Pre-registered study protocol

TITLE:

Journal policies related to research transparency and reproducibility in psychology: A cross-sectional assessment of influential journals and randomly-selected journals

CONTRIBUTORS:

Tom E. Hardwicke* (University of Amsterdam) & Brian A. Nosek (Center for Open Science)

CONTACT:

*tom.hardwicke@uva.nl

DATE OF REGISTRATION:

4th September, 2020

Aims & rationale

Serious concerns have been raised about the credibility, utility, and reproducibility of psychology research (Nelson et al., 2018; Pashler & Wagenmakers, 2012). One approach to address these concerns is to leverage academic journal policies to promote potentially beneficial research activities, such as research transparency, whilst discouraging potentially deleterious research activities, such as selective reporting (Nosek et al., 2012). Meta-research has indicated that reforming journal policy may be effective; for example, several observational studies have documented marked increases in data availability following the introduction of journal policies that encourage or mandate data sharing (Hardwicke et al., 2018; Kidwell et al., 2016; Nuijten et al., 2017). In order to inform further efforts towards journal policy improvement, it will be helpful to systematically describe the broader landscape of journal policies as they relate to transparency and reproducibility. This will help to establish a baseline with which to assess future progress, and also to identify specific journals and policies that may be in most need of improvement.

In the present project[1] we intend to collate, summarise, and describe the policies of a selection of psychology journals in relation to ten pre-defined standards (Table 1) related to research transparency and reproducibility. Eight of the standards - known as the Transparency and Openness Promotion (TOP) guidelines - were originally devised in 2014 by the Center for Open

---

[1] This project is a contribution to an in-progress review article solicited by the journal *Annual Review of Psychology,* provisionally titled "Replicability, Robustness, and Reproducibility in Psychological Science".

Science (COS) in conjunction with a group of experts, largely from the social and behavioral sciences, including journal editors and funding agency representatives (Nosek et al., 2015), and are maintained by COS and a committee of stakeholders (http://cos.io/top/). Two additional standards ("Publication Bias" and "Open Science Badges") were added when COS introduced the "TOP Factor", a metric to quantify journal adherence to open science principles (https://perma.cc/7LSM-PHVL). Although the standards cover some core aspects of transparency and reproducibility, they are not comprehensive. For example, they do not address policies related to statistical power, conflict of interest statements, or some novel initiatives like the 'pottery barn replication rule' in which journals commit to publishing replications of studies previously published in their archives (https://perma.cc/SP2P-7Y5Q). The main reason for adopting the existing TOP standards as outcome variables in the present study (as opposed to developing new standards or incorporating additional standards) is efficiency - COS is already in the process of extracting and collating journal policies related to these standards, so we can (a) partly capitalize on their existing database; and (b) contribute additional information to that database.

We will attempt to assess the prevalence of the ten standards in journal policy across the field of psychology and the frequency of adoption by some highly influential psychology journals. To address the former goal, we will examine a sample of 40 journals randomly selected from amongst all journals belonging to the 'psychology' category defined by Web of Science (Clarivate Analytics). To address the latter goal, we will examine the top-5 journals according to their 2019 Journal Impact Factor (JIF) in each of ten sub-fields of psychology (i.e., 50 journals). Whilst the use of any single quantitative metric as a proxy for journal influence has limitations, we only intend to only use JIF as a practical selection criterion that captures a set of highly influential journals. In total across both samples we will examine 90 journals.

Table 1. TOP Factor rubric (obtained from https://perma.cc/CT2T-5G3F?type=image on August 25, 2020).

	Policy level			
Policy	**0**	**1**	**2**	**3**
Data citation	No mention of data citation.	Journal describes citation of data in guidelines to authors with clear rules and examples.	Article requires appropriate citation for data and materials used consistent with the journal's author guidelines.	Article is not published until providing appropriate citation for data and materials following journal's author guidelines.

Data transparency	Data sharing is encouraged, or not mentioned.	Articles must state whether or not data are available. Requiring a data availability statement satisfies this level.	Articles must have publicly available data, or an explanation why ethical or legal constraints prevent it.	Articles must have publicly available data and must be used to computationally reproduce or confirm results prior to publication.
Analytical code transparency	Code sharing is encouraged, or not mentioned.	Articles must state whether or not code is available. Requiring a code availability statement satisfies this level.	Articles must have publicly available code, or an explanation why ethical or legal constraints prevent it.	Articles must have publicly available code and must be used to computationally reproduce or confirm results prior to publication.
Materials transparency	Materials sharing is encouraged, or not mentioned.	Articles must state whether or not materials are available. Requiring a materials availability statement satisfies this level.	Articles must have publicly available materials, or an explanation why ethical or legal constraints prevent it.	Articles must have publicly available materials and must be used to computationally reproduce or confirm results prior to publication.
Reporting guidelines	No mention of reporting guidelines.	Journal articulates design transparency standards.	Journal requires adherence to design transparency standards for review and publication.	Journal requires and enforces adherence to design transparency standards for review and publication.
Study prereg	Journal says nothing.	Articles will state if work was preregistered.	Article states whether work was preregistered and, if so, journal verifies adherence to preregistered plan.	Journal requires that confirmatory or inferential research must be preregistered.
Analysis prereg	Journal says nothing.	Articles will state if work was preregistered with an analysis plan.	Article states whether work was preregistered with an analysis plan and, if so, journal verifies adherence to preregistered plan.	Journal requires that confirmatory or inferential research must be preregistered with an analysis plan.
Replication	Journal says nothing.	Journal encourages submission of replication studies.	Journal will review replication studies blinded to results.	Registered Reports for replications as a regular submission option.
Publication Bias	Journal says nothing.	Journal states that significance or novelty are not a criteria for publication decisions.	Journal will review studies blinded to results.	Journal accepts Registered Reports for novel studies as a regular submission option.
Open science badges	Journal says nothing.	Journal awards 1 or 2 open science badges.	Journal awards all 3 open science badges.	

Method

Design

Retrospective observational study with a cross-sectional design. Outcome variables are shown in Table 1.

Obtaining extant data and bibliographic information

On August 13th, 2020, we downloaded the databases described in Table 2. Note that Web of Science classifies journals into subject categories, ten of which pertain to the field of psychology (https://perma.cc/9V5T-PH2U). A journal can belong to more than one subject category.

Existing data in TOP-d pertaining to the journals in the sample has been extracted. An R script documenting this procedure (https://osf.io/kdnwg/) and the data (https://osf.io/hz987/) are available.

Table 2. Assorted online databases. The copies we downloaded are available from the links in the "access link" column.

Database	Description	Acronym	Source	Date last updated#	Access link
TOP Factor database	Extant data extracted for assessment of journal TOP Factor	TOP-d	Open Science Framework (https://osf.io/qatkz/)	August 11, 2020	https://osf.io/5sbg7/
Web of Science Social Sciences Citation Index master list (psychology journals only)	List of all journals included in the Web of Science Social Sciences Citation filtered to	WOS-psych-d	Clarivate Analytics (https://mjl.clarivate.com/collection-list-downloads)	July 21, 2020	https://osf.io/6wsj3/

	remove all entries that did not contain the term 'psychology' in the subject category field				
Clarivate Analytics Journal Citation Reports	List of journals ranked by 2019 Journal Impact Factor for each of the ten Web of Science subject categories pertaining to the field of psychology*	JCR-psych-d	Clarivate Analytics Journal Citation Reports (https://perma .cc/T2CT-3QE6)	Jun 29, 2020	https://osf.io/v de48/

According to the database website.

* These categories are "Psychology, Applied", "Psychology, Biological", "Psychology, Clinical", "Psychology, Developmental", "Psychology, Educational", "Psychology, Experimental", "Psychology, Mathematical", "Psychology, Multidisciplinary", "Psychology, Psychoanalysis", and "Psychology, Social".

Sample

Definition of samples

There will be two samples of journals:

(1) Random sample: 40 journals randomly selected from amongst all psychology journals indexed in the Web of Science Social Sciences Citation Index (WOS-psych-d).

(2) High-impact sample: The top-5 journals according to 2019 Journal Impact Factors in each of ten subfields of psychology defined by Web of Science (JCR-psych-d; i.e., 50 journals).

Analysis scripts documenting the sampling and screening process are available (https://osf.io/4fqzj/).

Sample size justification

The purpose of the high impact sample is to enable us to gauge the frequency of the outcome variables at some of the most influential psychology journals. To achieve this, we have selected the top-5 journals in each subject category of psychology as we intuitively believe this to be informative and within our workload capacity.

The purpose of the random sample is to enable us to estimate the prevalence of the outcome variables across the field of psychology. We have therefore performed a precision analysis to aid our decision about target sample size. A precision analysis enables one to examine how precision (specifically confidence interval margin of error) varies with sample size for a given effect size (Gelman & Hill, 2007; Rothman & Greenland, 2018). To perform a precision analysis, one can take the relevant confidence interval formula, and specify two of the following parameters, solving for the third: effect size, sample size, and margin of error. Additionally, one can plot a precision curve (see Figure 1) to illustrate the relationship between sample size and precision across a range of values. In our particular context, we are interested in estimating proportions, so we use the following confidence interval formula (Gelman & Hill, 2007, p. 442):

$$MOE = z \, x \sqrt{\frac{p(1-p)}{n}}$$

where *MOE* = margin of error, *z* = z-score corresponding to the desired confidence interval (e.g., 1.96 for a 95% confidence interval), *p* = expected proportion, and *n* = sample size.

For the precision curve shown in Figure 1 we have used an expected proportion of 0.5 as we do not have strong a priori expectations and this is the most conservative value (it always leads to the maximum sample size estimate for a given MOE; Gelman & Hill, 2007, p.442). The results of a precision analysis that specifically assumes a sample size of 40 is displayed: this would yield an expected MOE of 0.15. An analysis script documenting this precision analysis is available (https://osf.io/uj5xk/).
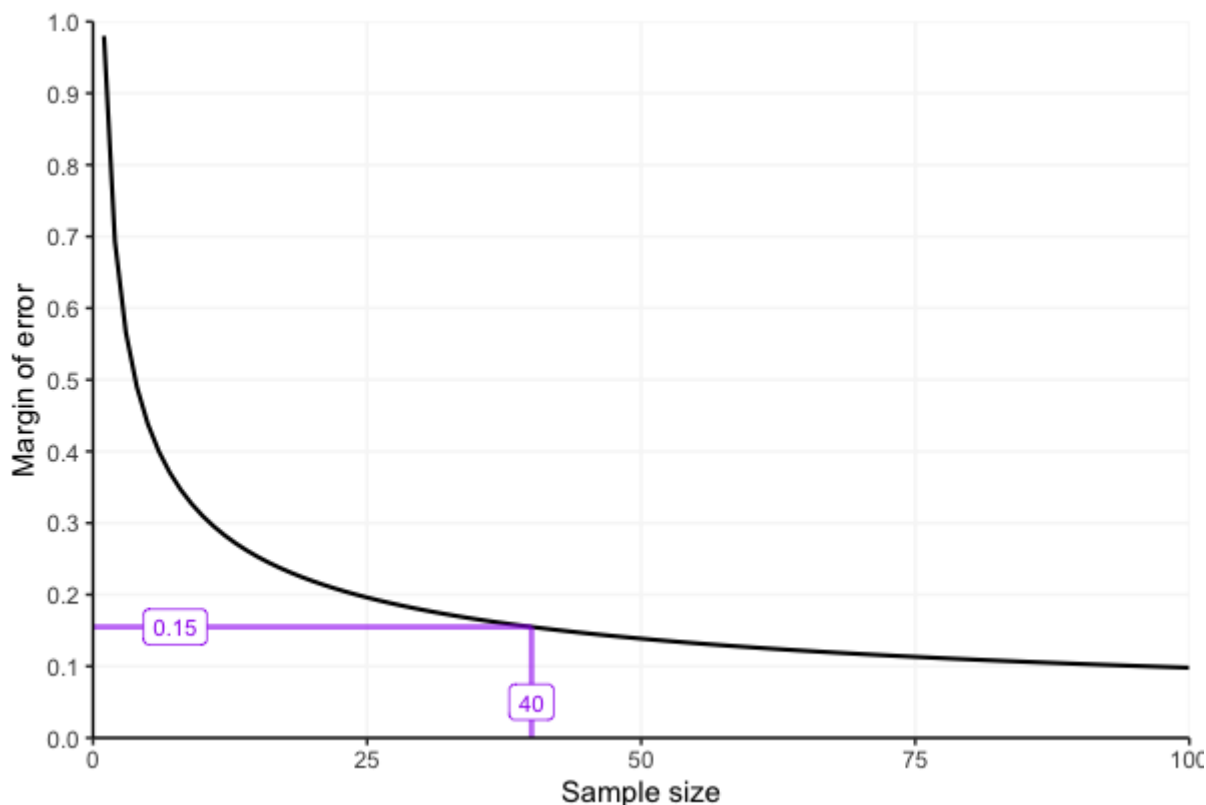
Figure 1. Precision curve showing margin of error for a 95% confidence interval as a function of sample size assuming a proportion of 0.5. Purple lines indicate the results of a precision analysis further assuming a sample size of 40, which results in an expected margin-of-error of 0.15.

Inclusion and exclusion criteria

- We will only include journals classified as English language in WOS-psych-d because we do not have the resources to achieve high-quality translation. 71 non-English language journals were excluded from WOS-psych-d prior to sampling.
- We will not include journals that do not typically publish empirical research involving primary data (from herein 'non-empirical journals') because the outcome variables are less applicable to these journals.
  - To exclude non-empirical journals, one contributor (T.E.H.) has screened all journals output during the sampling process and replaced any identified as non-empirical.
    - In the case of the random sample, replacement involved randomly drawing another journal from amongst remaining journals in the WOS-psych-d. Three such replacements were required.
    - In the case of the high-impact sample, replacement involved selecting the next ranked journal still available in the JCR-psych-d. Twenty-nine such replacements were required.
  - In order to identify non-empirical journals, T.E.H. manually examined each journal's website. If the website indicated that the journal only published non-empirical content

(e.g., news, opinions, narrative reviews, systematic reviews, meta-analyses) it was considered a non-empirical journal and excluded. If the journal's status was not clear from the journal website, T.E.H. examined the last ten articles published in the journal and if they were all non-empirical the journal was excluded.

- When journals included in the random sample appeared in the high impact sample, they were removed from the high impact sample and replaced. Four such replacements were required.
- If journals in the high impact sample appeared in multiple subfields, the highest ranked entry was retained and the lowest ranked entry replaced. Three such replacements were required.

Sampling procedure

The random sample was obtained first. We used R to randomly sample 40 journals from the WOS-psych-d, applied the inclusion and exclusion criteria (see above), and randomly sampled additional replacement journals as necessary.

The high-impact sample was obtained second. We used R to select the top-5 journals in each subject area of the JCR-psych-d, applied the inclusion and exclusion criteria (see above), and selected additional replacement journals from the next available rank in the relevant subfield as necessary.

Assignment to coders

Journals in both samples for which there was not already data in TOP-d (see "Obtaining extant data and bibliographic information") were randomly assigned to T.E.H. and B.A.N. such that each coder was assigned to half of the available journals in each sample.

Procedure

1. Each coder will complete the data extraction form (https://osf.io/fkvsp/) for each journal assigned to them in the high impact sample (https://osf.io/jyrh8/) and the random sample (https://osf.io/8qzhm/).
2. Information about journal policies will be obtained by manual inspection of journal websites. Particular attention will be paid to webpages related to instructions to authors and editorial policy.
3. Completed data extractions will be submitted to TOP-d. The stewards of TOP-d will conduct a limited quality control assessment. This process is not standardized and cannot be

considered an independent data extraction exercise. If any issues are detected during the quality control assessment, they will be referred back to us and addressed before the dataset is finalized.

Analysis plan

For each outcome variable (Table 1) we will report descriptive statistics (e.g., counts and percentages) and, for the random sample, 95% confidence intervals.

Conflict of interest statement

B.A.N. is co-founder and Executive Director of the Center for Open Science and was involved in the development, promotion, and administration of the Transparency and Openness Promotion Guidelines. T.E.H. declares no conflict of interest.

References

Gelman, A., & Hill, J. (2007). *Data analysis using regression and multilevel/hierarchical models*. Cambridge University Press.

Hardwicke, T. E., Mathur, M. B., MacDonald, K., Nilsonne, G., Banks, G. C., Kidwell, M. C., Hofelich Mohr, A., Clayton, E., Yoon, E. J., Henry Tessler, M., Lenne, R. L., Altman, S., Long, B., & Frank, M. C. (2018). Data availability, reusability, and analytic reproducibility: Evaluating the impact of a mandatory open data policy at the journal Cognition. *Royal Society Open Science*, *5*(8), 180448. https://doi.org/10.1098/rsos.180448

Kidwell, M. C., Lazarević, L. B., Baranski, E., Hardwicke, T. E., Piechowski, S., Falkenberg, L.-S., Kennett, C., Slowik, A., Sonnleitner, C., Hess-Holden, C., Errington, T. M., Fiedler, S., & Nosek, B. A. (2016). Badges to Acknowledge Open Practices: A Simple, Low-Cost, Effective Method for Increasing Transparency. *PLOS Biology*, *14*(5), e1002456. https://doi.org/10.1371/journal.pbio.1002456

Nelson, L. D., Simmons, J., & Simonsohn, U. (2018). Psychology's Renaissance. *Annual Review*

of Psychology, *69*(1), 511–534. https://doi.org/10.1146/annurev-psych-122216-011836

Nosek, B. A., Alter, G., Banks, G. C., Borsboom, D., Bowman, S. D., Breckler, S. J., Buck, S., Chambers, C. D., Chin, G., Christensen, G., Contestabile, M., Dafoe, A., Eich, E., Freese, J., Glennerster, R., Goroff, D., Green, D. P., Hesse, B., Humphreys, M., … Yarkoni, T. (2015). Promoting an open research culture. *Science*, *348*(6242), 1422–1425. https://doi.org/10.1126/science.aab2374

Nosek, B. A., Spies, J. R., & Motyl, M. (2012). Scientific Utopia: II. Restructuring Incentives and Practices to Promote Truth Over Publishability. *Perspectives on Psychological Science*, *7*(6), 615–631. https://doi.org/10.1177/1745691612459058

Nuijten, M. B., Borghuis, J., Veldkamp, C. L. S., Dominguez-Alvarez, L., Van Assen, M. A. L. M., & Wicherts, J. M. (2017). Journal Data Sharing Policies and Statistical Reporting Inconsistencies in Psychology. *Collabra: Psychology*, *3*(1), 31. https://doi.org/10.1525/collabra.102

Pashler, H., & Wagenmakers, E.-J. (2012). Editors' Introduction to the Special Section on Replicability in Psychological Science: A Crisis of Confidence? *Perspectives on Psychological Science*. https://doi.org/10.1177/1745691612465253

Rothman, K. J., & Greenland, S. (2018). Planning Study Size Based on Precision Rather Than Power. *Epidemiology*, *29*(5), 599–603. https://doi.org/10.1097/EDE.0000000000000876