Positive result rates in psychology: Registered Reports compared to conventional reports: a meta-analysis using Fanelli's method – Preregistration document

The following pre-registration document is based on Preregistration for Qualitative Research Template (“Preregistration for Qualitative Research Template,” 2018).

# Remarks regarding pilot studies

Before this study was preregistered, we piloted the method on a small subsample of Registered Reports and conventional reports. Details about the pilot, including the used dataset and methodology, are described below. In the Fanelli dataset pilot the researchers labelled 8 out of 11 papers identically. There were minor differences in coding between the researchers, but while discussing the results the researchers agreed on the labelling of the 3 conflictingly coded papers. In the RR dataset pilot researchers labelled 4 out of 10 papers differently. As will be discussed further on in this preregistration document, papers could be coded as having full, partial or no support for their first tested hypothesis. During final analysis the distinction is made between papers that had no support or at least partial support; this means that one researcher coding a paper as having partial support, while the other researcher coding it as having full support will not change the final proportion of papers that have at least partial support and papers that do not have support. When comparing the independent coding under the assumption that differences of partial vs. full support will not cause changes in the final analysis, the first pilot on Fanelli’s data resulted in 9 of 11 papers being labelled identically, while the second pilot on RRs resulted in 7 of 10 papers being labelled identically. While discussing the results the researchers agreed on the labelling of these conflicted papers. One of the conflicting papers was a Registered Replication Report that gauged the size and effect of the Replication studies that were performed, instead of declaring to have tested a specific hypothesis. In line with Fanelli’s coding method such papers (that do not state a hypothesis) will be excluded.

# General information

1. Project title: Positive result rates in psychology: Registered Reports compared to conventional reports: a meta-analysis using Fanelli's method
2. Authors and Affiliations:

* Researcher: M.R.M.J. Schijen [MS] (bachelor student at Eindhoven University of Technology)
* 1st supervisor: D. Lakens [DL] (professor at Eindhoven University of Technology)
* 2nd supervisor: A.M. Scheel [AS] (PhD student at Eindhoven University of Technology)

1. Abstract and research questions:

* Abstract: The aim of this study is to investigate the positive result rate in Registered Reports, compared to the positive result rate in conventional reports (published articles that are not Registered Reports) in psychology. The positive result rate is the proportion of papers in a scientific field that report to have found full or partial support for the hypothesis they set out to test. The positive result rate in different fields of science was previously investigated by Fanelli (2010). Fanelli found that Psychology had the largest positive result rate at 91.5%. These positive result rates are believed to be inflated by publication bias (amongst other factors, such as the ‘softness’ of the scientific field). Since this study was conducted, a new publication format for scientific papers has been introduced: the Registered Report (RR), where only the hypotheses and methods of a study are used to assess whether a paper is suitable for publication. In RRs, the decision to publish is made based on the quality of the research before data collection has occurred (see https://cos.io/rr/). This approach to publishing should decrease publication bias by allowing more studies to be published that were not able to find full or partial support for their tested hypotheses. In the present study, we aim to put this to the test by measuring and comparing the positive result rates in RRs and conventional reports in the field of psychology.
* Research questions:
  + Pilot study – Fanelli’s dataset: Are we able to apply Fanelli’s coding method well enough to reproduce his results with at least 90% accuracy on a subset of Fanelli’s original dataset? This pilot study was completed before the study was preregistered.
  + Pilot study – Registered Reports dataset: We apply Fanelli’s coding method to a small, random subset (10 items) of the list of published RRs. This pilot study will help uncover issues that might arise when performing the main research. The RRs have not been selected using the criterion that their abstract includes the sentence “test\* the hypothes\*”, therefore the wording of the hypothesis might be more ambiguous compared to Fanelli’s original sample. This pilot study was also completed before the study was preregistered.
  + Main research question: Is the positive result rate of Registered Reports in psychology statistically lower than the positive result rate of conventional reports in Psychology? This research has not been started at the time of preregistration.
  + Additional research question: One criterion for gathering data for Fanelli’s study entailed that papers could only be retrieved if a variation of the sentence “test\* the hypothes\*” was included somewhere in the paper. Are other key sentences also used in psychology papers that signal that the study tested a hypothesis? This research has not been started at the time of preregistration.

1. Project start date and end date:

* The project started on the 3rd of September, 2018. The final report that concludes the project is due on the 20th of January, 2019. At the time of preregistration the list of RRs is known to the researchers, and this list has been checked to make sure that the dataset will only contain published RRs in psychology. A new dataset of conventional reports in psychology has not yet been gathered at the time of preregistration. The pilot studies we performed made use of a subset of Fanelli’s original dataset and a subset of the list of RRs that we have verified. Full data collection will not start before the project is preregistered.

# Research design specifics

1. Please describe the data collection/generating process of the study:

* Data collection: pilots
  + Pilot study – Fanelli’s dataset: In personal communication with Fanelli (8th October 2018), we obtained a short list of coded data (11 psychology papers). AS found the original papers that corresponded with the coded data we received.
  + Pilot study – RR: We verified the items on the list of RRs curated by the Center for Open Science (https://www.zotero.org/groups/479248/osf/items/collectionKey/KEJP68G9?). For this subset, 10 items were randomly chosen from this list of verified RRs.
* Data collection: Main research  
  The dataset of RRs will consist of the set of papers in the list of RRs curated by the Center for Open Science (https://www.zotero.org/groups/479248/osf/items/collectionKey/KEJP68G9?), that remain after we have checked that they truly are published RRs .   
  We count papers as RRs if they are labelled as such by the journal itself and the journal submission guidelines make it clear that by “Registered Report” they mean that a submission gets reviewed and receives in-principle acceptance before the data collection or analyses for all studies in the paper have been conducted (in accordance with <https://cos.io/rr>). For journals that do not continuously offer RRs or papers that are not clearly labelled as RRs, we consult relevant editorial publications (e.g. for special issues) or consult the respective editors directly to determine whether these same criteria have been met. Lastly, we only include RRs that have been published by the journal (i.e., excluding study protocols and preprints) and not been retracted at the time of coding.  
  Additionally, the paper has to be published in a journal on psychology research, or it is classified as belonging to the field of psychology (for papers published in multi-disciplinary journals). The dataset of conventional reports will be determined using the same data-collection methodology that Fanelli used in his original research: “The sentence “test\* the hypothes\*” was used to search all 10837 journals in the Essential Science Indicators database, which classifies journals univocally in 22 disciplines. When the number of papers retrieved from one discipline exceeded 150, papers were selected using a random number generator.” (Fanelli, 2010, p. 8), with three additional constraints:
  + we only search journals from the ESI’s “psychiatry and psychology” category
  + the papers have to have been published between 2013 and 2018 (to match them to the RR dataset)
  + the papers must not overlap with the RR dataset.
* Data coding  
  The data will be coded using the method that Fanelli (2010) used in his study of positive result rates in science: “By examining the abstract and/or full- text, it was determined whether the authors of each paper had concluded to have found a positive (full or partial) or negative (null or negative) support. If more than one hypothesis was being tested, only the first one to appear in the text was considered. We excluded meeting abstracts and papers that either did not test a hypothesis or for which we lacked sufficient information to determine the outcome” (Fanelli, 2010, p. 8). “Insufficient information in the paper” is interpreted as the paper having an unclear abstract and the full text of the paper not being available for inspection (Fanelli, 2012), or, if the full text is available, an unclear abstract *and* unclear full text with regard to the tested hypotheses and/or conclusions.   
  That means that every paper in the sample will contribute one data point for the main dependent variable: Whether the first hypothesis was supported (fully or partially) or not.   
  For RRs, we will code the first hypothesis that relates to a study for which data collection or analysis had not yet been conducted when the study was submitted, meaning that we will exclude hypotheses that relate to e.g. a pilot study that already existed when the stage-1 RR was submitted to the journal.   
  Additionally, we will code if the paper reports original work or a replication of previous work.  
  For the pilot study, MS and AS independently coded the pilot samples and compared their results. The dataset used in the main and additional research will be primarily coded by MS, papers that were difficult to code will be marked. AS will double code the papers that were indicated as being difficult to code; AS will also double code a randomly selected subset of the dataset. Double coded data will be compared between AS and MS, after discussion the final coding will be decided.

1. Please describe the unit of analysis of the study:

* The unit of analysis for our research is a journal article, each article will contribute as one data point in our dataset.

1. Is this project hypothesis-testing or hypothesis-generating (or both)?

* This project is both hypothesis-testing and hypothesis-generating.

1. If hypothesis-testing, please name and describe the hypotheses below:

* Main research: The positive result rate in Registered Reports in psychology will be statistically lower than the positive result rate in conventional reports in psychology.

1. If hypothesis-testing, please describe the methodology you employ for the test:

* The study will make use of a set of RRs and a set of conventional reports. (CRs) These datasets will be collected as follows:
  + Data collection and coding is described in section 5.
  + For each paper in the RR and CR dataset, we will code if the first reported hypothesis will be labelled as either having full support, partial support or no support. “Full support” and “partial support” will be combined to one category (“support”), resulting in a binary dependent variable (“support” vs. “no support”). A one-sided proportion test with an alpha level of 5% will be performed to test whether the positive result rate (full or partial support) of Registered Reports in psychology is statistically lower than the positive result rate of conventional reports in psychology.  
    In addition to testing if there is a statistically significant difference between RRs and conventional reports, we will test if the difference is smaller than our smallest effect size of interest using an equivalence test for proportion tests with an alpha level of 5% (Lakens, Scheel, & Isager, 2018). We determined our smallest effect size of interest to be the difference between the positive result rate in psychology (91.5%) and the positive result rate in general social sciences (85.5%) as reported by Fanelli (2010), i.e. a difference of 91.5% - 85.5% = 6%. The rationale for choosing general social sciences as a comparison is that this discipline had the lowest positive result rate amongst the "soft" sciences (Fanelli, 2010). The exact percentage for general social sciences was extracted from Figure 1 in Fanelli (2010) using the software WebPlotDigitizer (Rohatgi, 2018).

1. If hypothesis-generating, please name and describe possible outcome hypotheses:

* Additional research: We will identify alternative key phrases that are used to signal that a hypothesis was tested. Possible outcome hypotheses include:
  + Conventional reports using these alternative key phrases rather than the phrase that was used to determine the sample in Fanelli’s study (2010) might have a different positive result rate than the literature examined by Fanelli.
  + Different key phrases signaling a hypothesis test are associated with different types of research questions or research areas.

1. If hypothesis-generating, please describe the methodology you employ to generate hypotheses:

* Additional research: While coding the RR dataset, we will code whether a sentence other than a variation on “test\* the hypothes\*” is used to indicate that the paper tested a hypothesis.

1. Please define and describe the outcome variable or concept in your hypotheses:

* Main research: The outcome variable is whether the paper reports (full or partial) support for the first stated hypothesis or not (binary variable).
* Additional research: The outcome variable consists of a list of alternative phrases that were found to be used in the RR dataset to indicate that the paper tested a hypothesis.

1. Please define and describe the independent variables or concepts in your hypotheses:

* The independent variable is what category the paper belongs to (RR or conventional report).

1. Does this project involve fieldwork?

* No

1. Does this project involve archival work?

* Yes, the data collection of the RR dataset is guided by the RRs listed by the COS. The data collection of both RRs and conventional reports will be performed using online libraries.

1. Does this project involve interviews or focus groups?

* No

1. Does this project involve case studies?

* No

# Analysis

1. Please describe the analytical tools you use for this project:

* The RR dataset will be stored and verified using Zotero (https://www.zotero.org/)
* Data will be coded in Microsoft Excel
* Analysis will be performed using either R (R Core Team, 2018) or Stata (StataCorp, 2015).

1. Does this project involve computer assisted data analysis?
   1. If yes, what software package do you use?

* Analysis of the main and additional research will be conducted using either R (R Core Team, 2018) or Stata (StataCorp, 2015).
  1. If yes, what are the analytical categories and codes used for the analysis?
* The main research will use several statistical tests, either the Chi-squared test or the Fisher’s Exact Test, and the TOST procedure (Lakens, Scheel, & Isager, 2018), on the dependent and independent variables described in sections 12 and 13.

1. Does this project involve sensitive data you cannot share?

* All data is gathered from published papers, therefore no sensitive data is used in this analysis.

# References

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

Fanelli, D. (2012). Negative results are disappearing from most disciplines and countries. *Scientometrics*, *90*(3), 891–904. https://doi.org/10.1007/s11192-011-0494-7

Lakens, D., Scheel, A. M., & Isager, P. M. (2018). Equivalence Testing for Psychological Research: A Tutorial. *Advances in Methods and Practices in Psychological Science*, *1*(2), 259–269. https://doi.org/10.1016/j.ympev.2015.01.015

Preregistration for Qualitative Research Template. (2018). Retrieved from https://osf.io/b6xmd/

R Core Team. (2018). R: A language and environment for statistical computing. Vienna: R Foundation for Statistical Computing. Retrieved from https://www.r-project.org/

Rohatgi, A. (2018). WebPlotDigitizer - Web Based Plot Digitizer. Austin, Texas, USA. Retrieved from https://automeris.io/WebPlotDigitizer

StataCorp. (2015). Stata. College Station: StataCorp. Retrieved from http://www.stata.com