Transparency of Research on Philanthropy

René Bekkers, June 5, 2022[1]

How transparent is research on philanthropy, and how has it changed in the past decades? How are characteristics of research on philanthropy related to the level of transparency about data and methods? How can the degree of transparency be increased? These are the questions we seek to answer in the current project.

Why philanthropy research transparency is important

Transparency of data and methods is important because it is a necessary condition for the evaluation of the quality of research. Users of research in the practice of philanthropy and academic peers can only evaluate the quality of research when original research materials are documented, data are available, and when procedures used to obtain the conclusions from the analyzed data can be traced. Without these, low quality data, methods and conclusions wrongfully inform decisions about philanthropic funding, fundraising, the governance, management and regulation of philanthropic organizations, and tax incentives and other policies affecting philanthropy.

Increasingly, funders of research require public availability of research publications, data and materials from grantees. The European Research Council (ERC) requires grantees in the Horizon Europe scheme to publish not only research publications with an open access (CC-BY) license, but also deposit research data in a trusted repository, following the principle 'as open as possible, as closed as necessary'.[2]

Recently, the Office of Science and Technology Policy (OSTP) of the US government issued a memo announcing that all federally funded research will require that "publications and their supporting data resulting from federally funded research [should be made] publicly accessible without an embargo on their free and public release" by the year 2026.[3]

Exactly how transparent is research on philanthropy? To the best of our knowledge, no study to date has explored the level of transparency of research on philanthropy, and its trends and correlates. Still we can derive some expectations from the field of meta science

(Peterson & Panofsky, 2023): research on research, particularly in the social and behavioral sciences.

The distinction between data and methods transparency

Transparency about data and methods refer to two different aspects of the research process. Data transparency means that it is clear which data were analyzed, and who, when, where, and how collected the data. Full data transparency does not mean that all data collected are publicly shared. It does include access to the processed data in which observations are anonymized to protect the privacy of individuals. Full data transparency also includes access to the research instruments that were used to produce the data, such as questionnaires, instructions for participants, topic lists, and so on.

Methods transparency means that it is clear which methods were applied to derive conclusions from the data: the anonymization of raw data to processed data, and the treatment of processed data to obtain quotes, tables, and figures. In research reporting statistical analyses obtained with software full methods transparency includes access to the code that produces the results reported. In research reporting interpretations of data without software full methods transparency is more difficult to achieve, because authors may not be fully aware or able to describe all the criteria and private knowledge they used for the interpretation of data.

Data transparency does not imply methods transparency: publications may identify the data source, but not disclose the methods used to analyze the data. Both are essential for a thorough evaluation of the validity of claims.

Availability of research data

A recent estimate by Serghiou et al. (2021), showed that only 8% of 27,000 articles from the Social Sciences included in the PubMed database provided access to data. In a smaller sample of 250 publications in Scopus-indexed outlets from the period 2014-2017, Hardwicke et al. (2020) find that 7% provided access to data. A study of 200 publications from the social sciences at a university in the Netherlands (Bekkers, 2023) found that 7% of empirical publications in 2018 referred to the underlying dataset with a link to a repository or other information about the data, and 6% of empirical publications in 2021 did so. Taken together, these estimates suggest that social science research publications only rarely provide access to the underlying data.

Research on philanthropy is published in a variety of disciplines. Most studies on philanthropy appear in the social sciences, including public administration, (nonprofit) management and business administration, economics, and sociology. In addition, research on philanthropy also appears in the humanities (e.g., law, philosophy, linguistics) and the natural sciences (e.g., biology). Over time, the number of publications on philanthropy has increased, and also the variety of disciplines in which the research gets published.

Some authors who do not make data and materials available to users and reviewers through a repository promise that they will make data and materials available upon (reasonable) request. In a study among economists who had indicated in their publications that data and materials were available upon request (Krawczyk & Reuben, 2012), only 44% actually complied with such a request. Among psychologists who had published in the top journals of the field and promised data and materials would be available upon request, only 26% complied (Wicherts et al., 2006). In a study among researchers who published in *Nature* and *Science*, which both require authors to promise they will give access to research data, still only 40% of psychologists and social scientists complied with a request to access the data (Tedersoo et al., 2021). In other words, the promise that "data are available upon request" usually means that the data are not made available.

Availability of code

There is no reliable baseline for the prevalence of sharing code in the social sciences. It is likely to be even lower than the prevalence of data sharing. A study in epidemiology reports that only 2% of systematic reviews in PubMed shared code while 13% provided access to the data (Page et al., 2022). For meta-analyses with individual participant data in the health sciences, 0.5% shared code while 2% shared data (Hamilton et al., 2023). A study in the field of ecology, where the base rate of sharing research data (79%) is much higher than in the social sciences, sharing code is still relatively uncommon (27%), even when it is mandatory or encouraged by journal policies (Culina et al., 2020).

When publications do not share the code that produces the results reported, it is difficult to assess whether methods transparency in the methods section of research reports is complete. In practice, the number of decisions that researchers make to derive conclusions from data is much higher than the word limit in a regular article permits to describe.

Prevalence of significant results in statistical tests

An indicator of a lack of methods transparency is an excess of significant p-values. P-values tend to be more common when they are just above the critical value of 1.96 for a statistically significant finding at the conventional p-value of .05. This pattern has been documented in sociology (Gerber & Malhotra, 2008a), political science (Gerber & Malhotra, 2008b), psychology (Simonsohn, Nelson & Simmons, 2014), communication science (Vermeulen et al., 2015), and economics (Brodeur, Lé, Sagnier, & Zylberberg, 2016). The excess of significant p-values is a sign of p-hacking (Simmons, Nelson & Simonsohn, 2011): researchers strategically make choices in the selection of observations and the statistical analyses to obtain findings that are statistically significant. The prevalence of p-hacking is particularly high in online experiments in marketing conducted through MTurk (Brodeur, Cook & Hayes, 2022).

Low statistical power

Tests of hypotheses are not informative if they have insufficient statistical power. For effect sizes in the social sciences, which tend to be small, a relatively large number of observations is required to achieve high statistical power. There are no studies documenting the prevalence of a priori power analyses in social science research. In health education and behavior studies between 2000 and 2003, 9% of studies reporting on survey data include an a priori power analysis (Price et al., 2005). In recently published motor behavior research, 13% of studies report a power analysis (McKay et al., 2022). In plastic surgery randomized control trials between 1990 and 2010, 12% of publications report a power analysis (Ayeni et al., 2012). For bone and joint surgery RCTs between 1988 and 2002 the prevalence of power analysis reporting is 6% (Bhandari et al., 2002).

From test statistics and the number of observations reported in publications, statistical power can be computed post hoc. Using this approach, studies in psychology (Maxwell, 2004), economics (Ioannidis Stanley & Doucouliagos, 2017) and political science (Arel-Bundock et al., 2022) tend to be underpowered.

A cost-benefit approach to correlates of transparency

What characteristics of data, methods, authors, journals, and disciplines are likely to be correlated with transparency? We approach this question from a cost-benefit approach. For studies, authors, journals and disciplines reporting on data and methods that are less costly and more beneficial to document, data and methods transparency will be higher.

Individual benefits

The currency of academia is prestige, often measured with the number of citations to research publications. Publications with publicly available data and methods are cited better than publications that are less transparent (Piwowar & Vision, 2013; Colavizza et al., 2020). The citation advantage of data transparency creates an individual benefit to the users of publicly available data. However, the creators of data and code do not usually reap the rewards of these citations, because impact factors and tenure and promotion decisions typically do not reward citations of data and code (Buneman et al., 2020). Only when data creators write a citable research paper describing the data (e.g., Wilhelm, 2006) the advantage can materialize.

Individual costs

Data transparency requires little effort for authors reporting on existing data sets: they can simply include a reference to the source data. This is also their duty. By default, usage of data deposited in a data archive such as the ICPSR Archive or Dataverse, is only possible upon the condition of adequate citation of the data. Code books and user manuals for

surveys that include data on philanthropy, such as the European Social Survey and the Giving in the Netherlands Panel Survey, require users of the data to refer to the data.

Authors reporting on newly collected data face higher costs of data transparency: they will have to create a repository for the codebook or user manual at a data archive, or make a citable description of the data available on a website. Extensive documentation of data and methods requires more effort than a quick and dirty flash report.

When detected mistakes are intentional, for instance to obtain a certain result that helps secure promotion or future funding, they can endanger the reputation of authors if it is made public in the course of a research integrity complaint or journal article retraction. The higher the chance of detection of mistakes, the more effective transparency requirements will be in improving the quality of research.

Violations of research integrity are like crime: when the probability of being detected is high enough, potential perpetrators will not engage in violations. Data from the Netherlands Survey of Research Integrity shows that a higher likelihood of being detected by a reviewer or collaborator for data fabrication is associated with a lower likelihood of engaging in questionable research practices (Gopalakrishna et al., 2022a) and a higher likelihood to engage in responsible research practices such as sharing data and materials (Gopalakrishna et al., 2022b).

Collective benefits

Data and methods transparency improve the reliability of published research (Smaldino, Turner & Contreras Kallens, 2019). Data and methods transparency increase the likelihood that users of research identify mistakes (Nosek, Spies & Motyl, 2012). The identification of honest mistakes creates a collective benefit in the form of improved reliability and validity of research. The level of transparency also has a communication benefit in interaction with users of research: open science research is generally viewed as more credible than closed science (Song et al., 2022). Scrutiny of transparent data and methods can also identify sloppy science such as incorrect reporting of statistical tests and insufficient adherence to rules of good practice.

Studies with high statistical power, a preregistration, and full data transparency through publicly available data and code are more reliable and replicate well (Protzko et al., 2020). When studies are more replicable with the same methods and new data from the same target population, they are also more generalizable to other populations across time and place (Delios et al., 2022).

While replication is a cornerstone for nomothetic social science research, i.e. research that seeks to identify and test regularities in human behavior, not all scholars in nonprofit and philanthropic studies seek such knowledge. An important type of research in our field seeks idiographic knowledge, i.e. detailed description and interpretation of particular cases without claiming generalizability to populations or individuals in other locations and times.

Collective costs

The collective benefits of transparency come at a price. Journals, associations of researchers, and institutions can only reap the benefits of data and methods transparency if they invest in an infrastructure that verifies and incentivizes it. Journal editors and reviewers do not require authors of manuscripts under review to be fully transparent about all the choices they have made. As a result, mistakes are rarely detected in the classical peer review process (Altman, 1994; Smith, 2006, 2010). Without extensive training, peer reviewers are bad at catching mistakes in manuscripts (Schroter et al., 2008). Even at the journals with the highest impact factors, the review process does not successfully keep out bad science (Brembs, 2018). In the past decade, however, significant progress has been made. The number of journals that encourage data and methods transparency through a data and code availability statement or with a requirement to submit data and code has increased.

While journal requirements to submit data and code can enhance the reliability of the published record of research (Smith, 2010; Munafo et al., 2017), they are not a guarantee that authors actually do share data and code. A data and code sharing policy is ineffective if it is not enforced (Stodden, Seiler, & Ma, 2018; Christensen et al., 2019). As the evidence on the prevalence of data and code sharing mentioned above illustrates, compliance with policies is typically low. The low compliance is likely to be the result of the lack of enforcement by journal editors and publishers. Authors should not only be required to share data and code, but a data editor should also verify the computational reproducibility of the data and code. Few journals are willing to incur the costs of appointing such an editor. For studies relying on experiments, a mandatory data and code sharing policy was introduced in July 2020 (Bekkers, 2020). Computational verification by the area editor ensured that the data and code produced the results reported in the manuscript. For other types of studies, however, data and code are not yet required.

The policy for experiments at NVSQ was the result of a workshop at the 2019 ARNOVA conference (Bekkers, 2019). With an eye to the epistemic diversity (Leonelli, 2022) of our field, the journal editors formed working groups to establish guidelines for the evaluation of submissions of manuscripts reporting on different types of research. The indicators of research quality for hypothesis testing research differ from indicators that hold for research that does not test hypotheses and does not seek knowledge on regularities in behavior. Hence it is important that communities of researchers working with different types of data and with different research goals specify which indicators they consider to be relevant in the evaluation of research quality. Guidance on data and methods transparency for authors submitting manuscripts and for reviewers evaluating them should discriminate between types of research.

Using computational social science to quantify data and methods transparency

The increasing pace of research publication in the field of philanthropic studies (Ma & Konrath, 2018) makes it increasingly difficult to keep track of trends. Computational social science techniques can help automate labor intensive tasks such as human coding (Ma et al., 2021). Advancements in natural language processing have enabled software engineers

to build tools that automatically screen full texts of articles and extract information about ethics statements, randomization, sample sizes, sharing of data and code, and other indicators of research quality (Menke et al., 2020; Riedel, Kip & Bobrov, 2020; Serghiou et al., 2021; Zavalis & Ioannidis, 2022). With the current project, we contribute to an infrastructure for peer review in which manuscripts can be screened automatically and scored with respect to transparency indicators.

Just like artificial intelligence facilitates plagiarism detection, it can also enable the evaluation of data and methods transparency in the peer review process by screening manuscripts for errors and the presence of information about relevant indicators of research quality. While peer review should not be automated, artificial intelligence will certainly help improve peer review (Schulz et al., 2022). One example of a useful tool is StatCheck, which helps reviewers check the consistency between reported p-values and the test-statistics (http://statcheck.io; Nuijten & Polanin, 2020). Another example of a useful tool is the p-curve app, which quickly provides reviewers with relevant information about the evidentiary value of a set of experiments (https://shinyapps.org/apps/p-checker/; see Simonsohn, Nelson & Simmons, 2014a, 2014b).

Data and methods

In the current version, we report on the current state of research on philanthropy in five journals: Nonprofit & Voluntary Sector Quarterly (NVSQ), Voluntas, Nonprofit Management & Leadership (NML), Journal of Philanthropy and Marketing (JPM) and Voluntary Sector Review (VSR). We consulted these journals' webpages and analyzed the content of the most recent issues.[4] For each publication, René Bekkers (RB) manually coded:

1. The number of authors;
2. The disciplinary affiliation of authors;
3. Whether the research is empirical or not;

   For empirical publications, RB determined

   1) whether the name of the dataset is provided or a reference to it;
   2) whether a link to the data analyzed is provided;
   3) whether a link to a code file producing the results is provided;
   4) how many observations are analyzed (n);
   5) whether the study is experimental vs cross-sectional vs longitudinal;
   6) whether philanthropy is the dependent or independent variable;
   7) whether giving behavior or intentions are reported;
   8) for giving behavior: whether it is observed or self-reported;
   9) what population the study participants is assumed to represent;

---

[4] NVSQ: https://journals.sagepub.com/toc/nvsb/52/3
Voluntas: https://link.springer.com/journal/11266/volumes-and-issues/34-2
NML: https://onlinelibrary.wiley.com/toc/15427854/2023/33/4
JPM: https://onlinelibrary.wiley.com/toc/26911361/2023/28/2
VSR: https://bristoluniversitypressdigital.com/view/journals/vsr/14/1/vsr.14.issue-1.xml

10) which mechanisms in giving behavior (Bekkers & Wiepking, 2011) are
    analyzed.

In future versions of this paper, we will extend the analyses with samples from the
Knowledge Infrastructure for Nonprofit and Philanthropic Studies (KINPS, Ma et al. 2022;
described in Ma et al., 2021) and other bibliometric databases such as dimensions.ai
(https://www.dimensions.ai/who/researchers/) and OpenAlex
(https://github.com/ourresearch).

Results

Table 1 shows data and methods transparency indicators for articles in the most recent
issues of NVSQ, Voluntas, and NML, as well as some other characteristics. The average
number of authors of articles in NVSQ (2.5) is somewhat higher than in Voluntas (2.3), with
NML scoring in between (2.4). Articles in JPM have the lowest number of authors (1.9). In
NVSQ and Voluntas Public Administration is the most common affiliation of contact authors.
In NML Public Administration and Communication Research tied for best representation. In
JPM Business Administration was best represented. The modal discipline in the most recent
issue of VSR is (mental) health research.

More than three quarters of articles in the most recent issues of the four journals are
empirical studies (see Table 1). In NML and VSR all articles reported on empirical data.

Data transparency

The findings on data transparency are encouraging. In the most recent issues of the five
journals publishing research on philanthropy the prevalence of data sharing is higher than in
the social sciences at large. Data transparency is highest in the most recent issue of NML.
20% of empirical articles (2/10) link to the source data analyzed, and another 20% (2/10)
refer to the name of the dataset without providing a bibliographic reference or link. Of the
empirical articles in VSR, 45% (5/11) provide details about the data analyzed. One article
provides a link to the source data, another to the name of the dataset, and three provide a
bibliographic reference. Among the empirical articles in NVSQ, 18% refer to the dataset with
a link to a repository. None of the other empirical articles refer to the data source with either
a reference to the data or the name of the dataset. None of the NVSQ articles that shared
data and code were accepted under the data and code submission policy for experiments.

Of the empirical articles in Voluntas, 13% (2/15) provide the name of the dataset analyzed,
and one article links to the source data.

Linking to original research materials is more rare. Around one fifth of articles in NML and
NVSQ do so. In JPM this is one out of seven. In Voluntas one out of fifteen. None of the
articles in VSR linked to original research materials. In NML and JPM articles, some of the
articles provided some of the instructions for participants in experiments in appendices.

Table 1. Data and methods transparency indicators in five nonprofit and philanthropy journals

	NVSQ	Voluntas	NML	JPM	VSR	All
Volume, issue	52 (3)	34 (2)	33 (4)	28 (2)	14 (1)	
number of articles	12	17	10	8	11	58
# authors	2.5	2.3	2.4	1.9	3.1	2.5
Most common discipline	PA	PA	PA/CR	BUS	HLTH	PA
% empirical	92	88	100	75	100	91
% name of dataset	0	13	20	0	9	9
+ reference to dataset	0	0	0	14	27	8
+ link to dataset	18	7	20	0	9	11
link to materials	18	7	20	14	0	11
link to code	9	0	0	0	0	2
median n	577	166	817	332	1367	611
% experiment	27	0	10	33	0	11
% cross-sectional	55	93	80	67	100	79
% longitudinal	9	13	10	0	0	8
Most common population	China	USA	USA	USA	UK	USA
% philanthropy outcome	50	13	20	50	36	28
% philanthropy input	13	13	0	0	18	8
% on giving behavior	100	100	100	33	100	83
% on giving intentions	0	0	0	67	0	17
% giving is observed	67	0	0	0	100	25
% giving is self-reported	33	100	100	100	0	75
Most common mechanism	psychological benefits	NA	costs	efficacy	costs	costs

Transparency and properties of methods

Sharing code is rare in the five journals. Only one of the 58 empirical articles (2%) provided access to the code for the analyses. This article appeared in NVSQ.

The median number of observations is much higher in NVSQ, NML and VSR than in Voluntas and JPM. This is likely to be the result of a higher proportion of studies reporting on administrative data and survey data in NVSQ, NML and VSR. Interviews are a more common source of data for articles in Voluntas. JPM publishes the highest proportion of experiments.

The most common research design in the five journals is cross-sectional (79%), with experiments and longitudinal designs having about equal representation. Between journals there are considerable differences. VSR only published cross-sectional designs. About 30% of articles in JPM and NVSQ are experiments.

The most commonly studied populations in the empirical articles are from the USA. The second most commonly studied populations are from China.

Aspects of philanthropy

36% of articles (19/53) are about philanthropy in the broadest sense - including studies of fundraising, volunteering, and blood donation. Philanthropy is more commonly studied as an outcome variable (28%) than as an input (8%). Most studies on philanthropy in the narrow sense of donating money measure giving behavior (83%, 10/12). JPM is an exception, with a majority of studies measuring giving intentions. Charitable giving is mostly self-reported (75%) rather than observed (25%). Among the mechanisms that drive charitable giving, the material costs of donating are most commonly studied, such as tax incentives for charity donors, perks for blood donors, and time pressure for volunteers.

Planned data collection

In future versions, we will determine:

a. whether an a priori statistical power analysis is reported;
b. what the statistical power is achieved by the number of observations and the research design;
c. which population the data was sampled from (e.g., general population, student sample, Mturk workers, Prolific participants, volunteers)
d. which sample selection strategy was used;
e. whether the dataset provides annotated metadata for reproducibility;
f. whether the data analyzed were collected through registers, surveys, or interviews;
g. whether behavior in experiments is incentivized or not;
h. how transparent data and methods are documented in the paper;
i. for hypothesis testing studies:
    i. what are the test statistics?

      ii.   are the reported p-values accurate / consistent with test statistics?

      iii.   is it a replication study?

Using the available data,

1. We will attempt to train a classifier for publications indicating whether the research is empirical or not, based on classifications of publications of VU Amsterdam (Bekkers, 2023). Only empirical research publications will be retained. These classifiers can be based on manually produced classifications of publications of VU Amsterdam.

2. We will develop rule-based algorithms to determine whether data presented is from a survey or not, is from a register or not, and contains an experiment or not. These classifiers can be based on manually produced classifications of publications of VU Amsterdam.

3. We will use rule-based algorithms from Serghiou et al. (2021) that classify whether data and code are shared or not.

4. We will attempt to train a classifier for publications indicating whether giving is self-reported or observed, based on classifications of publications included in the Science of Generosity review article by Bekkers & Wiepking (2007).

5. We will attempt to train classifiers for research quality indicators, based on data from Soderbergh et al. (2021) posted at https://osf.io/aj4zr/.

6. We will develop rule-based algorithms to determine in which academic disciplines researchers are employed and in which research has been published, based on classifications of publications included in the database for the Science of Generosity review article by Bekkers & Wiepking (2007).

7. We will develop rule-based algorithms to determine the country or countries in which the data were collected.

8. We will manually enter information about the data and methods used in research when extracting this information cannot be automated. This probably goes for the number of observations, the target population, test-statistics, and which mechanisms in giving behavior are studied.

Once the dataset has been constructed, the correlates of data and methods transparency can be studied.

Discussion

In the current study, we measure the transparency of philanthropy research among publications in journal articles. It is well-known that publications are a highly selective sample of all research conducted. Authors are less likely to submit research for publication when they expect the chances of acceptance to be lower. As a result, many studies never get published – a problem known as the 'file drawer'. Because of the reliance on published research, the current results are not representative for all philanthropy research.

The published record of research is selective with respect to the significance of findings and the degree of support for hypotheses. The restriction to published journal articles also implies that internal research within nonprofit organizations is not well-represented. Three of the authors in the current set of publications did not have an academic affiliation.

We focused on data and methods transparency. In a fully transparent workflow, not only data and methods are publicly available for review and reproduction. Fully transparent research also provides access to the reviews and revisions made in response to the reviews. It describes the sources of funding for research, who contributed what, the origins of research ideas, and the order in which the research was conducted: did the study start with a data set or with a research question, did theories or findings motivate hypotheses?

References

Altman, D. G. (1994). The scandal of poor medical research. *British Medical Journal*, 308(6924), 283-284. https://doi.org/10.1136/bmj.308.6924.283

Arel-Bundock, V., Briggs, R. C., Doucouliagos, H., Mendoza Aviña, M., & Stanley, T. D. (2022). Quantitative Political Science Research is Greatly Underpowered. I4R Discussion Paper Series, No. 6. http://hdl.handle.net/10419/265531

Ayeni, O., Dickson, L., Ignacy, T. A., & Thoma, A. (2012). A systematic review of power and sample size reporting in randomized controlled trials within plastic surgery. *Plastic and Reconstructive Surgery*, 130(1), 78e-86e. https://doi.org/10.1097/PRS.0b013e318254b1d1

Bekkers, R. (2023). Data Transparency at the Faculty of Social Sciences at VU Amsterdam.

Bekkers, R. (2022). Ten Meta Science Insights to Deal With the Credibility Crisis in the Social Sciences. December 3, 2022. https://renebekkers.wordpress.com/2022/12/03/ten-meta-science-insights-to-deal-with-the-credibility-crisis-in-the-social-sciences/

Bekkers, R. (2020). A Data Transparency Policy for Results Based on Experiments. July 8, 2020. https://renebekkers.wordpress.com/2020/07/08/a-data-transparency-policy-for-results-based-on-experiments/

Bekkers, R. (2019). A Conversation About Data Transparency. November 15, 2019. https://renebekkers.wordpress.com/2019/11/15/a-conversation-about-data-transparency/

Bekkers, R. & Wiepking, P. (2007). Generosity and Philanthropy: A Literature Review. Report commissioned by the John Templeton Foundation. https://renebekkers.files.wordpress.com/2011/08/bekkers_wiepking_generosity_07.pdf Supplementary Materials: https://understandingphilanthropy.wordpress.com/materials/

Bhandari, M., Richards, R.R., Sprague, S., & Schemitsch, E.H. (2002). The Quality of Reporting of Randomized Trials in The Journal of Bone and Joint Surgery from 1988 through 2000. *Journal of Bone & Joint Surgery*, 84(3): 388-396.

Brembs, B. (2018). Prestigious Science Journals Struggle to Reach Even Average Reliability. *Frontiers of Human Neuroscience*, 12 (37): 1-7. https://doi.org/10.3389/fnhum.2018.00037

Brodeur, A., Cook, N. & Heyes, A. (2022). We Need to Talk about Mechanical Turk: What 22,989 Hypothesis Tests Tell Us about Publication Bias and p-Hacking in Online Experiments. IZA Discussion Paper No 15478. https://www.econstor.eu/bitstream/10419/265699/1/dp15478.pdf

Brodeur, A., Cook, N. & Neisser, C. (2022). P-Hacking, Data Type and Data-Sharing Policy, IZA Discussion Papers, No. 15586. http://hdl.handle.net/10419/265807

Brodeur, A., Lé, M., Sangnier, M., & Zylberberg, Y. (2016). Star wars: The empirics strike back. *American Economic Journal: Applied Economics*, 8(1), 1-32. https://doi.org/10.1257/app.20150044

Buneman, P., Christie, G., Davies, J. A., Dimitrellou, R., Harding, S. D., Pawson, A. J., Sharman, J.L., & Wu, Y. (2020). Why data citation isn't working, and what to do about it. *Database*, 2020, baaa022. https://doi.org/10.1093/databa/baaa022

Christensen, G., Dafoe, A., Miguel, E., Moore, D.A., & Rose, A.K. (2019). A study of the impact of data sharing on article citations using journal policies as a natural experiment. *PLoS ONE* 14(12): e0225883. https://doi.org/10.1371/journal.pone.0225883

Colavizza, G., Hrynaszkiewicz, I., Staden, I., Whitaker, K., & McGillivray, B. (2020). The citation advantage of linking publications to research data. *PloS one*, 15(4), e0230416.

Culina, A., Van den Berg, I., Evans, S, & Sánchez-Tójar, A. (2020). Low availability of code in ecology: A call for urgent action. *PLoS Biology*, 18(7): e3000763. https://doi.org/10.1371/journal.pbio.3000763

Delios, A., Clemente, E. G., Wu, T., Tan, H., Wang, Y., Gordon, M., ... & Uhlmann, E. L. (2022). Examining the generalizability of research findings from archival data. *Proceedings of the National Academy of Sciences*, 119(30), e2120377119. https://doi.org/10.1073/pnas.2120377119

Franco, A., Malhotra, N., & Simonovits, G. (2014). Publication bias in the social sciences: Unlocking the file drawer. *Science*, 345(6203), 1502-1505. https://doi.org/10.1126/science.1255484

Gerber, A. S., & Malhotra, N. (2008a). Publication bias in empirical sociological research: Do arbitrary significance levels distort published results? *Sociological Methods & Research*, 37(1), 3-30. https://doi.org/10.1177/0049124108318973

Gerber, A., & Malhotra, N. (2008b). Do statistical reporting standards affect what is published? Publication bias in two leading political science journals. *Quarterly Journal of Political Science*, 3(3), 313-326. http://dx.doi.org/10.1561/100.00008024

Gopalakrishna, G., Ter Riet, G., Vink, G., Stoop, I., Wicherts, J. M., & Bouter, L. M. (2022a). Prevalence of questionable research practices, research misconduct and their potential explanatory factors: A survey among academic researchers in The Netherlands. *PloS one*, 17(2), e0263023. https://doi.org/10.1371/journal.pone.0263023

Gopalakrishna, G., Wicherts, J. M., Vink, G., Stoop, I., van den Akker, O. R., ter Riet, G., & Bouter, L. M. (2022b). Prevalence of responsible research practices among academics in The Netherlands. *F1000Research*, 11(471), 471. https://doi.org/10.12688/f1000research.110664.2

Hamilton, D.G. Hong, K., Fraser, H., Rowhani-Farid, A. Fidler, F., Page, M.J. (2023). Rates and predictors of data and code sharing in the medical and health sciences: A systematic review with meta-analysis of individual participant data. medRxiv 2023.03.22.23287607. https://doi.org/10.1101/2023.03.22.23287607

Hardwicke, T.E., Wallach, J.D., Kidwell, M.C., Bendixen, T., Crüwell, S. & Ioannidis, J.P.A. (2020). An empirical assessment of transparency and reproducibility-related research

practices in the social sciences (2014–2017). *Royal Society Open Science*, 7: 190806. https://doi.org/10.1098/rsos.190806

Ioannidis, J., Stanley, T. D., & Doucouliagos, H. (2017). The Power of Bias in Economics Research. *Economic Journal*, 127(605): F236-F265. https://doi.org/10.1111/ecoj.12461

Krawczyk, M. & Reuben, E. (2012). (Un)Available upon Request: Field Experiment on Researchers' Willingness to Share Supplementary Materials. *Accountability in Research*, 19: 175–186. https://doi.org/10.1080/08989621.2012.678688

Leonelli, S. (2022). Open Science and Epistemic Diversity: Friends or Foes? *Philosophy of Science*, 89, 991–1001. https://doi.org/10.1017/psa.2022.45

Ma, J., Ebeid, I.A., De Wit, A., Xu, M., Yang, Y., Bekkers, R. & Wiepking, P. (2021). Computational Social Science for Nonprofit Studies: Developing a Toolbox and Knowledge Base for the Field. *Voluntas, 34:* 52-63. https://osf.io/g9d8u/

Ma, J., et al., (2022). Knowledge Infrastructure for Philanthropic Studies. https://osf.io/nyt5x/

Ma, J., & Konrath, S. (2018). A Century of Nonprofit Studies: Scaling the Knowledge of the Field. *Voluntas,* 29, 1139–1158. https://doi.org/10.1007/s11266-018-00057-5

Maxwell, S. E. (2004). The Persistence of Underpowered Studies in Psychological Research: Causes, Consequences, and Remedies. *Psychological Methods, 9*(2), 147–163. https://doi.org/10.1037/1082-989X.9.2.147

McKay, B., Corson, A., Vinh, M. A., Jeyarajan, G., Tandon, C., Brooks, H., Hubley, J., & Carter, M. J. (2022). Low prevalence of a priori power analyses in motor behavior research. *Journal of Motor Learning and Development*, 11(1), 15-18. https://doi.org/10.1123/jmld.2022-0042

Menke, J., Roelandse, M., Ozyurt, B., Martone, M., & Bandrowski, A. (2020). The rigor and transparency index quality metric for assessing biological and medical science methods. iScience, 23(11): 101698. https://doi.org/10.1016/j.isci.2020.101698.

Menkveld, A. J., Dreber, A., Holzmeister, F., Huber, J., Johannesson, M., Kirchler, M., ... & Weitzel, U. (2021). Non-standard errors. https://dx.doi.org/10.2139/ssrn.3961574

Munafò, M. R., Nosek, B. A., Bishop, D. V., Button, K. S., Chambers, C. D., Percie du Sert, N., Simonsohn, U., Wagenmakers, E.-J., Ware, J. J. & Ioannidis, J. (2017). A manifesto for reproducible science. *Nature Human Behaviour*, 1(1), 1-9. https://doi.org/10.1038/s41562-016-0021

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. & Polanin, J.R. (2020). "statcheck": Automatically detect statistical reporting inconsistencies to increase reproducibility of meta-analyses. *Research Synthesis Methods*, 11(5): 574–579. https://doi.org/10.1002/jrsm.1408

Open Science Collaboration. (2015). Estimating the reproducibility of psychological science. *Science*, 349(6251). https://doi.org/10.1126/science.aac4716

Page, M. J., Nguyen, P. Y., Hamilton, D. G., Haddaway, N. R., Kanukula, R., Moher, D., & McKenzie, J. E. (2022). Data and code availability statements in systematic reviews of interventions were often missing or inaccurate: a content analysis. *Journal of Clinical Epidemiology*, 147, 1-10. https://doi.org/10.1016/j.jclinepi.2022.03.003

Peterson, D., Panofsky, A. (2023). Metascience as a Scientific Social Movement. *Minerva,* 61, 147–174. https://doi.org/10.1007/s11024-023-09490-3

Piwowar, H. A., & Vision, T. J. (2013). Data reuse and the open data citation advantage. *PeerJ*, 1, e175. https://doi.org/10.7717/peerj.175

Protzko, J., Krosnick, J., Nelson, L. D., Nosek, B. A., Axt, J., Berent, M., ... & Schooler, J. (2020). High replicability of newly-discovered social-behavioral findings is achievable. PsyArxiv preprint. https://psyarxiv.com/n2a9x/

Riedel, N., Kip, M., & Bobrov, E. (2020). ODDPub—a text-mining algorithm to detect data sharing in biomedical publications. *Data Science Journal*, 19:42. http://doi.org/10.5334/dsj-2020-042

Schäfer, T. & Schwarz, M.A. (2019). The Meaningfulness of Effect Sizes in Psychological Research: Differences Between Sub-Disciplines and the Impact of Potential Biases. *Frontiers of Psychology*, 10:813. https://doi.org/10.3389/fpsyg.2019.00813

Schroter, S., Black, N., Evans, S., Godlee, F., Osorio, L., & Smith, R. (2008). What errors do peer reviewers detect, and does training improve their ability to detect them? *Journal of the Royal Society of Medicine*, 101(10), 507-514. https://doi.org/10.1258/jrsm.2008.080062

Schulz, R., Barnett, A., Bernard, R., Brown, N. J., Byrne, J. A., Eckmann, P., ... & Weissgerber, T. L. (2022). Is the future of peer review automated? *BMC Research Notes*, 15(1), 1-5. https://doi.org/10.1186/s13104-022-06080-6

Serghiou, S., Contopoulos-Ioannidis, D.G., Boyack, K.W., Riedel, N., Wallach, J.D., & Ioannidis, J.P.A. (2021). Assessment of transparency indicators across the biomedical literature: How open is open? *PLoS Biology*, 19(3): e3001107. https://doi.org/10.1371/journal.pbio.3001107

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*, 1359-1366. https://doi.org/10.1177/0956797611417632

Simonsohn, U., Nelson, L.D. & Simmons, J.P. (2014a). P-Curve: A Key to the File Drawer. *Journal of Experimental Psychology: General*, 143 (2): 534–547. https://doi.org/10.1037/a0033242

Simonsohn, U., Nelson, L. D., & Simmons, J. P. (2014b). p-curve and effect size: Correcting for publication bias using only significant results. *Perspectives on Psychological Science*, 9(6), 666-681. https://doi.org/10.1177/1745691614553988

Smaldino, P. E., Turner, M. A., & Contreras Kallens, P. A. (2019). Open science and modified funding lotteries can impede the natural selection of bad science. *Royal Society Open Science*, 6(7), 190194. https://doi.org/10.1098/rsos.190194

Smith, D.H. (1994). Determinants of Voluntary Association Participation and Volunteering: A Literature Review. *Nonprofit and Voluntary Sector Quarterly*, 23 (3): 243-263. https://doi.org/10.1177%2F089976409402300305

Smith, R. (2006). Peer review: a flawed process at the heart of science and journals. *Journal of the Royal Society of Medicine*, 99: 178–182. https://dx.doi.org/10.1177/014107680609900414

Smith, R. (2010). Classical peer review: an empty gun. *Breast Cancer Research*, 12 (Suppl 4), S13. https://doi.org/10.1186/bcr2742

Soderberg, C. K., Errington, T. M., Schiavone, S. R., Bottesini, J., Thorn, F. S., Vazire, S., … & Nosek, B. A. (2021). Initial evidence of research quality of registered reports compared with the standard publishing model. *Nature Human Behaviour*, 5(8), 990-997. https://doi.org/10.1038/s41562-021-01142-4

Song, H., Markowitz, D. M., & Taylor, S. H. (2022). Trusting on the shoulders of open giants? Open science increases trust in science for the public and academics. *Journal of Communication*, 72(4), 497-510. https://doi.org/10.1093/joc/jqac017

Stodden, V., Seiler, J. & Ma, Z. (2018). An empirical analysis of journal policy effectiveness for computational reproducibility. *Proceedings of the National Academy of Sciences*, 115(11): 2584-2589. https://doi.org/10.1073/pnas.1708290115

Tedersoo, L., Küngas, R., Oras, E., Köster, K., Eenmaa, H., Leijen, Ä., Pedaste, M., Raju, M., Astapova, A., Lukner, H., Kogermann, K. & Sepp, T. (2021). Data sharing practices and data availability upon request differ across scientific disciplines. *Scientific Data*, 8(1), 1-11. https://doi.org/10.1038/s41597-021-00981-0

Vermeulen, I., Beukeboom, C. J., Batenburg, A., Avramiea, A., Stoyanov, D., van de Velde, B., & Oegema, D. (2015). Blinded by the light: How a focus on statistical "significance" may cause p-value misreporting and an excess of p-values just below. 05 in communication science. *Communication Methods and Measures*, 9(4), 253-279. https://doi.org/10.1080/19312458.2015.1096333

Wicherts, J. M., Borsboom, D., Kats, J., & Molenaar, D. (2006). The poor availability of psychological research data for reanalysis. *American Psychologist*, 61(7), 726–728. https://doi.org/10.1037/0003-066X.61.7.726

Wilhelm, M. O. (2006). New data on charitable giving in the PSID. *Economics Letters*, 92(1), 26-31. https://doi.org/10.1016/j.econlet.2006.01.009

Zavalis, E.A., & Ioannidis, J.P.A. (2022) A meta-epidemiological assessment of transparency indicators of infectious disease models. *PLoS ONE* 17(10): e0275380. https://doi.org/10.1371/journal.pone.0275380