Challenges and new trends in pre-registration

Johannes Algermissen

Donders Institute, Radboud University

09-12-2020

j.Algermissen@donders.ru.nl

Set-up

- I aim for a 60 minutes talk
- We have a full program
- There is no designated time for discussion at the end; if you have questions, ask them straight-away, but be prepared that I might postpone the answer
- Slides will be shared under https://openscience-nijmegen.nl/_pages/os_nijmegen/ (under “Resources”)

Overview

- What is pre-registration (recap)?
- When is pre-registration a good idea (and when not)?
  - Distinguishing confirmatory and exploratory research
  - Transparent workflow documentation
  - A pedagogical tool stimulating clearer thinking
- Applications:
  - Pre-registration without theory
  - Pre-registration on secondary data
  - Pre-registration on many hypotheses/ tests
- Summary

What I promised for today

What: Challenges and new trends in pre-registration - by Johannes Algermissen (Donders Institute)

Have you maybe downloaded one of the OSF templates for pre-registration, started filling it in, and noticed that it does not work for you? Because you have already analyzed bits of your data? Because you cannot predict certain analysis steps in advance? Or because you have too many hypotheses? Is it still worth pre-registering at all? In this event, Johannes will first give a general introduction into pre-registration and then cover its most recent developments, including secondary analysis of data, pre-registration with many hypotheses, and the use of theoretical constraints. There will also be time for questions & answers.

What I will offer

- History of how pre-registration evolved in the last 10 years
- What are its problems, how might they be solved
- Overview over some recent papers, “perspicuous presentation”
- What can we learn for our own research?
- Major perspective: (social) psychology, cognitive neuroscience
- Similar implications for other social/ life sciences like economics, medicine, …

Acknowledgments

- Most relevant papers:
  - Baldwin, J., Pingault, J. B., Schoeler, T., Sallis, H., & Munafo, M. (2020). Protecting against researcher bias in secondary data analysis: Challenges and solutions. *psyArXiv*. https://psyarxiv.com/md5pe/
  - Chambers, C. D. (2013). Registered reports: a new publishing initiative at Cortex. *Cortex*, *49*(3), 609-610. https://www.sciencedirect.com/science/article/abs/pii/S0010945212003735
  - Cramer, A. O., van Ravenzwaaij, D., Matzke, D., Steingroever, H., Wetzels, R., Grasman, R. P., ... & Wagenmakers, E. J. (2016). Hidden multiplicity in exploratory multiway ANOVA: Prevalence and remedies. *Psychonomic Bulletin & Review*, *23*(2), 640-647. https://link.springer.com/article/10.3758/s13423-015-0913-5
  - Gelman, A., Hill, J., & Yajima, M. (2012). Why we (usually) don't have to worry about multiple comparisons. *Journal of Research on Educational Effectiveness*, *5*(2), 189-211. https://www.tandfonline.com/doi/full/10.1080/19345747.2011.618213
  - Gelman, A., & Loken, E. (2013). The garden of forking paths: Why multiple comparisons can be a problem, even when there is no "fishing expedition" or "p-hacking" and the research hypothesis was posited ahead of time. *Department of Statistics, Columbia University*. https://stat.columbia.edu/~gelman/research/unpublished/p_hacking.pdf
  - Gentili, C., Cecchetti, L., Handjaras, G., Lettieri, G., & Cristea, I. A. (2020). The case for preregistering all region of interest (ROI) analyses in neuroimaging research. https://onlinelibrary.wiley.com/doi/full/10.1111/ejn.14954
  - 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://journals.sagepub.com/doi/full/10.1177/0956797611430953
  - Kerr, N. L. (1998). HARKing: Hypothesizing after the results are known. *Personality and Social Psychology Review*, *2*(3), 196-217. https://journals.sagepub.com/doi/abs/10.1207/s15327957pspr0203_4
  - Lin, W., & Green, D. P. (2016). Standard operating procedures: A safety net for pre-analysis plans. *PS: Political Science & Politics*, *49*(3), 495-500. https://www.cambridge.org/core/journals/ps-political-science-and-politics/article/standard-operating-procedures-a-safety-net-for-preanalysis-plans/5C5750CD150DC4DAD964263437DB7FA2
  - McIntosh, R. D. (2017). Exploratory reports: A new article type for Cortex. *Cortex*, *96*, A1-A4. https://www.sciencedirect.com/science/article/pii/S0010945217302393
  - Navarro, D. (2020). Paths in strange spaces: A comment on preregistration. *psyArXiv*. https://psyarxiv.com/wxn58
  - Nosek, B. A., Beck, E. D., Campbell, L., Flake, J. K., Hardwicke, T. E., Mellor, D. T., ... & Vazire, S. (2019). Preregistration is hard, and worthwhile. *Trends in Cognitive Sciences*, 23(10), 815-818. https://www.sciencedirect.com/science/article/pii/S1364661319301846
  - Nosek, B. A., Ebersole, C. R., DeHaven, A. C., & Mellor, D. T. (2018). The preregistration revolution. *Proceedings of the National Academy of Sciences*, *115*(11), 2600-2606. https://www.pnas.org/content/115/11/2600.short
  - Rozin, P. (2009). What kind of empirical research should we publish, fund, and reward?: A different perspective. *Perspectives on Psychological Science*, *4*(4), 435-439. https://journals.sagepub.com/doi/full/10.1111/j.1745-6924.2009.01151.x
  - Simmons, J. P., Nelson, L. D., & Simonsohn, U. (2011). False-positive psychology: Undisclosed flexibility in data collection and analysis allows presenting anything as significant. *Psychological Science*, *22*(11), 1359-1366. https://journals.sagepub.com/doi/full/10.1177/0956797611417632
  - Steegen, S., Tuerlinckx, F., Gelman, A., & Vanpaemel, W. (2016). Increasing transparency through a multiverse analysis. *Perspectives on Psychological Science*, *11*(5), 702-712. https://journals.sagepub.com/doi/full/10.1177/1745691616658637
  - Simonsohn, U., Simmons, J. P., & Nelson, L. D. (2020). Specification curve analysis. *Nature Human Behaviour*, 1-7. https://www.nature.com/articles/s41562-020-0912-z
  - Szollosi, A., Kellen, D., Navarro, D. J., Shiffrin, R., van Rooij, I., Van Zandt, T., & Donkin, C. (2019). Is preregistration worthwhile. *Trends in Cognitive Sciences*, *24*(2), 94-95. https://www.cell.com/trends/cognitive-sciences/fulltext/S1364-6613(19)30285-2
  - Van den Akker, O., Weston, S. J., Campbell, L., Chopik, W. J., Damian, R. I., Davis-Kean, P., ... & Ritchie, S. J. (2019). Preregistration of secondary data analysis: A template and tutorial. *psyArXiV*. https://psyarxiv.com/hvfmr/
  - van Rooij, I., & Baggio, G. (2020). Theory before the test: How to build high-verisimilitude explanatory theories in psychological science. *psyArXiV*. https://psyarxiv.com/7qbpr/
  - Vul, E., Harris, C., Winkielman, P., & Pashler, H. (2009). Puzzlingly high correlations in fMRI studies of emotion, personality, and social cognition. *Perspectives on Psychological Science*, *4*(3), 274-290. https://journals.sagepub.com/doi/full/10.1111/j.1745-6924.2009.01125.x
- Figure sources:
  - Mariella Paul: https://osf.io/4enz9/
  - Rima-Maria Rahal: https://rimamrahal.wordpress.com/
  - https://bayesianspectacles.org

Ready?

What is pre-registration?

How it all started…

Journal of Personality and Social Psychology
2011, Vol. 100, No. 3, 407–425

0022-3514/11/$12.00 DOI: 10.1037/a0021524

Feeling the Future: Experimental Evidence for Anomalous Retroactive Influences on Cognition and Affect

Daryl J. Bem
Cornell University

The term *psi* denotes anomalous processes of information or energy transfer that are currently unexplained in terms of known physical or biological mechanisms. Two variants of psi are *precognition* (conscious cognitive awareness) and *premonition* (affective apprehension) of a future event that could not otherwise be anticipated through any known inferential process. Precognition and premonition are themselves special cases of a more general phenomenon: the anomalous retroactive influence of some future event on an individual's current responses, whether those responses are conscious or nonconscious, cognitive or affective. This article reports 9 experiments, involving more than 1,000 participants, that test for retroactive influence by "time-reversing" well-established psychological effects so that the individual's responses are obtained before the putatively causal stimulus events occur. Data are presented

Table 1. Likelihood of Obtaining a False-Positive Result

Researcher degrees of freedom	Significance level $p < .1$	$p < .05$	$p < .01$
Situation A: two dependent variables ($r = .50$)	17.8%	9.5%	2.2%
Situation B: addition of 10 more observations per cell	14.5%	7.7%	1.6%
Situation C: controlling for gender or interaction of gender with treatment	21.6%	11.7%	2.7%
Situation D: dropping (or not dropping) one of three conditions	23.2%	12.6%	2.8%
Combine Situations A and B	26.0%	14.4%	3.3%
Combine Situations A, B, and C	50.9%	30.9%	8.4%
Combine Situations A, B, C, and D	81.5%	60.7%	21.5%

General Article

Psychological Science
22(11) 1359–1366
© The Author(s) 2011
Reprints and permission:
sagepub.com/journalsPermissions.nav
DOI: 10.1177/0956797611417632
http://pss.sagepub.com

False-Positive Psychology: Undisclosed Flexibility in Data Collection and Analysis Allows Presenting Anything as Significant

Joseph P. Simmons[1], Leif D. Nelson[2], and Uri Simonsohn[1]
[1]The Wharton School, University of Pennsylvania, and [2]Haas School of Business, University of California, Berkeley

Abstract

In this article, we accomplish two things. First, we show that despite empirical psychologists' nominal endorsement of a low rate of false-positive findings (≤ .05), flexibility in data collection, analysis, and reporting dramatically increases actual false-positive rates. In many cases, a researcher is more likely to falsely find evidence that an effect exists than to correctly find evidence that it does not. We present computer simulations and a pair of actual experiments that demonstrate how unacceptably easy it is to accumulate (and report) statistically significant evidence for a false hypothesis. Second, we suggest a simple, low-cost, and straightforwardly effective disclosure-based solution to this problem. The solution involves six concrete requirements for authors and four guidelines for reviewers, all of which impose a minimal burden on the publication process.

Who did it?

Research Article

Psychological Science
23(5) 524–532

DOI: 10.1177/0956797611430953
http://pss.sagepub.com

Measuring the Prevalence of Questionable Research Practices With Incentives for

Loewenstein[2], and Drazen Prelec[3]
...ool; [2]Department of Social & Decision Sciences, Carnegie Mellon University; Departments of Economics and Brain & Cognitive Sciences, Massachusetts

Item	Self-admission rate (%)		Odds ratio (BTS/control)	Two-tailed *p* (likelihood ratio test)	Defensibility rating (across groups)
	Control group	BTS group			
1. In a paper, failing to report all of a study's dependent measures	63.4	66.5	1.14	.23	1.84 (0.39)
2. Deciding whether to collect more data after looking to see whether the results were significant	55.9	58.0	1.08	.46	1.79 (0.44)
3. In a paper, failing to report all of a study's conditions	27.7	27.4	0.98	.90	1.77 (0.49)
4. Stopping collecting data earlier than planned because one found the result that one had been looking for	15.6	22.5	1.57	.00	1.76 (0.48)
5. In a paper, "rounding off" a *p* value (e.g., reporting that a *p* value of .054 is less than .05)	22.0	23.3	1.07	.58	1.68 (0.57)
6. In a paper, selectively reporting studies that "worked"	45.8	50.0	1.18	.13	1.66 (0.53)
7. Deciding whether to exclude data after looking at the impact of doing so on the results	38.2	43.4	1.23	.06	1.61 (0.59)
8. In a paper, reporting an unexpected finding as having been predicted from the start	27.0	35.0	1.45	.00	1.50 (0.60)
9. In a paper, claiming that results are unaffected by demographic variables (e.g., gender) when one is actually unsure (or knows that they do)	3.0	4.5	1.52	.16	1.32 (0.60)
10. Falsifying data	0.6	1.7	2.75	.07	0.16 (0.38)

How does it affect research?

Journal of Personality and Social Psychology
1996, Vol. 71, No. 2, 230–244

0022-3514/96/$3.00

Automaticity of Social Behavior: Direct Effects of Trait Construct and Stereotype Activation on Action

John A. Bargh, Mark Chen, and Lara Burrows
New York University

Previous research has shown that trait concepts and stereotypes become active automatically in the presence of relevant behavior or stereotyped-group features. Through the use of the same priming procedures as in previous impression formation research, Experiment 1 showed that participants whose concept of rudeness was primed interrupted the experimenter more quickly and frequently than did participants primed with polite-related stimuli. In Experiment 2, participants for whom an elderly stereotype was primed walked more slowly down the hallway when leaving the experiment than did control participants, consistent with the content of that stereotype. In Experiment 3, participants for whom the African American stereotype was primed subliminally reacted with more hostility to a vexatious request of the experimenter. Implications of this automatic behavior priming effect for self-fulfilling prophecies are discussed, as is whether social behavior is necessarily mediated by conscious choice processes.

Journal of Personality and Social Psychology
1998, Vol. 74, No. 4, 865–877

0022-3514/98/$3.00

The Relation Between Perception and Behavior, or How to Win a Game of Trivial Pursuit

Ap Dijksterhuis and Ad van Knippenberg
University of Nijmegen

The authors tested and confirmed the hypothesis that priming a stereotype or trait leads to complex overt behavior in line with this activated stereotype or trait. Specifically, 4 experiments established that priming the stereotype of professors or the trait *intelligent* enhanced participants' performance on a scale measuring general knowledge. Also, priming the stereotype of soccer hooligans or the trait *stupid* reduced participants' performance on a general knowledge scale. Results of the experiments revealed (a) that prolonged priming leads to more pronounced behavioral effects and (b) that there is no sign of decay of the effects for at least 15 min. The authors explain their results by claiming that perception has a direct and pervasive impact on overt behavior (cf. J. A. Bargh, M. Chen, & L. Burrows, 1996). Implications for human social behavior are discussed.

OPEN ACCESS Freely available online

PLoS one

Behavioral Priming: It's all in the Mind, but Whose Mind?

Stéphane Doyen[1,2,3]*, Olivier Klein[2], Cora-Lise Pichon[1], Axel Cleeremans[1]

1 Consciousness, Cognition and Computation Group, Université Libre de Bruxelles, Brussels, Belgium, **2** Social Psychology Unit, Université Libre de Bruxelles, Brussels, Belgium, **3** Social and Developmental Psychology Department, University of Cambridge, Cambridge, United Kingdom

Abstract

The perspective that behavior is often driven by unconscious determinants has become widespread in social psychology. Bargh, Chen, and Burrows' (1996) famous study, in which participants unwittingly exposed to the stereotype of age walked slower when exiting the laboratory, was instrumental in defining this perspective. Here, we present two experiments aimed at replicating the original study. Despite the use of automated timing methods and a larger sample, our first experiment failed to show priming. Our second experiment was aimed at manipulating the beliefs of the experimenters: Half were led to think that participants would walk slower when primed congruently, and the other half was led to expect the opposite. Strikingly, we obtained a walking speed effect, but only when experimenters believed participants would indeed walk slower. This suggests that both priming and experimenters' expectations are instrumental in explaining the walking speed effect. Further, debriefing was suggestive of awareness of the primes. We conclude that unconscious behavioral priming is real, while real, involves mechanisms different from those typically assumed to cause the effect.

OPEN ACCESS Freely available online

PLOS ONE

Priming Intelligent Behavior: An Elusive Phenomenon

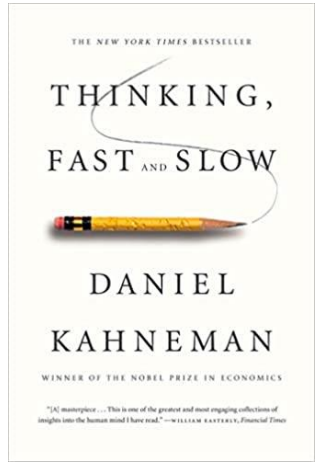
David R. Shanks[1]*, Ben R. Newell[2], Eun Hee Lee[1], Divya Balakrishnan[1], Lisa Ekelund[1], Zarus Cenac[1], Fragkiski Kavvadia[1], Christopher Moore[2]

1 Division of Psychology and Language Sciences, University College London, London, United Kingdom, **2** School of Psychology, University of New South Wales, Sydney, Australia

Abstract

Can behavior be unconsciously primed via the activation of attitudes, stereotypes, or other concepts? A number of studies have suggested that such priming effects can occur, and a prominent illustration is the claim that individuals' accuracy in answering general knowledge questions can be influenced by activating intelligence-related concepts such as *professor* or *soccer hooligan*. In 9 experiments with 475 participants we employed the procedures used in these studies, as well as a number of variants of those procedures, in an attempt to obtain this intelligence priming effect. None of the experiments obtained the effect, although financial incentives did boost performance. A Bayesian analysis reveals considerable evidential support for the null hypothesis. The results conform to the pattern typically obtained in word priming experiments in which priming is very narrow in its generalization and unconscious (subliminal) influences, if they occur at all, are extremely short-lived. We encourage others to explore the circumstances in which this phenomenon might be obtained.

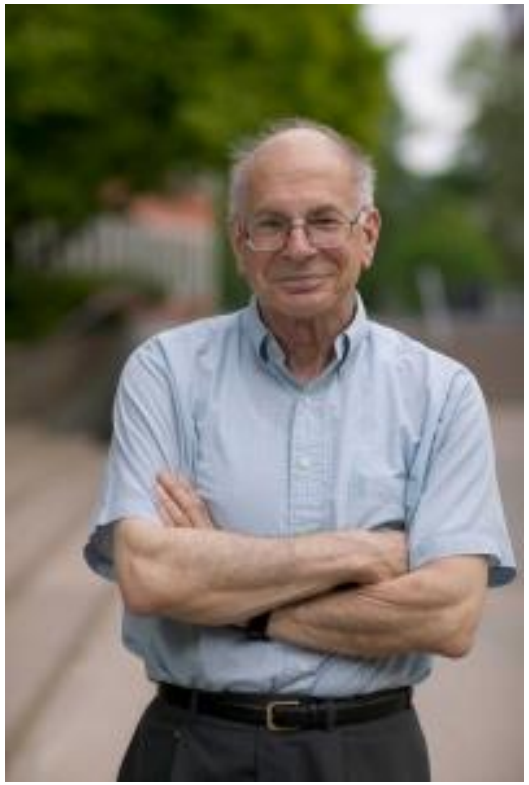
A “train wreck”?



- Open letter by Daniel Kahneman on September 26, 2012

For all these reasons, right or wrong, your field is now the poster child for doubts about the integrity of psychological research. Your problem is not with the few people who have actively challenged the validity of some priming results. It is with the much larger population of colleagues who in the past accepted your surprising results as facts when they were published. These people have now attached a question mark to the field, and it is your responsibility to remove it.

My reason for writing this letter is that I see a train wreck looming. I expect the first victims to be young people on the job market. Being associated with a controversial and suspicious field will put them at a severe disadvantage in the competition for positions. Because of the high visibility of the issue, you may already expect the coming crop of graduates to encounter problems. Another



https://www.nature.com/news/nobel-laureate-challenges-psychologists-to-clean-up-their-act-1.11535

More systematic replication failures

RESEARCH ARTICLE

PSYCHOLOGY

Estimating the reproducibility of psychological science

Open Science Collaboration*†

Reproducibility is a defining feature of science, but the extent to which it characterizes current research is unknown. We conducted replications of 100 experimental and correlational studies published in three psychology journals using high-powered designs and original materials when available. Replication effects were half the magnitude of original effects, representing a substantial decline. Ninety-seven percent of original studies had statistically significant results. Thirty-six percent of replications had statistically significant results; 47% of original effect sizes were in the 95% confidence interval of the replication effect size; 39% of effects were subjectively rated to have replicated the original result; and if no bias in original results is assumed, combining original and replication results left 68% with statistically significant effects. Correlational tests suggest that replication success was better predicted by the strength of original evidence than by characteristics of the original and replication teams.

ECONOMICS

Evaluating replicability of laboratory experiments in economics

Colin F. Camerer,[1*†] Anna Dreber,[2†] Eskil Forsell,[2†] Teck-Hua Ho,[3,4†] Jürgen Huber,[5†] Magnus Johannesson,[2†] Michael Kirchler,[5,6†] Johan Almenberg,[7] Adam Altmejd,[2] Taizan Chan,[8] Emma Heikensten,[2] Felix Holzmeister,[5] Taisuke Imai,[1] Siri Isaksson,[2] Gideon Nave,[1] Thomas Pfeiffer,[9,10] Michael Razen,[5] Hang Wu[4]

The replicability of some scientific findings has recently been called into question. To contribute data about replicability in economics, we replicated 18 studies published in the *American Economic Review* and the *Quarterly Journal of Economics* between 2011 and 2014. All of these replications followed predefined analysis plans that were made publicly available beforehand, and they all have a statistical power of at least 90% to detect the original effect size at the 5% significance level. We found a significant effect in the same direction as in the original study for 11 replications (61%); on average, the replicated effect size is 66% of the original. The replicability rate varies between 67% and 78% for four additional replicability indicators, including a prediction market measure of peer beliefs.

Evaluating the replicability of social science experiments in *Nature* and *Science* between 2010 and 2015

Colin F. Camerer[1,16], Anna Dreber[2,16], Felix Holzmeister[3,16], Teck-Hua Ho[4,16], Jürgen Huber[3,16], Magnus Johannesson[2,16], Michael Kirchler[3,5,16], Gideon Nave[6,16], Brian A. Nosek[7,8,16*], Thomas Pfeiffer[9,16], Adam Altmejd[2], Nick Buttrick[7,8], Taizan Chan[10], Yiling Chen[11], Eskil Forsell[12], Anup Gampa[7,8], Emma Heikensten[2], Lily Hummer[8], Taisuke Imai[13], Siri Isaksson[2], Dylan Manfredi[6], Julia Rose[3], Eric-Jan Wagenmakers[14] and Hang Wu[15]

A Multi-Site Preregistered Paradigmatic Test of the Ego Depletion Effect

Kathleen D. Vohs, University of Minnesota

Brandon J. Schmeichel, Texas A&M University

Sophie Lohmann, Max Planck Institute for Demographic Research and University of Illinois at Urbana-Champaign

Quentin F. Gronau, University of Amsterdam

Anna Finley, University of Wisconsin-Madison

...

E.-J. Wagenmakers, University of Amsterdam

Dolores Albarracín, University of Illinois at Urbana-Champaign

(Appendix A contains the entire author list.)

aps ASSOCIATION FOR PSYCHOLOGICAL SCIENCE

Registered Replication Report: Strack, Martin, & Stepper (1988)

E.-J. Wagenmakers*, T. Beek*, L. Dijkhoff*, Q. F. Gronau,* A. Acosta, R. B. Adams, Jr., D. N. Albohn, E. S. Allard, S. D. Benning,

Perspectives on Psychological Science
2016, Vol. 11(6) 917–928
© The Author(s) 2016
Reprints and permissions:
sagepub.com/journalsPermissions.nav
DOI: 10.1177/1745691616674458
pps.sagepub.com
SAGE

aps ASSOCIATION FOR PSYCHOLOGICAL SCIENCE

Many Labs 2: Investigating Variation in Replicability Across Samples and Settings

Richard A. Klein[1], Michelangelo Vianello[2], Fred Hasselman[3,4], Byron G. Adams[5,6], Reginald B. Adams, Jr.[7], Sinan Alper[8],

Advances in Methods and Practices in Psychological Science
2018, Vol. 1(4) 443–490
© The Author(s) 2018
Article reuse guidelines:
sagepub.com/journals-permissions
DOI: 10.1177/2515245918810225
www.psychologicalscience.org/AMPPS
SAGE

eLIFE elifesciences.org FEATURE ARTICLE

REPRODUCIBILITY IN CANCER BIOLOGY

Making sense of replications

REPRODUCIBILITY PROJECT CANCER BIOLOGY

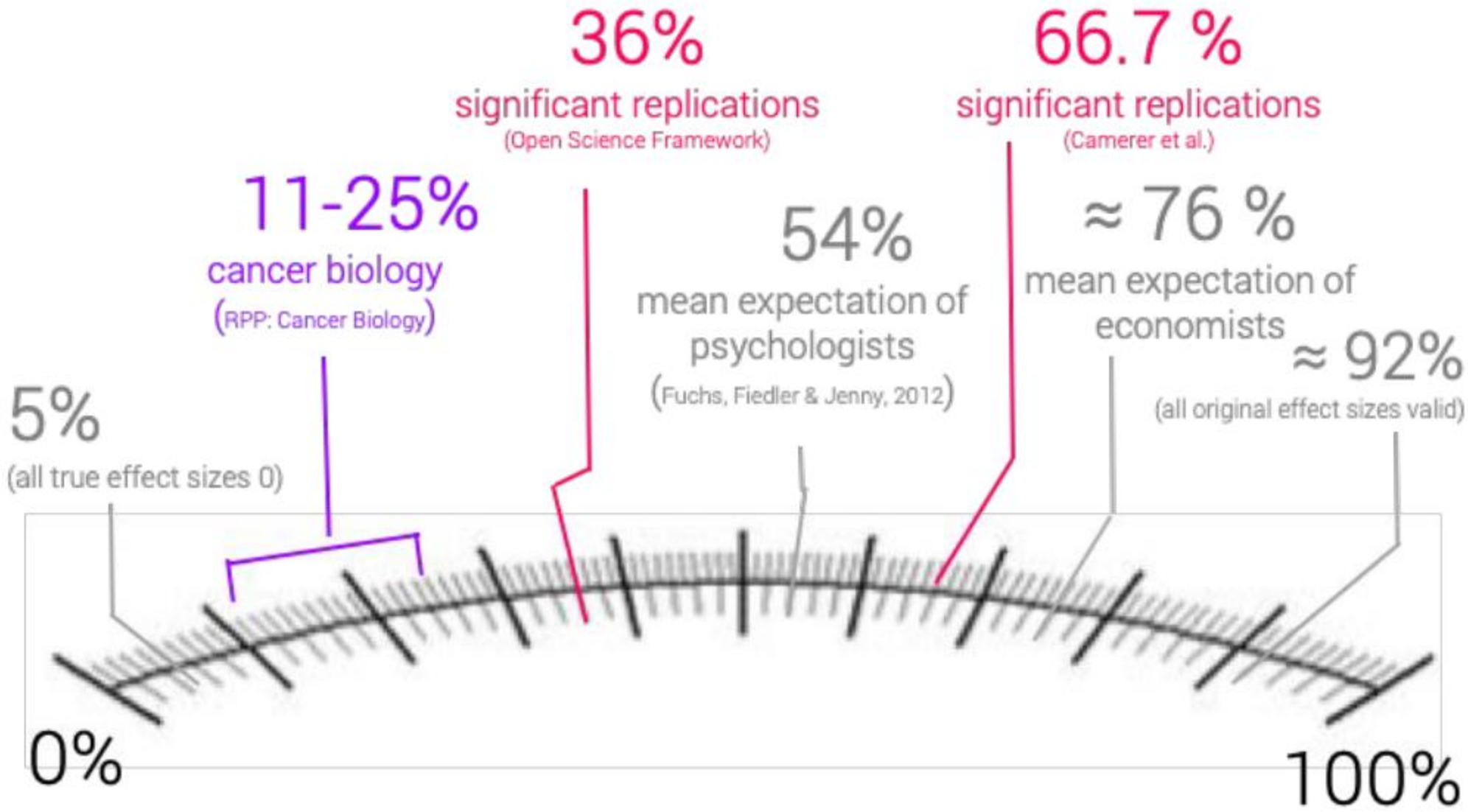
Abstract The first results from the Reproducibility Project: Cancer Biology suggest that there is scope for improving reproducibility in pre-clinical cancer research.

DOI: 10.7554/eLife.23383.001

BRIAN A NOSEK AND TIMOTHY M ERRINGTON*

Empirical examination of the replicability of associations between brain structure and psychological variables

Shahrzad Kharabian Masouleh[1,2]*, Simon B Eickhoff[1,2], Felix Hoffstaedter[1,2], Sarah Genon[1]*, Alzheimer's Disease Neuroimaging Initiative


36%
significant replications
(Open Science Framework)
66.7 %
significant replications
(Camerer et al.)
11-25%
cancer biology
(RPP: Cancer Biology)
54%
mean expectation of
psychologists
(Fuchs, Fiedler & Jenny, 2012)
≈ 76 %
mean expectation of
economists
≈ 92%
(all original effect sizes valid)
5%
(all true effect sizes 0)
0%
100%

Conclusion

Journal of Personality and Social Psychology
2011, Vol. 100, No. 3, 426–432

0022-3514/11/$12.00 DOI: 10.1037/a0022790

Why Psychologists Must Change the Way They Analyze Their Data: The Case of Psi: Comment on Bem (2011)

Eric–Jan Wagenmakers, Ruud Wetzels, Denny Borsboom, and Han L. J. van der Maas
University of Amsterdam

aps ASSOCIATION FOR PSYCHOLOGICAL SCIENCE

An Agenda for Purely Confirmatory Research

Eric-Jan Wagenmakers, Ruud Wetzels, Denny Borsboom, Han L. J. van der Maas, and Rogier A. Kievit
University of Amsterdam, The Netherlands

Perspectives on Psychological Science
7(6) 632–638

DOI: 10.1177/1745691612463078
http://pps.sagepub.com
SAGE

Summary: Problems to solve

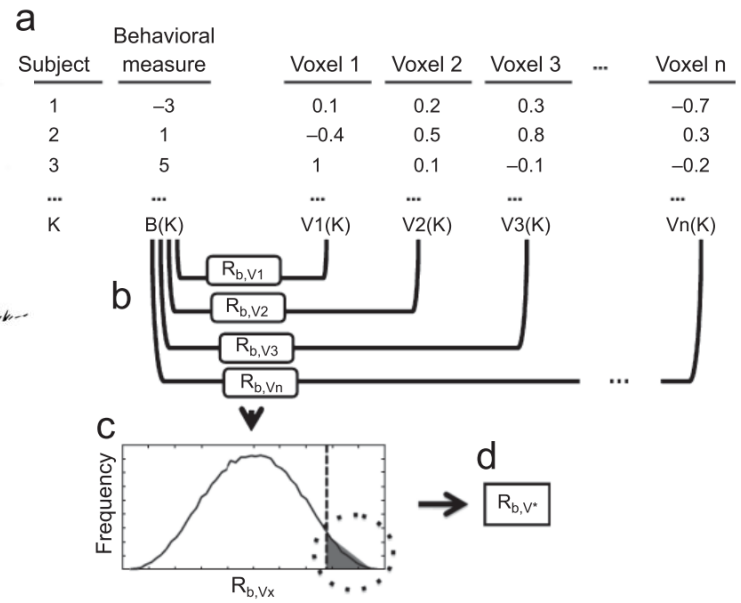
- Selective outcome reporting
  - Selective reporting of experiments/ subjects
  - Selective reporting of outcome measures
  - Selective reporting of experimental groups

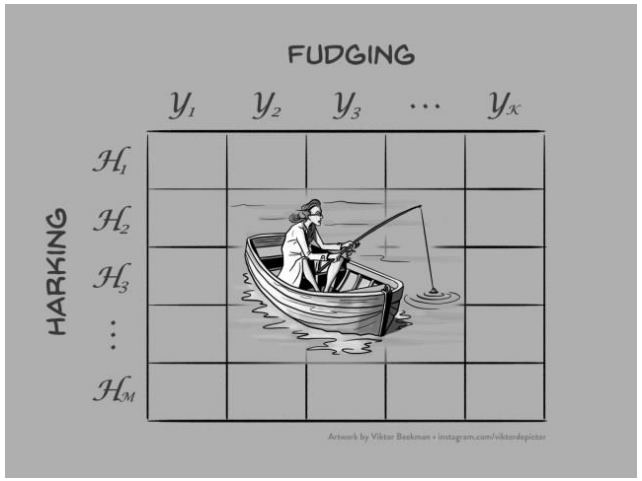
1. "P-hacking"/ cherry picking/ fudging
   - Choose analyses/ analysis parameters based on results
2. Hypothesizing after results are known ("HARK"ing)

Personality and Social Psychology Review
1998, Vol. 2, No. 3, 196–217

HARKing: Hypothesizing After the Results are Known

Norbert L. Kerr
Department of Psychology
Michigan State University





Kerr, 1998, *PSPR*; Vul et al., 2009, *PoPS*

Possible solutions

- Many attempts, no perfect solution

- Pre-registration
- Data and code sharing
- Reproducibility
- Replication

- Why *focus* on preregistration?
- Temporal precedence:
  If the original study findings are (likely) false-positives, other means (e.g. reproducibility, replicability) lose their value to start with

Definition of pre-registration

- To be on the same page
- **Public**, **time-stamped** registration in **institutional registration system**
- Registration **predates** data collection
- Contents:
  1. Procedures: sample size, trial number, randomization procedure
  2. Exclusion: under what conditions will data be excluded
  3. Variable construction: independent and dependent variables
  4. Tests and models: hypotheses, effects, statistical model, covariates, correction for multiple comparisons
  5. Procedures in event of foreseeable problems (attrition, non-compliance, not enough subjects, …)

https://osf.io/tvyxz/wiki/1.%20View%20the%20Badges/; Navarro, 2020, https://psyarxiv.com/wxn58

Popular options

- Open Science Framework (OSF): https://www.cos.io/initiatives/prereg
- As-predicted: https://aspredicted.org/
- Clinicaltrials.gov: http://clinicaltrials.gov/
- Declare Design: https://declaredesign.org/
- ….
- Registered reports

CORTEX 49 (2013) 609–610

Available online at www.sciencedirect.com

SciVerse ScienceDirect

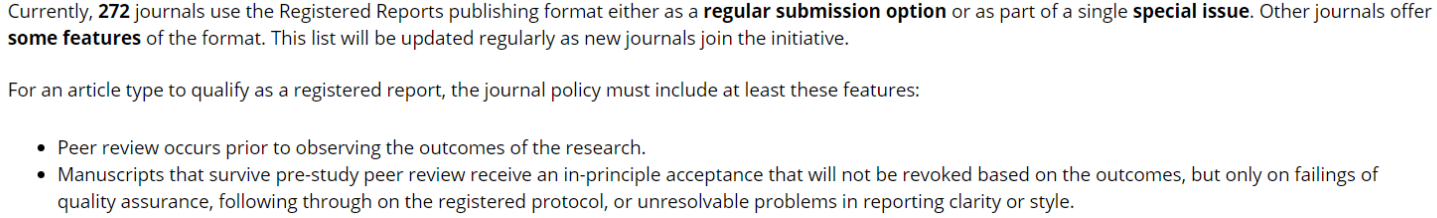
Journal homepage: www.elsevier.com/locate/cortex

Editorial

Registered Reports*: A new publishing initiative at *Cortex

Christopher D. Chambers

Cardiff University Brain Research Imaging Centre (CUBRIC), School of Psychology, Cardiff University, United Kingdom



Currently, **272** journals use the Registered Reports publishing format either as a **regular submission option** or as part of a single **special issue**. Other journals offer **some features** of the format. This list will be updated regularly as new journals join the initiative.

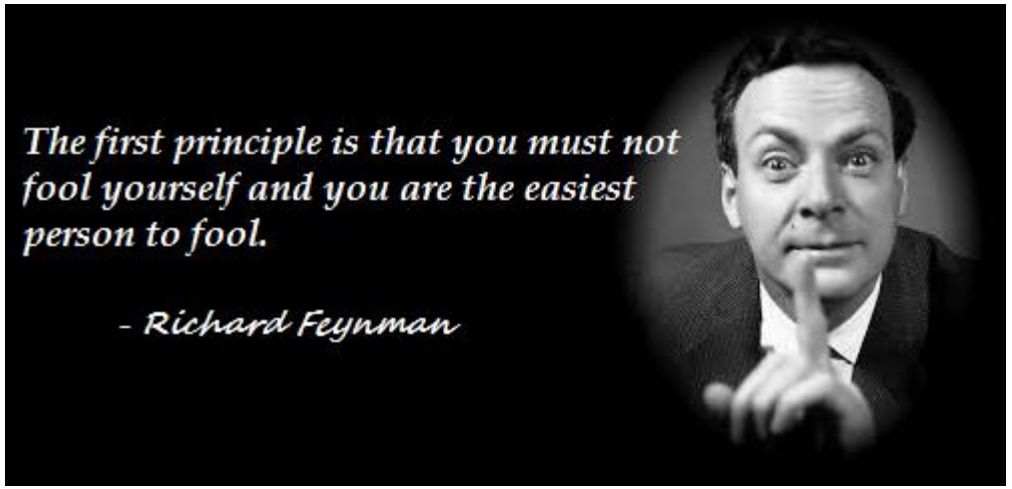
For an article type to qualify as a registered report, the journal policy must include at least these features:

- Peer review occurs prior to observing the outcomes of the research.
- Manuscripts that survive pre-study peer review receive an in-principle acceptance that will not be revoked based on the outcomes, but only on failings of quality assurance, following through on the registered protocol, or unresolvable problems in reporting clarity or style.

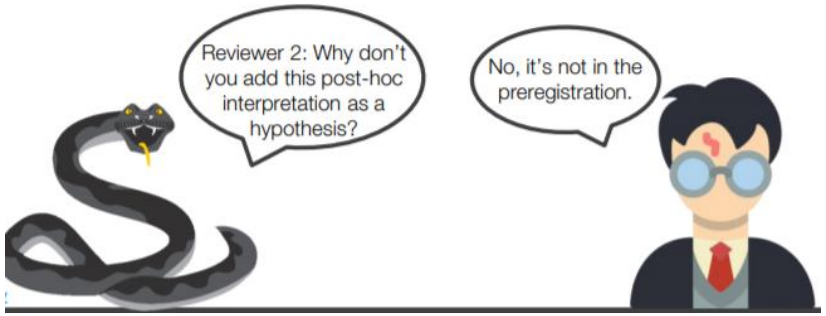
Chambers, 2013, *Cortex*; https://www.cos.io/initiatives/registered-reports

Advantages for pre-registration

- Test a-priori hypotheses, no HARKing
- Maintain error control, no fudging
- Make study discoverable for others



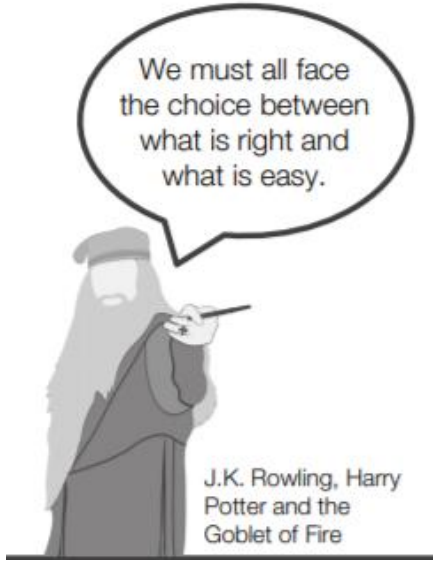
- Don’t fool other people/ get fooled
- Don’t fool yourself!
- Make stronger claims:
  - “As pre-registered a-priori, we found …”
    - High-risk research
  - Badges
- Reduce publication bias
  - Make every study publicly discoverable
  - Publish null findings
- Required by several agencies:
  - Required by International Committee of Medical Journal Editors policy
  - Required by signees of Transparency and Openness Promotion (TOP) Guidelines
  - Required by grant agencies, funding bodies, study programs, …



Nosek et al., 2018, *PNAS*

Disadvantages for pre-registration

- Commitment: Always report confirmatory results as pre-registered
  - Even if it “clutters” your paper…
- Need to put null findings out
  - Even if they do not support your theory…
- Might expose weaknesses in pre-registration:
  - Changes in exclusion criteria
  - Changes in fitted models (e.g. due to violation of assumptions)



Success stories

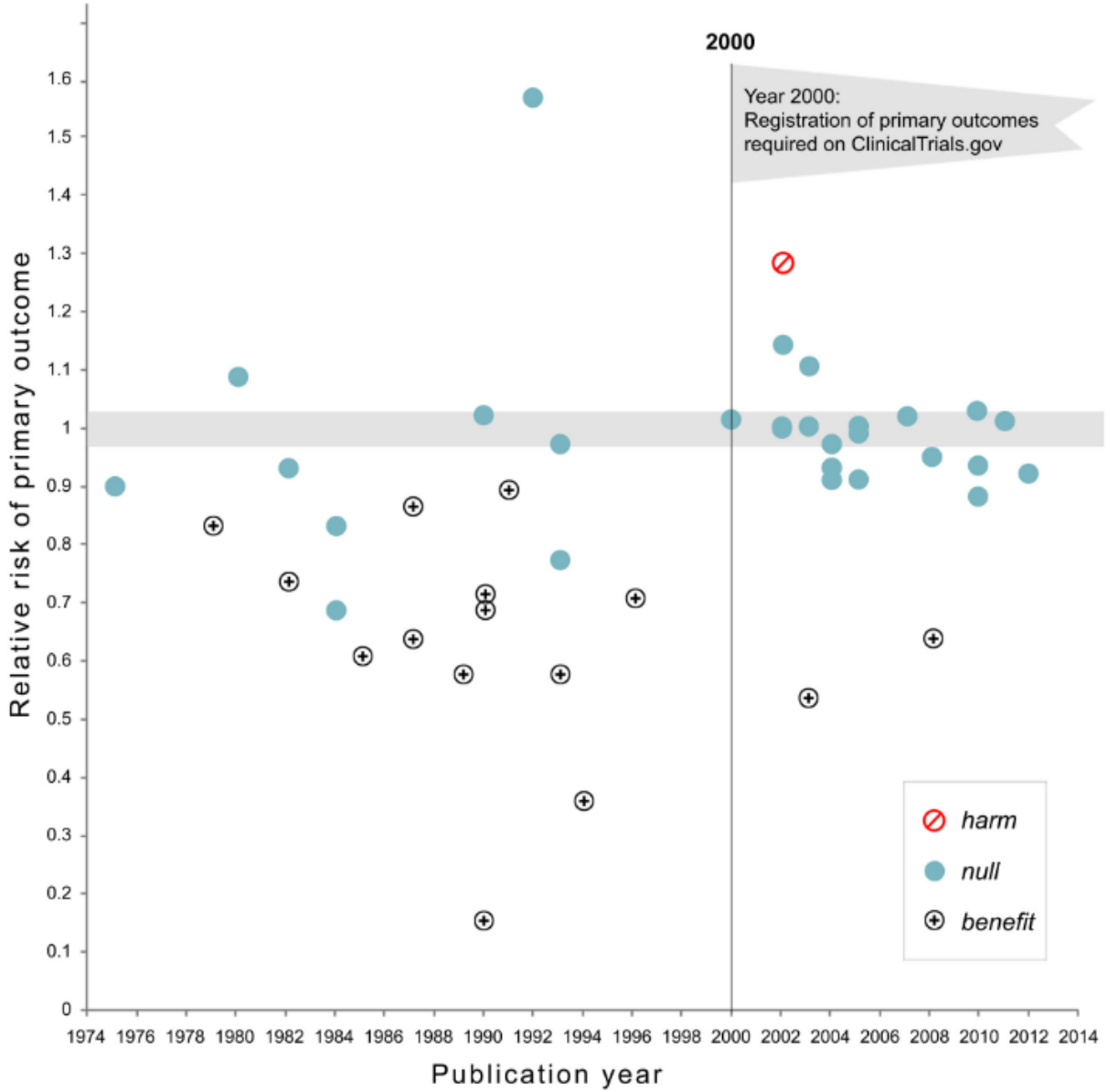


Fig 1. Relative risk of showing benefit or harm of treatment by year of publication for large NHLBI trials on pharmaceutical and dietary supplement interventions. Positive trials are indicated by the plus signs while trials showing harm are indicated by a diagonal line within a circle. Prior to 2000 when trials were not registered in clinical trials.gov, there was substantial variability in outcome. Following the imposition of the requirement that trials preregister in clinical trials.gov the relative risk on primary outcomes showed considerably less variability around 1.0.

Kaplan & Irvin, 2015, *PLoS One*

Divergent opinions…

theguardian

News | Sport | Comment | Culture | Business | Money | Life & style

News › Science › Peer review and scientific publishing

Trust in science would be improved by study pre-registration

Open letter: We must encourage scientific journals to accept studies before the results are in

Chris Chambers, Marcus Munafo and more than 80 signatories
theguardian.com, Wednesday 5 June 2013 12.45 BST

Jump to comments (43)

The quest: a better understanding of nature. Photograph: Sebastian Kaulitzki/Alamy

THE WORLD UNIVERSITY RANKINGS

PROFESSIONAL JOBS SUMMITS RANKINGS STU

Pre-registration would put science in chains

The pre-registration of study designs must be resisted, says Sophie Scott

July 25, 2013

Science is not well served by people deciding that their methodology is the only legitimate one

Recent discussions

Scientific Life

Preregistration Is Hard, And Worthwhile

Brian A. Nosek,[1,*] Emorie D. Beck,[2] Lorne Campbell,[3] Jessica K. Flake,[4] Tom E. Hardwicke,[5] David T. Mellor,[1] Anna E. van 't Veer,[6] and Simine Vazire[7]

Letter

Is Preregistration Worthwhile?

Aba Szollosi,[1,*] David Kellen,[2] Danielle J. Navarro,[1] Richard Shiffrin,[3] Iris van Rooij,[4] Trisha Van Zandt,[5] and Chris Donkin[1]

Preregistration is redundant, at best

Preregistration in Complex Contexts: A Preregistration Template for the Application of Cognitive Models

Sophia Crüwell[1,2] and Nathan J. Evans[2,3]

Japanese Psychological Review
2019, Vol. 62, No. 3, 221–230

The value of preregistration for psychological science: A conceptual analysis

Daniël LAKENS
Eindhoven University of Technology

Preregistration of secondary data analysis: A template and tutorial

Olmo R. van den Akker[1]
Sara J. Weston[2]
Lorne Campbell[3]
William J. Chopik[4]
Rodica Ioana Damian[5]
Pamela E. Davis-Kean[6]
Andrew N. Hall[7]
Jessica E. Kosie[8]
Elliott Kruse[9]
Jerome Olsen[10,11]
Stuart J. Ritchie[12]
K.D. Valentine[13]
Anna E. van 't Veer[14]
Marjan Bakker[1]

Aba Szollosi[1]*, David Kellen[2], Danielle J. Navarro[1], Richard Shiffrin[3], Iris van Rooij[4], Trisha Van Zandt[5], and Chris Donkin[1]

Ensuring the quality and specificity of preregistrations

Short title: the quality and specificity of preregistrations

Marjan Bakker[1*], Coosje L. S. Veldkamp[2*], Marcel A. L. M. van Assen[13], Elise A. V. Crompvoets[14], How Hwee Ong[5], Brian A. Nosek[67], Courtney K. Soderberg[6], David Mellor[6], & Jelte, M. Wicherts[1]

Paths in strange spaces: A comment on preregistration

Danielle J. Navarro

Original: November 2019; PsyArXiv: Sept

Preregistration: A Solution to Undisclosed Analytic Flexibility in ERP Research

Mariella Paul[a,b,c,§], Gisela H. Govaart[a,b,d,§], Antonio Schettino[e,f,§,*]

EDITORIAL

EJN European Journal of Neuroscience FENS WILEY

The case for preregistering all region of interest (ROI) analyses in neuroimaging research

When is pre-registration a good idea (and when not)?

Goals of preregistration

1. Distinguish confirmatory and exploratory research
2. Transparent workflow documentation
3. Pedagogical tool stimulating clearer thinking

Preregistration clarifies the distinction between planned and unplanned research by reducing unnoticed flexibility. This improves credibility of findings and calibration of uncertainty. However, mak-

Confirmatory vs. exploratory research

- Planned vs. unplanned
- A-priori vs. post-hoc
- Prediction vs. postdiction
- Hypothesis-testing vs. hypothesis-generating
- Data-independent vs. data-contingent analysis

Preregistration in Practice

Preregistration does not favor prediction over postdiction; its purpose is to make clear which is which. There are practical

Preregistration does not eliminate the possibility of poor statistical practices, but it does make them detectable.

Nosek et al., 2018, *PNAS*

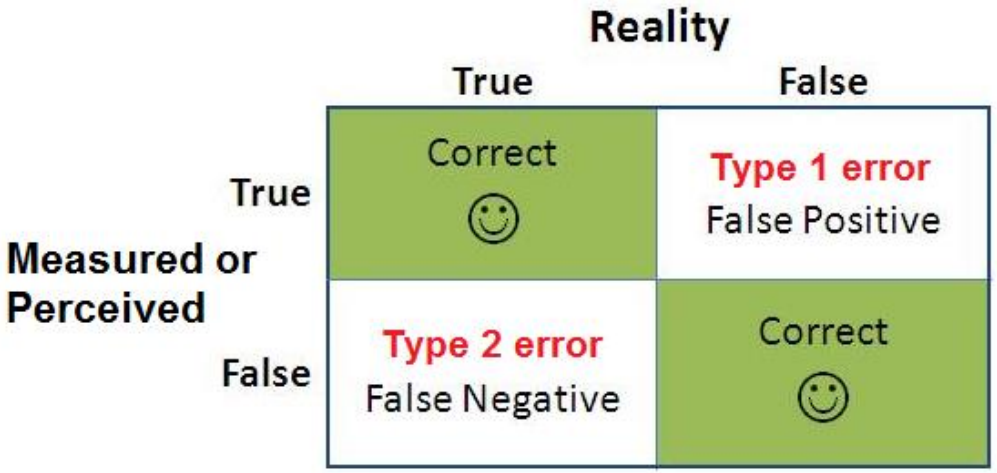
Confirmatory research

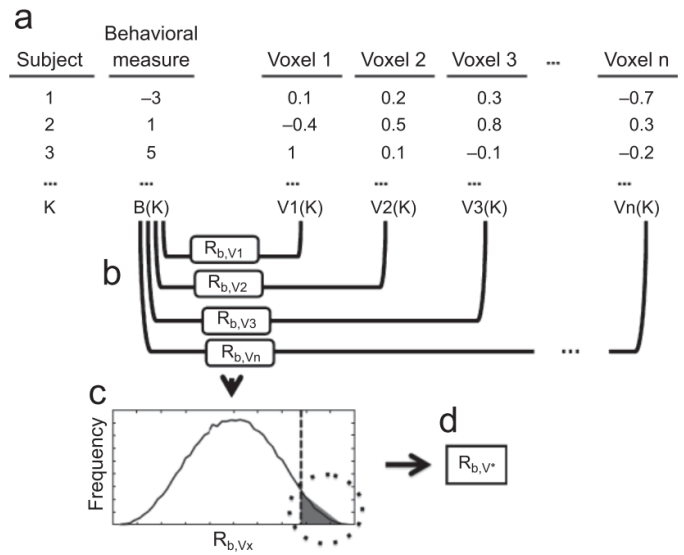
Psychon Bull Rev (2016) 23:640–647
DOI 10.3758/s13423-015-0913-5

Hidden multiplicity in exploratory multiway ANOVA: Prevalence and remedies

Angélique O. J. Cramer[1] · Don van Ravenzwaaij[2] · Dora Matzke[1] · Helen Steingroever[1] · Ruud Wetzels[3] · Raoul P. P. P. Grasman[1] · Lourens J. Waldorp[1] · Eric-Jan Wagenmakers[1]

1. Why do confirmatory research?
    1. Test hypotheses
    2. Reject null hypotheses by comparing observed p-values with set alpha levels
    3. Error control (set an overall alpha level)



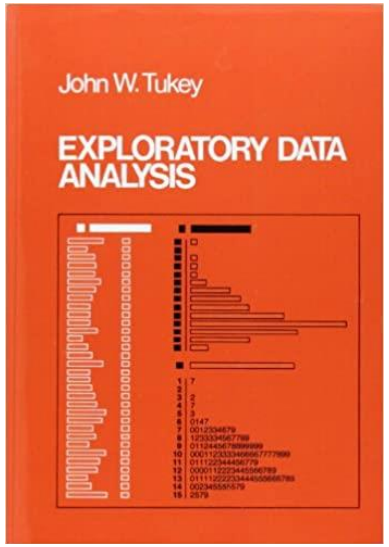


unplanned analyses. Statistical tests always involve uncertainty: for example, at a criterion value of 5%, the presumed Type I error rate is 5%. However, this holds only if data meet the assumptions for a single statistical test. Undisclosed and unplanned analyses undermine the assumptions. Preregistration makes transparent the uncertainty of statistical tests by showing how many statistical tests were conducted and enabling an accurate familywise error rate.

Cramer et al., 2016, *PBR;* Nosek et al., 2019, *TICS*

Exploratory research

- Data-dependent analyses, hypothesis generation, postdiction
- No error control (meaning of p-values unclear)
- Tukey (1977): Exploratory Data Analysis: description, visualization
- Parameter estimation (effect sizes)
- Exploratory reports
- Followed-up by confirmatory research (replication, cross-validation)



CORTEX 96 (2017) A1–A4

Available online at www.sciencedirect.com

ScienceDirect

Journal homepage: www.elsevier.com/locate/cortex

ELSEVIER

Cortex

Editorial

Exploratory reports: A new article type for *Cortex*

*Robert D. McIntosh**

quality of theory, are paramount. The overall bar for publication will be as high as ever, but different aspects of the work may be weighted more heavily than others, depending on the nature of the work. If the research question is open-ended, and the results provisional, then the discussion should be outstanding in its theoretical analysis and generation of novel testable predictions. Or, where an empirically-

Tukey, 1977; Cumming, 2014, *Psych Science*; McIntosh, 2017, *Cortex*

Requirements for confirmatory research

- Pre-registration must be complete, as specific as possible
  - Sample size/ power
  - All data selection/ exclusion, variable constructions, included variables
    - Prevent fudging
  - All fitted models and tested hypotheses
    - Prevent HARKing
- Analysis according to pre-registration must be reported as "*the* confirmatory results"
  - Even if preferred analysis pipeline changed in the meantime (?)
  - Even if a-priori pipeline doesn't make sense in retrospect (?)

What is suited for confirmatory research?

- Not all research
- Actually fairly little…
- Well suited:
  - When effect size is known, possible to estimate power, sample size
  - When analysis pipeline is established
  - When hypotheses are clear (best case: one-sided)
- That includes:
  - Direct replications
  - “Conceptual” replications
    - Depends on how conceptual

All other cases:

- things get fairly complicated
- high chance of “things going wrong” along the way…

Conclusion: Pre-registration is complicated—thus when in doubt: **Don’t do “it”**!?!

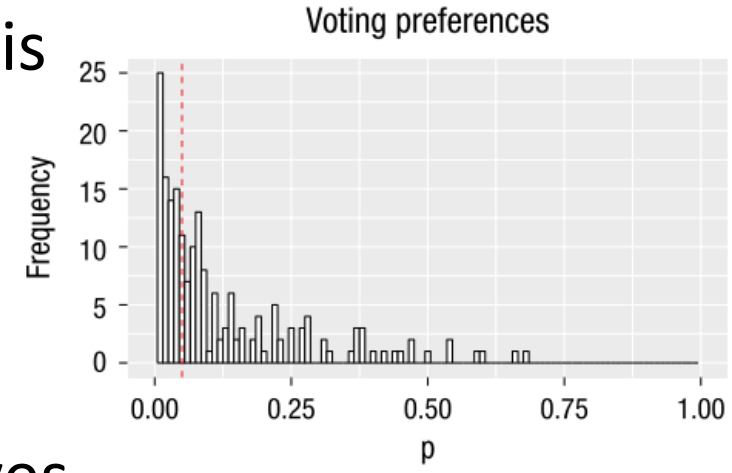
The following slides:

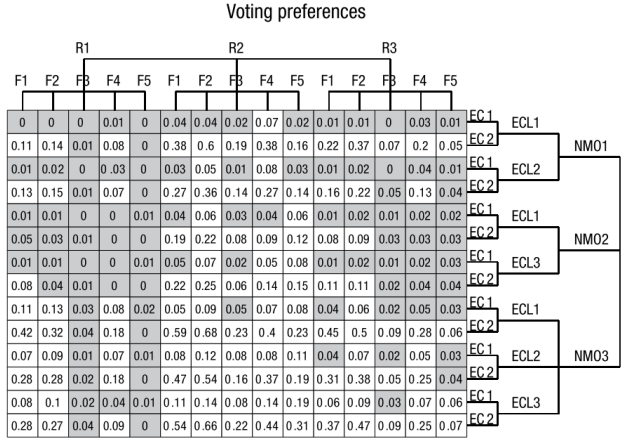
- Alternatives without pre-registration
- Alternatives “with” pre-registration

unplanned analyses. Statistical tests always involve uncertainty: for example, at a criterion value of 5%, the presumed Type I error rate is 5%. However, this holds only if data meet the assumptions for a single statistical test. Undisclosed and unplanned analyses undermine the assumptions. Preregistration makes transparent the uncertainty of statistical tests by showing how many statistical tests were conducted and enabling an accurate familywise error rate.

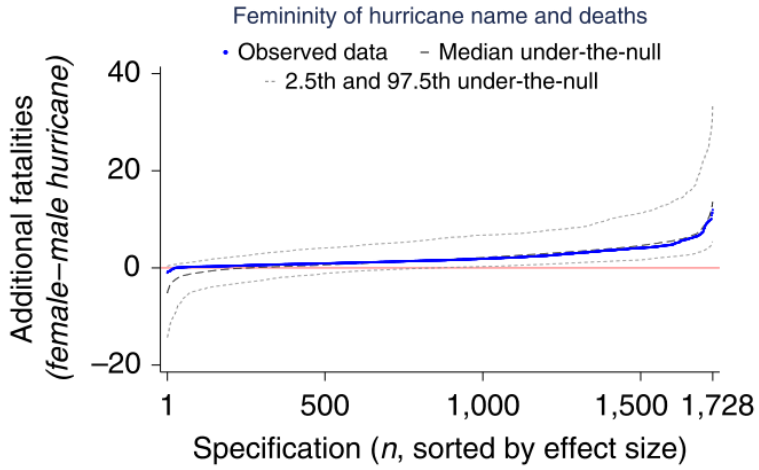
Alternatives without pre-registration

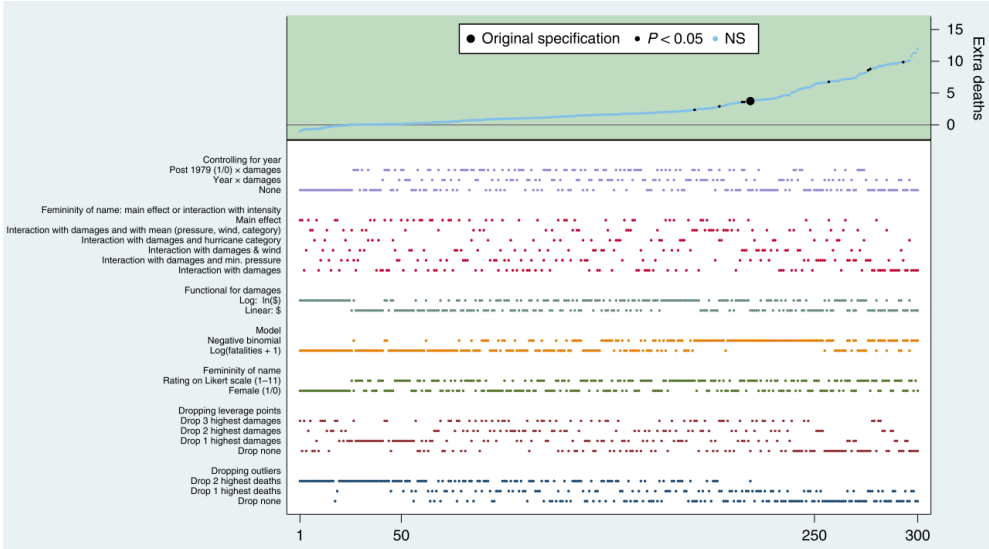
- Goal: be transparent about analytic decisions, error control
- Multiverse analysis





- Specification curves





Nosek et al., 2019, *TICS*; Simonsohn, Simmons, & Nelson, 2020, *NHB*; Steegen, Tuerlinckx, Gelman, Vanpaemel, 2014, *PoPS*

A plan or a prison?

Preregistration: A Plan, Not a Prison

May 23rd, 2017, Alexander DeHaven

Posted in: Preregistration

can I deviate without nullifying the preregistration?" Fortunately, a preregistration is a document of your pre-specified research plans before seeing the results; it is not a prison sentence. Depending on what stage of the research project you are in, you can either update the preregistration, or rely on transparency to give context to any unanticipated decisions. As you'll see, this transparency gives others the information they need to evaluate key decisions made to the planned research.

It is important to recognise this, I think, because there is a very real danger that preregistration systems will ossify scientific practices in an undesirable fashion. That is to say, while advocates of preregistration often claim that "*preregistration is a plan, not a prison*", they will also claim that preregistration is necessary to prevent p-hacking. Unfortunately, you cannot have it both ways: in my opinion these two claims are in direct opposition to one another unless you *also* stipulate an unreasonable level of foreknowledge on behalf of the experimenter.

- *Strict.* Researcher specifies the complete analysis plan in advance and does not deviate even if the data turn out to be highly surprising. This does indeed prevent p-hacking in the conventional sense but forces the researcher to use inappropriate statistics: under the strict interpretation, preregistration is in fact a prison, not a plan.
- *Flexible.* Researcher specifies the analysis plan, but is willing to deviate if in retrospect the planned analysis seems inappropriate for the data. This satisfies the claim that preregistration is a plan not a prison, but does so by opening the door to p-hacking once more. A researcher may unwittingly decide to apply greater scrutiny to data that they find disappointing, and thereby become more likely discover reasons to justify departures from the plan.
- *Oracular.* There is a third possibility: the researcher writes a preregistered plan that covers every possible eventuality, listing exactly how each case will be handled and then – because this composite if/then decision making procedure is no longer a specific hypothesis test – derives an appropriate decision policy for that plan which satisfies Neyman's admissability criteria. This would indeed allow us to have the best of both worlds: because the prespecified "deviation plan" is incorporated into the design of the subsequent decision policy, we can have the flexibility we desire (our preregistration is indeed a plan and not a prison) while ensuring that our overall Type I error rate remains bounded at its nominal level (no p-hacking allowed). Perfect . . . and all it requires is godlike planning abilities! I'll confess I don't personally have the intellect to construct that kind of analysis plan for the kinds of experiments I do, but perhaps someone smarter than me can figure it out.

https://www.cos.io/blog/preregistration-plan-not-prison ; Navarro, 2020, https://psyarxiv.com/wxn58; Nosek et al., 2019, *TICS*

II. Transparent workflow documentation

- Be transparent about deviations
- Be transparent about sequential decisions

planned analyses. Most often, planned analyses are equated with hypothesis testing or confirmatory research, but one can also preregister analysis plans when there are no *a priori* hypotheses to test (i.e., a planned exploratory analysis).

Deviations from data collection and analysis plans are common, even in the most predictable investigations. Deviations do not necessarily rule out testing predictions effectively. If the outcomes have not yet been observed, Jolene can document the changes to her preregistration without undermining diagnosticity. However, even if the data have been observed, preregistration provides substantial benefit. Jolene can transparently report changes that were made and why. Most of the design and analysis plan is still preserved, and deviations are reported transparently, making it possible to assess their impact. Compared with the situation in which Jolene did not preregister at all, preregistration with reported deviations provides substantially greater confidence in the resulting statistical inferences.

Nosek et al., 2018, *PNAS*; Nosek et al., 2019, *TICS*

“Update” pre-registration

Preregistration is a plan, not a prison (http://cos.io/blog/preregistration-plan-not-prison/). When deviations from the plan will improve the quality of the research, deviate from the plan. Many studies will have some deviations between the preregistered plan and what actually occurs. Planned analyses may contain errors [10]. Deviations inevitably make it harder to interpret with confidence what occurred in relation to what was planned. Transparency is key – all deviations from the plan should be acknowledged.

coherence, and reviewer expectations. If possible, report what occurs following the original plan alongside what occurs with the deviations, and share the materials, data, and code so that others can evaluate the reported outcomes and what would have occurred with alternative approaches.

- Update pre-registration (even before analyzing data)
- Mention all planned analyses (even if results not reported)
- Document all deviations from the original analysis plan (table, flowchart)
- Use supplementary materials
- Keep logging updates about analysis pipeline during the actual analyses

Pre-registration “light”?

- Any pre-registration is better than no pre-registration
- Incremental preregistrations (multiple stages)
  - Need to ensure analyst stays “blinded”
- Pre-register decision tree
  - Use Standard Operating Procedures (SOPs)

cies are most important to anticipate. This might lead researchers to shy away from preregistration for worries about imperfection. Embrace incrementalism. Pre-registration is a methodological skill that takes time to develop. Having some plans is better than having no plans, and sharing those plans in advance is better than not sharing them. With experience, planning will improve and the benefits will increase for oneself and for consumers of the research.

For some kinds of analysis, it is possible to define stages and preregister incrementally. For example, a researcher could define a preregistration that evaluates distributional forms of variables to determine data exclusions, transformations, and appropriate model assumptions that do not reveal anything about the research outcomes. After that, the researcher preregisters the model most appropriate for testing the outcomes of interest. Effective application of sequential preregistration is difficult in many research applications. If an earlier stage reveals information about outcomes to be tested at a subsequent stage, then the preregistration is compromised.

Lin & Green, 2015, *PS*; MacCoun & Permutter, 2018, *Nature*; Nosek et al., 2018, *PNAS*; Nosek et al., 2019, *TICS*

Other approaches to documentation

With this in mind, I don't think that (for example) the current OSF registration system provides the right toolkit. To produce the fine-grained document trail that shows precisely what I did, I would need to create a great many registrations for every project (dozens, at the very least). This is technically possible within the OSF system, of course, but there are much better ways to do it. Because what I'm really talking about here is something closer to an "open notebook" approach to research, and there are other excellent tools that can support this. For my own part I try to use git repositories to leave an auditable trail of commit logs that can be archived on any number of public servers (e.g., GitHub, BitBucket, GitLab), and I use literate programming methods such as R Markdown and Jupyter notebooks to allow me to document my thinking on the fly during the model building process. Other researchers might have different approaches.

Navarro, 2020, https://psyarxiv.com/wxn58

III. Pedagogical tool stimulating clearer thinking

- Pre-registration is hard….
- … but many of its questions are worth thinking about!

Scientific Life

Preregistration Is Hard, And Worthwhile

Brian A. Nosek,[1,*]
Emorie D. Beck,[2]
Lorne Campbell,[3]
Jessica K. Flake,[4]
Tom E. Hardwicke,[5]
David T. Mellor,[1]
Anna E. van 't Veer,[6] and
Simine Vazire[7]

analysis is converted to an uncomfortable mental simulation of what decisions will need to be made eventually. Moreover, research rarely goes precisely according to plan. Data collection can take longer than anticipated; a skewed data distribution may require adjustments to the planned analysis; unanticipated outliers may not be addressed by prespecified exclusion criteria. When the outcomes are known, the universe of contingencies is small; when the outcomes are unknown, the universe of contingencies is much larger.

Preregistration clarifies the distinction between planned and unplanned research by reducing unnoticed flexibility. This improves credibility of findings and calibration of uncertainty. However, making decisions before conducting analyses requires practice. During report writing, respecting both what was planned and what actually happened requires good judgment and humility in making claims.

These benefits are evident with rapid growth in use of preregistration to improve research rigor [6]. However, preregistration is a skill that requires experience to hone. Getting the most out of preregistration requires practice because the previous training of many scientists has involved making some important design and analysis decisions during analysis [4,7].

Nosek et al., 2019, *TICS*

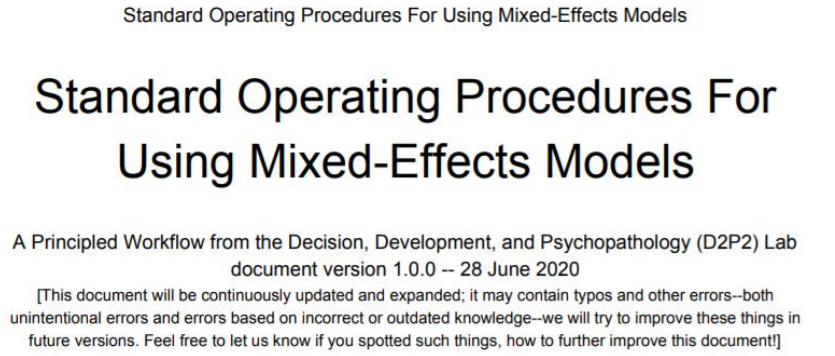
Tools to reduce uncertainty in advance

- Simulate data
  - Simulate to estimate power : https://github.com/debruine/lmem_sim
- Split data into training and test data set
  - First exploratory, than confirmatory
  - Direct replication
- Use stopping rules (maximum rules)
- Registered report: get peer review feedback
- Develop standard operating procedures (SOPs)


THE PROFESSION

Standard Operating Procedures:
A Safety Net for Pre-Analysis Plans

Winston Lin, *Columbia University*
Donald P. Green, *Columbia University*


Standard Operating Procedures For Using Mixed-Effects Models

Standard Operating Procedures For Using Mixed-Effects Models

A Principled Workflow from the Decision, Development, and Psychopathology (D2P2) Lab
document version 1.0.0 -- 28 June 2020
[This document will be continuously updated and expanded; it may contain typos and other errors--both unintentional errors and errors based on incorrect or outdated knowledge--we will try to improve these things in future versions. Feel free to let us know if you spotted such things, how to further improve this document!]

https://www.cos.io/blog/preregistration-plan-not-prison ; Nosek et al., 2018, *PNAS;* http://decision-lab.org/wp-content/uploads/2020/07/SOP_Mixed_Models_D2P2_v1_0_0.pdf ; Lin & Green, 2015, *PS*

Conclusion:

- thinking about pre-registration can lead to **important insights** for your research
- thinking about pre-registration can teach you **relevant skills** more broadly

Pre-registration “light”^2?

- “Private“ preregistration:
  - Fine to pre-register and not tell anyone (as-predicted)
  - Pre-register only for yourself (save pdf on your hard drive)
  - Will not help against fighting the file drawer

HOW DOES IT WORK?

- One author creates the pre-registration.
- Participating authors are emailed, requesting approval.
- If all approve, it is saved but remains private until an author makes it public; or remains private forever. (Why?)
- Authors may share an anonymous version of the pre-registration with reviewers.
- If made public, the final .pdf (sample) is automatically stored in the web-archive.

https://aspredicted.org/

Some applications

Pre-registration without theory

1. The Garden of Forking Paths
2. Getting rid of multiple comparisons
3. Postdiction

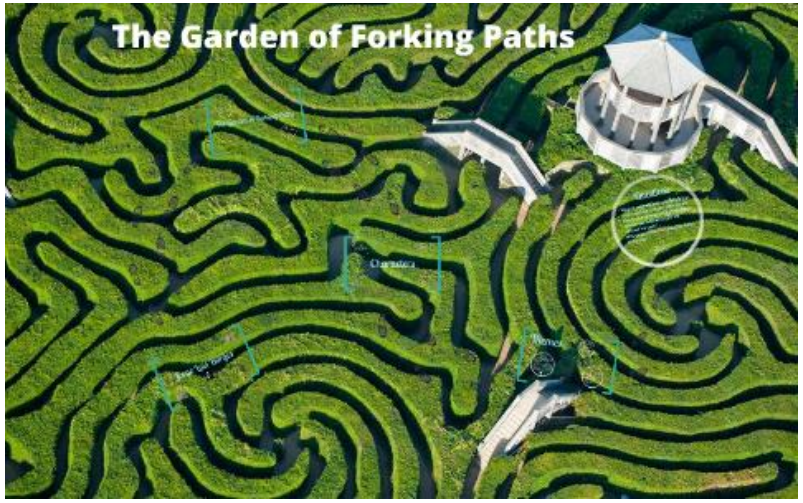
The garden of forking paths: Why multiple comparisons can be a problem, even when there is no "fishing expedition" or "p-hacking" and the research hypothesis was posited ahead of time*

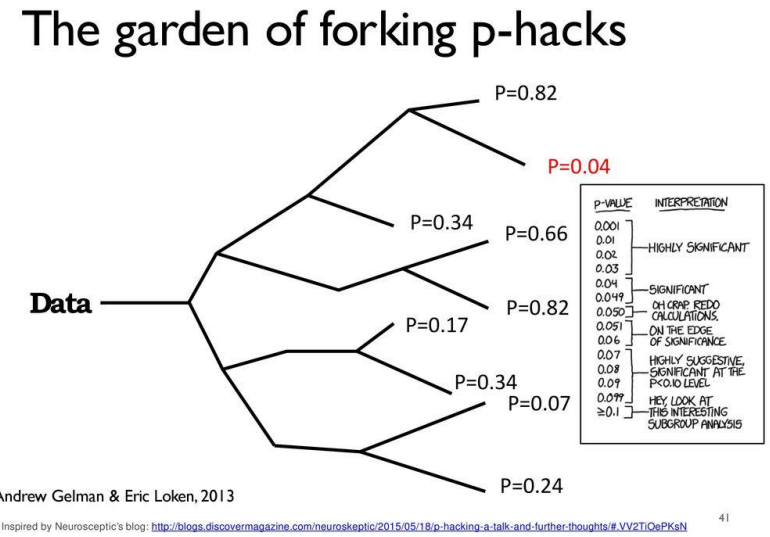
Andrew Gelman† and Eric Loken‡

14 Nov 2013

But we are starting to feel that the term "fishing" was unfortunate, in that it invokes an image of a researcher trying out comparison after comparison, throwing the line into the lake repeatedly until a fish is snagged. We have no reason to think that researchers regularly do that. We think the real story is that researchers can perform a reasonable analysis given their assumptions and their data, but had the data turned out differently, they could have done other analyses that were just as reasonable in those circumstances.

Our key point here is that it is possible to have multiple *potential* comparisons, in the sense of a data analysis whose details are highly contingent on data, without the researcher performing any conscious procedure of fishing or examining multiple p-values.





- Many steps seem reasonable post-hoc…
- Many things **need to be learned from the data**, we simply don't know them in advance…
- How can we even know them a-priori?

Gelman & Loken, 2013, http://www.stat.columbia.edu/~gelman/research/unpublished/p_hacking.pdf

Is pre-registration worthwhile?

Letter

Is Preregistration Worthwhile?

Aba Szollosi,[1,*] David Kellen,[2] Danielle J. Navarro,[1] Richard Shiffrin,[3] Iris van Rooij,[4] Trisha Van Zandt,[5] and Chris Donkin[1]

Preregistration is redundant, at best

Aba Szollosi[1]*, David Kellen[2], Danielle J. Navarro[1], Richard Shiffrin[3], Iris van Rooij[4], Trisha Van Zandt[5], and Chris Donkin[1]

Proponents of preregistration argue that, among other benefits, it improves the diagnosticity of statistical tests [1]. In the strong version of this argument, preregistration does this by solving statistical problems, such as family-wise error rates. In the weak version, it nudges people to think more deeply about their theories, methods, and analyses. We argue against both: the diagnosticity of statistical tests depend entirely on how well statistical models map onto underlying theories, and so improving statistical techniques does little to improve theories when the mapping is weak.

Szollosi et al., 2020, *TICS*

Deducing predictions

- Knowledge = true, *justified* belief (Plato, Theaetetos)
- Many analysis steps seem to be *without strong justification*:
  - What measures to pick?
  - What subjects/ trials/ data points to exclude?
  - We can make such choices, but they often appear like arbitrary picks between multiple reasonable options
- What is the value of knowing that a phenomenon can be observed under *coincidentally* chosen conditions?

Getting rid of multiple comparison by theory?

> Some problems arising from the statistical assumptions in such abstractions, such as family-wise error rates, only exist when hypotheses and statistical comparisons are effectively chosen at random. When statistical inference is used in scientific arguments, statistical models are just tools to test implications derived from theory. Therefore, such statistical problems become irrelevant because theories, not random selection, dictate what comparisons are necessary [12].

- If theory predicts all analysis steps (pre-processing, test selection): any fishing expedition exposed by theory itself
- If theory strong enough: pre-registration obsolete?

Descriptive science

- p-values as descriptive statistics (effect size relative to sample)?
- "Observations constrain future theories"
- Is this a good way to start science?

There are two aspects or types of research in the natural sciences. The first type of research describes phenomena, and the second involves the creation and testing of theories to explain these phenomena (Haig, 2005). Phenomena can be defined as "relatively stable, recurrent, general features of the world" (Haig, 2005, p. 374). Historically, in the mature natural sciences, description of phenomena, including invariant functional relations between variables (such as pressure and temperature in gasses, degree of selection pressure and rapidity of evolutionary change, and time in the dark- and light-detection threshold), precedes and becomes the basis for theory and hypothesis testing. In general, in spite of the example of better

the point home. In a sense, trying to build theories on collections of effects is much like trying to write novels by collecting sentences from randomly generated letter strings. Indeed, each novel ultimately consists of strings of letters, and theories should ultimately be compatible with effects.

Still, the majority of the (infinite possible) effects are irrelevant for the aims of theory building, just as the majority of (infinite possible) sentences are irrelevant for writing a novel.[4] Also, many of the *relevant* effects (sentences) we may never happen upon by chance, given the vast space of possibilities.[5] How can we know which effects are relevant and informative, and which ones are distractors? We first need to build candidate theories and determine which effects they imply.

Rozin, 2009, *POPS*; Szollosi et al., 2020, *TICS*; van Rooij & Baggio, 2020, *PoPS*

Post-hoc theorizing

- Fine, but needs “as rigorous test”
  - Replication
  - More predictions = stronger test
    - Explaining more data points
    - Including old data sets

problematic about *post hoc* scientific inference when theories are strong. The crucial difference is that strong scientific inference requires that *post hoc* explanations be tested just as rigorously as ones generated before an experiment; for example, by a collection of *post hoc* tests that evaluate the many regularities implied by a novel theory [7]. There is no reason not

Conclusion: We are not only missing knowledge about “the data”, but about data generating processes itself!
Stronger theory would make obsolete pre-registration obsolete…

Szollosi & Donkin, 2019, *psyArXiv*; Szollosi et al., 2020, *TICS*

Pre-registration using secondary data

Retaining blinding

The preregistration revolution

Brian A. Nosek[a,b,1], Charles R. Ebersole[b], Alexander C. DeHaven[a], and David T. Mellor[a]

Protecting against researcher bias in secondary data analysis:

Challenges and solutions

Jessie R. Baldwin[1,2], PhD, Jean-Baptiste Pingault[1,2], PhD, Tabea Schoeler,[1] PhD,

Hannah Sallis[3,4,5], PhD & Marcus Munafò[3,4,6], PhD

Preregistration of secondary data analysis: A template and tutorial

Olmo R. van den Akker[1]

Sara J. Weston[2]

Lorne Campbell[3]

William J. Chopik[4]

Rodica Ioana Damian[5]

Pamela E. Davis-Kean[6]

Andrew N. Hall[7]

Jessica E. Kosie[8]

Elliott Kruse[9]

Jerome Olsen[10,11]

Stuart J. Ritchie[12]

K.D. Valentine[13]

Anna E. van 't Veer[14]

Marjan Bakker[1]

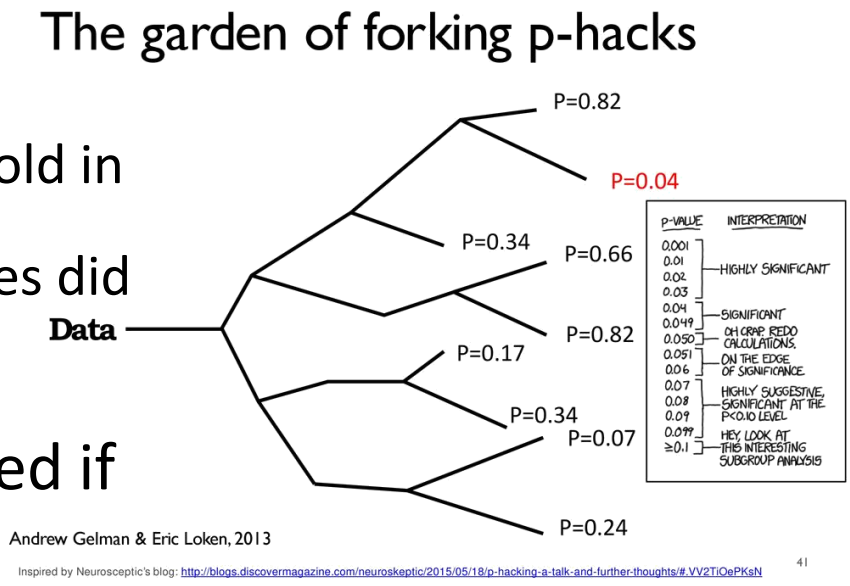
This highlights how partial blinding creates a gray area between prediction and postdiction. Once definitive blindness is killed, the diagnosticity of statistical inference is maximized by registering analysis plans and transparently reporting what was and was not known in advance about the dataset. This transparency provides insight about potential biasing influences for, at minimum, subjective assessment of credibility. Otherwise, nothing is preregistered, and there is no basis to assess credibility. An effective preregistration will account for any loss of blinding and what impact that could have on the reported results.

The extent to which testing predictions is possible on preexisting data depends on whether decisions about the analysis plan are blind to the data. "Pure" preregistration is still possible if no one has observed the data. For example, a paleontologist can test predictions about what will be observed from fossils yet to be discovered, and an economist can create predictions of government data that exist but have not been released. However, once the data have been observed, there are inevitable risks for blinding. Questions to ask include: "Who has observed the data?" and "What observations, summaries, or findings have been communicated, and to whom?" A researcher could test predictions using a dataset that has been examined by hundreds of others if the new analyst is entirely blind to what others have observed and reported. However, there are lots of ways—direct and indirect—to be influenced by observed data. If the new analyst reads a summary report of the dataset or receives advice on how to approach the dataset by prior analysts, decisions might be undesirably influenced. Likewise, knowing some outcomes might influence decisions, even if the analysis is on different variables from the dataset. For example, a political scientist might preregister an analysis examining the relationship between religiosity and volunteerism using an existing dataset. She has never observed data for the variables of interest, but she has previously observed a relationship between political ideology and charitable giving. Even though she is blind to data from her selected variables, the likelihood of positive correlations between ideology and religiosity and between charitable giving and volunteerism damages blinding.

Baldwin et al., 2020, *psyArXiv*, Nosek et al., 2018, *PNAS;* van den Akker et al., 2020, *psyArXiv*

Difficult questions while retaining blindness

- Initial data quality checks: What is “too far”?
  - Would you proceed if data was all missing?
  - Would you proceed if data was non-normally distributed?
  - Would you proceed if the phenomenon is known to not hold in a subset of 10% of the people?
  - Longitudinal projects: Would you proceed if previous waves did not find a certain effect?
- Transparency and reproducibility can still be facilitated if blindness is broken!
  - Open notebook
  - Log of all data-dependent decisions
  - Keep on trying not to fool yourself!



Baldwin et al., 2020, *psyArXiv*, Nosek et al., 2018, *PNAS;* van den Akker et al., 2020, *psyArXiv*

Pre-registration with many hypotheses/ tests

Who's gonna right it? Who's gonna read it?

But if preregistration is going to be the solution, then we need to ensure that it is done right. After casually reviewing several recent preregistration attempts in published papers, we noticed that there is room for improvement. We saw two kinds of problems.

Problem 1. Not enough information

For example, we saw one "preregistration" that was simply a time-stamped abstract of the project; it contained almost no details about how data were going to be collected and analyzed. Others failed to specify one or more critical aspects of the analysis: sample size, rules for exclusions, or how the dependent variable would be scored (in a case for which there were many ways to score it). These preregistrations are time-stamped, but they lack the other critical ingredient: precise planning.

Problem 2. Too much information

A preregistration cannot allow readers and reviewers to distinguish between confirmatory and exploratory analyses if it is not easy to read or understand. Thus, a preregistration needs to be easy to read and understand. This means that it should contain only the information that is essential for the task at hand. We have seen many preregistrations that are just too long, containing large sections on theoretical background and on exploratory analyses, or lots of procedural details that on the one hand will definitely be part of the paper, and on the other, are not p-hackable. Don't forget that you will publish the paper also, not just the preregistration; you don't need to say in the preregistration everything that you will say in the paper. A hard-to-read preregistration makes preregistration less effective [4].

Thinking about evidence, and vice versa

http://datacolada.org/64

Templates

Preregistration in Complex Contexts:
A Preregistration Template for the Application of Cognitive Models

Sophia Crüwell[1,2] and Nathan J. Evans[2,3]

Preregistration: A Solution to Undisclosed Analytic Flexibility in ERP Research

Mariella Paul[a,b,c,§], Gisela H. Govaart[a,b,d,§], Antonio Schettino[e,f,§,*]

Preregistration of secondary data analysis: A template and tutorial

Olmo R. van den Akker[1]

Sara J. Weston[2]

Lorne Campbell[3]

William J. Chopik[4]

Rodica Ioana Damian[5]

Pamela E. Davis-Kean[6]

Andrew N. Hall[7]

Jessica E. Kosie[8]

Elliott Kruse[9]

Jerome Olsen[10,11]

Stuart J. Ritchie[12]

K.D. Valentine[13]

Anna E. van 't Veer[14]

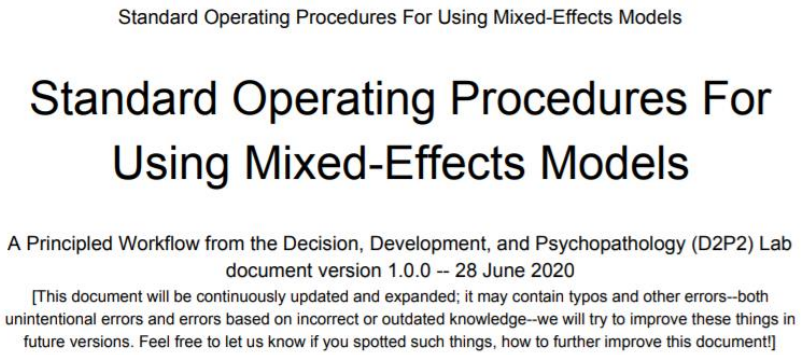
Marjan Bakker[1]

- Create your own lab template!
- Create your own Standard Operating Procedures (SOPs)!

Pre-registration using many hypotheses

- Create a template that is varied
  - Or use an existing one
  - Combine them via a decision tree
- Create SOPs to reuse in the future
  - Or use an existing one
- Pre-register only replications

Teams that run many experiments are often doing so in the context of a methodological paradigm in which each experiment varies some key aspects of a common procedure. In this situation, preregistration can be as efficient as the design of the experiments themselves. A preregistration template defines the variables and parameters for the protocol, and the preregistrations document which parameters will be changed or manipulated for each successive experiment.


Standard Operating Procedures For Using Mixed-Effects Models

Standard Operating Procedures For Using Mixed-Effects Models

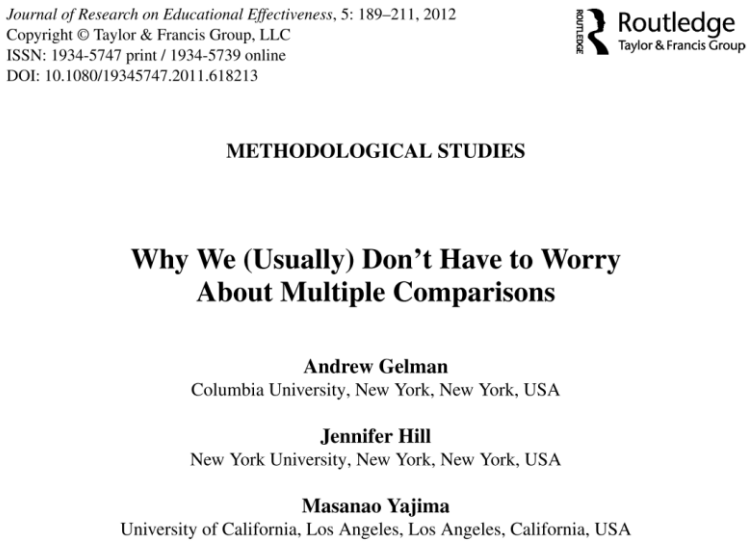
A Principled Workflow from the Decision, Development, and Psychopathology (D2P2) Lab
document version 1.0.0 -- 28 June 2020
[This document will be continuously updated and expanded; it may contain typos and other errors--both unintentional errors and errors based on incorrect or outdated knowledge--we will try to improve these things in future versions. Feel free to let us know if you spotted such things, how to further improve this document!]

In some cases, data acquisition is so simple that any documentation process interferes with efficiency. In such scenarios, researchers can achieve confirmatory research via replication. All initial experiments are treated as exploratory research. When something of interest is observed, then the initial design and analysis script become the preregistration for testing a prediction by running the experiment again. Easy data acquisition is a gift

http://decision-lab.org/wp-content/uploads/2020/07/SOP_Mixed_Models_D2P2_v1_0_0.pdf; Nosek et al., 2018, *PNAS*

Increase power by **avoiding** multiple tests

- Subset of data/ conditions, no omnibus tests (cluster-based tests)
  - Contrasts
  - Incorporate "prior knowledge" to restrict analyses (priors, regions of interest)
- Incorporate all analyses into one model that takes care of multiple comparisons



Journal of Research on Educational Effectiveness, 5: 189–211, 2012

ISSN: 1934-5747 print / 1934-5739 online
DOI: 10.1080/19345747.2011.618213

Routledge
Taylor & Francis Group

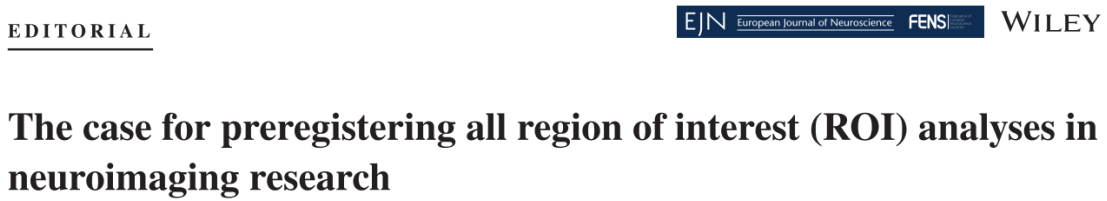
METHODOLOGICAL STUDIES

Why We (Usually) Don't Have to Worry About Multiple Comparisons

Andrew Gelman
Columbia University, New York, New York, USA

Jennifer Hill
New York University, New York, New York, USA

Masanao Yajima
University of California, Los Angeles, Los Angeles, California, USA



EDITORIAL

EJN European Journal of Neuroscience FENS WILEY

The case for preregistering all region of interest (ROI) analyses in neuroimaging research

Gelman, Hill, & Yajima, 2013; JREE ; Gentili et al., 2020, *EJN*

Summary

Summary

- Pre registered ≠ High quality
- Pre registered = Easier for readers to judge quality
- Not the only cure, cannot stop fraud, can be ignored
- Pre-registration is only suited when confirmatory research can be conducted
  - Very clear analysis plan, ideally direct/ conceptual replication
- “Light” versions of pre-registration can still facilitate transparency, reproducibility, error detection
- “Light versions of pre-registration can still make thus think about important aspects of our research and teach valuable skills
- “Light” versions suitable also for exploratory research, secondary data analysis, data analysis after blinding is broken…
  - Transparently distinguish what you know and what you don’t know while making decisions!
  - Don’t fool yourself

Summary

What: Challenges and new trends in pre-registration - by Johannes Algermissen (Donders Institute)

Have you maybe downloaded one of the OSF templates for pre-registration, started filling it in, and noticed that it does not work for you? Because you have already analyzed bits of your data? Because you cannot predict certain analysis steps in advance? Or because you have too many hypotheses? Is it still worth pre-registering at all? In this event, Johannes will first give a general introduction into pre-registration and then cover its most recent developments, including secondary analysis of data, pre-registration with many hypotheses, and the use of theoretical constraints. There will also be time for questions & answers.

There is no good and evil [p-value], there is only [reproducibility] and those too weak to seek it

J.K. Rowling, Harry Potter and the Philosopher's Stone [slightly paraphrased]

Credits for the slide idea go to Chris Brydges