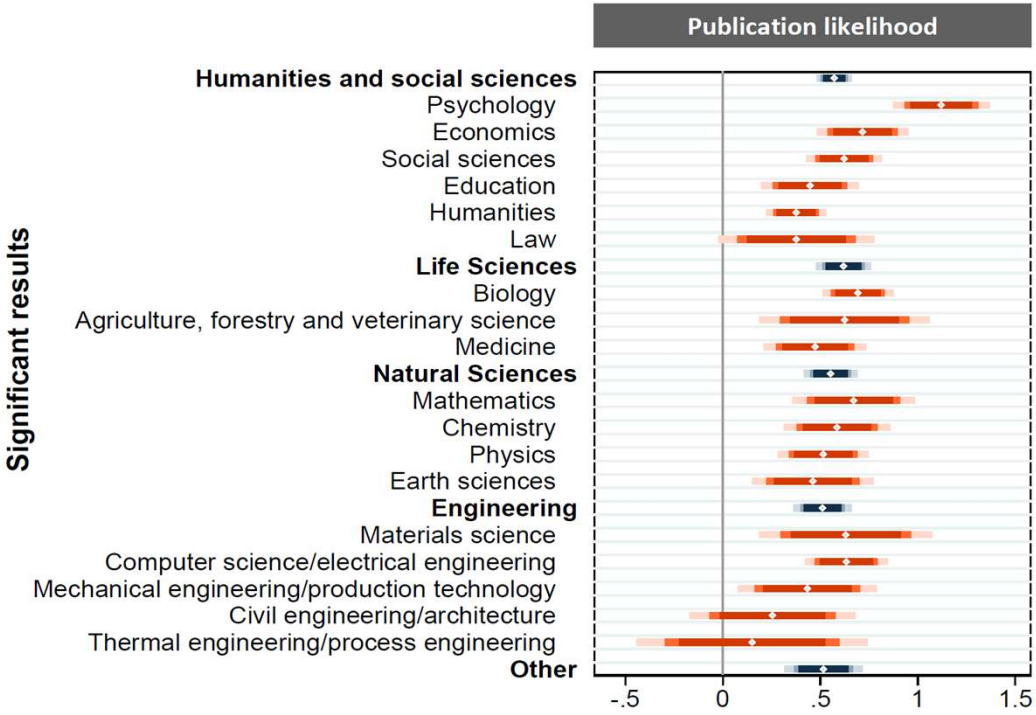
- Profs appear to differentiate the least with respect to the significance of the results



And what about differences between the disciplines?

Discipline-specific variation

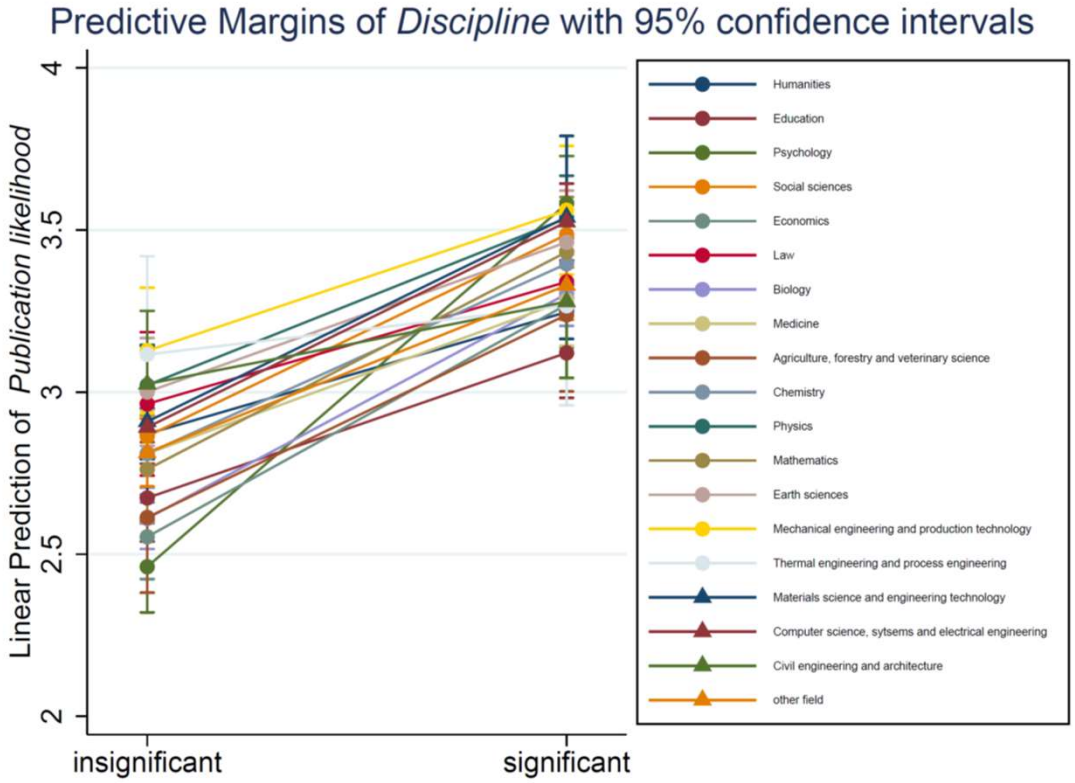
- Substantial effects observable in almost all disciplines
- With few exemptions ranging between 0.4 and 0.7
- Most remarkably: by far the largest bias among psychologists ($b = 1.12$)
- No such pronounced differences for Method and contribution rating and for the other treatment effects



And what about differences between the disciplines?

Looking at the predicted margins

- Psychology seems to be a clear exception
- Their predicted publication rating is the smallest of all disciplines
- Researchers from psychology on average seem to penalize insignificant results but also reward significant results more than researchers from any other discipline



Qualitative results

Qualitative analysis of the open answers

- We received 11’522 comments of which we already coded 8’367 comments (72.6%)

No. comments	Significant results	Insignificant results	*Total*
Publication likely	2’778 (84%)	1’850 (84%)	4’628 (84%)
Publication unlikely	2’784 (82%)	4’110 (84%)	6’894 (83%)
Total	5’562 (83%)	5’960 (84%)	11’522 (84%)

Qualitative analysis of the open answers

Resulting codes from the constant comparative method

Positive context (~in most cases publication *likely*)	**Negative context (~in most cases publication *unlikely*)**
Significant results	Insignficiant results
Reasonable results	Absurd results
Journal publication bias	Journal publication bias
Vogue topic	Uninteresting topic
Adequate methodology	Inappropriate methodology
Large sample	Small sample
Everything is publishable	-
Relevant study	Irrelevant study
Plausible story	Implausible story
Valid conclusions	Incoherent conclusions
Clear theory and hypothesis	Fuzzy theory and hypothesis
Transparent and good writing	Intransparent and bad writing

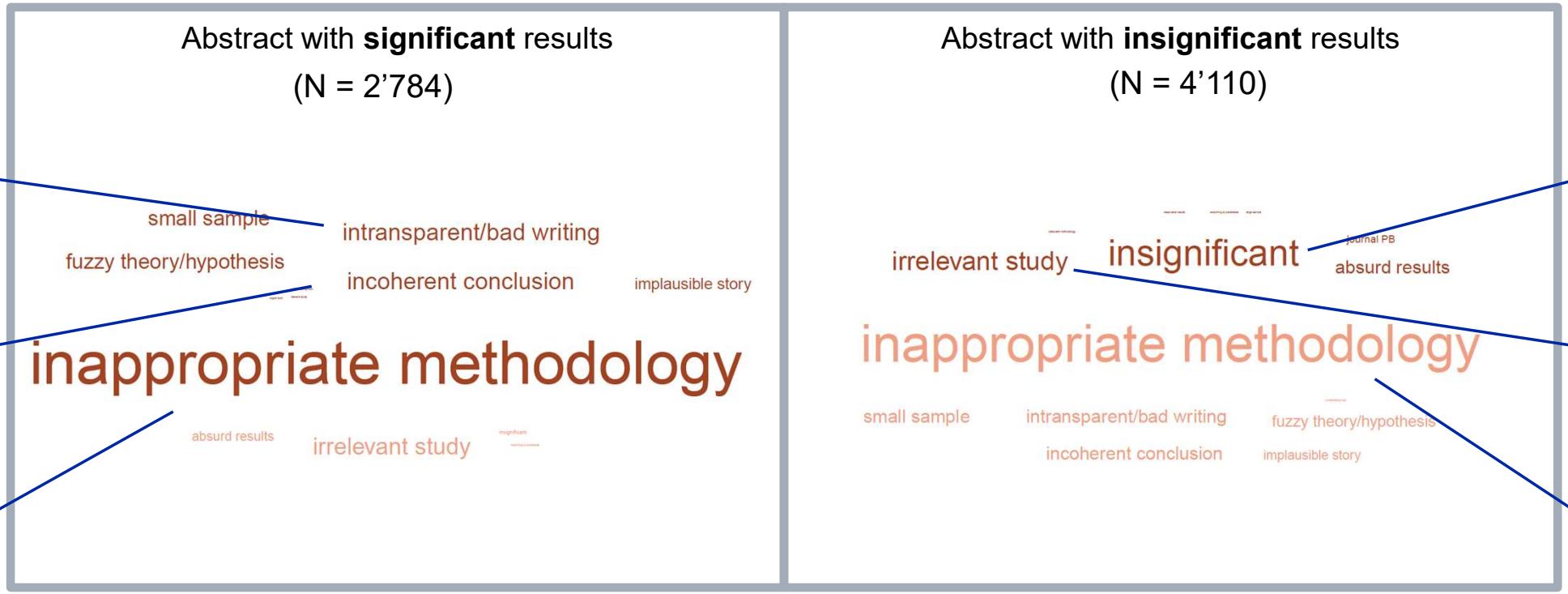
Most common reasons for a positive publication expectation*

Everything is publishable: more frequent as an explanation for the insignificant abstract

* Results are preliminary since the coding is not yet finished (8'367 comments)

Most common reasons for a negative publication expectation*

Statistical insignificance and irrelevance of study as frequent explanations for the insignificant abstract



* Results are preliminary since the coding is not yet finished (8'367 comments)

Relevance of the signficance of the results for the assessment*

- Around **12%** of the researchers explicitely mention the significance of the results as a reason for their assessment
- But taking into account their actual decision and the abstract that they received, we see some differences:

	Significant results	Insignificant results
Publication likely	**11.8%**	1.9%
Publication unlikely	0.2%	**24.1%**

➢ Significance as a necessary condition for the success of a study:
  ➢ significance of the results is **not specifically rewarded when it is present**
  ➢ but **penalized when it is not**

* Results are preliminary since the coding is not yet finished (8'367 comments)

Do we find an explanation for the strong differences by status from the quantitative results?

	Significant results		Insignificant results	
Publication likely	11.8%			
	$predoc = 13.9\%$	8% of them mention PB		
	$postdoc = 11.4\%$	5% of them mention PB		
	$prof = 8.1\%$	2.9% of them mention PB		
Publication unlikely			24.1%	
			$predoc = 26.7\%$	14.4% of them mention PB
			$postdoc = 23.3\%$	17.2% of them mention PB
			$prof = 20.3\%$	12.4% of them mention PB

«Significant results. It will find a place somewhere.»

«It's a negative result. Currently, I'm not aware of any venue that accepts or invites negative results, at least in my research field.»

*** translated from German

Do we also find an explanation for the strong differences by discipline from the quantitative results?

<table>
<tr><th></th><th colspan="2">Significant results</th><th colspan="2">Insignificant results</th></tr>
<tr><td rowspan="2">Publication likely</td><td colspan="2">11.8%</td><td colspan="2" rowspan="2"></td></tr>
<tr><td>$Psychology = \mathbf{27.9}\%$</td><td>Most other disciplines range between 5% and 15%</td></tr>
<tr><td rowspan="2">Publication unlikely</td><td colspan="2" rowspan="2"></td><td colspan="2">24.1%</td></tr>
<tr><td>$Psychology = \mathbf{38.3}\%$</td><td>Most other disciplines range between 15% and 30%</td></tr>
</table>

Limitations

- We do not measure actual behavior or actual publication bias
- Setup “only” mimics actual decision-making behavior in the field
- On the other hand:
    - This was not the purpose of our investigation!
    - And we are reassured by the extent and seriousness of the comments that were made by the researchers in our survey
- No clear distinction possible between whether the researchers themselves think that insignificant results should not be published or whether they just anticipate that journals would not publish them

Discussion (1)

- Significance of the results seems to clearly trigger associations with respect to the qualitiy and the publication chance of a study
- Meta-analytical studies show: correcting for bias afterwards won't solve the problem of biased research
- Instead: improving research quality right at where research starts
- It is necessary to change incentive structures and eventually the involved actors' behavior
- In the ***ZSoA***
  - only **17%** of the researchers say that they always write down their results
  - **28%** say that they tend to not write down results that do not meet their expectations
- Larger bias among junior scientists: they are either…

  … more prone to contribute to such a bias (**alarming**), or

  … more sensitized to this bias (**promising**)

Discussion (2)

- Implications: Initiatives to overcome the bias should focus on
    1. encouraging scientists to publish more frequently findings that allegedly are of lesser value, and
    2. on improving the image of negative results

- In the future:
    - linkage to bibiometric data possible → comparing «successful» researchers with less «successful» researchers (h-Index, IF of published papers)
    - Follow-up study with researchers in the ZSoA with more specific questions with respect to PB, publication behavior and the publishing system

University of Zurich[UZH]

Institute of Sociology

@SurveyAcademics

https://linktr.ee/ZSoA

Thank you!

Julia Jerke, Antonia Velicu, Heiko Rauhut
University of Zurich

jerke@soziologie.uzh.ch
velicu@soziologie.uzh.ch
rauhut@soziologie.uzh.ch