UNIVERSITY of CALIFORNIA PRESS | Collabra

Vanpaemel, W et al 2015 Are We Wasting a Good Crisis? The Availability of Psychological Research Data after the Storm. *Collabra,* 1(1): 3, pp. 1–5, DOI: http://dx.doi.org/10.1525/collabra.13

ORIGINAL RESEARCH REPORT

Are We Wasting a Good Crisis? The Availability of Psychological Research Data after the Storm

Wolf Vanpaemel*, Maarten Vermorgen*, Leen Deriemaecker* and Gert Storms*

To study the availability of psychological research data, we requested data from 394 papers, published in all issues of four APA journals in 2012. We found that 38% of the researchers sent their data immediately or after reminders. These findings are in line with estimates of the willingness to share data in psychology from the recent or remote past. Although the recent crisis of confidence that shook psychology has highlighted the importance of open research practices, and technical developments have greatly facilitated data sharing, our findings make clear that psychology is nowhere close to being an open science.

Keywords: Data-sharing; open science; research integrity; questionable research practices; replicability crisis

Introduction

Making research data publicly available has several benefits for the advancement of science. Data sharing facilitates, among others things, verification, replication, robustness check, reuse, follow up, and meta-analysis, and thus leads to a more reliable, less wasteful, less costly, more efficient and overall better science, as well as to an increased confidence in research findings and a greater trust in science [4, 12, 16, 19, 20, 22]. From the perspective of the individual scientist, advantages of sharing one's data include prevention of data loss and an increased visibility and citability [13]. Data sharing does not only accelerate scientific progress, but as publicly funded data can be considered a public good, sharing such data is sometimes regarded as a moral obligation [9].

The importance of openness of data has been recognized and highlighted by several learned societies, research institutes, and leading journals. For example, a condition of acceptance in a *Nature* journal is that authors "make materials, data, code, and associated protocols promptly available to readers without undue qualifications" (http://www.nature.com/authors/policies/availability.html). Similarly, *Science* requires that "[a]ll data necessary to understand, assess, and extend the conclusions of the manuscript must be available to any reader of *Science*. . . . After publication, all reasonable requests for data and materials must be fulfilled" (http://www.sciencemag.org/about/authors/prep/gen_info.xhtml).

The poor availability of psychological research data

The importance of open research practices is acknowledged in psychology as well. For example, the *Ethical Principles of Psychologists and Code of Conduct* from the *American Psychological Association* (APA) unambiguously state that "[a]fter research results are published, psychologists do not withhold the data on which their conclusions are based from other competent professionals who seek to verify the substantive claims through reanalysis and who intend to use such data only for that purpose" ([2], p. 234). When publishing in an APA journal, all authors whose research involved human participants or animal subjects are required to certify compliance with these ethical principles.

Despite the considerable scientific benefits of open data, psychological research data are rarely available. Upon contacting 37 authors who published in APA journals between 1959 and 1961, Wolins found that 9 shared their data under a reasonable set of conditions (24%) [26]. About a decade later, Craig and Reese obtained 20 data sets or summary analyses out of 53 requests (38%), with rates for individual journals ranging from 30% to 75% [3]. When Wicherts, Borsboom, Kats and Molenaar [25] contacted the corresponding authors of every article in the last two 2004 issues of four APA journals, only 38 of the 141 contacted authors sent the raw data (27%).

Since the study by Wicherts et al. [25], psychology as a science has gone through particularly turbulent times [11]. It has become increasingly clear that questionable research practices (QRPs) may be disturbingly common (e.g., [5, 17]), leading to a low replicability [27] and a decreased

* Faculty of Psychology and Educational Sciences, University of Leuven, Belgium
wolf.vanpaemel@kuleuven.be, maarten.vermorgen@student.kuleuven.be, leen.deriemaecker@student.kuleuven.be, gert.storms@kuleuven.be

Corresponding author: Wolf Vanpaemel

confidence in research findings. One reaction to this state of emergency has been a renewed call for open data, culminating in the foundation of the Center for Open Science, which hosts the Open Science Framework (OSF, osf.io) for data archiving and sharing [10]. In this paper, we evaluate whether the willingness to open up research data has increased to more acceptable levels.

Materials and Methods

The data we requested will be used to investigate whether a Bayesian analysis results in a different conclusion compared to a traditional (frequentist) analysis (see also [6, 23]). Adopting the Bayesian framework for data analysis is, besides embracing open research, another recommendation in response to the crisis of confidence. Advantages of the Bayesian approach to inference include strong conceptual appeal, intuitive interpretations, intuitive account of uncertainty, coherence, limitless flexibility, validity for small sample sizes, ability to incorporate prior knowledge, ability to quantify evidence in favor of the null hypothesis, and the ability to monitor evidence as the data come in (e.g., [7]).

We considered all papers published in 2012 in the following four APA journals: *Emotion* (155), *Experimental and Clinical Psychopharmacology* (56), *Journal of Abnormal Psychology* (98) and *Psychology and Aging* (115), which, respectively, represent the research domains of personality and social psychology, experimental psychology, clinical psychology and developmental psychology, totaling 424 papers. We requested data from papers published in an APA journal because the authors have certified compliance with the *APA Ethics Code* mentioned above, and are thus expected to share their data for reanalysis "provided that the confidentiality of the participants can be protected and unless legal rights concerning proprietary data preclude their release" ([2], p. 234).

From these 424 papers, we selected papers with at least one *p*-value and for which a Bayesian reanalysis seemed feasible. There were 25 papers without *p*-values, and five papers for which no easy Bayesian alternative seemed available. (For four of these five papers, the difficulty of doing a Bayesian analysis was noted only after we contacted the authors with a request for data. However, we will treat these as unrequested data sets. None of these four authors shared their data.) Our final selection included 394 papers, making the current study the largest published study to date on willingness to share in psychology.

In November 2013, MV and LD started to approach the corresponding authors of the remaining 394 papers, using a standardized email which can be found on https://osf.io/bqg6v/. When the email address of the corresponding author proved invalid, we first searched the internet for an updated email address. If we were unable to track down a working email address of the corresponding author, another author (usually the first or the last) was contacted. For all 394 papers, we were able to reach at least one author. If contacted authors had additional questions, our replies were standardized as much as possible. For example, if an author asked which data format we preferred, we always replied the same way: "*You can send us the data in any format you have, if we cannot convert them into a format that we can import in R we will get back to you*". Following a significant time lapse (ranging from weeks to months), MV and LD sent a reminder to the authors who did not respond to our initial request and to those who had replied to our email without sending their data. A reminder was also sent to authors who promised to share their data but had failed to do so after a considerable period of time (even when they had already received a reminder earlier on).

This study has been approved by the Ethics Committee of the Faculty of Psychology and Educational Sciences of the University of Leuven, under the restriction that we would not disclose who shared their data and who did not (see also [24]). Making the response to our data request public would constitute a breach of confidentiality, so not sharing the data seems in line with the *APA Ethics Code* mentioned above, as it serves to protect the confidentiality of the participants (though see [21], for a different perspective).

We were exempted to obtain informed consents from the contacted authors, because it was both impossible and unnecessary, as all authors had certified compliance with the APA ethical principles, which include clear stipulations on data sharing.

Results

Table 1 shows the percentages of reactions in the response categories used by Wicherts et al. [25], for each journal separately, as well as aggregated across journals. (In three cases, authors were willing to share their data, but under very strict conditions, such as co-authorship or payment. As we deemed these conditions unreasonable, we refused to accept these data and classified these authors as unwilling to share.)

The good news is that the overall response rate has gone up since the study by Wicherts et al. [25]. The bad news is that the response rate is nowhere close to 100%. Despite the growing awareness of QRPs in psychology, the increased emphasis on open data, and several initiatives facilitating data storing and sharing, we ended up with 148 positive responses only (38%).

There are marked differences between the journals. The highest sharing rate was found in *Emotion*, where 72 of the 149 contacted authors shared their data (48 %). The lowest willingness to share was found for *Journal of Abnormal Psychology*, with only 22 of the 89 contacted authors sharing their data (25%). In the remaining two journals, the response rates were 30% (16 out of 53) and 37% (38 out of 103), for *Experimental and Clinical Psychopharmacology* and *Psychology and Aging*, respectively.

We can of course only guess why the 161 contacted authors who did not reply to our e-mails preferred not to share their data. The responses of the 69 contacted authors who took the time to explain why they preferred not to share their data provide an interesting window on reasons for turning down our request. To our surprise, some authors are apparently willing to share, but have no easy access to their own data or have lost their data altogether,

Type of response	Emotion	Experimental and Clinical Psychopharmacology	Journal of Abnormal Psychology	Psychology and Aging	Overall
No reply	36%	55%	38%	44%	41%
Refused/unable to share data	10%	15%	31%	17%	18%
No data despite promise	6%	0%	6%	2%	4%
Data shared after reminder	21%	15%	7%	17%	16%
Data shared after first request	28%	15%	18%	20%	22%

Table 1: Percentages of Reactions in Different Response Categories.

due to computer crashes or collaborators having left the university. Many authors cite a lack of time as a reason not to share, and note that sharing their data would take too much effort, which is probably due to poor documentation and storage practices. Several authors refer to strict local privacy or data sharing policies and regulations, and one to unspecified security issues. Further, the fact that we did not offer monetary compensation was a reason not to share for some. With one author, our request came too late, as others had already started to perform a Bayesian reanalysis. Finally, some authors are clear and to the point, and were simply not interested. These reasons are likely to be distorted by social desirability. Not a single author mentioned reasons reflecting what Rouder [14] terms *professional vulnerability*. Raising the research curtains could potentially lead to uncovering mistakes, which in turn might lead to losing face and, in case of a retraction, a paper.

Discussion

Approximately two thirds of the authors did not share their data. Even in the journal with the highest sharing rate, less than half of the contacted authors practiced open research. Apparently, the crisis of confidence has not been sufficient to bring about a high willingness to share research data. Although the sharing rate has increased as compared to the study by Wicherts et al. [25], our findings are worrisome.

Even if we had observed a response rate of 100%, the situation would still be far from ideal. First, in an ideal implementation of data sharing, our request is superfluous. Researchers would make their data available without being prompted by any request for sharing, either upon publication of their paper or even immediately when the data come in – a practice referred to as *born-open* [14]. There are many third party public repositories available for data sharing, such as the Dataverse project (dataverse.org), Figshare (figshare.com), the Open Science Framework (osf.io) or GitHub (github.com). It is telling that in our study, only four authors shared their data by referring to an online repository where the data were publicly available.

Second, even if all research data were spontaneously made publicly available, a lot of research output is still hidden from scrutiny and unavailable for re-use. Ideally, researchers should not only make the raw and processed data available, but should also routinely share the research materials used in the study (i.e., the stimuli, the experimental instructions, and so on) and the code used in the processing and analysis of the data (see [18], and the associated OSF project page on https://osf.io/ivfu6/ for an example). With the pre-computer technological limitations gone, it strikes us as anachronistic to consider a dense research report the sole end product of a study.

Given the current poor availability of data, it is unlikely that the spontaneous public dissemination of data, material and code will happen naturally, or anytime soon. The strategy of celebrating the virtues of open research (e.g., [9, 10]) has not yet brought the anticipated success. Another strategy might involve convincing journals to adopt policies on open practices. But also this mechanism is probably not enough, as recent studies found that adherence to the data access instructions issued by the journals is low [1, 15]. One promising initiative is the recently launched Peer Reviewers' Openness Initiative ([8]; see also http://opennessinitiative.org/). Starting 1 June, 2016, signatories of the Initiative will withhold comprehensive review if data and research materials are not made publicly available on a comply or explain basis (note that in the present case, we explained why we could not comply with the sharing default). We hope that initiatives like these will lead to an updated publication standard, in which papers that do not share the data, the materials, and the code are considered as incomplete as papers that report their hypothesis and conclusion, but not the necessary statistical analyses.

Competing Interests

The authors declare that they have no competing interests.

References

1. **Alsheikh-Ali, AA, Qureshi, W, Al-Mallah, MH** and **Ioannidis, JPA.** Public availability of published research data in high-impact journals. *PLoS ONE* 2011; 6: e24357. DOI: http://dx.doi.org/10.1371/journal.pone.0024357
2. **American Psychological Association.** *Publication Manual of the American Psychological Association* (6th ed.). Washington, DC: American Psychological Association. 2010.
3. **Craig, JR** and **Reese, SC.** Retention of raw data: A problem revisited. *American Psychologist.* 1973; 28: 723. DOI: http://dx.doi.org/10.1037/h0035667
4. **Hrynaszkiewicz, I.** A call for BMC Research Notes contributions promoting best practice in data standardization, sharing, and publication. *BioMed*

Central Research Notes. 2010; 3: 235. DOI: http://dx.doi.org/10.1186/1756-0500-3-235

5. **John, L, Loewenstein, G** and **Prelec, D.** Measuring the prevalence of questionable research practices with incentives for truth-telling. *Psychological Science*. 2012; 23: 524–532. DOI: http://dx.doi.org/10.1177/0956797611430953
6. **Johnson, VE.** Revised standards for statistical evidence. *Proceedings of the National Academy of Sciences*. 2013; 110: 19313–19317. DOI: http://dx.doi.org/10.1073/pnas.1313476110
7. **Lee, MD** and **Wagenmakers, E-J.** Bayesian statistical inference in psychology: Comment on Trafimow (2003). *Psychological Review*. 2005; 112, 662–668. DOI: http://dx.doi.org/10.1037/0033-295X.112.3.662
8. **Morey, RD, Chambers, CD, Etchells, PJ, Harris, CR, Hoekstra, R, Lakens, D, Lewandowsky, S, Morey, CC, Newman, DP, Schönbrodt, F, Vanpaemel, W, Wagenmakers, E-J** and **Zwaan, RA.** The Peer Reviewers' Openness Initiative: Incentivising open research practices through peer review. Submitted.
9. **Nosek, BA** and **Bar-Anan, Y.** Scientific Utopia: I. Opening scientific communication. *Psychological Inquiry*. 2012; 23: 217–243. DOI: http://dx.doi.org/10.1080/1047840X.2012.692215
10. **Nosek, BA, Spies, JR** and **Motyl, M.** Scientific Utopia II. Restructuring incentives and practices to promote truth over publishability. *Perspectives on Psychological Science*. 2012; 7: 615–631. DOI: http://dx.doi.org/10.1177/1745691612459058
11. **Pashler, H** and **Wagenmakers, EJ.** Editors' Introduction to the Special Section on Replicability in Psychological Science: A Crisis of Confidence? *Perspectives on Psychological Science*. 2012; 7: 528–530. DOI: http://dx.doi.org/10.1177/1745691612465253
12. **Piwowar, HA.** Who shares? Who doesn't? Factors associated with openly archiving raw research data. *PLoS ONE*. 2011; 6: e18657. DOI: http://dx.doi.org/10.1371/journal.pone.0018657. PMid: 21765886; PMCid: PMC3135593.
13. **Piwowar, HA, Day, RS** and **Fridsma, DB.** Sharing detailed research data is associated with increased citation rate. *PLoS ONE*. 2007; 2: e308. DOI: http://dx.doi.org/10.1371/journal.pone.0000308
14. **Rouder, JN.** The what, why, and how of born-open data. *Behavior Research Methods*. in press. PMid: 25271090.
15. **Savage, CJ** and **Vickers, AJ.** Empirical study of data sharing by authors publishing in PLoS journals. *PLoS ONE*. 2009; 4: e70–78. DOI: http://dx.doi.org/10.1371/journal.pone.0007078
16. **Schofield, PN, Bubela, T, Weaver, T, Portilla, L, Brown, SD, Hancock, JM, Einhorn, D, Tocchini-Valentini, G, Hrabe de Angelis, M, Rosenthal, N.** and **CASIMIR Rome Meeting participants.** Post-publication sharing of data and tools. *Nature*. 2009; 461: 171–173. DOI: http://dx.doi.org/10.1038/461171a
17. **Simmons, JP, Nelson, LD** and **Simonsohn, U.** False-positive psychology: Undisclosed flexibility in data collection and analysis allows presenting anything as significant. *Psychological Science*. 2011; 22: 1359–1366. DOI: http://dx.doi.org/10.1177/0956797611417632
18. **Steegen, S, Dewitte, L, Tuerlinckx, F** and **Vanpaemel, W.** Measuring the crowd within again: A pre-registered replication study. *Frontiers in Psychology*. 2014; 5: 786. DOI: http://dx.doi.org/10.3389/fpsyg.2014.00786
19. **Stock, WA** and **Kulhavy, RW.** Reporting primary data in scientific articles: Technical solutions to a perennial problem. *American Psychologist*. 1989; 44: 741–742. DOI: http://dx.doi.org/10.1037/0003-066X.44.4.741
20. **Stodden, V.** Trust your science? Open your data and code [Blog post]. 2011, July 1. Retrieved from http://magazine.amstat.org/blog/2011/07/01/trust-your-science/
21. **Tractenberg, R.** The "responsible conduct of research" is not limited to properly obtained consent [Blog post]. 2011, November 2. Retrieved from http://blogs.plos.org/everyone/2011/11/02/the-responsible-conduct-of-research-is-not-limited-to-properly-obtained-consent/.
22. **Vision, TJ.** Open data and the social contract of scientific publishing. *BioScience*. 2010; 60: 330–331. DOI: http://dx.doi.org/10.1525/bio.2010.60.5.2
23. **Wetzels, R, Matzke, D, Lee, MD, Rouder, JN, Iverson, GJ** and **Wagenmakers, E-J.** Statistical evidence in experimental psychology: An empirical comparison using 855 t tests. *Perspectives on Psychological Science*. 2011; 6: 291–298. DOI: http://dx.doi.org/10.1177/1745691611406923
24. **Wicherts, JM, Bakker, M** and **Molenaar, D.** Willingness to share research data is related to the strength of the evidence and the quality of reporting of statistical results. *PLoS ONE*. 2011; 6: e26828. DOI: http://dx.doi.org/10.1371/journal.pone.0026828
25. **Wicherts, JM, Borsboom, D Kats, J** and **Molenaar, D.** The poor availability of psychological research data for reanalysis. *American Psychologist*. 2006; 61: 726–728. DOI: http://dx.doi.org/10.1037/0003-066X.61.7.726
26. **Wolins, LL.** Responsibility for raw data. *American Psychologist*. 1962; 17: 657–658. DOI: http://dx.doi.org/10.1037/h0038819
27. **Yong, E.** Bad copy. *Nature*. 2012; 485: 298–300. DOI: http://dx.doi.org/10.1038/485298a

Peer review comments
The author(s) of this paper chose the Open Review option, and the peer review comments are available at: http://dx.doi.org/10.1525/collabra.13.opr

How to cite this article: Vanpaemel, W, Vermorgen, M, Deriemaecker, L and Storms, G 2015 Are We Wasting a Good Crisis? The Availability of Psychological Research Data after the Storm. *Collabra,* 1(1): 3, pp. 1–5, DOI: http://dx.doi.org/10.1525/collabra.13

Submitted: 04 April 2015 **Accepted:** 27 August 2015 **Published:** 09 October 2015

UNIVERSITY of CALIFORNIA PRESS | Collabra

Collabra is a peer-reviewed open access journal published by University of California Press.

OPEN ACCESS