

Scoggins, Bermond; Robertson, Matthew P.

Working Paper

Measuring Transparency in the Social Sciences: Political Science and International Relations

I4R Discussion Paper Series, No. 14

Provided in Cooperation with:

The Institute for Replication (I4R)

Suggested Citation: Scoggins, Bermond; Robertson, Matthew P. (2023) : Measuring Transparency in the Social Sciences: Political Science and International Relations, I4R Discussion Paper Series, No. 14, Institute for Replication (I4R), s.l.

This Version is available at:

<https://hdl.handle.net/10419/268345>

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.



No. 14
I4R DISCUSSION PAPER SERIES

Measuring Transparency in the Social Sciences: Political Science and International Relations

Bermond Scoggins
Matthew P. Robertson

January 2023

I4R DISCUSSION PAPER SERIES

I4R DP No. 14

Measuring Transparency in the Social Sciences: Political Science and International Relations

Bermond Scoggins¹, Matthew P. Roberson¹

*¹School of Politics and International Relations, Australian National University,
Canberra/Australia*

JANUARY 2023

Any opinions in this paper are those of the author(s) and not those of the Institute for Replication (I4R). Research published in this series may include views on policy, but I4R takes no institutional policy positions.

I4R Discussion Papers are research papers of the Institute for Replication which are widely circulated to promote replications and meta-scientific work in the social sciences. Provided in cooperation with EconStor, a service of the [ZBW – Leibniz Information Centre for Economics](#), and [RWI – Leibniz Institute for Economic Research](#), I4R Discussion Papers are among others listed in RePEc (see IDEAS, EconPapers). Complete list of all I4R DPs - downloadable for free at the I4R website.

I4R Discussion Papers often represent preliminary work and are circulated to encourage discussion. Citation of such a paper should account for its provisional character. A revised version may be available directly from the author.

Editors

Abel Brodeur
University of Ottawa

Anna Dreber
Stockholm School of Economics

Jörg Ankel-Peters
RWI – Leibniz Institute for Economic Research

Measuring Transparency in the Social Sciences: Political Science and International Relations

Bermond Scoggins* Matthew P. Robertson†

First version: 16 January, 2022

Latest version: 15 January, 2023

Abstract

The scientific method is predicated on transparency – yet the pace at which transparent research practices are being adopted by the scientific community is slow. The replication crisis in psychology showed that published findings employing statistical inference are threatened by undetected errors, data manipulation, and data falsification. To mitigate these problems and bolster research credibility, open data and preregistration practices have gained traction in the natural and social sciences. However, the extent of their adoption in different disciplines are unknown. We introduce procedures to identify the transparency of a research field using large-scale text analysis and machine learning classifiers. Using political science and international relations as an illustrative case, we examine 93,931 articles across the top 160 political science and international relations journals between 2010 and 2021. We find that approximately 21% of all statistical inference papers have open data and 5% of all experiments are preregistered. Despite this shortfall, the example of leading journals in the field shows that change is feasible and can be effected quickly.

*PhD Candidate, School of Politics and International Relations, Australian National University. Corresponding author. Email: bermond.scoggins@anu.edu.au.

†PhD Candidate, School of Politics and International Relations, Australian National University. Email: m.p.roberston@anu.edu.au

1 Introduction

The Royal Society has as its motto the injunction *Nullius in verba*: “Take nobody’s word for it.” Yet a large portion of published studies in the social sciences demand of the reader exactly this.

Beginning in the 2010s, the open science movement began advocating for the routinization of open science practices such as posting data and code upon a paper’s publication and the preregistration of experiments. These reforms were prompted by large-scale replication failures of prominent psychological studies which highlighted the widespread presence of false positive findings.

Open science practices bolster the credibility of a field and its findings by allowing readers to evaluate the methods by which researchers reach their conclusions. While trust is the currency of every epistemic community, the demand for trust alone weakens credibility. If data and code are available, interested researchers can ensure a finding’s results are computationally reproducible, robust to alternate model specifications, and error free. For experiments, preregistration allows the reader to determine whether there was the selective exclusion of hypotheses, measurements, or statistical analyses that run counter to the author’s favored hypotheses.

Concern for research transparency has become more salient over the past decade as scholars recognize that the accumulation of false positives can drive unsuccessful decision-making and interventions. This leads to inefficient resource allocation and weakens the credibility of a field. In fields like medicine, open science practices have been strongly advocated in recognition of the direct harm that false positives can cause (National Academies of Sciences and Medicine 2021; Baggerly and Coombes 2009). Leading journals in political science and international relations are increasingly mandating the provision of data and code, as well as encouraging the preregistration of experiments.

Political science and international relations appear to have taken open science practices seriously, with high-profile journals and academics endorsing initiatives like the Data Access and Research Transparency (DA-RT) statement. This has led some scholars to believe that the open data problem has mostly been solved. Yet current assessments of the field's progress have been based on relatively small samples.

In a 2020 paper, Grossman and Pedahzur (2020, 1) analyze 92 articles published in the Fall 2019 issues of six journals and argue that the field is now approaching a “replicability utopia,” wherein “virtually every evidence-based article in *leading* (emphasis added) political science journals offers a methodological appendix, a downloadable repository of replication data, or both.” They argue that an increasing number of journals will adopt such practices over the next few years. Their findings show an improvement from the results of studies by Key and Stockemer et al. in previous years (2016; 2018).

Our paper presents the largest-scale study of open science practices in political science and international relations thus far; it is also the first systematic study of the prevalence of preregistration in experiments in these fields. Our study spans the years 2010 to 2021 and includes population-level data, allowing us illustrate trends in specific journals. Documenting such trends is important given the key role played by journals in promulgating and enforcing transparent research norms.

We ask two questions: (1) What proportion of papers that rely on statistical inference make their data and code public? (2) What proportion of experimental studies were preregistered? We gather 93,931 published articles from the top 160 journals ranked by Clarivate's Journal Citation Reports (2020) and use machine learning classifiers to identify either statistical inference or experimental papers.¹ We identify which had open data and preregistration using public application programming interfaces (API), text analysis, and web scraping.

¹A complete list of the journals can be found in the appendix.

In 2020, the percentage of statistical inference papers with open data in political science journals was approximately 34% and 23% for international relations. For all experiments published in 2020, about 11% could be matched to a corresponding preregistration document. While this may be far from the replicability utopia we all hope for, we also show that open science practices have been on an upward trajectory and that significant changes can be made in just a few years.

2 The state of open science practices

Since the onset of the replication crisis, how much of the literature dependent on data and statistical inference still relies solely on reader trust? Extant research on the prevalence of open data practices in political science paints a sobering picture. Stockemer, Koehler, and Lentz's (2018) analysis of 145 quantitative studies published in three journals during 2015 found that only 55% provided original data and 56% provided code.² An earlier analysis, conducted on 494 quantitative articles in six leading political science journals between 2013 and 2014, found that full replication materials (data and code) were available for only 58% of papers (Key 2016).³ In their effort to reproduce studies using instrumental variable strategies in three political science journals, Lal et al. (2021) found that of the 115 papers slated for analysis only 65 (or 56%) had all the files necessary for computational reproducibility.

The trend holds in other disciplines. A random sample of 250 psychology papers published between 2014 and 2017 estimated that 14% of papers shared research materials, 2% provided original data, and 1% shared their code (Hardwicke et al. 2021). Preregistration was rare (3%). In ecological research, while 79% of articles used original data, only 27% posted code and data, despite three quarters of ecology journals mandating or encouraging code sharing (Culina et al. 2020). In the Reproducibility Project: Cancer Biology, which investigated

²The three journals are *Electoral Studies*, *Party Politics*, *Journal of Elections*, *Public Opinion*, and *Parties* were analysed.

³The six journals analysed were *American Political Science Review*, *American Journal of Political Science*, *British Journal of Political Science*, *International Organization*, *Journal of Politics*, *Political Analysis*.

the replicability of pre-clinical cancer biology research, a major challenge faced by the team was accessing the data and code necessary to conduct a direct replication. Out of the 193 experiments targeted for replication, data from 68% of the experiments could not be obtained by the authors (Errington et al. 2021). This was a major factor that contributed to the team only replicating 50 of the 193 planned experiments.

A tonic for many of these problems is straightforward: replication materials for all quantitative studies and preregistration for experiments. Replication materials and preregistration militates against questionable research practices (QRPs) that lead to false positives by constraining researcher degrees of freedom and ensuring that key decisions made in the analysis process are transparent to peers.

In the behavioural sciences, false positives can arise from decisions that are rationalised as legitimate by authors: failing to report all dependent variables in a study, collecting more data after seeing whether the results were statistically significant, failing to report all conditions, stopping data collection after achieving the desired result, rounding down p-values, selectively reporting studies that ‘worked’, selectively excluding observations, and claiming an unexpected finding was predicted (or hypothesising after results are known). However, these practices obfuscate the uncertainty around a particular set of claims and mislead readers into being overconfident about a study’s conclusions.

The use of QRPs appears to be widespread in many of the social sciences. In psychology, a survey of researchers by John, Loewenstein, and Prelec (2012) showed that 78% admitted to not reporting all dependent variables, 72% admitted to collecting more data after peeking at the results, and 67% admitted to selectively reporting significant studies. The story is similar for criminology – 53% of researchers admitted to underreporting results, 43% admitted to omitting non-significant studies or variables, and 39% admitted to switching analysis selectively to reach statistical significance (Chin et al. 2021). Other methods of detecting publication bias, such as analysing sets of studies or literatures using a p-curve or z-curve,

reveal extensive clustering of p-values (z-scores) just past $p < 0.05$ (Simonsohn, Nelson, and Simmons 2014; Bartoš and Schimmack 2020). Examples of these problems in the behavioural and social sciences range from the power posing literature (J. P. Simmons and Simonsohn 2017) to economic research using instrumental variables and difference-in-differences (Brodeur, Cook, and Heyes 2020).

In recognition of these problems, professional organisations in political science and international relations, including the American Political Science Association (APSA), have led efforts to increase the availability of data and code that accompany published papers. The DA-RT statement developed by the APSA council in 2014 involved a commitment by journal editor signatories to increase the availability of data “at the time of publication through a trusted digital repository”, as well as require authors to “delineate clearly the analytic procedures upon which their published claims rely, and where possible to provide access to all relevant analytic materials” (Statement 2015).

While there was an intramural debate about how DA-RT standards would affect qualitative work, given the heterogeneity of interview data and other forms of qualitative analysis, we bypass these arguments in this paper by focusing exclusively on papers relying on statistical inference.⁴ It is relatively straightforward for researchers using statistical inference to release the very data and code that were necessary to produce the results in their papers. As Key (2016) notes, the internet has reduced the cost for journals to set up Dataverse repositories and made it easier for researchers to share their data and code. Rising usage of free statistical programming software, such as R and its desktop application RStudio, also reduces barriers to computational reproducibility.

The 27 journal editors who adopted the statement agreed to implement reforms by January 2016. Of the 16 DA-RT signatory journals in our dataset, two made no change in

⁴Summaries of these debates can be found in Lupia and Elman (2014) and on the Dialogue on DART website (“Perspectives on DA-RT,” n.d.).

practice and a further four have data and code that is difficult to accurately estimate.⁵

3 The need for open data

3.1 Uncovering data errors and misinterpretation

Errors in data or the misreporting of p-values or test statistics invariably occur in research and can go undetected by an article’s authors or peer reviewers. These problems, if addressed, may substantively alter an article’s conclusions or produce null rather than positive results. Reporting errors in regression coefficients or test statistics occur frequently.

Using the R package *statcheck*, which extracts statistical results from published articles, Nuijten et al. (2016) found that over half of the articles published in eight major psychology journals between 1985 and 2013 reported at least one p-value that that was inconsistent with its test statistic and degrees of freedom. One in eight papers reported a “grossly inconsistent” p-value that threatened the article’s statistical conclusions. Such p-values were also more likely to be statistically significant than not. It is important to note that implausible statistical results do not necessarily imply data falsification.

Access to the original data can help determine whether errors are trivial, and contribute to retraction efforts if they are not (“Retraction Notice” 2020). In some cases, access to the data allows for detailed concerns with a paper’s analysis to be illustrated without the journal believing a retraction is warranted – as occurred for instance with Joseph Hilgard’s (2020, 2021) interactions with *Aggressive Behavior*.

3.2 Identifying model misspecification

Researchers have tremendous flexibility in deciding how to collect data and which statistical models should be specified to analyse them. Andrew Gelman has termed this process the

⁵We discuss these issues further in the results section.

‘garden of the forking paths’ (2014): some set of decisions might yield a positive result, while another set of equally justifiable decisions might lead to a null result. The replication crisis has shown that it is a mistake to view a single study or set of statistical analyses as a definitive answer to a given theory or claim — the scientific process should instead be iterative, exploratory, and cumulative (Tong 2019). Further, modelling involves assumptions about the underlying data generating process that researchers are not always able to meet; often they cannot even know whether they have met those assumptions (Neumayer and Plümper 2017). An example of such difficult assumptions are how the concept of interest should be measured and how errors in its measurement should be minimized (Amrhein, Trafimow, and Greenland 2019).

Open data can address the problem of model misspecification. Since researchers cannot anticipate changes to methodological best practices, replication materials allow scholars to make adjustments if best practices change. Lenz and Sahn (2021) find, for instance, that over 30% of observational studies published in the *American Journal of Political Science* rely on suppression effects to achieve statistical significance. This means that a single control variable, or combination of them, included in a regression may increase the predictive validity of the main independent variable and result in it being statistically significant when it otherwise would not have been in a bivariate specification. Lenz and Sahn argue that while there may be good theoretical reasons for suppression effects, none of articles they examined justified or disclosed their presence. Open data, in this instance, allowed Lenz and Sahn to uncover the prevalence of suppression effects and thus contribute to improving quantitative methodology.

The model specification of many studies using instrumental variables have also been questioned. Lal et al. (2021) examine this by reproducing and analyzing 65 papers in three leading political science journals.⁶ The authors show that the papers often overestimate the strength of the instrumental variables and make other misspecifications that result in biased estimates.

⁶*American Political Science Review*, the *American Journal of Political Science*, and the *Journal of Politics*. 10

3.3 Extension and learning

Open data can contribute to extending scientific knowledge by allowing researchers to more easily build off existing datasets. For example, researchers can use open data to better perform a direct replication of an experiment or undertake a systematic review or metaanalysis. The process of reproducing published statistical results can also aid student education and training (Janz 2016).

3.4 Exposing data falsification

In the most egregious cases, open data allows researchers to investigate and expose data falsification. High-profile exposures of data falsification include the LaCour and Green (2014) case in political science, and the Shu et al. (2012) case in psychology. Both rested on investigator access to the original data. While presumably data falsification is exceedingly rare, there is no way to know its extent given the general absence of replication materials in the first place.

In the case of Shu et al. (2012), the exposure of falsified data in a field experiment only arose after the original authors conducted a series of failed direct and conceptual replications and posted the anomalies alongside their replication materials (in Kristal et al. 2020). This then allowed a team of anonymous researchers to scrutinize the data and publicize their findings via Nelson, Simonsohn, and Simmons (2021). In this instance a highly influential study took nine years to be exposed as fraudulent, an ordeal that might have been avoided entirely, or taken place much sooner, if the data was available for scrutiny at the time of publication.⁷

4 The need for preregistration

⁷Kristal et al. (2020) make note that their original study influenced policy.

4.1 Distinguishing confirmatory from exploratory analysis

Preregistration means that researchers specify their hypotheses, measurements, and analytic plans prior to running an experiment. This commits researchers to making theoretical predictions before they can view the data and be influenced by observing the outcomes (J. P. Simmons, Nelson, and Simonsohn 2011; J. Simmons, Nelson, and Simonsohn 2021). By temporally separating predictions from the data that tests their accuracy, there is much less flexibility for both post hoc theorising and alterations of statistical tests to fit the prediction.

Post hoc theorising, also known as hypothesising after the results are known (HARKing), is an example of circular logic — the researcher conducts many tests when exploring a dataset, the data reveals a relationship that can be made into a hypothesis, and that hypothesis is ‘tested’ on the data that generated it (Nosek et al. 2018). But the diagnosticity of a p-value is in part predicated on knowing how many tests were performed: when an exploratory finding is reported as a prediction, the normal methods employed to evaluate the validity of a hypothesis — such as whether the p-value is less than 0.05 (i.e. null hypothesis significance testing) — no longer hold. P-values in that case have unknown diagnosticity (Nosek et al. 2018). Thus, post hoc theorising and selective reporting greatly contribute to false positives.

4.2 Reducing the selective reporting of results

The selective reporting of statistical tests and results can occur for a variety of reasons. There are numerous legitimate ways of analyzing data, and this makes selective reporting seem justifiable. Danger arises when researchers convince themselves that the measures and tests lending evidence to their claims are the ‘right’ ones, while unjustifiably failing to report measures and tests that did not support the favored hypothesis.

A rare testimonial illustrating these problems is found in Carney (2015), whose studies on power posing led to multiple failed replications and the detection of selective reporting (Carney, Cuddy, and Yap 2010; Carney 2015; J. P. Simmons and Simonsohn 2017). Aside

from the fact that the studies were statistically underpowered, Carney writes that her team excluded observations deemed outliers and “ran subjects in chunks and checked the effect along the way” (Carney 2015, 1–2). To her team, this “did not seem like p-hacking” and just “a way of saving money” (Carney 2015, 1). She outlines how her team reported a statistical test that yielded the desired p-value of 0.05 despite being less appropriate than another test where the p-value was 0.052. Her team also peeked at different measures of the dependent variable and reported only the ones that ‘worked’ (Carney 2015, 2).

Selectively reported experimental studies often result in overconfident theoretical claims and inflated effect sizes when compared to replications. The Open Science Collaboration (2015) and Many Labs studies (2014, 2018) have shown that the effect sizes in highly powered replications are much smaller than those in the original studies. In medicine, Kaplan and Irvin (2015) discuss how the number of null results markedly rose after the National Heart, Lung, and Blood Institute (NHLBI) instituted a requirement in 2000 that randomised control trials preregister their primary outcomes in advance on ClinicalTrials.gov. They attribute this fall to the role preregistration plays in limiting researcher flexibility to pick and choose outcomes. Of the 25 registered NHLBI trials published after 2000, researchers identified positive, statistically significant effects for cardiovascular-related variables that were not the primary outcome in 12 studies – had preregistration not been required readers might not have this important piece of information (Kaplan and Irvin 2015, 8). Similarly, the registered report publication format, aimed at diminishing the incentives for selective reporting by initiating peer review and agreeing to publish *prior* to data collection, appear to have far fewer positive results than standard articles (Scheel, Schijen, and Lakens 2021).

The primary purpose of preregistration is to provide journal reviewers and readers the ability to transparently evaluate predictions and the degree of flexibility researchers had to arrive at their conclusions (Lakens 2019; Claesen et al. 2019; Franco, Malhotra, and Simonovits 2014). It is up to the reader to determine whether preregistered studies followed

their preregistration plan and adequately justified deviations.

Insufficiently detailed preregistration reports are a problem in some of those registered with Experiments in Governance and Politics (EGAP) and the American Economic Association (AEA). Ofofu and Posner (2020, 10) outline four requirements necessary for preregistration to serve its intended purpose of adequately ‘tying the hands’ of researchers – “specifying a clear hypothesis, specifying the primary dependent variable(s) sufficiently so as to prevent post-hoc adjustments, specifying the treatment or main explanatory variable so as to prevent post-hoc adjustments, and spelling out the precise statistical model to be tested including the functional forms and estimator”. A representative sample of 195 preregistration reports collected by Ofofu and Posner (2020) find that approximately half met all four of the key requirements. A further third contained three of the four.

The replication crisis has altered best practices and changed the habits of many researchers in the behavioural sciences. As we show below, preregistration is not yet the norm in political science and international relations. The conclusions from many studies relying on statistical inference, even some that have been preregistered on a registry, remain exposed to the statistical pitfalls described above.

5 Methods

Our study design called for a comprehensive analysis of population-level data, yet our populations — (1) papers using data and statistics, and (2) original experiments — were embedded in a larger population of *all* political science and international relations publications in target journals. We downloaded all of the journals’ papers from 2010 to 2021. Once we had these papers locally, we identified the data, statistical, and experimental papers through dictionary-based feature engineering and machine learning. We then used public APIs, web scraping, and text analysis to identify which of the studies had replication materials. We outline this process below.

5.1 Phase one: gathering and classifying the papers

We used Clarivate’s 2021 Journal Citation Report to identify target journals. We filtered for the top 100 journals in both political science and international relations, and combined the two lists for a total of 176 journals.⁸

With this list, we used the Crossref API to download all publication metadata. We were able to obtain records for 162 journals. This resulted in over 445,000 papers, which we then filtered on Crossref’s `published.print` field for 2010 and onwards, resulting in 109,553 papers. As of January 15, 2023 we were able to obtain 93,931 of these PDFs, and we use this as the denominator in the study. We converted the PDFs to plaintext using a combination of UNIX command line utilities and optical character recognition software.

Identifying the papers that relied on data, statistical analysis, and experiments was an iterative process. In each case we read target papers and devised a dictionary of terms meant to uniquely identify others like them. We extensively revised these dictionaries to arrive at terms that seemed to maximally discriminate for target reports. The dictionaries eventually comprised 52, 180, and 133 strings, symbols, or regular expressions for the three categories respectively.

The dictionaries were then used with custom functions to create document feature matrices (DFM), where each paper is an observation, each column a dictionary term, and each cell a count of that term.⁹ The DFM format made the papers amenable to large-scale analysis. In machine learning parlance, this process is known as feature engineering.

For the first research question – examining the presence of replication code and data in papers involving statistical inference – we hand-coded a total of 1,624 papers with boolean categories and identified 585 that relied on statistical inference. We defined statistical

⁸As some journals publish both political science and international relations articles, the top 100 journals in each category overlapped.

⁹A custom function was preferable to existing text analysis libraries like `quanteda` because of our need to capture regular expressions and asterisks.

inference papers as any that involved mathematical modeling of data. This definition is meant to capture a simple idea: mathematical modeling requires computer instructions that perform functions on numbers. In the absence of replication materials, these transformations cannot be exactly reproduced by readers. We also developed a dictionary of 35 terms for formal theory papers, because we wished to exclude papers that did not apply a model to real-world data.

For the second question — examining what proportion of experiments were preregistered — we hand-coded 518 papers with a single boolean category: whether the paper reported one or more original experiments. We defined this as any article containing an experiment where the researchers had control over treatment.

We then trained two machine learning models — the Support Vector Machine (SVM) and Naive Bayes (NB) binary classifiers — to arrive at estimates for the total number of statistical inference and experimental papers.¹⁰ SVMs are a pattern recognition algorithms that give binary classifications to variables in high-dimensional feature space by finding the optimal separating boundary between labeled training data (James et al. 2021, 337–72; Cristianini and Shawe-Taylor 2000). The NB family of algorithms calculate the posterior probability of a given classified input based on the independent probability of all the values of its features; it then applies this trained algorithm to classify new inputs (Rhys 2020, 135–67).

We report the SVM model results both for their greater accuracy and due to our theoretical prior that the model would be more suitable for a high-dimensional classification problem. For the first research question, our SVM model achieved 92.35% accuracy for statistical papers. For the classifying experiments, the accuracy was 86.05%. In Appendix 1 we report the confusion matrices, hyperparameter tuning data, NB models, and initial exploratory analysis using logistic regression to create predictive probability plots.

¹⁰As an additional robustness check to predict open data and statistical inference papers, we estimated a series of bivariate logistic regressions using the same DFMs. The predicted probability plots can be found in the appendix. These plots give a lower estimate than the machine learning models, though they are in the same broad range.

The application of the SVM model to the full dataset of 93,931 publications leads to an estimate of 24,026 using statistical inference.

The identification of experimental papers proceeded slightly differently. Rather than beginning with the full corpus, we first filtered for only the papers that included the word “experiment” over five times (4,835). We then ran the SVM classifier on this subset. The resulting estimate was 2,552 papers reporting experiments.

5.2 Phase two: Identifying open data and preregistrations

We attempted to identify open data resources in seven ways.

1. Using the Harvard Dataverse API, we downloaded all datasets held by all journals in our corpus who maintained their own, named dataverse (n=20);
2. We queried the Dataverse for the titles of each of the 109,553 papers in our corpus and linked them to their most likely match with the aid of a custom fuzzy string matching algorithm. We validated these matches and manually established a string-similarity cut-off, setting aside the remainder;
3. We extracted from the full text of each paper in our corpus the link to its dataset on the Dataverse (1,142; note this had significant overlap with the results of the first and second queries);
4. We downloaded the metadata listing the contents of these datasets, to confirm firstly that they had data in them, and secondly that it did not consist of only pdf or doc files. In cases where a list of metadata was not available via the Dataverse API, we scraped the html of the dataset entry and searched for text confirming the presence of data files;
5. We used regular expressions to extract from the full text papers references to “replication data,” “replication materials,” “supplementary files” and similar terms, then searched in the surrounding text for any corresponding URLs or mentions of author websites;
6. We searched all of the full text papers for references to other repositories, including

Figshare, Dryad, and Code Ocean.

7. As additional validation for DA-RT signatory journals specifically, we downloaded the html file corresponding to each article and/or the html file hosting supplemental material (n=2,284), then extracted all code and data-related file extensions to establish their open data status.

We attempted to identify preregistration of experiments in the following ways:

1. We used regular expressions to extract from all of the experimental papers sentences that referred to “prereg” or “pre-reg”, as well as any references to commonly used preregistration servers (OSF, EGAP, and AsPredicted), and then searched for the availability of the corresponding link to validate that the preregistration had taken place. Parts of this process — for instance, searching author names in the Experiments in Governance and Politics (EGAP) registry to look for the corresponding paper — involved time-consuming detective work;
2. We downloaded all EGAP preregistration metadata in JSON format from the Open Science Foundation Registry (<https://osf.io/registries/discover>), extracted from this file all osf.io links and unique EGAP registry IDs, and used command line utilities to search for them through the corpus of all the papers.

We did not examine whether the published report conformed to the preregistration plan.

6 Results

Statistical inference papers are infrequently accompanied by the datasets or code that generated their findings. For the 12 year period under observation, we were able to match 21% of statistical inference articles to data repositories (overwhelmingly the Harvard Dataverse). As Figure 1 shows, the proportion of articles from political science journals with open data is higher than those from international relations. Encouragingly, the percentage of open data in

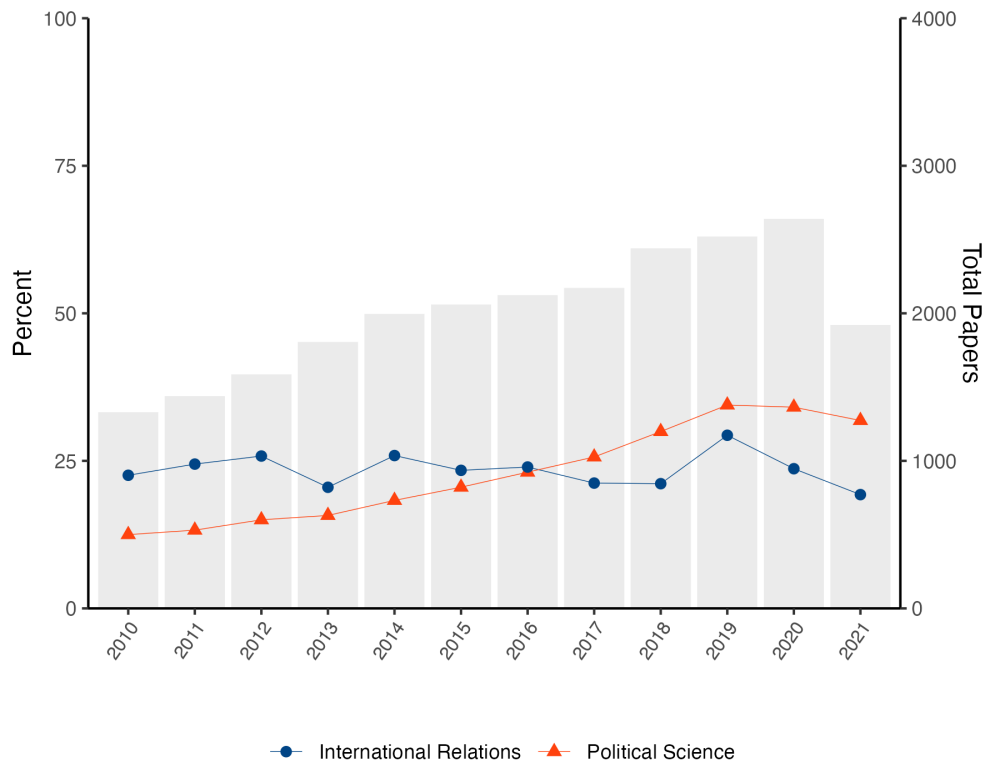


Figure 1: Open data in statistical inference papers by year

political science has increased between 2010 and 2021 – rising steadily from about 12% to 34% during this period. International relations papers, however, have remained stable and open data percentages have fluctuated between 22 and 29%.

The total number of statistical inference papers have gradually increased during the 12 year period. In 2010, we found 1,329 papers and 2,640 in 2020 – the last year with complete data. This supports King’s (1990) observation that political science and international relations have long been disciplines increasingly concerned with quantitative methods.¹¹ While the percentage of papers with open data have increased, so too have the absolute number of statistical papers without it. There are simply more published papers making inferences based on hidden data.

¹¹Gary King illustrated that by 1988 almost half of publications in the *American Political Science Review* were quantitative.

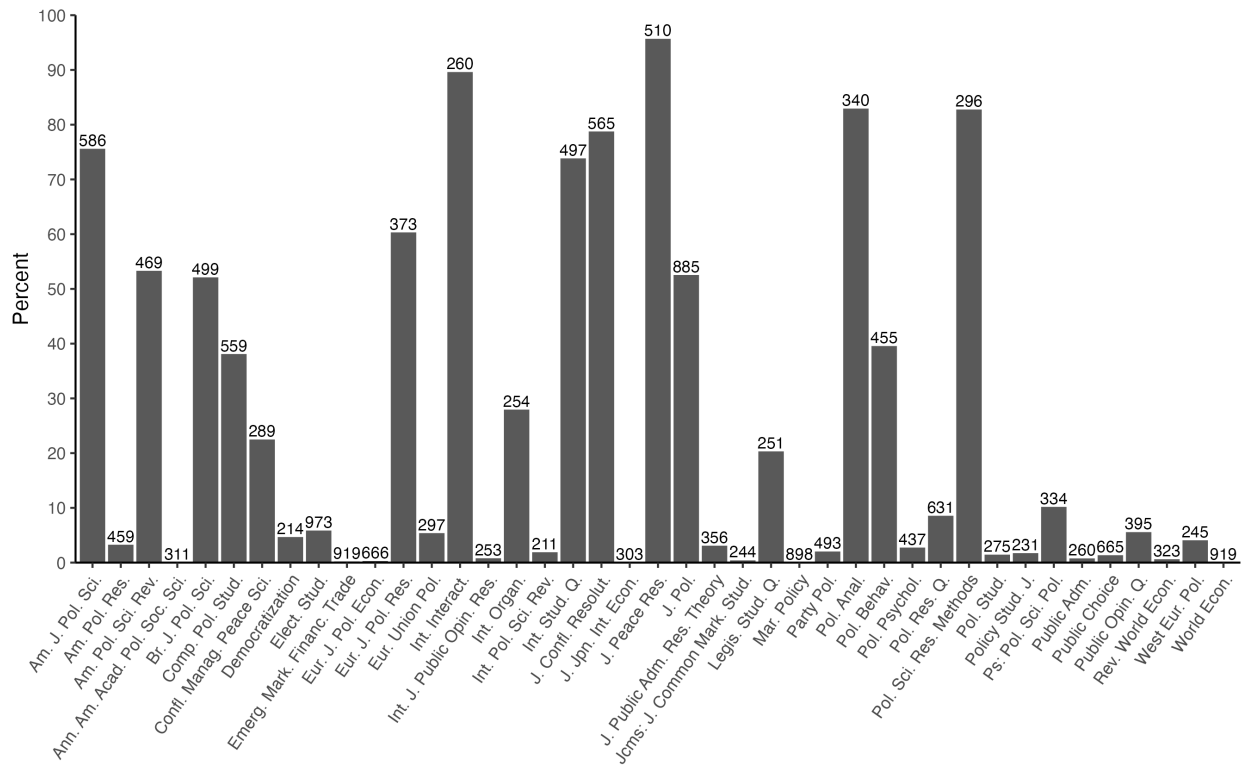


Figure 2: Open data in statistical inference papers by journal (with over 200 papers)

There are significant differences in open data practices between journals. Figure 2 displays the percentage of statistical inference papers with open data in the 41 journals with over 200 such papers.¹² The number above each journal’s bin represents the number of statistical inference papers detected by the support vector machine classifier. Of the 41 journals, 11 have over 50% open data, and 16 have over 20%.¹³

The effectiveness of the DA-RT statement on journal open data practices is illustrated in Figure 3, which displays the percentage of statistical inference papers with open data by

¹²The cutoff was established to focus on journals who publish more quantitative papers and for ease of viewing – the graph with all 158 journals with at least one statistical inference paper is very large.

¹³The journals with over 50% open data are the *American Journal of Political Science*, the *American Political Science Review*, the *British Journal of Political Science*, *European Journal of Political Research*, *International Interactions*, *International Studies Quarterly*, *Journal of Conflict Resolution*, *Journal of Peace Research*, *Journal of Politics*, *Political Analysis*, and *Political Science Research and Methods*. Those with over 20% open data include the aforementioned journals as well as *Comparative Political Studies*, *Conflict Management and Peace Science*, *International Organisation*, *Legislative Studies Quarterly*, and *Political Behaviour*.

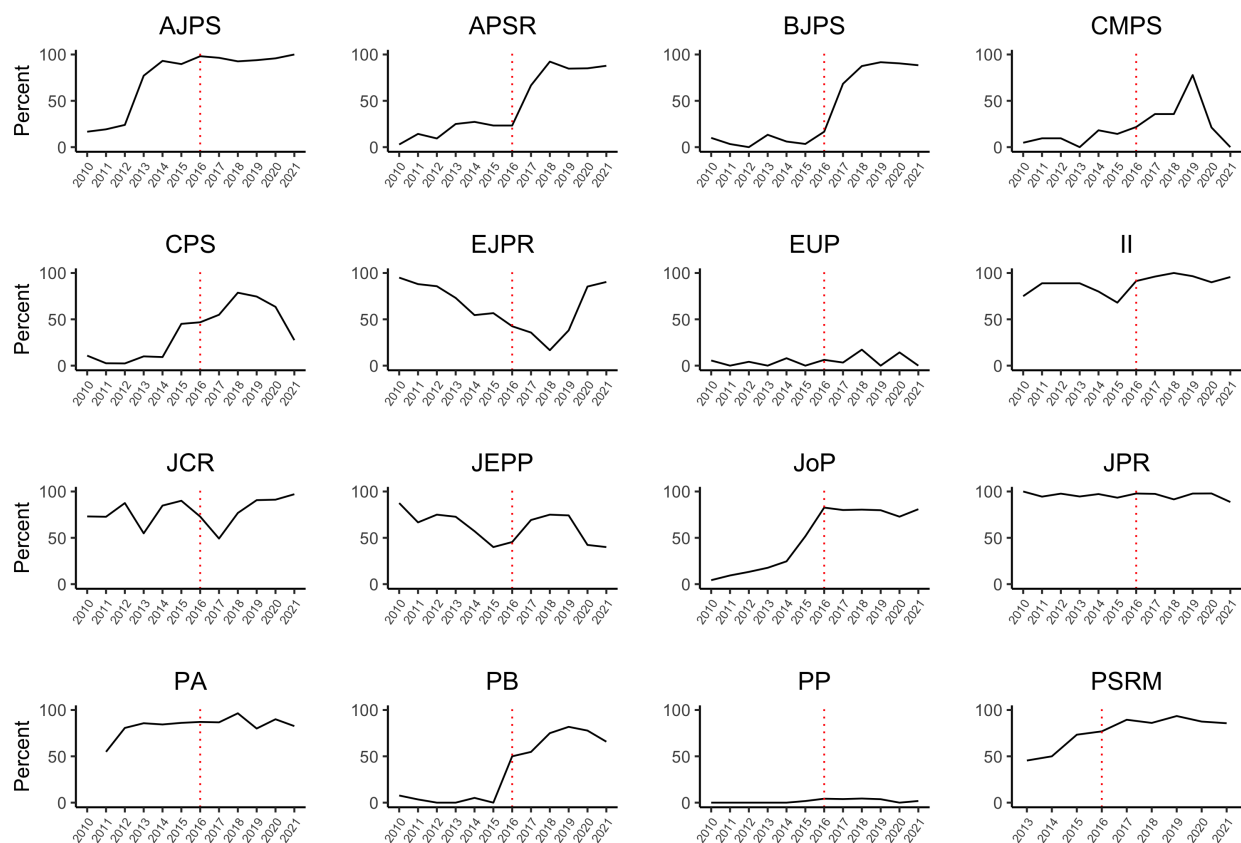


Figure 3: Open data in statistical inference papers by year published in 16 of the 27 journals signatory to the DA-RT statement

year in each of the 16 DA-RT signatory journals we consider.¹⁴

Four journals – *American Journal of Political Science*, *International Interactions*, *Political Analysis*, and *Political Science Research and Methods* – already made significant progress prior to the release of the initial DA-RT guidelines in 2014. Many of the remaining journals either made significant progress in 2016 or shortly thereafter.

One caveat is that is that 2 of the 16 journal signatories have consistently low levels of open data even after DA-RT reforms were agreed to commence on January 15, 2016. The extent of transparent practices in three other journals – *Journal of European Public Policy*,

¹⁴A total of 27 journals signed the DA-RT statement. The majority of these journals publish quantitative research (as can be seen in Figure 2). Note that there are actually 20 DA-RT signatory journals in our dataset, but four of them have an insignificant number of statistical inference publications and so we omit them from the analysis.

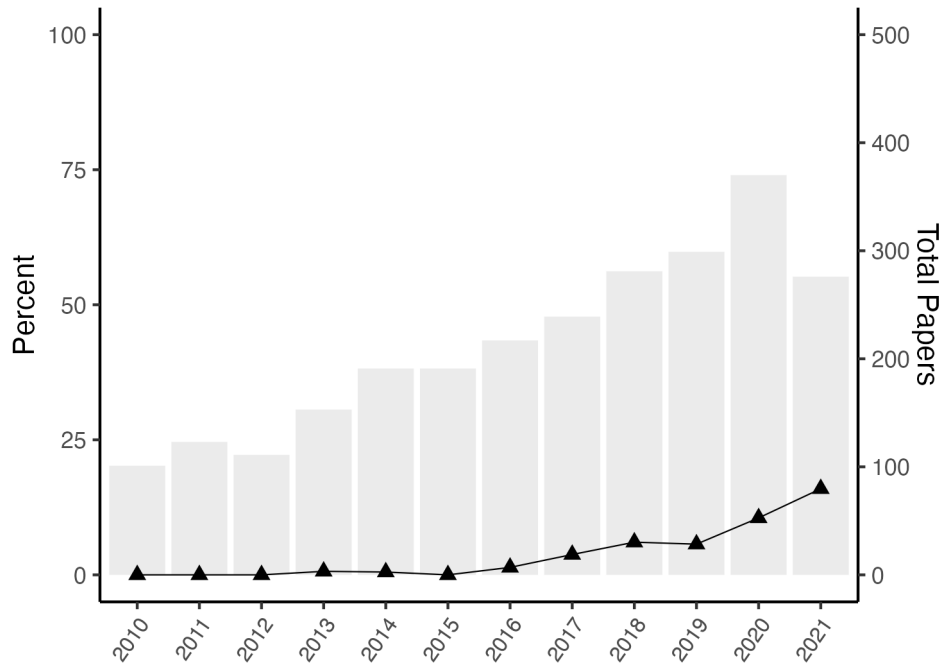


Figure 4: Preregistration in experiments by year

European Journal for Political Research, and *Conflict Management and Peace Science* – was more difficult to determine, given they did not use the Harvard Dataverse. Our attempt to estimate data and code availability for such journals, noted in point seven of phase two of the methods section, appears to produce unreliable and puzzling results.¹⁵

The preregistration of experiments is rare in political science and international relations journals. Figure 4 shows that the first preregistered study in the dataset was in 2013, and that the rate of preregistration only began climbing in 2016. The proportion of experiments that were preregistered for the entire period is approximately 5%; the annual rate has slowly risen to 16% in 2021.

Figure 5 shows the percentage of experiments that were preregistered in the 29 journals with more than 20 experiments. Only the *American Political Science Review* exceeds 20%.

¹⁵These results for the three DA-RT journals, as well as all journals more broadly, should be treated as preliminary and may change as we revise and improve our methods. We are in the process of contacting the three journals to arrive at a more accurate assessment.

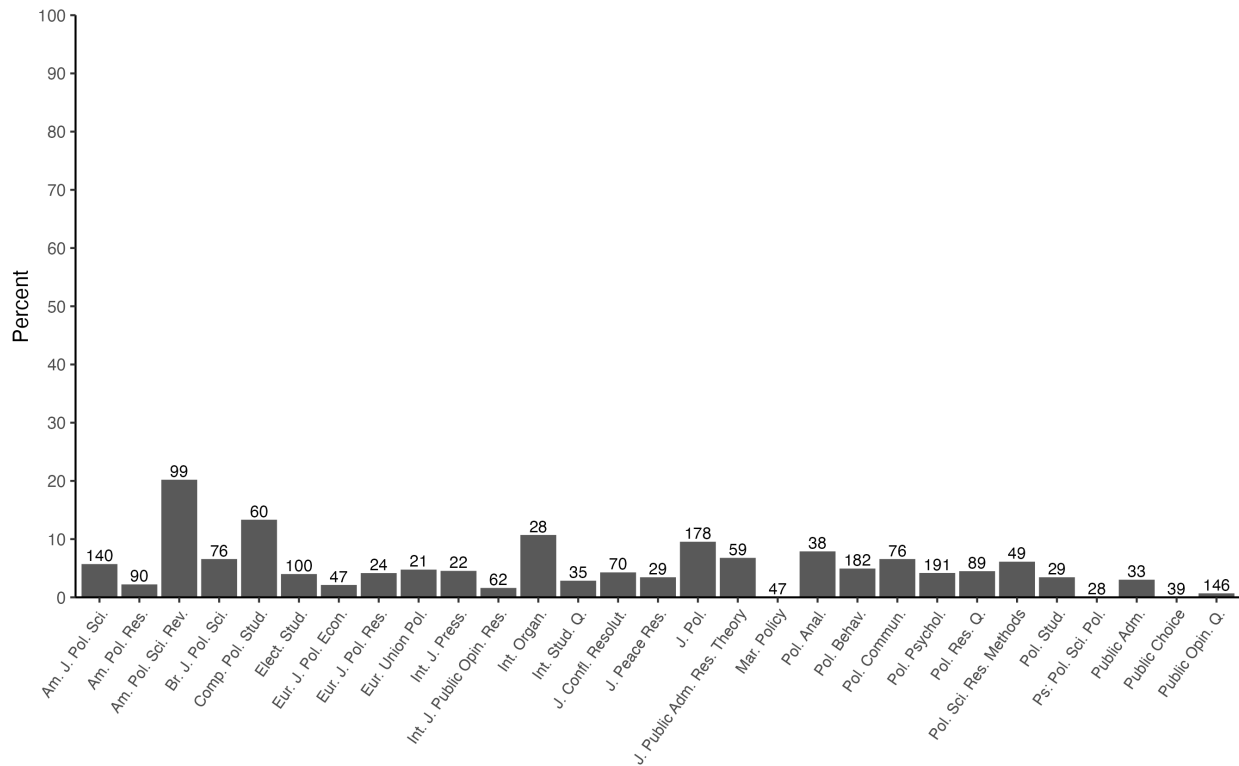


Figure 5: Preregistration in experiments by journal (with over 20 papers)

Unlike with open data, when it comes to preregistration the differences between journals are small. Of the experiments published in *Political Psychology* and *Political Behaviour*, the two journals with the most experiments that bridge the gap between political science and psychology, only four and five percent respectively are preregistered.

Prior to the replication crisis at the beginning of the 2010s, there were no organized attempts at enforcing preregistration or using registered reports as a way of curbing researcher flexibility and its attendant QRPs. As psychology was among the first of the sciences to reckon with its methodological issues, brought to light in part by such articles as Simmons, Nelson, and Simonsohn’s (2011), it is logical that it took several years for these new practices to be adopted in contiguous disciplines like political science and international relations. But our data illustrate that significant improvements must be made in order for experiments in these fields to meet current methodological best practices.

7 Discussion

Scientists must carry out their work while simultaneously signalling and vouchsafing for its credibility. For the pioneers of the scientific method in 17th century Europe, this included an ensemble of rhetorical and social practices, including the enlistment of trusted witnesses to testify that experiments had in fact taken place as claimed – this is what Shapin refers to as the moral economy of science (Shapin 2018, 84, 107–8; 1995).

In the digital age, we argue that the credibility of social science must largely rest on computational reproducibility. The same goes for preregistration and experiments. Adhering to these practices ensures other social scientists can check and reproduce the findings, that the findings are valid, and also demonstrates a commitment to the norm of science as a shared enterprise.

The chief reason for depositing code and data is not for signalling: Open science practices provide the reader with an opportunity to transparently evaluate the evidence for a set of claims and scrutinise an article for any of the myriad problems that plague the use of data and statistical models. An interested reader could investigate an article’s data and code for errors, determine whether results are robust to different model specifications, or, in rare cases, detect data falsification. For experiments, the published paper can be compared to the preregistration document to determine whether there were any unjustified deviations.

Our findings show that political science and international relations are not currently living up to these best practices. For the approximately 25,000 statistical inference papers in the dataset, we could only identify approximately 21% that had a corresponding data repository. Despite improvement in most years, change has not been uniform across the discipline — most of the progress has been made by a handful of the highest impact factor journals. In 2020, for example, 16 out of the 52 journals with over 20 statistical inference papers had an open data percentage over 50% (see Figure A6) – 20 journals had an open data percentage over 20%. We could not locate data or code for two of the 16 DA-RT signatories

in our dataset.

Universal open data is a collective action problem, and it is the responsibility of journals to foster and enforce these disciplinary norms. In the absence of that, individual researchers do not always share data, and requesting it can sometimes be mistaken as a gesture of challenge rather than collegiality. As Simonsohn (2013) notes, the modal response to his requests for original data was that the data was no longer available. We suspect that variation in open data practices between journals reflects differences in journal editors' views of its importance for research credibility.

The DA-RT initiative sparked spirited debate in the field about the provision of data and code — but the same cannot be said for preregistration. Experiments are rarely preregistered. Of the roughly 2,552 experiments in our dataset, 5% are preregistered. Given that the use of experiments only began to take off in 2014, as shown in Figure 3, the proportion of preregistered experiments in the literature is understandably low. Fortunately, the trend is positive. Two journals of 26 with more than 5 published experiments had a preregistration percentage of over 30% in 2020 (see Figure A7).

Identifying whether an experiment had a corresponding preregistration report was at times difficult. Numerous experiments made no mention of their preregistration report in the manuscript despite having one listed in a repository. Locating it was also difficult given changing manuscript titles and authors. Their omission in the manuscript is likely due to the fact that many journal editors do not determine whether an experiment has a preregistration or pre-analysis plan or request their disclosure.¹⁶

The difficulty of matching an experiment with its preregistration report is far smaller than matching a manuscript to a concealed preregistration report. A unique and unanticipated problem we found were authors publishing a study where they omitted any reference to a preregistered experiment — ostensibly due to null findings. Byun, Kim, and Li (2021) use their

¹⁶Journals like the *Journal of Politics* require authors to disclose a preregistration report or justify why they did not preregister their experiment.

survey data to make descriptive claims while failing to discuss the design or results of their experimental manipulation (Kim, Byun, and Li 2021). It is not clear whether their results failed to further their own argument or were possibly disconfirmatory. In either situation, readers are not permitted to transparently evaluate the strength of their claims.

Peer reviewers and readers of published works routinely examine whether a theory or explanation has appropriate evidence; whether the measurements are valid and reliable; whether the model has been appropriately specified. Here, we prompt referees and readers to also begin asking: (1) Are the replication materials on the Harvard Dataverse or some other reliable repository? (2) Is the paper computationally reproducible based on those materials? (3) If an experiment, was it preregistered? (4) Does the analysis in the experimental paper follow the preregistration plan and are deviations from that plan justified?¹⁷ We hope that evaluating scientific research in this manner will help move readers away from trusting research in the absence of open science practices to a more informed trust in their presence.

¹⁷For experiments, we acknowledge that these are by no means definitive criteria on which to judge the trustworthiness of a paper or finding. These practices should accompany efforts to build confidence in a finding through direct and conceptual replications.

8 Acknowledgements

Both authors acknowledge support from the Australian Government Research Training Program Scholarship, the ANU Library, Taylor & Francis, and TeamViewer AG. Matthew Robertson acknowledges support from the Google Cloud Research Credits program (award GCP19980904), and a Victims of Communism Memorial Foundation fellowship.

9 References

- Amrhein, Valentin, David Trafimow, and Sander Greenland. 2019. “Inferential Statistics as Descriptive Statistics: There Is No Replication Crisis If We Don’t Expect Replication.” *Am. Stat.* 73 (sup1): 262–70.
- Baggerly, Keith A., and Kevin R. Coombes. 2009. “Deriving chemosensitivity from cell lines: Forensic bioinformatics and reproducible research in high-throughput biology.” *The Annals of Applied Statistics* 3 (4): 1309–34. <https://doi.org/10.1214/09-AOAS291>.
- Bartoš, František, and Ulrich Schimmack. 2020. “Z-Curve 2.0: Estimating Replication Rates and Discovery Rates.” <https://psyarxiv.com/urgtn/download?format=pdf>.
- Brodeur, Abel, Nikolai Cook, and Anthony Heyes. 2020. “Methods Matter: P-Hacking and Publication Bias in Causal Analysis in Economics.” *Am. Econ. Rev.* 110 (11): 3634–60.
- Byun, Joshua, DG Kim, and Sichen Li. 2021. “The Geopolitical Consequences of COVID-19: Assessing Hawkish Mass Opinion in China.” *Political Science Quarterly* 136 (4): 641–65.
- Carney, Dana. 2015. “My Position on ‘Power Poses.’” Website.
- Carney, Dana, Amy Cuddy, and Andy Yap. 2010. “Power Posing: Brief Nonverbal Displays Affect Neuroendocrine Levels and Risk Tolerance.” *Psychol. Sci.* 21 (10): 1363–68.
- Chin, Jason M, Justin T Pickett, Simine Vazire, and Alex O Holcombe. 2021. “Questionable Research Practices and Open Science in Quantitative Criminology.” *J. Quant. Criminol.*, August.
- Claesen, Aline, Sara L B T Gomes, Francis Tuerlinckx, and Wolf Vanpaemel. 2019. “Preregistration: Comparing Dream to Reality.”
- Clarivate. 2020. “Journal Citation Reports.” 2020.
- Cristianini, Nello, and John Shawe-Taylor. 2000. *An Introduction to Support Vector Machines and Other Kernel-Based Learning Methods*. Cambridge University Press.
- Culina, Antica, Ilona van den Berg, Simon Evans, and Alfredo Sánchez-Tójar. 2020. “Low Availability of Code in Ecology: A Call for Urgent Action.” *PLoS Biol.* 18 (7): e3000763.
- Errington, Timothy M, Alexandria Denis, Nicole Perfito, Elizabeth Iorns, and Brian A Nosek. 2021. “Challenges for Assessing Replicability in Preclinical Cancer Biology.” *Elife* 10 (December).
- Franco, Annie, Neil Malhotra, and Gabor Simonovits. 2014. “Unlocking the File Drawer.” *Science* 345 (6203): 1502–5.
- Gelman, Andrew, and Eric Loken. 2014. “The Statistical Crisis in Science: Data-Dependent Analysis—a ‘Garden of Forking Paths’—Explains Why Many Statistically Significant Comparisons Don’t Hold Up.” *Am. Sci.* 102: 460+.
- Grossman, Jonathan, and Ami Pedahzur. 2020. “Can We Do Better? Replication and Online Appendices in Political Science.” *Perspectives on Politics*, 1–6.
- Hardwicke, Tom E, Robert T Thibault, Jessica E Kosie, Joshua D Wallach, Mallory C Kidwell, and John P A Ioannidis. 2021. “Estimating the Prevalence of Transparency and Reproducibility-Related Research Practices in Psychology (2014–2017).” *Perspect. Psychol. Sci.*, March.
- Hilgard, Joseph. 2020. “Curious Features of Data in Zhang Et Al. (2019).” OSF.
- . 2021. “I Tried to Report Scientific Misconduct. How Did It Go?” 2021. <http://crystalprisonzone.blogspot.com/2021/01/i-tried-to-report-scientific-misconduct.html>.

- James, Gareth Michael, Daniela Witten, Trevor Hastie, and Robert Tibshirani. 2021. *An Introduction to Statistical Learning: With Applications in R*. Springer Nature.
- Janz, Nicole. 2016. “Bringing the Gold Standard into the Classroom: Replication in University Teaching.” *Int Stud Perspect* 17 (4): 392–407.
- John, Leslie K, George Loewenstein, and Drazen Prelec. 2012. “Measuring the Prevalence of Questionable Research Practices with Incentives for Truth Telling.” *Psychol. Sci.* 23 (5): 524–32.
- Kaplan, Robert M, and Veronica L Irvin. 2015. “Likelihood of Null Effects of Large NHLBI Clinical Trials Has Increased over Time.” *PLoS One* 10 (8).
- Key, Ellen M. 2016. “How Are We Doing? Data Access and Replication in Political Science.” *PS Polit. Sci. Polit.* 49 (02): 268–72.
- Kim, D. G., Joshua Byun, and Sichen Li. 2021. “Foreign Policy Revisionism in the Era of COVID-19: Theory and Evidence from Chinese Public Opinion.” <https://osf.io/r9dn7>.
- King, Gary. 1990. “On Political Methodology.” *Polit. Anal.* 2: 1–29.
- Klein, Richard A, Kate A Ratliff, Michelangelo Vianello, Reginald B Adams Jr, Štěpán Bahmík, Michael J Bernstein, Konrad Bocian, et al. 2014. “Investigating Variation in Replicability: A ‘Many Labs’ Replication Project.” *Soc. Psychol.* 45 (3): 142–52.
- Klein, Richard A, Michelangelo Vianello, Fred Hasselman, Byron G Adams, Reginald B Adams, Sinan Alper, Mark Aveyard, et al. 2018. “Many Labs 2: Investigating Variation in Replicability Across Samples and Settings.” *Advances in Methods and Practices in Psychological Science* 1 (4): 443–90.
- Kristal, Ariella S, Ashley V Whillans, Max H Bazerman, Francesca Gino, Lisa L Shu, Nina Mazar, and Dan Ariely. 2020. “Signing at the Beginning Versus at the End Does Not Decrease Dishonesty.” *Proc. Natl. Acad. Sci. U. S. A.* 117 (13): 7103–7.
- LaCour, Michael J, and Donald P Green. 2014. “Political Science. When Contact Changes Minds: An Experiment on Transmission of Support for Gay Equality.” *Science* 346 (6215): 1366–69.
- Lakens, Daniel. 2019. “The Value of Preregistration for Psychological Science: A Conceptual Analysis.” *PsyArXiv*.
- Lal, Apoorva, Mackenzie William Lockhart, Yiqing Xu, and Ziwen Zu. 2021. “How Much Should We Trust Instrumental Variable Estimates in Political Science? Practical Advice Based on over 60 Replicated Studies.”
- Lenz, Gabriel S, and Alexander Sahn. 2021. “Achieving Statistical Significance with Control Variables and Without Transparency.” *Polit. Anal.* 29 (3): 356–69.
- Lupia, Arthur, and Colin Elman. 2014. “Openness in Political Science: Data Access and Research Transparency: Introduction.” *PS Polit. Sci. Polit.* 47 (1): 19–42.
- National Academies of Sciences, Engineering, and Medicine. 2021. *Developing a Toolkit for Fostering Open Science Practices: Proceedings of a Workshop*. Edited by Thomas Arrison, Jennifer Saunders, and Emi Kameyama. Washington, DC: The National Academies Press. <https://www.nap.edu/catalog/26308/developing-a-toolkit-for-fostering-open-science-practices-proceedings-of>.
- Nelson, Leif D, Uri Simonsohn, and Joseph P Simmons. 2021. “[98] Evidence of Fraud in an Influential Field Experiment about Dishonesty.” <http://datacolada.org/98>.
- Neumayer, Eric, and Thomas Plümper. 2017. *Robustness Tests for Quantitative Research*. Cambridge University Press.

- Nosek, Brian A, Charles R Ebersole, Alexander C DeHaven, and David T Mellor. 2018. “The Preregistration Revolution.” *Proc. Natl. Acad. Sci. U. S. A.* 115 (11): 2600–2606.
- Nuijten, Michèle B, Chris H J Hartgerink, Marcel A L M van Assen, Sacha Epskamp, and Jelte M Wicherts. 2016. “The Prevalence of Statistical Reporting Errors in Psychology (1985–2013).” *Behav. Res. Methods* 48 (4): 1205–26.
- Ofosu, George K, and Daniel N Posner. 2020. “Pre-Analysis Plans: An Early Stocktaking.” *Perspectives on Politics*, 1–17.
- Open Science Collaboration. 2015. “PSYCHOLOGY. Estimating the Reproducibility of Psychological Science.” *Science* 349 (6251): aac4716.
- “Perspectives on DA-RT.” n.d. <https://dialogueondart.org/perspectives-on-da-rt/>.
- “Retraction Notice.” 2020. *Youth Soc.* 52 (2): 308–8.
- Rhys, Hefin. 2020. *Machine Learning with r, the Tidyverse, and Mlr*. Simon; Schuster.
- Scheel, Anne M, Mitchell R M J Schijen, and Daniël Lakens. 2021. “An Excess of Positive Results: Comparing the Standard Psychology Literature with Registered Reports.” *Advances in Methods and Practices in Psychological Science* 4 (2).
- Shapin, Steven. 1995. “Trust, Honesty, and the Authority of Science.” In *Society’s Choices: Social and Ethical Decision Making in Biomedicine*, edited by Ruth Ellen Bulger, Elizabeth Meyer Bobby, and Harvey V Fineberg, 388–408. National Academy Press.
- . 2018. *The Scientific Revolution*. University of Chicago Press.
- Shu, Lisa L, Nina Mazar, Francesca Gino, Dan Ariely, and Max H Bazerman. 2012. “Signing at the Beginning Makes Ethics Salient and Decreases Dishonest Self-Reports in Comparison to Signing at the End.” *Proceedings of the National Academy of Sciences*.
- Simmons, Joseph P, Leif D Nelson, and Uri Simonsohn. 2011. “False-Positive Psychology: Undisclosed Flexibility in Data Collection and Analysis Allows Presenting Anything as Significant.” *Psychol. Sci.* 22 (11): 1359–66.
- Simmons, Joseph P, and Uri Simonsohn. 2017. “Power Posing: P-Curving the Evidence.” *Psychol. Sci.* 28 (5): 687–93.
- Simmons, Joseph, Leif Nelson, and Uri Simonsohn. 2021. “Pre-registration: Why and How.” *J. Consum. Psychol.* 31 (1): 151–62.
- Simonsohn, Uri. 2013. “Just Post It: The Lesson from Two Cases of Fabricated Data Detected by Statistics Alone.” *Psychol. Sci.* 24 (10): 1875–88.
- Simonsohn, Uri, Leif D Nelson, and Joseph P Simmons. 2014. “P-Curve and Effect Size: Correcting for Publication Bias Using Only Significant Results.” *Perspect. Psychol. Sci.* 9 (6): 666–81.
- Statement, Journal Editors’ Transparency. 2015. “Data Access and Research Transparency (DA-RT): A Joint Statement by Political Science Journal Editors.” *Political Science Research and Methods* 3 (3): 421–21.
- Stockemer, Daniel, Sebastian Koehler, and Tobias Lentz. 2018. “Data Access, Transparency, and Replication: New Insights from the Political Behavior Literature.” *PS Polit. Sci. Polit.* 51 (4): 799–803.
- Tong, Christopher. 2019. “Statistical Inference Enables Bad Science; Statistical Thinking Enables Good Science.” *Am. Stat.* 73 (sup1): 246–61.

10 Appendix

Table 1: Journals analysed as ranked by Journal Citation Report 2020

Journal	Category	Citations	JIF	JIF Quartile
Political Analysis	PS	5,581	8.6	Q1
Annual Review Of Political Science	PS	5,418	8.091	Q1
International Affairs	IR	4,123	7.91	Q1
Political Communication	PS	3,692	7.859	Q1
American Political Science Review	PS	19,880	7.828	Q1
Review Of International Organizations	PS	1,287	7.795	Q1
International Security	IR	4,312	7.486	Q1
Journal Of European Public Policy	PS	6,982	7.339	Q1
Journal Of Public Administration Research				
And Theory	PS	8,156	7	Q1
Environmental Politics	PS	3,989	6.71	Q1
International Journal Of Press-Politics	PS	1,607	6.592	Q1
International Organization	PS	9,433	6.276	Q1
Political Behavior	PS	4,207	6.172	Q1
American Journal Of Political Science	PS	17,512	6.081	Q1
Regulation & Governance	PS	1,795	5.4	Q1
Journal Of Democracy	PS	4,879	5.326	Q1
British Journal Of Political Science	PS	6,110	5.174	Q1
Comparative Political Studies	PS	7,381	5.143	Q1
Policy Studies Journal	PS	3,908	5.141	Q1
European Journal Of Political Research	PS	5,458	4.943	Q1
Foreign Affairs	IR	4,766	4.791	Q1
New Political Economy	PS	2,369	4.681	Q1
Review Of International Political Economy	PS	3,647	4.659	Q1
Socio-Economic Review	PS	2,584	4.443	Q1
Political Psychology	PS	6,702	4.333	Q1
Policy And Society	PS	1,252	4.231	Q1
Marine Policy	IR	12,969	4.173	Q1
Public Opinion Quarterly	PS	7,999	4.154	Q1
European Political Science Review	PS	1,137	4.143	Q1
Geopolitics	PS	2,172	4.117	Q1

Journal	Category	Citations	JIF	JIF Quartile
Global Environmental Politics	PS	2,005	4.055	Q1
Journal Of Peace Research	PS	6,610	4.054	Q1
European Journal Of International Relations	IR	3,302	4.023	Q1
Jcms-Journal Of Common Market Studies	PS	4,774	3.99	Q1
West European Politics	PS	4,788	3.96	Q1
Territory Politics Governance Governance-An International Journal Of	PS	770	3.878	Q1
Policy Administration And Institutions	PS	3,292	3.838	Q1
Policy And Internet	PS	798	3.8	Q1
Political Science Research And Methods	PS	1,170	3.798	Q1
Perspectives On Politics	PS	3,444	3.776	Q1
South European Society And Politics	PS	1,160	3.771	Q1
Policy And Politics	PS	1,653	3.75	Q1
Public Administration	PS	5,750	3.72	Q1
International Political Sociology	PS	1,234	3.673	Q1
Political Geography	PS	5,342	3.66	Q1
Chinese Journal Of International Politics	IR	623	3.649	Q1
Cooperation And Conflict	PS	1,141	3.579	Q1
Journal Of Conflict Resolution	PS	6,921	3.53	Q1
Security Dialogue	IR	2,282	3.459	Q1
Journal Of Politics	PS	11,991	3.458	Q1
World Politics	PS	5,539	3.444	Q1
European Union Politics	PS	2,168	3.391	Q1
Government And Opposition	PS	1,979	3.322	Q2
Common Market Law Review	IR	1,286	3.257	Q1
Political Studies Review	PS	755	3.241	Q2
African Affairs	PS	2,102	3.203	Q2
Research & Politics	PS	1,080	3.141	Q2
Cambridge Review Of International Affairs	PS	1,024	3.096	Q2
American Journal Of International Law	IR	2,173	3.091	Q1
Politics & Society	PS	2,194	3.089	Q2
Democratization	PS	2,617	3.055	Q2
International Peacekeeping	IR	1,088	3	Q1
Post-Soviet Affairs	PS	837	2.98	Q2

Journal	Category	Citations	JIF	JIF Quartile
International Studies Quarterly	PS	5,856	2.936	Q2
Millennium-Journal Of International Studies	IR	1,968	2.93	Q2
Party Politics	PS	3,652	2.829	Q2
International Theory	PS	737	2.778	Q2
Terrorism And Political Violence	PS	2,334	2.741	Q2
Review Of International Studies	IR	2,934	2.73	Q2
Local Government Studies	PS	1,521	2.726	Q2
Political Science Quarterly	PS	1,586	2.675	Q2
International Studies Review	PS	1,647	2.658	Q2
International Environmental Agreements-Politics Law And Economics	PS	1,131	2.649	Q2
Contemporary Security Policy	PS	710	2.64	Q2
Mediterranean Politics	PS	683	2.588	Q2
Conflict Management And Peace Science	IR	1,496	2.563	Q2
Political Research Quarterly	PS	5,025	2.556	Q2
British Politics	PS	464	2.54	Q2
Journal Of Public Policy	PS	1,533	2.513	Q2
International Studies Perspectives	IR	871	2.5	Q2
Politics	PS	1,047	2.492	Q2
Ps-Political Science & Politics	PS	3,244	2.472	Q2
Publius-The Journal Of Federalism	PS	1,077	2.472	Q2
Security Studies	IR	1,368	2.464	Q2
American Politics Research	PS	1,772	2.451	Q2
Journal Of Strategic Studies	PS	1,103	2.44	Q2
Pacific Review	IR	1,261	2.432	Q2
British Journal Of Politics & International Relations	PS	1,642	2.422	Q2
Acta Politica	PS	1,044	2.404	Q2
Political Studies	PS	4,647	2.396	Q2
European Journal Of Political Economy	PS	2,865	2.366	Q2
Journal Of Chinese Governance	PS	274	2.333	Q2
International Relations Of The Asia-Pacific	IR	387	2.324	Q2
Emerging Markets Finance And Trade	IR	2,849	2.315	Q2
Social Movement Studies	PS	1,428	2.266	Q2
Journal Of Information Technology & Politics	PS	905	2.224	Q2

Journal	Category	Citations	JIF	JIF Quartile
Journal Of Political Philosophy	PS	1,645	2.224	Q2
Journal Of International Relations And Development	PS	695	2.2	Q2
Journal Of Intervention And Statebuilding	IR	669	2.2	Q2
German Politics	PS	673	2.159	Q2
Legislative Studies Quarterly	PS	1,720	2.159	Q2
Studies In Comparative International Development	PS	1,384	2.159	Q2
Globalizations	IR	1,762	2.155	Q2
Annals Of The American Academy Of Political And Social Science	PS	7,473	2.15	Q2
International Relations	IR	892	2.135	Q2
Problems Of Post-Communism	PS	761	2.127	Q2
International Journal Of Public Opinion Research	PS	1,819	2.11	Q2
Europe-Asia Studies	PS	2,006	2.102	Q2
Critical Policy Studies	PS	876	2.098	Q2
Bulletin Of The Atomic Scientists	IR	749	2.092	Q2
Politics & Gender	PS	1,350	2.088	Q2
Global Policy	PS	1,402	2.084	Q2
International Feminist Journal Of Politics	PS	1,093	2.083	Q2
Electoral Studies	PS	4,461	2.07	Q3
Politics And Governance	PS	978	2.061	Q3
International Political Science Review	PS	2,023	2.049	Q3
Public Choice	PS	5,970	2.019	Q3
New Left Review	PS	2,917	2.015	Q3
Comparative European Politics	PS	1,025	2.01	Q3
Journal Of Women Politics & Policy	PS	476	2	Q3
Quarterly Journal Of Political Science	PS	916	2	Q3
Review Of Policy Research	PS	1,290	2	Q3
European Security	IR	545	1.942	Q2
Asia Europe Journal	IR	368	1.846	Q3
European Journal Of International Law	IR	1,852	1.833	Q3
Ethics & International Affairs	IR	749	1.825	Q3

Journal	Category	Citations	JIF	JIF Quartile
Foreign Policy Analysis	IR	829	1.776	Q3
Peacebuilding	IR	346	1.75	Q3
Survival	IR	1,208	1.669	Q3
World Trade Review	IR	476	1.596	Q3
Human Rights Law Review	IR	639	1.569	Q3
Journal Of The Japanese And International Economies	IR	931	1.559	Q3
International Journal Of Transitional Justice	IR	816	1.55	Q3
Ocean Development And International Law	IR	388	1.541	Q3
Review Of World Economics	IR	1,286	1.517	Q3
Washington Quarterly	IR	833	1.5	Q3
Journal Of European Integration	IR	1,436	1.483	Q3
World Economy	IR	3,972	1.45	Q3
Australian Journal Of International Affairs	IR	602	1.411	Q3
Chinese Journal Of International Law	IR	326	1.395	Q3
International Interactions	IR	1,404	1.372	Q3
Stanford Journal Of International Law	IR	167	1.357	Q3
Journal Of Contemporary European Studies	IR	455	1.355	Q3
Intelligence And National Security	IR	903	1.326	Q3
Contemporary Southeast Asia	IR	513	1.3	Q3
Latin American Politics And Society	IR	800	1.255	Q3
Space Policy	IR	427	1.231	Q4
Revista Brasileira De Politica Internacional	IR	307	1.114	Q4
Alternatives	IR	686	1.095	Q4
Communist And Post-Communist Studies	IR	869	1.062	Q4
Journal Of World Trade	IR	452	0.977	Q4
Global Governance	IR	1,123	0.877	Q4
International Politics	IR	653	0.874	Q4
International Journal	IR	513	0.836	Q4
Asian Perspective	IR	324	0.8	Q4
Cornell International Law Journal	IR	373	0.72	Q4
Journal Of Human Rights	IR	535	0.694	Q4

Journal	Category	Citations	JIF	JIF Quartile
International Journal Of Conflict And Violence	IR	390	0.643	Q4
Asian Journal Of Wto & International Health Law And Policy	IR	86	0.611	Q4
War In History	IR	209	0.558	Q4
Pacific Focus	IR	163	0.553	Q4
Columbia Journal Of Transnational Law	IR	331	0.515	Q4
Journal Of Cold War Studies	IR	345	0.48	Q4
Middle East Policy	IR	462	0.475	Q4
Korean Journal Of Defense Analysis Current History	IR	104	0.413	Q4
Uluslararası İlişkiler-International Relations	IR	357	0.386	Q4
Korea Observer	IR	132	0.338	Q4
Diplomacy & Statecraft	IR	150	0.286	Q4
Internasjonal Politikk	IR	236	0.264	Q4
Asia-Pacific Review	IR	54	0.264	Q4
Foro Internacional	IR	96	n/a	n/a
Global Society	IR	130	n/a	n/a
Ipri Journal	IR	501	n/a	n/a
Revista Unisci	IR	30	n/a	n/a
Strategic Analysis	IR	57	n/a	n/a
	IR	265	n/a	n/a

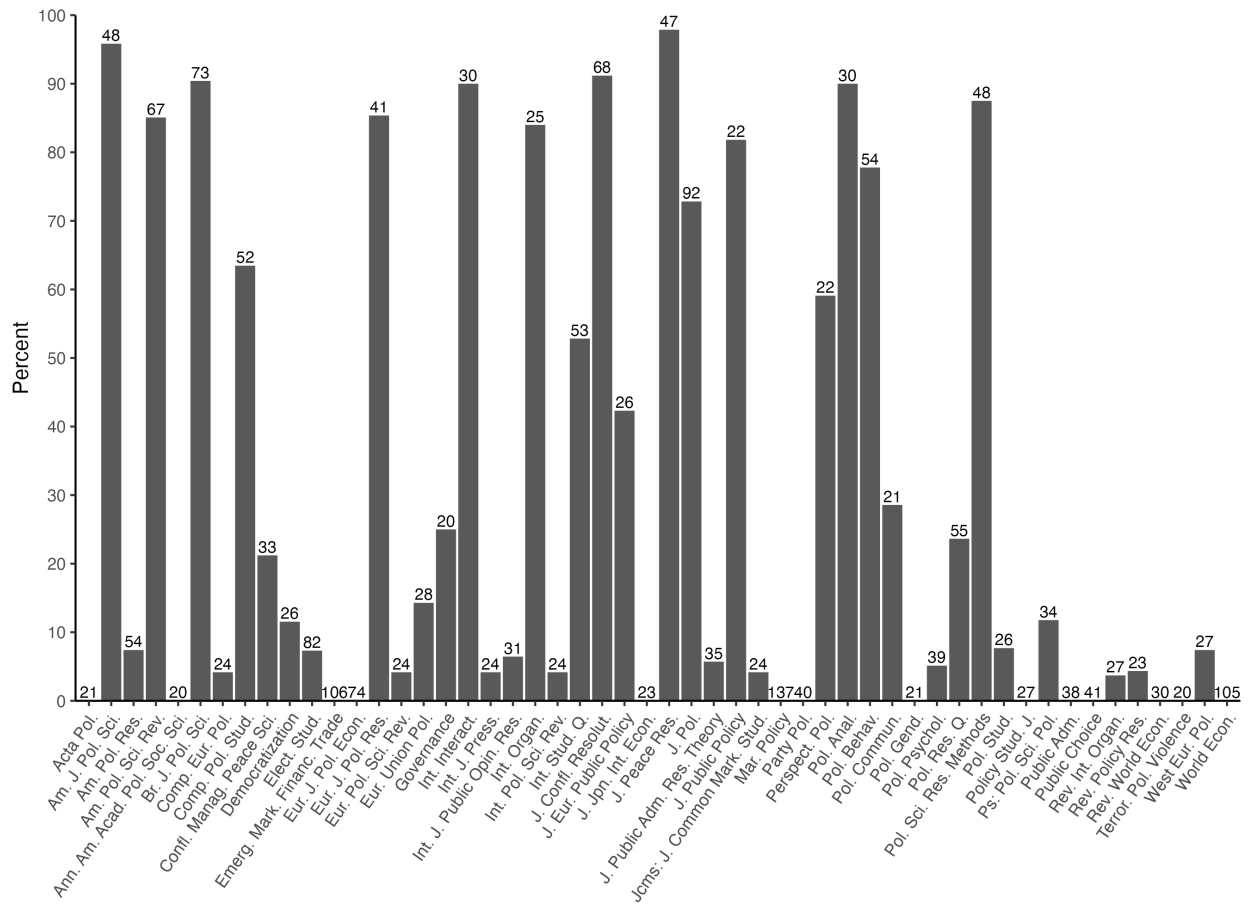


Figure 6: Open data by journal (with over 20 statistical inference papers in 2020)

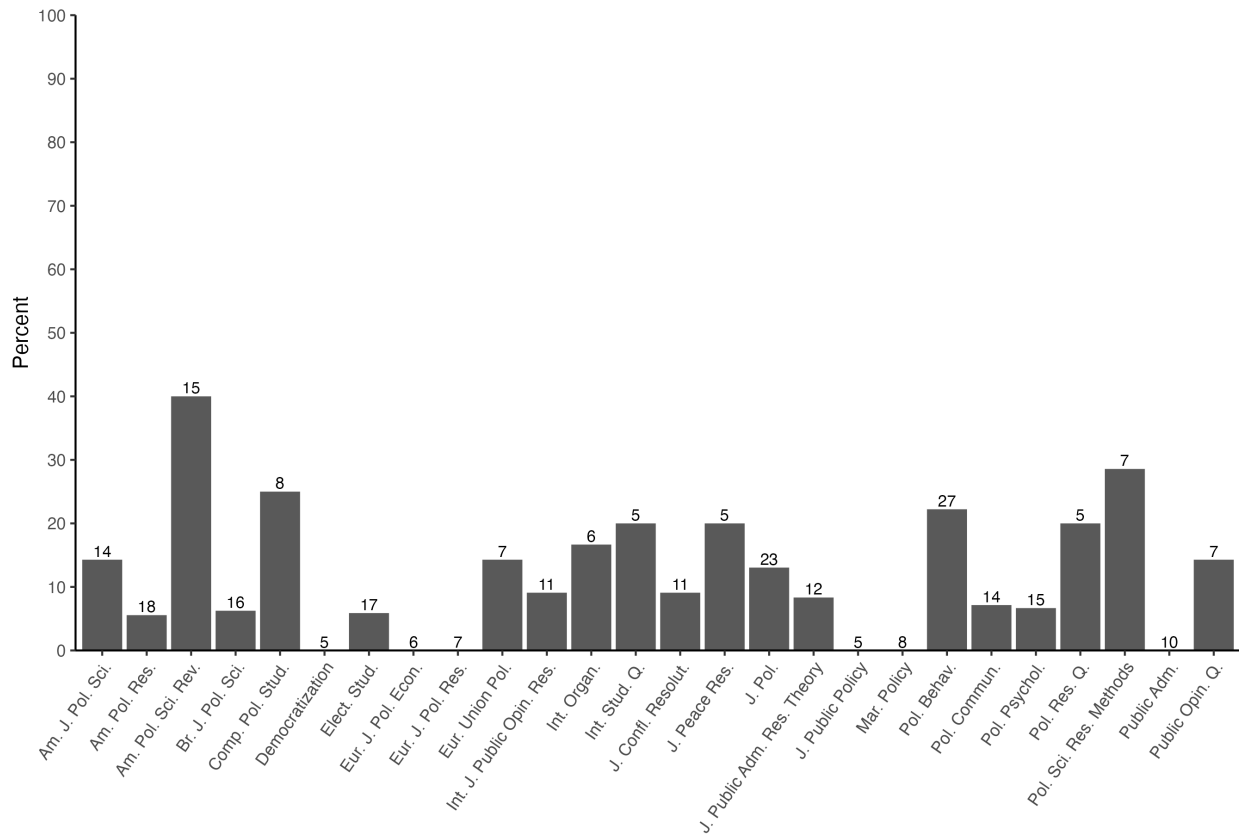


Figure 7: Preregistration by journal (with over 5 experimental papers in 2020)

10.1 Identifying papers relying on data analysis

We defined data analysis papers as those that made any display or presentation of numerical data, most commonly in tables and graphs. Maps that included data-rich overlays and required software to produce were included in this category.