Evaluating Registered Reports: A Naturalistic Comparative Study of Article Impact

Short Title: Impact of Registered Reports

Lilian T. Hummer¹, Felix Singleton Thorn², Brian A. Nosek^{1,3}, Timothy M. Errington^{1*}¹ Center for Open Science, Charlottesville, Virginia, USA² University of Melbourne, Parkville, Victoria, Australia³ University of Virginia, Charlottesville, Virginia, USA

* Correspondence can be addressed to Timothy Errington, tim@cos.io.

Abstract

Registered Reports (RRs) is a publishing model in which initial peer review is conducted prior to knowing the outcomes of the research. In-principle acceptance of papers at this review stage combats publication bias and provides a clear distinction between confirmatory and exploratory research. Some editors raise a practical concern about adopting RRs. By reducing publication bias, RRs may produce more negative or mixed results and, if such results are not valued by the research community, receive less citations as a consequence. If so, by adopting RRs, a journal's impact factor may decline. Despite known flaws with impact factor, it is still used as a heuristic for judging journal prestige and quality. Whatever the merits of considering impact factor as a decision-rule for adopting RRs, it is worthwhile to know whether RRs are cited less than other articles. We will conduct a naturalistic comparison of citation and altmetric impact between published RRs and comparable empirical articles from the same journals.

Keywords: Registered Reports, Impact Factor, Altmetrics

Introduction

An accumulating body of evidence across disciplines suggests that findings from the published literature are not as reproducible as anticipated or desired (Begley and Ellis, 2012; Camerer et al., 2016; Ebersole et al., 2016; Klein et al., 2014; Nosek & Errington, 2017; Open Science Collaboration, 2015; Prinz et al., 2011). A variety of investigations identify human decision-making factors to irreproducibility such as selective reporting and questionable research practices such as *p*-hacking (Button et al., 2013; Greenwald, 1975; Ioannidis, 2005; Rosenthal, 1979; Simmons et al., 2011), and that these influences may be exacerbated by motivated reasoning and pressures to achieve positive, novel, clean results in order to maximize publishability and, ultimately, career advancement (Giner-Sorolla, 2012; Nosek et al., 2012).

Two factors believed to be particularly influential are (1) the tendency to report positive results and ignore negative results (Dickersin, 1990; Franco et al., 2014; Greenwald, 1975; Mahoney, 1977; Rosenthal, 1979; Sterling, 1959; Wagenmakers et al., 2012), and (2) the ease with which exploratory findings could be mistaken as confirmatory tests of pre-existing questions or hypotheses (Forstmeier et al., 2017; Nosek et al., 2018; Wagenmakers et al., 2012). Both of these potential challenges can be addressed with preregistration (Nosek et al., 2012; Nosek et al., 2018; Wagenmakers et al., 2012). If all studies and their results are registered then it is, in principle, possible to discover all results related to a particular phenomenon. This would eliminate the bias of non-discoverability of negative results compared to positive results. Likewise, if analysis plans for all studies were preregistered in advance then it is, in principle, possible to know whether any particular finding was a pre-specified confirmatory test of a phenomenon, or the outcome of an exploratory analysis with greater uncertainty in statistical inference. In some countries, preregistration is required by law in clinical trials research, and the practice is gaining popularity in basic and preclinical sciences (Nosek & Lindsay, 2018). For

example, multiple registries for basic science data have launched in the last 10 years (e.g., AEA Registry; eGAP; RIDIE; OSF), and some are demonstrating accelerating growth.

Preregistration, on its own, does not address the incentive pressures for researchers seeking publishable results to advance their careers. This could limit adoption of preregistration, and fail to reduce publication bias favoring positive results. A solution is to integrate preregistration with publishing incentives with a model called Registered Reports (Chambers, 2013; Nosek & Lakens, 2014).

The Registered Report (RR) publishing model is comprised of two key features: (1) peer review by the journal in advance of observing the research outcomes, and (2) if successful in that initial peer review, pre-commitment to publishing the results as long as all outcome-irrelevant quality control criteria are met. That is, preregistration is built into the RR publication process with expert peer review and editors giving in-principle acceptance to a preregistered protocol for publication. Then in the final report, authors clearly distinguish between confirmatory analyses specified in the accepted protocol and exploratory analyses that were conducted afterwards. Also, RRs address publication bias by committing to publish the results regardless of outcome eliminating privileging of positive results. Papers can be rejected in the second stage of peer review because of failures to follow through with the specified protocol, or evidence on outcome irrelevant criteria that the experimental context was not executed as planned (e.g., failure of a manipulation check), not because the results are different than desired or deemed unexciting.

Detailed treatments of the RR publishing model are available elsewhere (Chambers, 2013; Chambers et al., 2014; Nosek & Lakens, 2014; <http://cos.io/rr/>). The purpose of the present article is to evaluate a common concern of journal editors for adopting the RR publishing format. The bias against publishing null results has a variety of sources (Greenwald,

1975), one of which is concern that null results are less interesting and, therefore, less citable. Some journal editors like the RR format in principle, but resist adopting it because of concern that a commitment to publish null results will lead to fewer citations of articles in their journal, and a lower Journal Impact Factor (JIF). Despite well-known weaknesses of JIF as a means of evaluating the quality of journals or articles (Brembs et al., 2013; Larivière et al., 2016; Pulverer, 2015), it still retains power over the decision-making of some journal editors. Indeed, JIF may be partly responsible for the dominance of positive results in published journals in the first place (Brembs et al., 2013; Easterbrook et al., 1991). However, it is not clear whether RRs are, in fact, less likely to be cited than other articles. For example, it may be that RRs are more likely to contain null results, but they may also be higher quality research benefitting from design improvements during the first stage of peer review. Also, during the first stage of peer review, reviewers and editors do not know what the outcomes will be. As such, their decision to accept cannot be based on the results. Instead, reviewers and editors apply different decision-making criteria to decide whether to accept the RR such as “Is this question important?” and “Is the methodology an effective test of the research question?”. Because of the change in acceptance criteria, it may be that RRs are more likely to contain results that are deemed important, and citable, regardless of their outcome.

The use of anticipated citations is a questionable criterion for adopting RRs or accepting individual articles. Nevertheless, because it is used by journal editors for decision-making, it is useful to know whether their intuition that RRs will be less cited than other articles is correct. If so, then RR advocates would need to make the case that RRs should be adopted because of other values despite the potential negative impact on citation rates. If RRs are similarly cited or cited more, then the presumed negative effects on JIF would not persist as a barrier to adopting RRs.

To assess the current influence of RRs on citation impact, we will conduct a naturalistic evaluation of the existing published RRs and a comparison sample of similar non-RR empirical articles. We will compare these articles on citation counts from Web of Science, and on social media and news coverage from Altmetric data. This will provide an initial evidence base for assessing whether RRs have weaker, similar, or stronger citation impact than other empirical articles published near the same time in the same journals.

Materials and Methods

Description of Proposed Sample Characteristics

Journals and articles are not randomly assigned to be RRs. As a consequence, the value of this study hinges on identifying sensible comparison articles to RRs without introducing selection biases. The Center for Open Science maintains a database of all known published RRs (<https://www.zotero.org/groups/479248/osf/items/collectionKey/KEJP68G9>). On 2018-04-16 there was a population of 117 published RRs. We planned to use as many RRs as possible and to oversample comparison articles to increase overall precision of the comparison dataset.

To identify comparison articles, we considered our key outcome variables for assessing article impact: (1) citation counts (Web of Science) and (2) Altmetric attention scores (consolidation of social, digital, and news media coverage and discussion of the article). These impact indicators can be influenced by a variety of factors including: date of publication (impact accumulates with time), publishing journal (some journals get more attention than others), article type (e.g., empirical, review, editorial, commentary), discipline (Seglen, 1997), article length (Moed et al., 1996), and open access (Eysenbach, 2006).

The first inclusion rule was that all comparison articles must report empirical studies to match RRs on article type. The second inclusion rule was that the comparison articles must

have been published in the same journal as the RRs. This meant that 20 RRs published in the journal *Comprehensive Results in Social Psychology* would have no comparison articles because the journal publishes only RRs. This reduced the effective sample size to 97 RRs. The remainder of the inclusion and exclusion criteria maximized similarity of comparison articles and RRs on date of publication. We also set the minimum number of comparison articles for each RR at 9, so that at least 10 articles would be compared for each measure. As a result of this rule, 27 RRs were excluded. Of the included RRs, we selected all empirical articles from the same publishing journal that were published online (ePub date) in the period 14 days prior and 14 days after publication of the RR. This rule was applied directly for 27 RRs. For the remaining 43 RRs, this selection rule was complicated by the fact that more than one RR appeared in the same 28-day window. As a consequence, blindly applying the rule would result in the same article(s) serving as comparisons for multiple RRs. To solve this, RRs with ePub dates within 14 days of one another were combined into a cluster. This created 9 clusters of >1 RR (6 had 2 RRs, 1 had 3 RRs, 1 had 5 RRs, and 1 had 15 RRs). Then, we applied the selection rule for comparison articles to be 14 days prior to the first published and 14 days after the last published in the cluster. With this approach, it was still possible for comparison articles to be selected for more than one cluster/individual RR. There were 6 cases (2 clusters with 2 RRs in each cluster and 4 individual RRs) of potential overlap. To avoid using the same article twice, we split the overlapping articles evenly between the comparison samples. In total, the 31 individual RRs and 11 clusters of >1 RR will provide 42 independent assessments of the citation and altmetric impact of the 70 RRs included in this study compared with similar empirical reports from the same journals such that each RR will have 9 to 116 comparison articles ($Mdn = 21$, $IQR = 12 - 84.25$, $M = 45.0$, $SD = 39.4$). The R script with iterative conditional statements used to identify

clusters and date ranges for comparison articles can be found at <https://osf.io/phav4/> and the outcome of this selection process is summarized in Table 1.

Data and Extraction Procedure

All data and materials used in this study will be made available at <https://osf.io/gty97> (Hummer et al., 2017) and will include the following information for every RR and comparison article defined in our study: journal, publication title, publication authors, publication doi, ePub date, open access, number of pages, Web of Science citation counts and Altmetric attention scores. The WOS and Altmetric databases will be accessed for non-commercial use and was not accessed for the articles identified in this study until after this design was preregistered.

Publication doi and ePub date were manually extracted from the published article on each journal's site. Article citation counts were extracted from Web of Science for each article on a single day using the advanced search tool. Additionally, Altmetric attention scores were collected using an advanced search for each article on a single day in Altmetric Explorer database. The researchers handling raw attention score data during this study applied individually to the Altmetric Researcher Data Access Program and were granted provisional access to the Altmetric database.

Following preregistration and prior to data collection, we performed a preliminary programmatic query through the WOS and Altmetric Explorer databases using the DOIs of every article in our sample. This preliminary query further informed our data collection procedure for each database and allowed us to explicitly identify the articles in our sample that did not return data in one or both databases. We confirmed with Altmetric that articles missing from the Altmetric Explorer database have not accumulated mentions that contribute to an attention score and thus were assigned an Altmetric attention score of zero. Concerning missing data

from the WOS database, we performed iterative manual searches for all missing articles using the following format: first search: article title; second search: all publishing authors; third search: publication title, publication year, and first author; and fourth search: publishing journal and publication year. This additional step recovered a portion of the missing articles, which listed erroneous DOIs in the WOS database. We accounted for this during data collection by performing an additional programmatic query of WOS-assigned Accession Numbers for the specific articles listing incorrect DOIs in the WOS database.

On the predetermined day of data collection, we queried every indexed article in our sample using the advanced search functions of the WOS and Altmetric Explorer databases. We formatted the article DOIs in our advanced search by using the field tag "DO=" followed by article DOI (ex. "DO = 10.7554/eLife.32340"), and the boolean operator "OR" in between each DOI to enter a continuous string of up to 500 DOIs in the search field at a time. We kept other advanced search parameters on default, to include all languages, all document types, and all years. Once the results of a search were returned, we selected "Add to marked list" to create an output of the maximum available records from the search. When preparing the output records of our marked list, we selected the following custom settings beyond the default: Funding Information, Open Access, and Page Count/Chapter Count. To assign a destination for the output records, we selected "Save to Other File Formats" before choosing BibTeX reference software. This process was performed iteratively until data for the entire sample was collected. To account for articles with incorrect DOIs in the WOS database, we used the Accession Number assigned by WOS to these articles in place of article DOI during a final programmatic search once all other searches were completed (field tag: "UT="; boolean operator: "OR") to input all DOIs into the "Scholarly Identifiers" field. Once the results of the search were returned, we exported the results as a .csv file.

Planned analysis

For each cluster of RRs and comparison articles the modified z scores (Iglewicz and Hoaglin, 1993) was computed by modifying the *outliers* function of the *spatialEco* package and R software (R Core Team, 2018). The modified z score was calculated using the median absolute deviation, unless the median absolute deviation was zero, in which case the mean absolute deviation was used (IBM Knowledge Center, 2018). This was done separately for each outcome variable (Web of Science citation counts and Altmetric attention scores). If outcome data for an article was missing it was excluded for that specific outcome only. In the cases where there were multiple RRs in a cluster, we determined the modified z scores for each RR in the cluster independent from the other RRs, and then averaged them to generate a single modified z score for each article in the cluster. For example, if a cluster contained 2 RRs, we calculated the modified z scores for the comparison articles and one of the RRs, then we calculated the modified z scores for the comparison articles and the other RR. This generated a single modified z score for each RR relative to the comparison articles; however, each comparison article had two scores, which we combined into an average score. When necessary, we normalized the scores for all the articles in the cluster (RRs and comparison articles) to make the median value zero (e.g. subtract the median value from each score). Once we determined the modified z scores for each article we then weighted the articles in each cluster by a weighting factor that took into account the number of comparison articles in that cluster relative to the other clusters. The weighting factor for each cluster was determined by taking the number of comparison articles in the cluster, divided by the average number of comparison articles per cluster, and taken to its square root. That is, every article (RR and comparison articles) of a cluster was multiplied by the weighting factor for that cluster.

For each outcome variable, the weighted modified z scores for RRs and comparison articles across all clusters were combined into arithmetic medians, quartiles, and full presentations of individual data points with visualizations as scatter and box plots for the cumulative data across clusters and for the clusters individually. Reporting will emphasize descriptive statistics and effect sizes in the units of measurement (citations, Altmetric attention score). For the specific question of whether RR's have different impact than the comparison articles, we will conduct 3 one-sample Wilcoxon signed rank tests of the RR weighted modified z scores compared with zero, one for each outcome measure -- Web of Science citations, and Altmetric attention scores. Our emphasis of reporting and interpretation will be on the overall effect sizes, precision of estimation, and observed variability between indicators and within RRs.

As a secondary analysis, we will possess additional coding of the RRs on variables such as whether the study tested a novel research question or was a replication and whether the article was part of a special issue. Coding all of the comparison articles on these variables is outside the scope of this project. But, we will examine these other variables on an exploratory basis for correlations with impact within the sample of RRs. For example, it will be interesting to observe whether RRs of replications are cited more or less than RRs of novel research, or whether the weighted modified z scores of RR impact are declining or increasing over time as a potential indicator of interest in the format.

Results

Preliminary results (6/18/2018)

We identified published Registered Reports and comparison articles (matched on journal and time) to determine the relative impact of Registered Reports via two measures, Altmetric Attention Scores and Web of Science Citation rates. Raw data were high skewed

(**Supplemental Figures S1,S2**) so weighted modified z scores for each article were calculated to make the median article score 0 for each measure. We found Registered Reports, on average, attracted higher Altmetric Attention Scores (**Figures 1,2 & Table 2**) and similar to slightly greater numbers of citations (**Figures 3,4 & Table 3**) than the matched comparison articles. The main driver for the higher Attention Scores may well be the format's novelty, nevertheless these results provide preliminary evidence to counter the claim that Registered Reports will be less widely discussed and impactful than typical research articles.

Discussion

[To be added when study has been completed.]

Acknowledgments

We thank Steve Lindsay, Anne Scheel, and Bobbie Spellman for helpful comments during the planning of this research.

Ethical Statement

NA.

Funding Statement

All authors and resources for this study have been funded by the James S. McDonnell Foundation.

Data Accessibility

All data files and scripts for this study can be found at Hummer et al., 2017 (<https://osf.io/gty97>).

Competing Interests

LTH, BAN, TME: Employed by the Center for Open Science, which advocates for Registered Reports as a journal submission option. BAN, TME: Are authors of published Registered Reports included in this evaluation. BAN: Was a guest editor of a special issue of *Social Psychology*, which introduced and published Registered Reports (Nosek & Lakens, 2014).

Authors' Contributions

- 1) Substantial contributions to conception and design, or acquisition of data, or analysis and interpretation of data: LTH, FST, BAN, TME.
- 2) Drafting the article or revising it critically for important intellectual content: LTH, FST, BAN, TME.
- 3) Final approval of the version to be published: LTH, FST, BAN, TME.
- 4) Agreement to be accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved: LTH, FST, BAN, TME.

References

1. Begley, C.G., Ellis, L.M. 2012. Drug development: Raise standards for preclinical cancer research. *Nature* **483**, 531-533. (<http://dx.doi.org/10.1038/483531a>)
2. Brembs, B., Button, K., Munafò, M. 2013. Deep impact: unintended consequences of journal rank. *Front Hum Neurosci* **7**. (<https://doi.org/10.3389/fnhum.2013.00291>)
3. Button, K.S., Ioannidis, J.P.A., Mokrysz, C., Nosek, B.A., Flint, J., Robinson, E.S.J., Munafò, M.R. 2013. Power failure: why small sample size undermines the reliability of neuroscience. *Nature Rev Neurosci* **14**, 365-376. (<https://doi.org/10.1038/nrn3475>)
4. Camerer, C.F., Dreber, A., Forsell, E., Ho, T.-H., Huber, J., Johannesson, M., Kirchler, M., Almenberg, J., Altmejd, A., Chan, T., Heikensten, E., Holzmeister, F., Imai, T., Isaksson, S., Nave, G., Pfeiffer, T., Razen, M., Wu, H. 2016. Evaluating replicability of laboratory experiments in economics. *Science* **351**, 1433-1436. (<https://doi.org/10.1126/science.aaf0918>)
5. Chambers, C.D. 2013. Registered Reports: A new publishing initiative at Cortex. *Cortex* **49**, 609-610. (<https://doi.org/10.1016/j.cortex.2012.12.016>)
6. Chambers, C.D., Feredoes, E., Muthukumaraswamy, S.D., Etchells, P.J. 2014. Instead of “playing the game” it is time to change the rules: Registered Reports at AIMS Neuroscience and beyond. *AIMS Neurosci.* **1**, 4-17. (<https://doi.org/10.3934/Neuroscience.2014.1.4>)
7. Dickersin, K. 1990. The existence of publication bias and risk factors for its occurrence. *JAMA*. **263**, 1385-1389. (<https://doi.org/10.1001/jama.1990.03440100097014>)

8. Easterbrook, P.J., Gopalan, R., Berlin, J.A., Matthews, D.R. 1991. Publication bias in clinical research. *The Lancet* **337**, 867-872.
([https://doi.org/10.1016/0140-6736\(91\)90201-Y](https://doi.org/10.1016/0140-6736(91)90201-Y))
9. Ebersole, C.R., Atherton, O.E., Belanger, A.L., Skulborstad, H.M., Allen, J.M., Banks, J.B., Baranski, E., Bernstein, M.J., Bonfiglio, D.B.V., Boucher, L., Brown, E.R., Budiman, N.I., Cairo, A.H., Capaldi, C.A., Chartier, C.R., Chung, J.M., Cicero, D.C., Coleman, J.A., Conway, J.G., Davis, W.E., Devos, T., Fletcher, M.M., German, K., Grahe, J.E., Hermann, A.D., Hicks, J.A., Honeycutt, N., Humphrey, B., Janus, M., Johnson, D.J., Joy-Gaba, J.A., Juzeler, H., Keres, A., Kinney, D., Kirshenbaum, J., Klein, R.A., Lucas, R.e., Lustgraaf, C.J.N., Martin, D., Menon, M., Metzger, M., Moloney, J.M., Morse, P.J., Prislín, R., Razza, T., Re, D.E., Rule, N.O., Sacco, D.F., Sauerberger, K., Shrider, E., Shultz, M., Siensen, C., Sobocko, K., Weylin Sternglanz, R., Summerville, A., Tskhay, K.O., van Allen, z., Vaughn, L.A., Walker, R.J., Weinberg, A., Wilson, J.P., Wirth, J.H., Wortman, J., Nosek, B.A. 2016. Many Labs 3: Evaluating participant pool quality across the academic semester via replication. *J. Exp. Psychol.* **67**, 68-82.
(<https://doi.org/10.1016/j.jesp.2015.10.012>)
10. Eysenbach, G. (2006). Citation Advantage of Open Access Articles. *PLOS Biology* 4(5):e157. (<https://doi.org/10.1371/journal.pbio.0040157>)
11. Forstmeier, W., Wagenmakers, E.-J., Parker, T.H. 2017. Detecting and avoiding likely false-positive findings – a practical guide: Avoiding false-positive findings. *Biol Rev Camb Philos Soc* **92**, 1941-1968. (<https://doi.org/10.1111/brv.12315>)
12. Franco, A., Malhotra, N., Simonovits, G. 2014. Publication bias in the social sciences: Unlocking the file drawer. *Science* **345**, 1502-1505.
(<https://doi.org/10.1126/science.1255484>)

13. Giner-Sorolla, R. 2012. Science or Art? How aesthetic standards grease the way through the publication bottleneck but undermine science. *Perspect Psychol Sci* **7**, 562-571.
(<https://doi.org/10.1177/1745691612457576>)
14. Greenwald, A.G. 1975. Consequences of prejudice against the null hypothesis. *Psychol Bull* **82**, 1-20. (<https://doi.org/10.1037/h0076157>)
15. Hummer, L.T., Thorn, F.S., Nosek, B.A., Errington, T.M. 2017. Evaluating Registered Reports: A naturalistic comparative study of article impact. Open Science Framework.
(<https://doi.org/10.17605/OSF.IO/GTY97>)
16. IBM Knowledge Center. (2018). Retrieved June 13, 2018, from
https://www.ibm.com/support/knowledgecenter/en/SSWLVY_1.0.0/com.ibm.spss.analytic_catalyst.help/analytic_catalyst/modified_z.html
17. Iglewicz, B., Hoaglin, D.C., 1993. How to detect and handle outliers, ASQC basic references in quality control. ASQC Quality Press, Milwaukee, Wis.
18. Ioannidis, J.P.A. 2005. Why most published research findings are false. *PLoS Medicine* **2**, e124. (<https://doi.org/10.1371/journal.pmed.0020124>)
19. Klein, R.A., Ratliff, K.A., Vianello, M., Adams, R.B., Bahník, Š., Bernstein, M.J., Bocian, K., Brandt, M.J., Brooks, B., Brumbaugh, C.C., Cemalcilar, Z., Chandler, J., Cheong, W., Davis, W.E., Devos, T., Eisner, M., Frankowska, N., Furrow, D., Galliani, E.M., Hasselman, F., Hicks, J.A., Hovermale, J.F., Hunt, S.J., Huntsinger, J.R., IJzerman, H., John, M.-S., Joy-Gaba, J.A., Barry Kappes, H., Krueger, L.E., Kurtz, J., Levitan, C.A., Mallett, R.K., Morris, W.L., Nelson, A.J., Nier, J.A., Packard, G., Pilati, R., Rutchick, A.M., Schmidt, K., Skorinko, J.L., Smith, R., Steiner, T.G., Storbeck, J., Van Swol, L.M., Thompson, D., van't Veer, A.E., Anne Vaughn, L., Vranka, M., Wichman, A.L., Woodzicka, J.A., Nosek, B.A. 2014. Investigating variation in replicability: A “many labs”

replication project. *Soc Psychol* **45**, 142-152.

(<https://doi.org/10.1027/1864-9335/a000178>)

20. Larivière, V., Kiermer, V., MacCallum, C.J., McNutt, M., Patterson, M., Pulverer, B., Swaminathan, S., Taylor, S., Curry, S. 2016. A simple proposal for the publication of journal citation distributions. *BioRxiv* **062109**. (<https://doi.org/10.1101/062109>)
21. Mahoney, M.J. 1977. Publication prejudices: An experimental study of confirmatory bias in the peer review system. *Cognit Ther Res* **1**, 161-175.
(<https://doi.org/10.1007/BF01173636>)
22. Moed, H.F., Van Leeuwen, Th.N., Reedijk, J., 1996. A critical analysis of the journal impact factors of *Angewandte Chemie* and the journal of the American Chemical Society in inaccuracies in published impact factors based on overall citations only. *Scientometrics* **37**, 105–116. <https://doi.org/10.1007/BF02093487>
23. 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.
<https://doi.org/10.1073/pnas.1708274114>
24. Nosek, B.A., Errington, T.M. 2017. Making sense of replications. *eLife* **6**.
(<https://doi.org/10.7554/eLife.23383>)
25. Nosek, B.A., Lakens, D. 2014. Registered reports: A method to increase the credibility of published results. *Soc Psychol* **45**, 137-141.
(<https://doi.org/10.1027/1864-9335/a000192>)
26. Nosek, B. A., & Lindsay, D. S. (2018). Preregistration becoming the norm in psychological science. *APS Observer*, (31)3.

27. Nosek, B.A., Spies, J.R., Motyl, M. 2012. Scientific utopia: II. Restructuring incentives and practices to promote truth over publishability *Perspect Psychol Sci* **7**, 615-631.
(<https://doi.org/10.1177/1745691612459058>)
28. Open Science Collaboration. 2015. Estimating the reproducibility of psychological science. *Science* **349**, aac4716-aac4716. (<https://doi.org/10.1126/science.aac4716>)
29. Prinz, F., Schlange, T., Asadullah, K. 2011. Believe it or not: how much can we rely on published data on potential drug targets? *Nat Rev Drug Discov* **10**, 712.
(<https://doi.org/10.1038/nrd3439-c1>)
30. Pulverer, B. 2015. Dora the brave. *EMBO J* **34**, 1601-1602.
(<https://doi.org/10.15252/emboj.201570010>)
31. R Core Team. 2018. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria
32. Rosenthal, R. 1979. The file drawer problem and tolerance for null results. *Psychol Bull* **86**, 638-641. (<https://doi.org/10.1037/0033-2909.86.3.638>)
33. Seglen, P.O., 1997. Why the impact factor of journals should not be used for evaluating research. *BMJ* **314**, 498–502. <https://doi.org/10.1136/bmj.314.7079.497>
34. Simmons, J.P., Nelson, L.D., Simonsohn, U. 2011. False-positive psychology: Undisclosed flexibility in data collection and analysis allows presenting anything as significant. *Psychol Sci* **22**, 1359-1366. (<https://doi.org/10.1177/0956797611417632>)
35. Sterling, T.D. 1959. Publication decisions and their possible effects on inferences drawn from tests of significance – or Vice versa. *J Am Stat Assoc* **49**, 108.
(<https://doi.org/10.1080/01621459.1959.10501497>)

36. Wagenmakers, E.-J., Wetzels, R., Borsboom, D., van der Maas, H.L.J., Kievit, R.A.
2012. An agenda for purely confirmatory research. *Perspect Psychol Sci* **7**, 632-638.
(<https://doi.org/10.1177/1745691612463078>)

Running Head: EVALUATING REGISTERED REPORTS: IMPACT

Table 1. Number of Registered Reports (# RR) included in each cluster with journal, RR doi(s), ePub dates (ePub), date range (before and after date) for identification of comparison articles (CA), and number of CAs (# CA).

cluster	# RR	Journal	RR doi(s)	ePub(s)	before	after	# CA
1	15	<i>Social Psychology</i>	10.1027/1864-9335/a000178; 10.1027/1864-9335/a000179; 10.1027/1864-9335/a000180; 10.1027/1864-9335/a000181; 10.1027/1864-9335/a000182; 10.1027/1864-9335/a000183; 10.1027/1864-9335/a000184; 10.1027/1864-9335/a000185; 10.1027/1864-9335/a000186; 10.1027/1864-9335/a000187; 10.1027/1864-9335/a000188; 10.1027/1864-9335/a000189; 10.1027/1864-9335/a000190; 10.1027/1864-9335/a000191; 10.1027/1864-9335/a000193	2014-01-01; 2014-01-01; 2014-01-01; 2014-01-01; 2014-01-01; 2014-01-01; 2014-01-01; 2014-01-01; 2014-01-01; 2014-01-01; 2014-01-01; 2014-01-01; 2014-01-01; 2014-01-01; 2014-01-01	2013-12-18	2014-01-15	34
2	1	<i>Frontiers in Psychology</i>	10.3389/fpsyg.2014.00259	2014-03-31	2014-03-17	2014-04-14	59
3	1	<i>Frontiers in Psychology</i>	10.3389/fpsyg.2014.00539	2014-06-04	2014-05-21	2014-06-18	92
4	1	<i>Frontiers in Psychology</i>	10.3389/fpsyg.2014.00786	2014-07-28	2014-07-14	2014-08-04	55
5	1	<i>Frontiers in Psychology</i>	10.3389/fpsyg.2014.00875	2014-08-12	2014-08-05	2014-08-26	54
6	2	<i>Frontiers in Psychology</i>	10.3389/fpsyg.2014.00987; 10.3389/fpsyg.2014.01035	2014-09-10; 2014-09-16	2014-08-27	2014-09-30	90
7	1	<i>Frontiers in Psychology</i>	10.3389/fpsyg.2014.01343	2014-11-21	2014-11-07	2014-12-05	102

Running Head: EVALUATING REGISTERED REPORTS: IMPACT

8	1	<i>Frontiers in Psychology</i>	10.3389/fpsyg.2014.01511	2015-01-05	2014-12-22	2015-01-19	59
18	1	<i>Cortex</i>	10.1016/j.cortex.2014.12.019	2015-02-13	2015-01-30	2015-12-27	11
9	2	<i>Frontiers in Psychology</i>	10.3389/fpsyg.2015.00314; 10.3389/fpsyg.2015.00335	2015-04-01; 2015-04-02	2015-03-18	2015-04-13	85
19	1	<i>Cortex</i>	10.1016/j.cortex.2015.03.010	2015-04-17	2015-04-03	2015-05-01	12
10	1	<i>Frontiers in Psychology</i>	10.3389/fpsyg.2015.00494	2015-04-24	2015-04-14	2015-05-08	106
11	1	<i>Frontiers in Psychology</i>	10.3389/fpsyg.2015.00672	2015-06-08	2015-05-25	2015-06-22	116
25	1	<i>Comparative Political Studies</i>	10.1177/0010414015621072	2016-01-14	2015-12-31	2016-01-28	10
13	1	<i>Perspectives on Psychological Science</i>	10.1177/1745691615605826	2016-01-27	2016-01-13	2016-02-10	10
20	1	<i>Cortex</i>	10.1016/j.cortex.2016.01.006	2016-02-06	2016-01-23	2016-02-20	13
26	1	<i>Comparative Political Studies</i>	10.1177/0010414015626436	2016-03-03	2016-02-18	2016-03-17	10
21	3	<i>Cortex</i>	10.1016/j.cortex.2016.03.004; 10.1016/j.cortex.2016.03.020; 10.1016/j.cortex.2016.03.019	2016-03-31; 2016-04-06; 2016-04-15	2016-03-17	2016-04-29	22
30	1	<i>Journal of Research in Personality</i>	10.1016/j.jrp.2016.05.003	2016-05-06	2016-04-22	2016-05-20	18
14	1	<i>Perspectives on Psychological Science</i>	10.1177/1745691616652873	2016-07-29	2016-07-15	2016-08-12	9
32	2	<i>Journal of Research in Personality</i>	10.1016/j.jrp.2016.08.005; 10.1016/j.jrp.2016.08.010	2016-08-21; 2016-09-04	2016-08-07	2016-09-18	13
15	1	<i>Perspectives on Psychological Science</i>	10.1177/1745691616664694	2016-09-29	2016-09-15	2016-10-13	10

Running Head: EVALUATING REGISTERED REPORTS: IMPACT

38	1	<i>Royal Society Open Science</i>	10.1098/rsos.160431	2016-11-23	2016-11-09	2016-12-07	62
45	5	<i>eLife</i>	10.7554/eLife.21253; 10.7554/eLife.18173; 10.7554/eLife.21634; 10.7554/eLife.17044; 10.7554/eLife.17584	2017-01-19; 2017-01-19; 2017-01-19; 2017-01-19; 2017-01-19	2017-01-05	2017-02-02	101
43	1	<i>Cognition and Emotion</i>	10.1080/02699931.2017.1292217	2017-02-20	2017-02-06	2017-03-06	11
49	1	<i>Attention, Perception, & Psychophysics</i>	10.3758/s13414-017-1304-y	2017-02-23	2017-02-09	2017-03-09	15
46	2	<i>eLife</i>	10.7554/eLife.25306; 10.7554/eLife.26030	2017-06-27; 2017-06-27	2017-06-13	2017-01-11	109
52	1	<i>Psychological Science</i>	10.1177/0956797617719730	2017-08-09	2017-07-26	2017-08-23	12
53	1	<i>Drug and Alcohol Dependence</i>	10.1016/j.drugalcdep.2017.08.022	2017-09-19	2017-09-05	2017-10-03	22
39	2	<i>Royal Society Open Science</i>	10.1098/rsos.160935; 10.1098/rsos.171172	2017-09-20; 2017-10-04	2017-09-06	2017-10-18	98
22	1	<i>Cortex</i>	10.1016/j.cortex.2017.09.026	2017-10-09	2017-09-25	2017-10-23	20
34	1	<i>Journal of Research in Personality</i>	10.1016/j.jrp.2017.10.005	2017-10-28	2017-10-14	2017-11-11	10
40	1	<i>Royal Society Open Science</i>	10.1098/rsos.170543	2017-12-20	2017-12-06	2018-01-03	69
47	2	<i>eLife</i>	10.7554/eLife.29747; 10.7554/eLife.30274	2018-01-09; 2018-01-09	2017-12-26	2018-01-23	82

Running Head: EVALUATING REGISTERED REPORTS: IMPACT

44	1	<i>Cognition and Emotion</i>	10.1080/02699931.2018.1429389	2018-01-31	2018-01-17	2018-02-14	12
23	2	<i>Cortex</i>	10.1016/j.cortex.2018.01.001; 10.1016/j.cortex.2018.01.006	2018-01-31; 2018-02-02	2018-01-17	2018-02-14	13
36	2	<i>Journal of Research in Personality</i>	10.1016/j.jrp.2016.06.022; 10.1016/j.jrp.2018.01.002	2018-02-02; 2018-02-04	2018-01-19	2018-02-18	9
41	1	<i>Royal Society Open Science</i>	10.1098/rsos.171678	2018-02-21	2018-02-07	2018-03-07	107
50	1	<i>Attention, Perception, & Psychophysics</i>	10.3758/s13414-018-1494-y	2018-02-21	2018-02-07	2018-03-07	9
55	1	<i>International Journal of Psychophysiology</i>	10.1016/j.ijpsycho.2018.01.012	2018-02-21	2018-02-07	2018-03-07	14
24	1	<i>Cortex</i>	10.1016/j.cortex.2018.02.009	2018-02-27	2018-02-15	2018-03-13	12
48	1	<i>eLife</i>	10.7554/eLife.25801	2018-03-13	2018-02-27	2018-03-27	115

Figure 1 - Aggregate Altmetric Attention Scores for Registered Reports and comparison articles

Box plots of Altmetric Attention Score weighted z-scores for comparison articles (black) and Registered Reports (blue, dots represent all Registered Reports' scores). Note: the y-axis has been truncated and excludes 85 comparison articles and 3 Registered Reports with modified z scores greater than 12.

Figure 2 - Altmetric Attention Scores for Registered Reports and comparison articles over time

Scatter plot of Altmetric Attention Score weighted z-scores for all articles by date. Note: the y-axis has been truncated and excludes 14 comparison articles with modified z scores of greater than 50.

Table 2: Summary stats for Altmetric Attention Score weighted z scores:

Stats	Registered Reports (n=70)	Comparison articles (n=1892)
Min	-0.60	-1.11
1st Quartile	0.15	-0.38
Median	1.48	0.00
Mean	3.07	2.38
3rd Quartile	3.68	1.11
Max	30.43	253.86

Figure 3 - Aggregate Web of Science Citations for Registered Reports and comparison articles

Box plots of Web of Science Citation weighted z-scores for comparison articles (black) and Registered Reports (blue, dots represent all Registered Reports' scores). Note: the y-axis has been truncated and excludes 55 comparison articles with modified z scores of greater than 20.

Figure 4 - Web of Science Citations for Registered Reports and comparison articles over time

Scatter plot of Web of Science Citation weighted z-scores for all articles by date. Note: the y-axis has been truncated and excludes 2 comparison articles and 1 Registered Report with modified z scores of greater than 10

Table 3: Summary stats for Web of Science Citation weighted z scores:

Stats	Registered Reports (n=64)	Comparison articles (n=1823)
Min	-1.37	-1.28
1st Quartile	-0.11	-0.18
Median	0.00	0.00
Mean	0.64	0.37
3rd Quartile	0.78	0.52
Max	17.30	13.04

Supplementary Figures:**Figure S1 - Histogram of raw Altmetric Attention Scores for all articles** **Figure S2 - Histogram of raw Web of Science Citations for all articles**

Figure S3 - Raw Altmetric Attention Scores for Registered Reports and comparison articles over time

Scatter plot of raw Altmetric Attention Scores for all articles by date. Note: the y-axis has been truncated and excludes 8 comparison articles with scores greater than 400.

Figure S4 - Raw Web of Science Citations for Registered Reports and comparison articles over time

Scatter plots of raw Web of Science Citations for all articles by date.

Alternative Figures by cluster: box and whisker plots for all articles with Registered Reports in red:

Running Head: EVALUATING REGISTERED REPORTS: IMPACT