Replication Package for JPEMicro entitled "Do Pre-Registration and Pre-Analysis Plans Reduce p-Hacking and Publication Bias?: Evidence from 15,992 Test Statistics and Suggestions for Improvement**"**

Abstract:
Pre-registration is regarded as an important contributor to research credibility. We investigate this by analyzing the pattern of test statistics from the universe of randomized controlled trials (RCT) studies published in 15 leading economics journals. We draw two conclusions: (a) Pre-registration frequently does not involve a pre-analysis plan (PAP), or sufficient detail to constrain meaningfully the actions and decisions of researchers after data is collected. Consistent with this, we find no evidence that pre-registration in itself reduces p-hacking and publication bias. (b) When pre-registration is accompanied by a PAP we find evidence consistent with both reduced p-hacking and publication bias.

Data availability:
All the data is collected by the authors. We provide all the data that we have in this folder. A full computational reproducibility is feasible.

We focus on leading economics journals for the years 2018 through 2021. We select the highest 15 journals as ranked using RePEc's Simple Impact Factor (2018 Simple Impact Factor, calculated over the last ten years) excluding any journal that did not publish at least one paper using RCTs.

Sample:
Our final sample includes 314 journal articles.

Data:
For each test statistic, we record how it is reported (e.g., *t*-statistic versus coefficient and standard error). We treat coefficient and standard error ratios as if they follow an asymptotically standard normal distribution. When articles report t-statistics or p-values, we transform them into equivalent z-statistics.

Finally, we collect various contextual data. For each article, we record: the journal and year of publication; the number of authors; gender of authors; the affiliations of authors at time of publication; when and from what institution they graduated; and whether they are editors of an academic journal at the time of publication. The latter information is collected from author websites and curriculum vitae available online. We code top institutions using the highest rated 20 in RePec's ranking of top institutions. The following 20 institutions were coded as top: Barcelona GSE, Boston University, Brown, Chicago, Columbia, Dartmouth, Harvard, LSE, MIT, Northwestern, NYU, Princeton, PSE, TSE, UC Berkeley, UCL, UCSD, UPenn, Stanford, and Yale.

The final codes were prepared by Abel Brodeur and Nikolai Cook December 2023.

Content of this package:

merged.dta	Primary dataset
completeness.dta	Secondary dataset
make_figures.do	Stata do file to make Figures 1-7, App Figures 1-5, 7-11, and prepare data for Appendix Figure A12
make_tables.do	Stata do file to make Tables 3-5, 7-8, Appendix Tables 2-19, and prepare data for Table 6,
figure_a12_input_to_elliot_prereg_0.csv	.cvs file to be used for Appendix Figure A12
figure_a12_input_to_elliot_prereg_1.csv	.cvs file to be used for Appendix Figure A12
figure_a12_elliot_code_modified	R scripts for Appendix Figure A12
ReadMe	This file

Versions and Run-Time

Computations were done using Stata 18.0.

make_tables.do takes 109 seconds to run on an imac using an M1 chip and equipped with 16GB ram.

make_figures.do takes 101 seconds to run on an imac using an M1 chip and equipped with 16GB ram.

Computational Reproducibility:
To computationally reproduce our results, please open the make_figures.do or make_tables.do Stata do-files. Change the path and run the programs. All the tables and figures will be generated, except Appendix Figure A12 and Table 6. Those do-files will clean and prepare the data necessary for those two exhibits.

Use the R scripts “figure_a12_elliot_code_modified” for Appendix Figure A12. See Andrews and Kasy for Table 6 and our do-file make_tables.do for more information.

Andrews, I. and Kasy, M.: 2019, Identification of and Correction for Publication Bias, American Economic Review 109(8), 2766–94