

**REGISTERED REPORTS AS A METHOD
TO INCREASE CREDIBILITY OF SCIENCE
– EXPERIMENTAL STUDY AMONG PSYCHOLOGY
STUDENTS**

PIOTR SOROKOWSKI

Institute of Psychology, University of Wrocław
Dawida 1, Wrocław, Poland

E-mail address: sorkokwskipoitr@yahoo.co.uk

ORCID: <https://orcid.org/0000-0001-9225-9965>

AGATA GROYECKA

Institute of Psychology, University of Wrocław
Dawida 1, Wrocław, Poland

E-mail address: agata.groyecka@gmail.com

ORCID: <https://orcid.org/0000-0002-1932-4828>

KAMIL BŁASZCZYŃSKI

Faculty of Psychology, University of Wrocław
Ul. Dawida 1, Wrocław, Poland

E-mail address: kamil.b.1988@wp.pl

ORCID: <https://orcid.org/0000-0003-4896-8371>

TOMASZ FRĄCKOWIAK

Faculty of Psychology, University of Wrocław
Ul. Dawida 1, Wrocław, Poland

E-mail address: tomasz.frackowiak@uwr.edu.pl

ORCID: <https://orcid.org/0000-0002-0244-2879>

ALEKSANDER KOBYLAREK

Faculty of Pedagogy, University of Wrocław
Ul. Dawida 1, Wrocław, Poland

E-mail address: aleksander.kobylarek@gmail.com

ORCID: <https://orcid.org/0000-0002-4562-9035>



ABSTRACT

Aims. The modern scientific publishing system suffers from many problems, amongst which one of the most important is the pressure to publish positive results. A potentially simple way to mitigate publication bias are reviews of manuscript, prior to the data collection and analyses, as well as results presentation and discussion (i.e. Registered Reports).

Methods. To test this prediction we conducted a quasi-experiment: two groups of students ($n=38$), as a part of their academic classes, were asked to design and conduct research projects. They were divided in two groups – Registered Reports- and control group. In both groups students have been encouraged to publish their papers in a local scientific journal.

Results. Analyses revealed significant differences in p levels between groups, suggesting more reliable scores for Registered Reports group.

Conclusions. Our study partially confirmed the stated hypothesis and suggested, that registered reports might mitigate publication bias. Future didactic and research recommendations are advised.

Key words: registered reports, methodology of science, reviews, p-hacking, publication bias

INTRODUCTION

The modern scientific publishing system suffers from many problems (e.g. Sorokowski et al., 2017), amongst which one of the most important is the pressure to publish positive results (Fanelli, 2010; Zwaan et al. 2017). For example, results published in psychology or psychiatry journals are in over 90% of the cases significant (Fanelli, 2010; Rosenthal, 1979; Sterling et al., 1995). Such high percentage of positive cases may suggest that a) researchers publish only significant results b) researchers publish dishonest results. The tendency to publish mostly significant results was shown in medical (Begley & Ellis, 2012), social science (Camerer et al., 2018) or psychological (Open Science Collaboration, 2015) studies. A report published by Open Science Collaboration project (2015) in which only 36% of psychological studies published in the top-notch scientific journals had been positively replicated can be seen as challenging for their reliability. Furthermore, the mean effect size in the replicated studies was two times lower than in the original ones (0.197 versus 0.403).

Although few researchers directly fabricate data, or at least admit to doing so (John, Loewenstein, & Prelec, 2012), many use other deceptive methods (i.e. “p-hacking” or “selective reporting”) to increase the publishability of their studies (e.g. John et al., 2012; Head, Holman, Lanfear, Kahn, & Jennions, 2015; Lakens, 2015; Simonsohn, Nelson, & Simmons, 2014b; Stahl & Pickles, 2018; Wicherts et al., 2016). For example, over 50% of the respondents participated in John, Loewenstein and Prelec (2012) study committed to have engaged in half of the study practices listed by the authors (i.e. “Failing to report all dependent measures”, “Collecting more data after seeing whether results were significant,” etc.).

Why do psychologists and other scientists engage in these questionable practices? Some authors tried to find predictors of publishing misconduct (Jackson, Levine, Furnham, & Burr, 2002; McCabe, Treviño, & Butterfield, 2001). However, in our opinion people are driven to use questionable research practices because their scientific quality (including salaries, promotion, etc.) is heavily judged based on their publication record (Everett & Earp, 2015; Fanelli, 2010; Powell, 2016). As discussed earlier, published papers contain mostly significant results confirming the presented hypothesis, (e.g. Leimu & Koricheva, 2004; Fanelli, 2010; Rosenthal, 1979) and replications are relatively seldom published (Makel, Plucker, & Hegarty, 2012; Zwaan et al., 2017). Consequently, researchers who wish to publish their studies in respected journals, try to present only statistically significant results.

Authors suggest that one possible way of reducing scientists' motivations to avoiding publishing false positive outcomes is to support so called Registered Reports (RR) (Munafò et al., 2017; Nosek, Spies, & Motyl, 2012; Nosek et al., 2018). The main idea of this initiative is that the success of their paper should be determined by the merit of the research question, the a priori derivation of sound hypothesis and the rigour and adequacy of methodology used to test these hypotheses, rather than the mere statistical significance of the results (Probst & Hagger, 2015). In other words, if there is a preliminary project of study which is well planned (it is reviewed), the authors have a guarantee of publishing results no matter what the outcomes (Nosek & Lakens, 2014; Nosek et al., 2018). A second review checks only the correctness of research implementation (See Figure 1.) To date, about 100 journals have adopted Registered Reports as a part of a regular submission option. For the full list of participating journals see: <https://cos.io/rr/#journals>. However, the majority of the pre-registered studies replicated previous findings (Nosek, & Lakens, 2014).

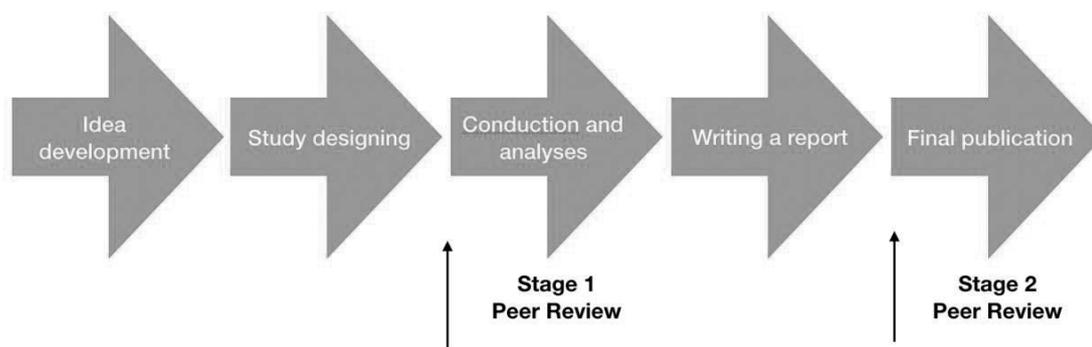


Figure 1.

Registered Reports publishing process.

Source: <https://osf.io/8mpji/wiki/home/>.

Although the number of published Registered Reports is high, it is not clear to what extent they increase data credibility. Studies have replicated the well-known psychological phenomena (Begley & Ellis, 2012; Camerer et al., 2018; Open Science Collaboration, 2015) but there is no evidence in regard to more recently reported findings. Therefore there is a debate about the degree to which RR may influence publishing research data in the future (Erlich, 2018; Findley, et al., 2016; da Silva, 2016).

Here, we report a quasi-experiment that refers to the aforementioned lack of data (e.g. Morgan, 2000) that refers to the aforementioned lack of data. Two groups of students were asked to create research projects as a part of their academic classes. One group was assigned to the Registered Reports condition, while the second used the standard procedure (control group). In both groups, students have been encouraged to publish their papers in a local scientific journal.

MATERIAL AND METHODS

Participants

38 students participated in the study (2nd year of psychology) (the total number of participants comprised 40 students, although 2 students did not finish their projects in time). Students took part in class titled "Research Project" during which they were instructed to conduct a research project on topic that has been indicated by the lecturer.

Methods

The study protocol and consent procedure received ethical approval from the Institutional Review Board (IRB).

Students were divided into two groups (with 15 class sessions each) that were led by the same lecturer. The first group was led by the use of standard procedure. The second group has been led with the use of registered reports standards (see below). Students were not informed of this fact before their signed up for the course. Students were not assigned randomly to the groups (quasi-experiment), although they were signed up to the same lecturer and they were on the same specialisation (psychology, full-time studies). The topic of the research project was: "Correlation between the number of used emoticons and psychological masculinity and femininity" (i.e.: Oleszkiewicz et al., 2017a,b; Bem, 1974). The topic uniformity was used to create equal research conditions in two class groups.

Each of the students taking part in the quasi-experiments was required to: 1) design and conduct the research, 2) perform the statistical analysis and 3) write brief report from the conducted study. Students have been informed during the first meeting, that the best report from each group will be published in a students' Journal *Ogrody Nauk i Sztuk* (<https://ogrodynauk.pl/index.php/onis>) and the authors of best three reports will receive a small reward (thermo cups).

Each of the students had to acquire data from 50 female and 50 male respondents. There were no strict procedures how students had to gather data about masculinity and femininity – although there is a very popular adaptation of Bem's questionnaire (Kuczyńska, 1992) – and about number of the emoticons (e.g. via survey, text analysis of the respondents etc.). There were also no demographical and age restrictions of the respondents, according to which students could include someone in their study sample.

Students could test an unlimited number of hypotheses and they could also include new variables. However, they were obliged to present general correlation between the number of emoticons and about masculinity and femininity of the respondents. Students have been informed that they do not need to attach any raw data.

The lecturer presented identical content for the students during classes. Those included, for example: correlation analysis, social science methodology, research ethics or presentation of data in form of short report. One class meeting in both groups was conducted by the editor of *Ogrody Nauk i Sztuk* journal who encouraged students to publish results from their study in the journal, pointing out the prestige that is connected with such a practice. However, he did not suggest that those results need to be positive or support any hypothesis.

Two groups (Registered Reports vs control) varied in the procedure of passing the course. Students from the control group were evaluated after handing over the final version of their work. Students from the registered report group were informed that 1.5 months before the ending of the semester, they are obliged to hand to the lecturer the first part of the project. Project should contain 1) preface 2) detailed description of the method and 3) anticipated description of the research sample. Students were informed, that their work will be evaluated after introducing these materials. The articles designated to be published in the *Ogrody Nauk i Sztuk* journal were chosen on the same basis in both groups. Students were informed that the grade received after the first stage is final, although the final acceptance is done after passing the second part of the report that contained: 4) results and 5) discussion section.

RESULTS

We compared the results of Registered Reports ($n = 20$) and control groups ($n = 18$) in regard to a) correlation coefficient (r value) between femininity/masculinity and number of emoticons; b) statistical significance (p value) of the correlation between femininity/masculinity and number of emoticons.

Our variables had non-normal distribution (Shapiro-Wilk test) which suggested further usage of non-parametric tests (Mann-Whitney test).

There was no significant difference in correlation "femininity – number of emoticons" between registered reports (Mdn=0.21, SD=0.15) and control (Mdn=0.24, SD=0.15) groups ($U=169$, $p=.76$). These results suggested that students from compared groups did not differ in strengths of reported associa-

tions between femininity and number of emoticons. Additionally, there was no significant difference in statistical significance presented by students from the two groups (registered reports: $Mdn=0.06$, $SD=0.28$; control group: $Mdn=0.08$, $SD=0.17$; $U=165$, $p=.66$). These results suggested that students from compared groups did not differ in level of reported p value.

Additionally, there was no significant difference in correlation “masculinity - number of emoticons” between Registered Reports ($Mdn=0.08$, $SD=0.08$) and control ($Mdn=0.11$, $SD=0.20$) groups ($U=173$, $p=.84$). These results suggested that students from compared groups did not differ in the level of reported correlations between masculinity and number of emoticons. However, there was significant difference in p level presented by students from two the groups (registered reports: $Mdn=0.18$, $SD=0.23$; control group: $Mdn=0.35$, $SD=0.22$; $U=104.5$, $p=.026$, $d_{Cohen}=.767$). These results suggested that students from Registered Reports group reported significantly higher (less significant) p value than students from control group.

DISCUSSION AND CONCLUSIONS

The presented study gave the first empirical evidence to verify earlier stated hypothesis which assumed that Registered Reports publishing method increases credibility of the published results. Although the results are unambiguous (no significant differences between groups in case of femininity), and should be treated with caution, it can be stated that registered reports may give possibility to generate more reliable results in certain circumstances. Students in the Registered Reports groups presented significantly higher p values for association between masculinity and emoticons, which suggests that in their studies results were more often less supportive for the hypothesis and not statistically significant. Such results in comparison to the control group, raise the question of: what influenced the results to differ in level of statistical significance in these two groups? Their motivation to p -hack might have differed depending on the condition. At the same time, our findings do not indicate clearly that students from the control group manipulate their data.

Students recruited to this study were not professional scientists. Thanks to that, results of this study may suggest that pressure to publish positive results is visibly active already in minds of second-year students. Future studies should concern real scientists (see, Findley et al., 2016). On the other hand, this experiment may give clear guidelines how to teach students to conduct experimental studies. Students should be informed about the importance of replicating research (i.e. Open Science Collaboration, 2015) and make them aware of dishonest publishing practices (Callaway, 2011). Additionally, it should be important to present students papers or books, with results which do not confirm assumed hypothesis.

To sum up, in the last years, many scholars (e.g. Ioannidis, 2005; John, Loewenstein, & Prelec, 2012; Simmons, Nelson, & Simonsohn, 2011) discussed

a crisis in the credibility of science. Registered Reports seem to be a promising solution to this problem – the pattern was partially confirmed by our study. Nevertheless, future studies in this area are necessary for support and the more in-depth understanding of the phenomenon.

REFERENCES

1. Begley, C. G., & Ellis, L. M. (2012). Drug development: Raise standards for preclinical cancer research. *Nature*, 483(7391), 531–533. <https://doi.org/10.1038/483531a>
2. Bem, S. L. (1974). The measurement of psychological androgyny. *Journal of consulting and clinical psychology*, 42(2), 155.
3. Callaway, E. (2011). Report finds massive fraud at Dutch universities. *Nature*, 479(7371), 15.
4. Camerer, C. F., Dreber, A., Holzmeister, F., Ho, T. H., Huber, J., Johannesson, M., ... & Altmejd, A. (2018). Evaluating the replicability of social science experiments in Nature and Science between 2010 and 2015. *Nature Human Behaviour*, 2(9), 637.
5. Enserink, M. (2012). Final Report on Stapel Also Blames Field As a Whole. *Science*, 338(6112), 1270–1271. <https://doi.org/10.1126/science.338.6112.1270>
6. Erlich, A. (2018). Pre-acceptance as a method to combat publication bias in area studies: a pilot in the Caucasus. *Caucasus Survey*, 6(3), 224–229.
7. Fanelli, D. (2009). How many scientists fabricate and falsify research? A systematic review and meta-analysis of survey data. *PLoS ONE*, 4(5), e5738.
8. Fanelli, D. (2010). “Positive” Results Increase Down the Hierarchy of the Sciences. *PLoS ONE*, 5(4), e10068. <https://doi.org/10.1371/journal.pone.0010068>
9. Findley, M. G., Jensen, N. M., Malesky, E. J., & Pepinsky, T. B. (2016). Can results-free review reduce publication bias? The results and implications of a pilot study. *Comparative Political Studies*, 49(13), 1667–1703.
10. Head, M. L., Holman, L., Lanfear, R., Kahn, A. T., & Jennions, M. D. (2015). The Extent and Consequences of PHacking in Science. *PLOS Biology*, 13(3), e1002106. <https://doi.org/10.1371/journal.pbio.1002106>
11. Jackson, C. J., Levine, S. Z., Furnham, A., & Burr, N. (2002). Predictors of Cheating Behavior at a University: A Lesson From the Psychology of Work. *Journal of Applied Social Psychology*, 32(5), 1031–1046. <https://doi.org/10.1111/j.1559-1816.2002.tb00254.x>
12. 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://doi.org/10.1177/0956797611430953>
13. Kuczyńska, A. (1992). *Inwentarz do oceny płci psychologicznej. Podręcznik*. [The Psychological Sex Inventory] Warszawa: Pracownia Testów Psychologicznych PTP.
14. Lakens, D. (2015). On the challenges of drawing conclusions from p-values just below 0.05. *PeerJ*, 3, e1142. <https://doi.org/10.7717/peerj.1142>
15. Leimu, R., & Koricheva, J. (2004). Cumulative meta-analysis: a new tool for detection of temporal trends and publication bias in ecology. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, 271(1551), 1961–1966.
16. Levine, T. R., Asada, K. J., & Carpenter, C. (2009). Sample Sizes and Effect Sizes are Negatively Correlated in Meta-Analyses: Evidence and Implications of a Publication Bias Against NonSignificant Findings. *Communication Monographs*, 76(3), 286–302. <https://doi.org/10.1080/03637750903074685>
17. Makel, M. C., Plucker, J. A., & Hegarty, B. (2012). Replications in psychology research: How often do they really occur? *Perspectives on Psychological Science*, 7(6), 537–542.
18. Morgan, G. A. (2000). Quasi-Experimental Designs. *Journal of the American Academy of Child & Adolescent Psychiatry*, 39, 794–796
19. McCabe, D. L., Treviño, L. K., & Butterfield, K. D. (2001). Cheating in academic institutions: A decade of research. *Ethics & Behavior*, 11(3), 219–232.

20. Munafò, M. R., Nosek, B. A., Bishop, D. V., Button, K. S., Chambers, C. D., Du Sert, N. P., Simonsohn U., Wagenmakers E.J., Ware J.J. & Ioannidis, J. P. (2017). A manifesto for reproducible science. *Nature Human Behaviour*, 1(1), 21. <https://doi.org/10.1038/s41562-016-0021>
21. Nosek, B. A., & Lakens, D. (2014). Registered Reports: A Method to Increase the Credibility of Published Results. *Social Psychology*, 45(3), 137–141. <https://doi.org/10.1027/1864-9335/a000192>
22. Nosek, B. A., Ebersole, C. R., DeHaven, A. C., & Mellor, D. T. (2018). The preregistration revolution. *Proceedings of the National Academy of Sciences*, 115, 2600-2606
23. 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>
24. Oleszkiewicz, A., Karwowski, M., Pisanski, K., Sorokowski, P., Sobrado, B., & Sorokowska, A. (2017a). Who uses emoticons? Data from 86 702 Facebook users. *Personality and Individual Differences*, 119, 289-295.
25. Oleszkiewicz, A., Frackowiak, T., Sorokowska, A., & Sorokowski, P. (2017b). Children can accurately recognize facial emotions from emoticons. *Computers in Human Behavior*, 76, 372-377.
26. Open Science Collaboration. (2015). Estimating the reproducibility of psychological science. *Science*, 349(6251), aac4716–aac4716. <https://doi.org/10.1126/science.aac4716>
27. Powell, K. (2016, November 4). Hard work, little reward: Nature readers reveal working hours and research challenges. Retrieved from <http://www.nature.com/news/hard-work-little-reward-nature-readers-reveal-working-hours-and-research-challenges-1.20933>
28. Probst, T. M., & Hagger, M. S. (2015). Advancing the Rigour and Integrity of Our Science: The Registered Reports Initiative: Editorial. *Stress and Health*, 31(3), 177–179. <https://doi.org/10.1002/smi.2645>
29. Rosenthal, R. (1979). The file drawer problem and tolerance for null results. *Psychological Bulletin*, 86(3), 638–641. <https://doi.org/10.1037/0033-2909.86.3.638>
30. Saunders, R., & Savulescu, J. (2008). Research ethics and lessons from Hwanggate: what can we learn from the Korean cloning fraud? *Journal of Medical Ethics*, 34(3), 214–221. <https://doi.org/10.1136/jme.2007.023721>
31. da Silva, J. A. T. (2016). Does the removal of results from a submitted paper reduce publication bias?. *Pacific Science Review B: Humanities and Social Sciences*, 2(1), 29-30
32. 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://doi.org/10.1177/0956797611417632>
33. Simonsohn, U. (2013). Just Post It: The Lesson From Two Cases of Fabricated Data Detected by Statistics Alone. *Psychological Science*, 24(10), 1875–1888. <https://doi.org/10.1177/0956797613480366>
34. 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>
35. 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.
36. Sorokowski, P., Kulczycki, E., Sorokowska A., Pisanski, K. (2017). Predatory journals recruit fake editor. *Nature*, 543, 481-483.
37. Stahl, D., & Pickles, A. (2018). Fact or fiction: reducing the proportion and impact of false positives. *Psychological Medicine*, 48(7), 1084-1091
38. Sterling, T. D., Rosenbaum, W. L., & Weinkam, J. J. (1995). Publication decisions revisited: The effect of the outcome of statistical tests on the decision to publish and vice versa. *The American Statistician*, 49(1), 108-112.
39. Stroebe, W., & Strack, F. (2014). The Alleged Crisis and the Illusion of Exact Replication. *Perspectives on Psychological Science*, 9(1), 59–71. <https://doi.org/10.1177/1745691613514450>
40. 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://doi.org/10.1111/j.1745-6924.2009.01125.x>
41. Wagenmakers, E., Wetzels, R., Borsboom, D., & van der Maas, H. L. J. (2011). Why psychologists must change the way they analyze their data: The case of psi: Comment on Bem

- (2011). *Journal of Personality and Social Psychology*, 100(3), 426–432. <https://doi.org/10.1037/a0022790>
42. Wicherts, J. M., Bakker, M., & Molenaar, D. (2011). Willingness to Share Research Data Is Related to the Strength of the Evidence and the Quality of Reporting of Statistical Results. *PLoS ONE*, 6(11), e26828. <https://doi.org/10.1371/journal.pone.0026828>
 43. Wicherts, J., Veldkamp, C., Augusteijn, H., Bakker, M., van Aert, R., & van Assen, M. (2016). Degrees of freedom in planning, running, analyzing, and reporting psychological studies A checklist to avoid p-hacking. *Frontiers in Psychology*, 7, 1832. <https://doi.org/10.3389/fpsyg.2016.01832>
 44. Zwaan, R. A., Pecher, D., Paolacci, G., Bouwmeester, S., Verkoeijen, P., Dijkstra, K., & Zeelenberg, R. (2017). Participant nonnaiveté and the reproducibility of cognitive psychology. *Psychonomic Bulletin & Review*, 25, 1968-1972