INSTITUTE for REPLICATION

No. 6
I4R DISCUSSION PAPER SERIES

Quantitative Political Science Research is Greatly Underpowered

Vincent Arel-Bundock
Ryan C. Briggs
Hristos Doucouliagos
Marco Mendoza Aviña
Tom D. Stanley

October 2022

I4R DISCUSSION PAPER SERIES

I4R DP No. 6

Quantitative Political Science Research is Greatly Underpowered

Vincent Arel-Bundock[1], Ryan C. Briggs[2], Hristos Doucouliagos[3], Marco Mendoza Aviña[4], Tom D. Stanley[3]

[1]Université de Montréal/Canada, [2]University of Guelph/Canada,
[3]Deakin University Melbourne/Australia,[4]Harvard University Cambridge/USA

OCTOBER 2022

ISSN: 2752-1931

Quantitative Political Science Research is Greatly Underpowered*

Vincent Arel-Bundock[†] Ryan C. Briggs[‡]
Hristos Doucouliagos[§] Marco Mendoza Aviña[¶] T.D. Stanley[⊠]

October 19, 2022

Abstract

The social sciences face a replicability crisis. A key determinant of replication success is statistical power. We assess the power of political science research by collating over 16,000 hypothesis tests from about 2,000 articles. Using generous assumptions, we find that the median analysis has about 10% power and that only about 1 in 10 tests have at least 80% power to detect the consensus effects reported in the literature. We also find substantial heterogeneity in tests across research areas, with some being characterized by high power but most having very low power. To contextualize our findings, we survey political methodologists to assess their expectations about power levels. Most methodologists greatly overestimate the statistical power of political science research.

*We thank Elissa Berwick, Aaron Erlich, Patrick Fournier, Bart Harvey, Andrew Philips, as well as the software authors who made this work possible (see our Software Bibliography, p. SI-15) and all researchers who shared data with us (see Appendix A). Sara Duran-Bettancour, Khaoula El Khalil, and Aden Morton-Ferguson provided research assistance. We would like to thank participants at the Toronto Data Workshop and APSA 2022 for feedback. We gratefully acknowledge funding from Canada's Social Sciences and Humanities Research Council. Replication materials will be made available on the OSF registry upon publication.

[†]Université de Montréal. vincent.arel-bundock@umontreal.ca
[‡]University of Guelph. rbriggs@uoguelph.ca
[§]Deakin University. chris.doucouliagos@deakin.edu.au
[¶]Harvard University. mmendozaavina@fas.harvard.edu
[⊠]Deakin University. tom.stanley1@deakin.edu.au

Statistical power is important to any discipline that practices null hypothesis significance testing, such as political science. As power increases, the probability of committing a false negative (e.g. type II error) decreases, and empirical findings are more likely to replicate (Altmejd et al. 2019). When power is low and one happens to find a statistically significant estimate, that estimate is likely to be many-fold greater than the "true" underlying effect, and it may well have the wrong sign (Gelman and Carlin 2014; Gelman and Tuerlinckx 2000; Ioannidis, Stanley, and Doucouliagos 2017).

We examine statistical power in political science by assembling a dataset of 16649 hypothesis tests, grouped in 351 meta-analyses, reported in 46 peer-reviewed meta-analytic articles.[1] We estimate power retrospectively by leveraging estimates of mean population effects from the meta-analyses. In essence, we calculate the power of each test to detect the consensus effect reported in its literature. Our results suggest that quantitative political science research is greatly underpowered. The median research result has about 10% power ($\alpha = 0.05$), and only about 1 in 10 statistical tests have at least 80% power.

These results are both dispiriting and surprising. To contextualize them, we conducted an original expert survey of all authors who published in the peer-reviewed methodology journal *Political Analysis* between 2010 and 2021. In total, 131 methodologists answered our survey, for a response rate of 27%.[2] We asked experts to consider all hypothesis tests published in the 50 peer-reviewed journals with the highest impact factors in political science and closely adjacent fields over the past two decades. They then guessed what share of those tests had at least 50% and 80% power to reject the null at the $\alpha = 0.05$ level.[3]

On average, political methodologists believe that 66% of studies have at least 50% power, and 43% have at least 80% power. We demonstrate that these expectations are overly optimistic. On average, experts overestimate the share of studies powered at the 50% level by 48 percentage points, and the share of studies powered at the 80% level by 32 percentage points. Political science research suffers from low power and this problem is not sufficiently appreciated.

[1] We define a "meta-analysis" as a grouping of at least 5 comparable estimates which researchers have aggregated to calculate meta-analytic effects. A single meta-analytic article often reports many meta-analyses. Some meta-analyses address closely related substantive questions using slight variations on the independent and dependent variables, or using different sets of estimates (e.g., experimental vs. observational).

[2] See Appendix C for details on the survey administration and pre-registration.

[3] To align the survey question with our empirical approach, we asked about the share of tests that had at least 50% or 80% "power to reject the null hypothesis." This guided respondents to think about power relative to likely effect sizes. For logical consistency, we exclude a few respondents who stated that more studies have at least 80% than 50% power. 42 respondents said that they did not know the answer to at least one of the questions.

METHOD

Our goal is to assess statistical power in recent quantitative political science research. In other words, we want to estimate the probability that any given statistical test will reject the null hypothesis, for a given "true" effect size. To achieve this, the ideal dataset would have a standard error along with a value for the population mean effect of interest for every estimate reported in peer-reviewed journals in political science and closely adjacent fields. Given these data, one could calculate *retrospective* power for each test at some α level such as 0.05.[4] A test for which the standard error is less than the population mean effect divided by 2.8 would have at least 80% power.

Of course, in general "true" effect sizes are unknown. Therefore, we estimate power retrospectively using meta-analytic estimates of population mean effects for various research questions. This approach follows previous work in neuroscience (Button et al. 2013), economics (Ioannidis, Stanley, and Doucouliagos 2017), and psychology (Stanley, Carter, and Doucouliagos 2018). Because meta-analyses encompass all relevant, reported hypothesis tests for specific research areas, meta-analysis provides the best and most widely informed estimate of the population mean effect. Perhaps the only exception are estimates from preregistered multi-lab replications (Klein et al. 2018; Open Science Collaboration 2015), but these are uncommon in political science.

To find meta-analyses, we searched through the archives of 141 peer-reviewed journals in political science and closely adjacent fields. In total, 82 articles met our inclusion criteria. We obtained data files for 46 of those articles from publicly available sources or by contacting authors directly. In the end, we were able to assemble a dataset of 16649 hypothesis tests grouped in 351 meta-analyses covering a broad cross-section of the discipline. Appendices A and B provide details on our data collection and cleaning.

The main drawback of using these data for retrospective power analysis is that estimates of the population mean effects are based on reported results, which may have been selected based on statistical significance. Examples of such selection include the file drawer problem, reporting bias, specification searching, p-hacking, and the garden of forking paths.[5] Selection on statistical significance will result in there being too few statistically non-significant estimates and too many statistically significant estimates, which will inflate the overall collection of estimates reported in the literature.[6] If meta-analyses aggregate inflated estimates,

[4] Throughout the paper, we use an α level of 0.05.

[5] The presence of publication bias is "one of the strongest findings across the sciences" (Berinsky, Druckman, and Yamamoto 2021, 370).

[6] As noted in the *Journal of Politics*' Pre-Registration Guidelines, "if power analyses and smallest effect sizes of interest are based on effect sizes reported in previous studies, authors should keep in mind that meta-scientific studies reliably report inflated effect sizes in the reported literature that often shrink in replication studies" (Journal of Politics 2022).

they will likely produce inflated estimates of population mean effects, which will in turn lead us to overstate power (Gelman 2019).

There is no universal agreement in the methodological literature on a best approach for estimating population mean effects from reported results. Thus, we use three alternative methods, which we introduce below from least to most aggressive in how they attempt to correct for publication bias.

The first method is the Unrestricted Weighted Least Squares (UWLS). The UWLS is a simple weighted average of the form $\hat{\mu}_w = \sum(1/\phi\sigma_i^2)y_i/\sum(1/\phi\sigma_i^2)$, where y_i are study-level estimates, σ_i^2 are within-study variances, and ϕ is a scaling factor. UWLS can be estimated by regressing the study-level estimates on a constant using weighted least squares with weights equal to $1/\sigma_i^2$.[7]

The second method is the Weighted Average of the Adequately Powered (WAAP), introduced by Stanley, Doucouliagos, and Ioannidis (2017). We compute the WAAP in three steps: (1) calculate the UWLS; (2) use the UWLS to estimate the power of each study; (3) calculate the UWLS again, but using only the subset of estimates that exceed 80% power.

The final method is our most aggressive strategy to unwind publication bias: the Precision-Effect Test and Precision-Effect Estimate with Standard Error (PET-PEESE) (Egger et al. 1997; Stanley 2017; Stanley and Doucouliagos 2014). The intuition that motivates this approach is that selection on statistical significance will result in a positive relationship between reported effect estimates and reported standard errors. One can thus regress estimates on standard errors or standard errors squared, interpreting the intercept of this regression as the estimated effect when the standard error is equal to zero. Our calculations for PET-PEESE follow the procedure outlined in Stanley and Doucouliagos (2014).

Each of these three methods produces alternative estimates of population mean effects for each meta-analysis in our sample. With these in hand, we then calculate the statistical power of each estimate based on its standard error, the population mean effect from the relevant meta-analysis, and an α level of 0.05. As we show below, results obtained using the three techniques are broadly consistent. For simplicity, the rest of this paper generally focuses on results obtained via UWLS, the most "generous" approach.

[7] UWLS produces the exact same point estimates as the more common Fixed Effects meta-analysis, but simulations suggest that the UWLS standard errors have better properties (Stanley and Doucouliagos 2022). Since our retrospective power calculations only require meta-analytic estimates of the population mean effects, and not their standard errors, the two approaches are functionally equivalent for our purposes.

RESULTS

We begin by examining selection on statistical significance. Figure 1 shows a histogram of the absolute value of z-statistics for all hypothesis tests.[8] The distribution is right-skewed: there are many significant tests. The large spikes and humps show that many z-statistics are concentrated at or just above two conventional thresholds of statistical significance: 95% and 99%. These results resemble those found by Gerber and Malhotra (2008) for political science and by Brodeur et al. (2016; 2020) and Gorajek and Malin (2021) for economics.[9] It is difficult to explain such a distribution of z-statistics in the absence of selection on statistical significance.

Figure 1: Distribution of z-statistics in the full sample of estimates. The red dashes highlight the conventional thresholds of statistical significance of 95% and 99%.

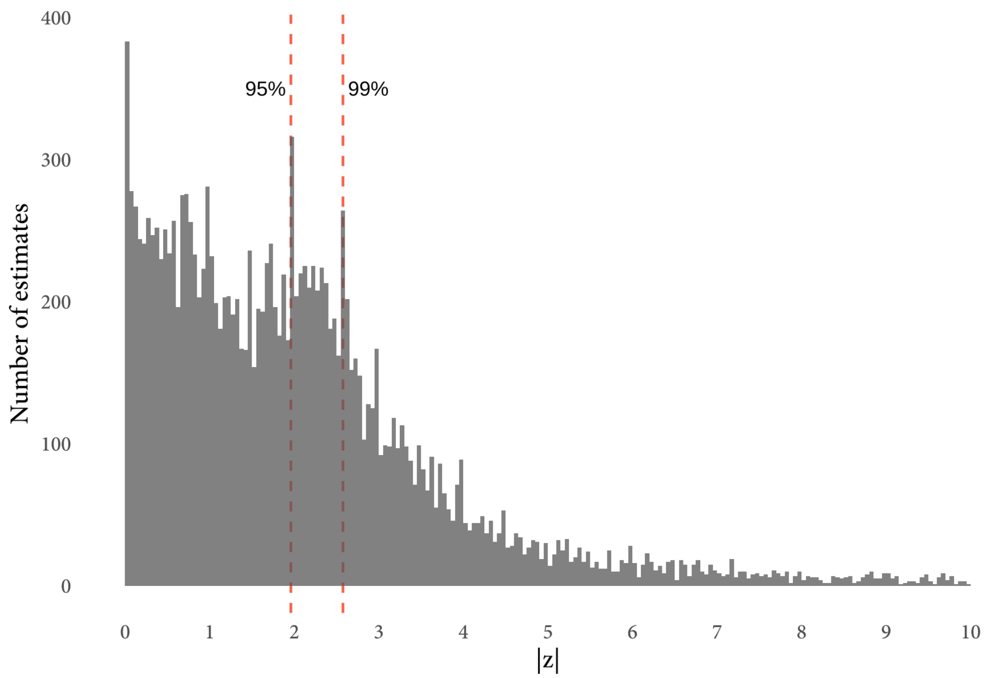


Figure 1 has important implications for our study of statistical power. Since reported findings are characterized by publication bias, our estimates of the "truth" and our power calculations are likely to be inflated. The UWLS results presented below should thus be interpreted as a best-case scenario for statistical power in the discipline.

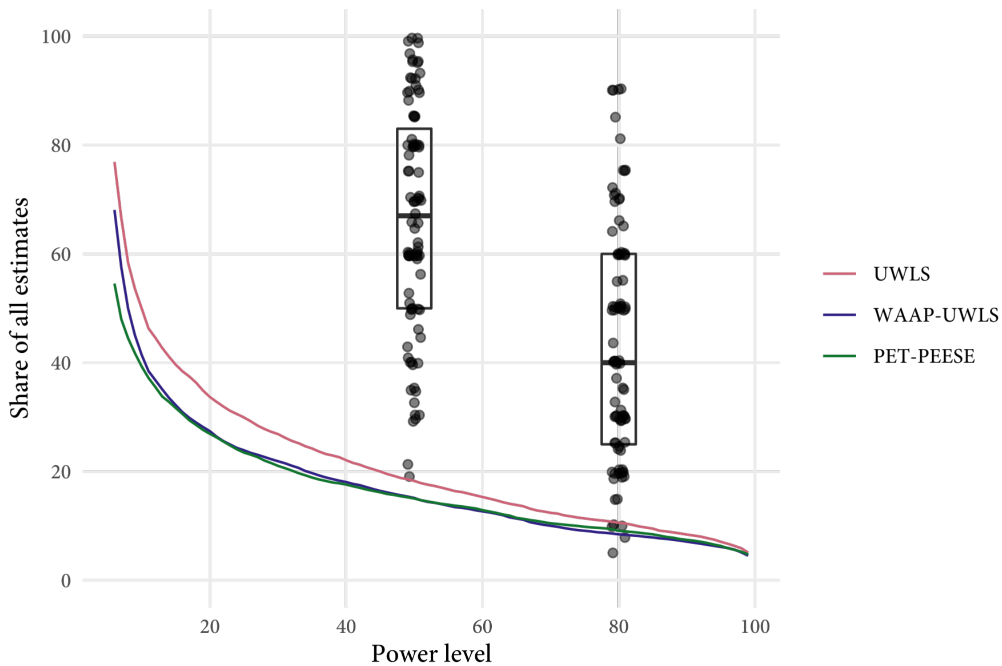
Our key results are presented in Figure 2. The results for each of the three methods for estimating population effects are displayed using differently colored lines,

[8] We lack information on degrees of freedom for most observations, so we calculate z-statistics instead of t-statistics. For visual clarity, we exclude z-statistics larger than 10.

[9] See especially Figure 1 by Brodeur et al. (2016). Vivalt (2019) shows similar results for quasi-experimental impact evaluations.

which represent the share of all hypothesis tests that reach a given power level from 6% to 99%.

Figure 2: Retrospective power analyses and the view from political methodology. Lines represent the share of estimates by power level, using three different approaches to estimate the population mean effect. Black dots and boxplots represent the distribution of responses in the expert survey.



All three techniques produce similarly shaped curves. UWLS, which addresses publication bias least aggressively, yields the highest power estimates. Yet power is low even under these generous conditions: only half of tests reach 10% power, a fifth reach 50% power, and a tenth reach 80%. Our two other methods for estimating population mean effects result in even lower power levels.

The black overlays in Figure 2 display the results of our expert survey. As a reminder, we asked respondents to estimate the share of hypothesis tests that achieve at least 50% power and at least 80% power within all papers published since 2000 in the top journals in political science and related fields. Their responses are shown in black dots and the boxplots mark the quartiles for each distribution. The curves for our power calculations barely overlap with the expert responses: methodologists overestimate the power of hypothesis tests by a large margin. Clearly, the extent of the problem of low power is not properly appreciated by the discipline.

We conducted several other analyses to refine the interpretation, add nuance, and test the robustness of our findings.

Minimum effects of interest. For this paper we calculated power retrospectively based on meta-analytic estimates of population mean effects. In practice, however, experimentalists in political science typically design their studies based on prospective power calculations anchored by the size of a "minimum effect of interest" (MEI). Our results can tell us if studies are sufficiently powered to detect consensus effects in the literature; they do not directly tell us if studies are well powered to detect the MEIs targeted by individual researchers. But even if we cannot directly characterize power with respect to MEIs, our results suggest useful informal bounds. Consider two cases.

First, if MEIs are systematically smaller than population mean effects, our analyses would overstate the power of political science. In that case, our results could be interpreted as an upper bound for the average power of the studies in our sample; the problems that we highlight in this paper would be even more worrying.

Second, if MEIs are systematically larger than meta-analytic estimates, our analyses would understate the level of power in political science research.[10] Appendix D probes the sensitivity of our conclusions to this potential issue. In it, we artificially inflate all population mean effect estimates by a factor of five and replicate our analyses.[11] We find that one third of tests still fail to reach 50% power. In other words, even if researchers were designing studies to target effect sizes five times larger than the consensus estimates reported in the literature, quantitative political science research would still be greatly underpowered.

Another way to circumvent the problem posed by the disconnect between population mean effects and MEIs is to shift the focus away from power, toward the magnitude and sign of point estimates.

Magnitude and sign. Low-powered studies that are subsequently filtered for statistical significance are more likely to report effects that are inflated or of the incorrect sign (Gelman and Carlin 2014; Gelman and Tuerlinckx 2000; Ioannidis, Stanley, and Doucouliagos 2017). Among our 16649 hypothesis tests, 7775 are statistically significant at the 0.05 level. Among these significant effects, 15% have a different sign than the consensus UWLS estimate.[12] To see if reported estimates tend to be larger than population mean effects, we simply divide individual estimates by the corresponding UWLS estimate of the mean. In the subset of estimates that are both statistically significant and in the "correct" direction, the median estimate-to-UWLS ratio is 3.0. Thus, the significant estimates in our sample are likely to be about 3.0 times too big or wrongly signed.[13]

[10] If MEIs are systematically larger than meta-analytic estimates, the vast majority of political science research should also yield null results, which we do not see in Figure 1.

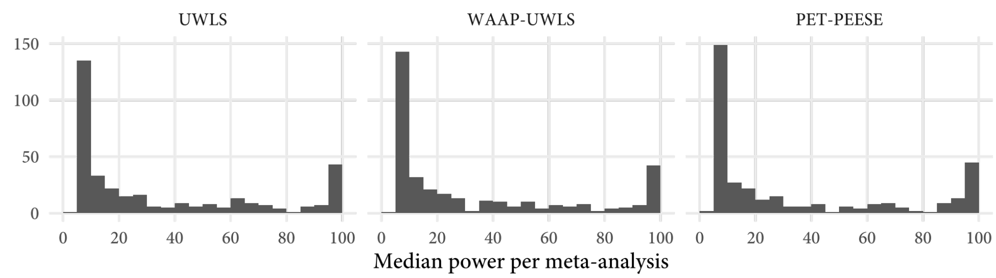
[11] This analysis is inspired by Kollepara et al. (2021).

[12] While low power interacted with significance filtering can lead to these "errors," they do not depend in any way on how we calculated power. These results come simply from comparing the meta-analytic population mean effect to each significant coefficient within the meta-analysis.

[13] The calculations in this paragraph can be thought of as an empirical analogue to the Type M and Type S error rates described in Gelman and Carlin (2014). These calculations are also

Sample composition. One potential objection relates to the composition of our dataset. In our expert survey, respondents stated expectations about power levels for all estimates published in well-ranked journals over the past two decades. However, our sample includes only a small share of those estimates because few studies ever get aggregated in meta-analyses (see Appendix A for the research questions and meta-analyses covered). To address this concern, we asked our experts if they believed estimates in meta-analyses to be better- or worse-powered than all other estimates. 75% of them think that studies included in meta-analyses (in our sample) have about the same power or higher power than other studies (out of our sample). Appendix C reports further details about this question.

Figure 3: Histogram of median power per meta-analysis, using three different approaches to estimate the population mean effect.



Heterogeneity across research questions. Figure 2 merges all our estimates together, but there might be variation in power levels across meta-analyses. We examine this by graphing the median power within each of the 351 meta-analyses in our dataset. Figure 3 reveals substantial heterogeneity. While most research areas have median power below 10%, a substantial share has median power above 80%. This finding is consistent with previous assessments of power in neuroscience (Button et al. 2013; Nord et al. 2017) and economics (Ioannidis, Stanley, and Doucouliagos 2017).[14]

Outliers One potential concern is that our findings may be driven by atypical observations or atypical meta-analyses.[15] In Appendix E, we redraw Figures 1 and 2 while excluding outliers based on the distribution of effect sizes, the distribution of standard errors, and the influence of individual observations. To examine the sensitivity of the results to which meta-analyses we include, in

related to the "Exaggeration Factor" which Ioannidis, Stanley, and Doucouliagos (2017) calculate as $|\frac{\bar{\beta}-\beta^{\text{UWLS}}}{\beta^{\text{UWLS}}}| + 1$, where $\bar{\beta}$ is the average of estimates in a meta-analysis, β^{UWLS} is the UWLS estimate for that meta-analysis. The median exaggeration factor in our dataset is 1.9, which is close to what these authors found in economics.

[14]Using a somewhat different method, Szucs and Ioannidis (2017) find similarly low power in psychology, cognitive neuroscience, and medically-oriented neuroscience.

[15]One atypical article in our dataset contributes 259 of our 351 meta-analyses (though only 3445 of our 16649 hypothesis tests). Median power (UWLS) overall is 0.1; excluding this paper brings median power to 0.09.

Appendix F we reproduce Figure 2 while sequentially withholding one of the 46 meta-analytic articles at a time.[16] Our main findings are not influenced by outliers or the exclusion of any single meta-analytic paper.

Power over time. In Appendix G, we examine how power has changed over time based on the publication year of each test. Power slowly increased from 1990 until about 2015. From 2015, power appears to have increased more quickly, but we have few tests after 2015 so this result should be interpreted cautiously.

CONCLUSION

Statistical power is an important and neglected aspect of quantitative political science research. This paper has shown that power is typically very low, and that this issue is not properly appreciated by the discipline.

While the problem of low power is clear, its origin is less clear. Most of the variation in power in our sample exists across rather than within meta-analyses, with some having either very high or very low median power (see Figure 3). Differences of this size across meta-analyses are more likely due to differences in the magnitude and stability of the effects under study rather than to, for example, differences in sample sizes. Put simply, some research areas may be studying larger and more consistent treatment effects than others. The presence of many small "true" effects may also explain why so many significant estimates had the "wrong" sign and why significant and "correctly-signed" effects were so inflated.

The problem of low power in political science should motivate the search for institutional, methodological, and theoretical remedies. Institutionally, journal editors, peer reviewers, and promotion committees should do more to incentivize researchers to replicate prior works, to conduct large scale team-based research, and to improve the publication process (Malhotra 2021; Williamson et al. 2022). Methodologically, we should seek ways to increase power by using innovative research designs and measurement strategies (Broockman, Kalla, and Sekhon 2017; Clifford, Sheagley, and Piston 2021). With respect to theory testing, we must recognize that the direct approach to answering some research questions, like trying to detect small differences between a limited number of states, are likely underpowered. In those cases, a useful response may be to derive new implications from our theories and test them where data are richer.

[16] As a reminder, individual articles can contain more than one meta-analysis. This is thus a stricter test than withholding one of the 351 meta-analyses at a time.

REFERENCES

Altmejd, Adam, Anna Dreber, Eskil Forsell, Juergen Huber, Taisuke Imai, Magnus Johannesson, Michael Kirchler, Gideon Nave, and Colin Camerer. 2019. "Predicting the Replicability of Social Science Lab Experiments." *PLOS ONE* 14(12): e0225826.

Belsley, David A., Edwin Kuh, and Roy E. Welsch. 1980. *Regression Diagnostics: Identifying Influential Data and Sources of Collinearity*. John Wiley & Sons.

Berinsky, Adam J., James N. Druckman, and Teppei Yamamoto. 2021. "Publication Biases in Replication Studies." *Political Analysis* 29(3): 370–84.

Brodeur, Abel, Nikolai Cook, and Anthony Heyes. 2020. "Methods Matter: P-Hacking and Publication Bias in Causal Analysis in Economics." *American Economic Review* 110(11): 3634–60.

Brodeur, Abel, Mathias Lé, Marc Sangnier, and Yanos Zylberberg. 2016. "Star Wars: The Empirics Strike Back." *American Economic Journal: Applied Economics* 8(1): 1–32.

Broockman, David E., Joshua L. Kalla, and Jasjeet S. Sekhon. 2017. "The Design of Field Experiments with Survey Outcomes: A Framework for Selecting More Efficient, Robust, and Ethical Designs." *Political Analysis* 25(4): 435–64.

Button, Katherine S., John Ioannidis, Claire Mokrysz, Brian A. Nosek, Jonathan Flint, Emma S. J. Robinson, and Marcus R. Munafò. 2013. "Power Failure: Why Small Sample Size Undermines the Reliability of Neuroscience." *Nature Reviews Neuroscience* 14(5): 365–76.

Clifford, Scott, Geoffrey Sheagley, and Spencer Piston. 2021. "Increasing Precision Without Altering Treatment Effects: Repeated Measures Designs in Survey Experiments." *American Political Science Review* 115(3): 1048–65.

Egger, Matthias, George Davey Smith, Martin Schneider, and Christoph Minder. 1997. "Bias in Meta-Analysis Detected by a Simple, Graphical Test." *BMJ* 315(7109): 629–34.

Gelman, Andrew. 2019. "Don't Calculate Post-Hoc Power Using Observed Estimate of Effect Size." *Annals of Surgery* 269(1): e9–10.

Gelman, Andrew, and John Carlin. 2014. "Beyond Power Calculations: Assessing Type S (Sign) and Type M (Magnitude) Errors." *Perspectives on Psychological Science* 9(6): 641–51.

Gelman, Andrew, and Francis Tuerlinckx. 2000. "Type S Error Rates for Classical and Bayesian Single and Multiple Comparison Procedures." *Computational Statistics* 15(3): 373–90.

Gerber, Alan, and Neil Malhotra. 2008. “Do Statistical Reporting Standards Affect What Is Published? Publication Bias in Two Leading Political Science Journals.” *Quarterly Journal of Political Science* 3(3): 313–26.

Gorajek, Adam, and Benjamin A. Malin. 2021. *Comment on “Star Wars: The Empirics Strike Back”*. Federal Reserve Bank of Minneapolis. https://doi.org/10.21034/sr.629.

Ioannidis, John, T.D. Stanley, and Hristos Doucouliagos. 2017. “The Power of Bias in Economics Research.” *The Economic Journal* 127(605): F236–65.

Journal of Politics. 2022. “Pre-Registration Guidelines.” https://www.journals.uchicago.edu/journals/jop/pre-registration.

Klein, Richard A., Michelangelo Vianello, Fred Hasselman, Byron G. Adams, Reginald B. Adams Jr., Sinan Alper, Mark Aveyard, Jordan R. Axt, Mayowa T. Babalola, and Štěpán Bahnik. 2018. “Many Labs 2: Investigating Variation in Replicability Across Samples and Settings.” *Advances in Methods and Practices in Psychological Science* 1(4): 443–90.

Kollepara, Pratyush K., Alexander F. Siegenfeld, Nassim Nicholas Taleb, and Yaneer Bar-Yam. 2021. “Unmasking the Mask Studies: Why the Effectiveness of Surgical Masks in Preventing Respiratory Infections Has Been Underestimated.” *Journal of Travel Medicine* 28(7): taab144.

Malhotra, Neil. 2021. “Threats to the Scientific Credibility of Experiments: Publication Bias and p-Hacking.” In *Advances in Experimental Political Science*, eds. James N. Druckman and Donald P. Green. Cambridge University Press, 354–68.

Nord, Camilla L, Vincent Valton, John Wood, and Jonathan P Roiser. 2017. “Power-up: A Reanalysis of ‘Power Failure’ in Neuroscience Using Mixture Modeling.” *Journal of Neuroscience* 37(34): 8051–61.

Open Science Collaboration. 2015. “Estimating the Reproducibility of Psychological Science.” *Science* 349(6251): aac4716.

Stanley, T.D. 2017. “Limitations of PET-PEESE and Other Meta-Analysis Methods.” *Social Psychological and Personality Science* 8(5): 581–91.

Stanley, T.D., Evan C. Carter, and Hristos Doucouliagos. 2018. “What Meta-Analyses Reveal about the Replicability of Psychological Research.” *Psychological Bulletin* 144(12): 1325–46.

Stanley, T.D., and Hristos Doucouliagos. 2014. “Meta-Regression Approximations to Reduce Publication Selection Bias.” *Research Synthesis Methods* 5(1): 60–78.

———. 2022. “Harnessing the Power of Excess Statistical Significance: Weighted and Iterative Least Squares.” *Psychological Methods.* Online ahead of print 12 May 2022. DOI: 10.1037/met0000502.

Stanley, T.D., Hristos Doucouliagos, and John Ioannidis. 2017. "Finding the Power to Reduce Publication Bias." *Statistics in Medicine* 36(10): 1580–98.

Szucs, Denes, and John Ioannidis. 2017. "Empirical Assessment of Published Effect Sizes and Power in the Recent Cognitive Neuroscience and Psychology Literature." *PLOS Biology* 15(3): e2000797.

Vivalt, Eva. 2019. "Specification Searching and Significance Inflation Across Time, Methods and Disciplines." *Oxford Bulletin of Economics and Statistics* 81(4): 797–816.

Williamson, Scott, Andrea Dillon, Jens Hainmueller, Dominik Hangartner, Michael Hotard, David D. Laitin, Duncan Lawrence, and Jeremy Weinstein. 2022. "Learning from Null Effects: A Bottom-up Approach." *Political Analysis*: 1–9.

Quantitative Political Science Research is Greatly Underpowered

Supporting Information

A DATA COLLECTION

To find meta-analyses, we searched through the archives of 141 peer-reviewed journals in political science and closely adjacent fields. The list of peer-reviewed journals in the scope of our data collection came from two sources. First, we selected all journals appearing in the social science subcategories “Political Science,” “Diplomacy & International Relations,” and “Public Policy & Administration” of Google Scholar Metrics’ top publications as of 2021. Second, we selected the 50 journals with the highest total citations for the year 2020 in the categories “political science,” “international relations,” and “public administration” categories of Clarivate’s Journal Citation Reports. The full list of journals is printed below. We conducted full-text searches for the keyword “meta” in the archives of all these journals, ignoring all results dating from before 2000. All keyword matches were checked manually to retain articles for which authors gathered data from other articles in order to run a quantitative meta-analysis.

We excluded qualitative meta-syntheses and articles in which authors combine multiple estimates from their own work, using meta-analytic techniques as a form of model averaging. We also excluded articles not focused on political topics, that is, articles where both the outcome and explanator variables are primarily the object of research in other disciplines, such as economics, criminology, management, and psychology. Ambiguous cases were reviewed by two different co-authors of the present paper. In total, 82 articles met our definition of a meta-analysis in political science.

We were able to obtain the data from 46 of those meta-analytic articles from public journal archives, research repositories, author websites, by transcribing results from published tables and figures, and by contacting authors directly. In total, we assembled a dataset with 16649 point estimates grouped 351 meta-analyses. The data for 13 of the meta-analytic articles were collected by Hristos Doucouliagos for a separate project. This external data file also included one relevant book and one article, both of which we kept in our sample even though they were not identified by our data collection strategy.

The year 2000 cutoff point is arbitrary and was chosen due to resource constraints and data availability. It is useful to note that data collection involved conducting full text searches on the full universe of articles published by a large sample of peer-reviewed journals (listed in the following sub-section). Through this process we found that very few meta-analyses had been published in political science before 2000 and that the data for these articles were generally unavailable.

Journals surveyed for meta-analyses

Administration and Society; African Affairs; American Journal of International Law; American Journal of Political Science; American Political Science Review; Annals of the American Academy of Political and Social Science; Annual Review of Political Science; Australian Journal of Public Administration; British Journal of Political Science; British Journal of Politics and International Relations; Bulletin of the Atomic Scientists; Cambridge Review of International Affairs; Canadian Public Administration; Canadian Public Policy; Citizenship Studies; Climate Policy; Common Market Law Review; Communist and Post-Communist Studies; Comparative Political Studies; Comparative Politics; Conflict Management and Peace Science; Contemporary Economic Policy; Cooperation and Conflict; Critical Policy Studies; Democratization; Electoral Studies; Emerging Markets Finance and Trade; Environment and Planning C: Government and Policy; Environmental Politics; Ethics and International Affairs; European Journal of International Law; European Journal of International Relations; European Journal of Political Economy; European Journal of Political Research; European Union Politics; Geopolitics; Global Environmental Politics; Global Governance; Global Policy; Globalizations; Governance: An International Journal of Policy Administration and Institutions; Human Rights Quarterly; Human Service Organizations Management Leadership and Governance; International Affairs; International Affairs; International Interactions; International Journal of Public Administration; International Journal of Transitional Justice; International Organization; International Peacekeeping; International Political Sociology; International Public Management Journal; International Relations; International Review of Administrative Sciences; International Security; International Studies Perspectives; International Studies Quarterly; International Studies Review; JCMS: Journal of Common Market Studies; Journal of Accounting and Public Policy; Journal of Accounting and Public Policy; Journal of Comparative Policy Analysis; Journal of Conflict Resolution; Journal of Democracy; Journal of European Integration; Journal of European Public Policy; Journal of European Social Policy; Journal of Homeland Security and Emergency Management; Journal of Peace Research; Journal of Policy Analysis and Management; Journal of Politics; Journal of Public Administration Research and Theory; Journal of Public Policy; Journal of Social Policy; Journal of Strategic Studies; Journal of the Japanese and International Economies; Latin American Politics and Society; Lex localis – Journal of Local Self-Government; Local Government Studies; Local Government Studies; Millennium: Journal of International Studies; New Political Economy; Nonprofit Management and Leadership; Pacific Review; Party Politics; Perspectives on Politics; Philosophy and Public Affairs; Policy and Politics; Policy and Politics; Policy and Society; Policy and Society; Policy Sciences; Policy Studies; Policy Studies Journal; Policy Studies Journal; Political Analysis; Political Behavior; Political Communication; Political Geography; Political Psychology; Political Research Quarterly; Political Studies; Political Theory; Politics; Politics and Society; Post-Soviet Affairs; PS: Political Science and Politics; Public Administration; Public Administration and Development; Public Administration Review; Public Choice; Public Management Review; Public Money and Management; Public Opinion Quarterly; Public Performance and Management Review; Public Personnel Management; Public Policy and Administration; Regulation and Governance; Review of

International Organizations; Review of International Political Economy; Review of International Studies; Review of Policy Research; Review of Public Personnel Administration; Review of World Economics; Science and Public Policy; Security Dialogue; Security Studies; Social Policy and Administration; Social Science Quarterly; Socio-Economic Review; Studies in Comparative International Development; Studies in Conflict and Terrorism; Survival; Terrorism and Political Violence; American Review of Public Administration; Third World Quarterly; Transylvanian Review of Administrative Sciences; VOLUNTAS: International Journal of Voluntary and Nonprofit Organizations; West European Politics; World Economy; World Politics

Meta-analyses included in the sample

Ahmadov, Anar K. 2014. "Oil, Democracy, and Context: A Meta-Analysis." *Comparative Political Studies* 47(9): 1238–67.

Arceneaux, Kevin, and David W. Nickerson. 2009. "Who Is Mobilized to Vote? A Re-Analysis of 11 Field Experiments." *American Journal of Political Science* 53(1): 1–16.

Askarov, Zohid, Hristos Doucouliagos, Martin Paldam, and T.D. Stanley. 2021. "Rewarding Good Political Behavior: US Aid, Democracy, and Human Rights." *European Journal of Political Economy*: 102089.

Awan, Sahar, Germà Bel, and Marc Esteve. 2020. "The Benefits of PSM: An Oasis or a Mirage?" *Journal of Public Administration Research & Theory* 30(4): 619–35.

Balliet, Daniel. 2010. "Communication and Cooperation in Social Dilemmas: A Meta-Analytic Review." *Journal of Conflict Resolution* 54(1): 39–57.

Balliet, Daniel, Joshua M. Tybur, Junhui Wu, Christian Antonellis, and Paul AM Van Lange. 2018. "Political Ideology, Trust, and Cooperation: In-Group Favoritism among Republicans and Democrats during a US National Election." *Journal of Conflict Resolution* 62(4): 797–818.

Barnhart, Joslyn N., Robert F. Trager, Elizabeth N. Saunders, and Allan Dafoe. 2020. "The Suffragist Peace." *International Organization* 74(4): 633–70.

Belle, Nicola, and Paola Cantarelli. 2017. "What Causes Unethical Behavior? A Meta-Analysis to Set an Agenda for Public Administration Research." *Public Administration Review* 77(3): 327–39.

Bhatti, Yosef, Jens Olav Dahlgaard, Jonas Hedegaard Hansen, and Kasper M. Hansen. 2019. "Is Door-to-Door Canvassing Effective in Europe? Evidence from a Meta-Study across Six European Countries." *British Journal of Political Science* 49(1): 279–90.

Blair, Graeme, Darin Christensen, and Aaron Rudkin. 2020. "Do Commodity Price Shocks Cause Armed Conflict? A Meta-Analysis of Natural Experiments." *American Political Science Review* 115(2): 709–116.

Broderstad, Troy Saghaug. 2018. "A Meta-Analysis of Income and Democracy." *Democratization* 25(2): 293–311.

Burke, Brian L., Spee Kosloff, and Mark J. Landau. 2013. "Death Goes to the Polls: A Meta-Analysis of Mortality Salience Effects on Political Attitudes." *Political Psychology* 34(2): 183–200.

Colagrossi, Marco, Domenico Rossignoli, and Mario A. Maggioni. 2020. "Does Democracy Cause Growth? A Meta-Analysis (of 2000 Regressions)." *European Journal of Political Economy* 61: 101824.

Dinesen, Peter Thisted, Merlin Schaeffer, and Kim Mannemar Sønderskov. 2020. "Ethnic Diversity and Social Trust: A Narrative and Meta-Analytical Review." *Annual Review of Political Science* 23(1): 441–65.

Ding, Fangda, Jiahuan Lu, and Norma M. Riccucci. 2021. "How Bureaucratic Representation Affects Public Organizational Performance: A Meta-Analysis." *Public Administration Review* 81(6): 1003–18.

Doucouliagos, Chris, and Mehmet Ali Ulubaşoğlu. 2006. "Economic Freedom and Economic Growth: Does Specification Make a Difference?" *European Journal of Political Economy* 22(1): 60–81.

———. 2008. "Democracy and Economic Growth: A Meta-Analysis." *American Journal of Political Science* 52(1): 61–83.

Efendic, Adnan, Geoff Pugh, and Nick Adnett. 2011. "Institutions and Economic Performance: A Meta-Regression Analysis." *European Journal of Political Economy* 27(3): 586–99.

Eshuis, Jasper et al. 2021. "The Effect of the EU-Brand on Citizens' Trust in Policies: Replicating an Experiment." *Public Administration Review* 81(4): 776–86.

Gerrish, Ed. 2016. "The Impact of Performance Management on Performance in Public Organizations: A Meta-Analysis." *Public Administration Review* 76(1): 48–66.

Green, Donald P., and Alan S. Gerber. 2019. *Get Out the Vote: How to Increase Voter Turnout*. Brookings Institution Press.

Greenberg, David H., Charles Michalopoulos, and Philip K. Robin. 2006. "Do Experimental and Nonexperimental Evaluations Give Different Answers about the Effectiveness of Government-Funded Training Programs?" *Journal of Policy Analysis & Management* 25(3): 523–52.

Heimberger, Philipp. 2020. "Does Economic Globalization Affect Government Spending? A Meta-Analysis." *Public Choice* 187(3): 349–74.

———. 2021. “Corporate Tax Competition: A Meta-Analysis.” *European Journal of Political Economy* 187(349–374).

Heinemann, Friedrich, Marc-Daniel Moessinger, and Mustafa Yeter. 2018. “Do Fiscal Rules Constrain Fiscal Policy? A Meta-Regression-Analysis.” *European Journal of Political Economy* 51: 69–92.

Holbein, John B., Marcos A. Rangel, Raeal Moore, and Michelle Croft. 2021. “Is Voting Transformative? Expanding and Meta-Analyzing the Evidence.” *Political Behavior*: 1–30.

Homberg, Fabian, Dermot McCarthy, and Vurain Tabvuma. 2015. “A Meta-Analysis of the Relationship between Public Service Motivation and Job Satisfaction.” *Public Administration Review* 75(5): 711–22.

Houck, Shannon C., and Lucian Gideon Conway. 2019. “Strategic Communication and the Integrative Complexity-Ideology Relationship: Meta-Analytic Findings Reveal Differences Between Public Politicians and Private Citizens in Their Use of Simple Rhetoric.” *Political Psychology* 40(5): 1119–41.

Incerti, Trevor. 2020. “Corruption Information and Vote Share: A Meta-Analysis and Lessons for Experimental Design.” *American Political Science Review* 114(3): 761–74.

Kalla, Joshua L., and David E. Broockman. 2018. “The Minimal Persuasive Effects of Campaign Contact in General Elections: Evidence from 49 Field Experiments.” *American Political Science Review* 112(1): 148–66.

Lau, Richard R., Lee Sigelman, and Ivy Brown Rovner. 2007. “The Effects of Negative Political Campaigns: A Meta-Analytic Reassessment.” *Journal of Politics* 69(4): 1176–1209.

Li, Quan, Erica Owen, and Austin Mitchell. 2018. “Why Do Democracies Attract More or Less Foreign Direct Investment? A Metaregression Analysis.” *International Studies Quarterly* 62(3): 494–504.

Lu, Jiahuan. 2016. “The Philanthropic Consequence of Government Grants to Nonprofit Organizations.” *Nonprofit Management & Leadership* 26(4): 381–400.

———. 2018. “Fear the Government? A Meta-Analysis of the Impact of Government Funding on Nonprofit Advocacy Engagement.” *American Review of Public Administration* 48(3): 203–18.

Lu, Jiahuan, Weiwei Lin, and Qiushi Wang. 2019. “Does a More Diversified Revenue Structure Lead to Greater Financial Capacity and Less Vulnerability in Nonprofit Organizations? A Bibliometric and Meta-Analysis.” *VOLUNTAS: International Journal of Voluntary & Nonprofit Organizations* 30(3): 593–609.

Matthes, Jörg et al. 2019. “A Meta-Analysis of the Effects of Cross-Cutting Exposure on Political Participation.” *Political Communication* 36(4): 523–42.

Merkle, Jessica S., and Michelle Andrea Phillips. 2018. "The Wage Impact of Teachers Unions: A Meta-Analysis." *Contemporary Economic Policy* 36(1): 93–115.

Munzert, Simon, and Sebastian Ramirez-Ruiz. 2021. "Meta-Analysis of the Effects of Voting Advice Applications." *Political Communication* 38(6): 1–16.

O'Brochta, William. 2019. "A Meta-Analysis of Natural Resources and Conflict." *Research & Politics* 6(1): 2053168018818232.

Owen, Erica, and Quan Li. 2020. "The Conditional Nature of Publication Bias: A Meta-Regression Analysis." *Political Science Research & Methods*: 1–11.

Philips, Andrew Q. 2016. "Seeing the Forest through the Trees: A Meta-Analysis of Political Budget Cycles." *Public Choice* 168(3): 313–41.

Schwarz, Susanne, and Alexander Coppock. 2021. "What Have We Learned About Gender From Candidate Choice Experiments? A Meta-Analysis of 67 Factorial Survey Experiments." *Journal of Politics*: 40.

Trinn, Christoph, and Thomas Wencker. 2020. "Integrating the Quantitative Research on the Onset and Incidence of Violent Intrastate Conflicts." *International Studies Review* (1): 115–39.

Walter, Nathan, Jonathan Cohen, R. Lance Holbert, and Yasmin Morag. 2020. "Fact-Checking: A Meta-Analysis of What Works and for Whom." *Political Communication* 37(3): 350–75.

de Wit, Arjen, and René Bekkers. 2017. "Government Support and Charitable Donations: A Meta-Analysis of the Crowding-out Hypothesis." *Journal of Public Administration Research & Theory* 27(2): 301–19.

Yesilyurt, Filiz, and M Ensar Yesilyurt. 2019. "Meta-Analysis, Military Expenditures and Growth." *Journal of Peace Research* 56(3): 352–63.

Zhang, Jiasheng, Wenna Chen, Nicolai Petrovsky, and Richard M. Walker. 2021. "The Expectancy-Disconfirmation Model and Citizen Satisfaction with Public Services: A Meta-Analysis and an Agenda for Best Practice." *Public Administration Review* 82(1): 147–59.

B DATA CLEANING

We gathered estimates and standard errors for all hypothesis tests included in the data files or published tables and figures of the 31 meta-analytic articles identified during data collection. When standard errors were missing, we calculated them from related information when available (e.g. from confidence intervals or variances). We merged this dataset with Doucouliagos', which comprised hypothesis tests for 15 meta-analytic publications.

We filtered out hypothesis tests based on a series of criteria, beginning with a missing estimate or standard error. We dropped observations containing invalid computations or implausible values; these often involved pure zero estimates or pure zero standard errors (both probably the result of data entry mistakes). We excluded standard errors smaller than 10^{-10}. While this was an arbitrary tolerance level, using different thresholds did not make a substantial difference. We checked for mistakes in the data such as p-values reported as standard errors, and we computed transformations when necessary (e.g., deriving partial correlations from t-statistics and degrees of freedom). We also checked whether excluding estimates with absolute values less than or equal to 10^{-5} made a difference (it did not). We visually examined funnel plots to identify variables in need of transformation or visually odd patterns in some meta-analyses, which were then double-checked for accuracy. Lastly, we dropped all observations drawn from meta-analyses with fewer than 5 aggregated hypothesis tests.

As a check on the faithfulness of our cleaning process, we cross-validated our cleaned dataset against authors' replication scripts when available. We ran meta-analyses using our cleaned dataset rather than the raw data originally found online or provided to us by the authors. We were generally able to produce results that matched what authors describe in their original meta-analytic articles, though in some cases we had fewer observations due to our filtering criteria. In some rare cases, we were unable to back out how authors computed their results from their raw data without their replication scripts.

All in all, our dataset of estimates and standard errors contains 16649 rows with unique identifiers for hypothesis tests, meta-analyses, and meta-analytic articles.

C EXPERT SURVEY

Our expert survey was conducted in June 2022. We emailed 478 methodologists (555 minus 77 for whom emails were undeliverable). 131 answered all of our questions, yielding a response rate of 27%. The design was approved by two different institutional review boards from the universities of two coauthors of the present paper. Our analysis of the survey data was preregistered following the AsPredicted template, and the pre-analysis plan is available at https://aspredicted.org/blind.php?x=G28_YCP.

Below, we present the following: the email template used to recruit respondents; the full survey questionnaire; and an assessment of whether or not tests in meta-analysis are biased according to our expert sample.

Recruitment email template

Dear Professor [LAST NAME],

We write to invite you to participate in a survey to gauge expectations around levels of statistical power in political science research. The survey consists of only 3 questions and should take about 2 minutes of your time.

We are emailing you because you published in Political Analysis since 2010, and so are working in political methodology and are part of our expert sample. The survey will close in one month. If you would like to do the survey, please click this link to see the consent form and take the survey

Or copy and paste the URL below into your internet browser: [URL]

If you click this link, you will be removed from our mailing list and we will not email you about this again

Please do not hesitate to contact us if you have any questions or concerns by sending an email to [EMAIL].

This project has been reviewed by the [IRB] for compliance with federal guidelines for research involving human participants ([CERTIFICATE NUMBER]).

Thank you,
[AUTHORS NAMES]

Questionnaire

Our questions focus on statistical power in political science research. As a reminder, higher statistical power means a better chance to detect an effect, if in fact an effect exists.

For the two questions on this page, consider all hypothesis tests reported in the 50 peer-reviewed journals with the highest impact factors in political science, international relations, and public administration over the past two decades.

What percent of the tests in these journals do you believe had at least **80% power** to reject the null with a significance level of 0.05? [Slider 0-100; Don't know]

What percent of the tests in these journals do you believe had at least **50% power** to reject the null with a significance level of 0.05? [Slider 0-100; Don't know]

Our final question is about statistical power in the subset of research articles that end up being included in meta-analyses. Not every published study or hypothesis test ends up being included in a meta-analysis. In some cases, for example, there are not enough comparable estimates to conduct a meta-analysis.

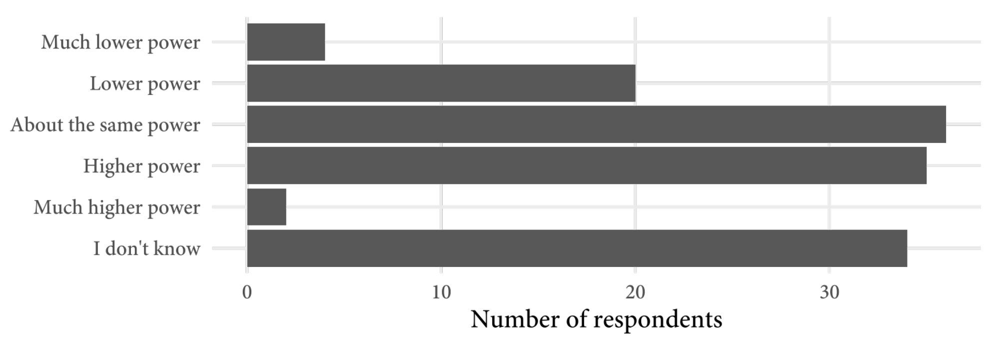
For this question, we are **not** interested in the meta-analyses themselves, but rather in the individual hypothesis tests that are included, aggregated, and summarized in meta-analyses.

Do you think that the individual hypothesis tests that end up being included in meta-analyses are likely to have lower, equal, or higher power than those that do not end up included in meta-analyses? [Much lower power; Lower power; About the same power; Higher power; Much higher power; Don't know]

Are tests in meta-analyses biased?

In our expert survey we asked the respondents if they thought that hypothesis tests in meta-analyses were likely to have lower or higher power than tests that do not end up in meta-analyses. This helps us understand if our strategy of starting with meta-analyses produces a biased picture of power in the overall discipline, and it helps us link the respondent's guesses about power to our core results. Respondents do not generally think that tests in meta-analyses will have lower power.

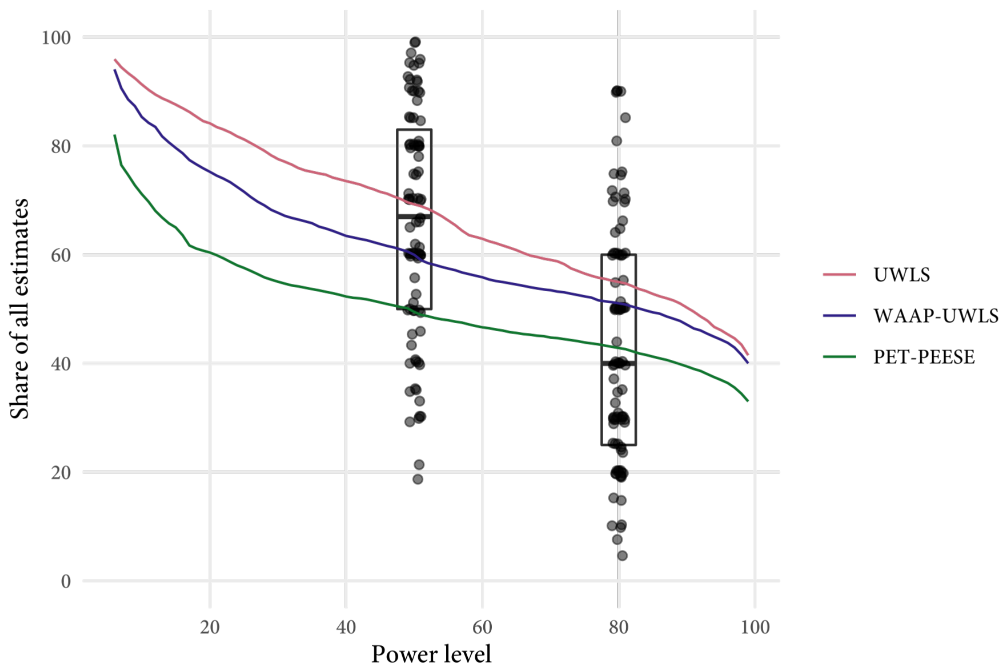
Figure C.1: Power of tests in meta-analyses compared to those not in meta-analyses



D Minimum effects of interest

Here we show that our core qualitative result that much of quantitative political science research is under-powered holds even if researchers are powering their studies to detect effects that are much larger than the consensus effects reported in the literature. We do this by artificially inflating our estimates of population mean effects by a factor of five and then re-running the analysis. Using the most generous UWLS population mean effect and this five-fold upward adustment, it is still the case that around 1 in 3 tests fails to reach 50% power.

Figure D.2: Share of estimates by power level, after each population mean effect estimate is inflated by a factor of 5



E OUTLIERS: ESTIMATES

To ensure that our main results are not driven by extreme observations, we replicate our core findings from Figures 1 and 2 while excluding outliers identified with Tukey's "fences."

For each meta-analysis, we first compute the interquartile range of estimates and standard errors. Then, we categorize an estimate or standard error as an outlier if it is 1.5 interquartile range above the 75th percentile or below the 25th percentile of estimates or standard errors within that meta-analysis.

To exclude influential observations, we calculate DFBeta statistics for each meta-analysis and reproduce our main figure when we exclude observations with DFBeta scores above 2 divided by the square root of the number of estimates in the relevant meta-analysis (Belsley, Kuh, and Welsch 1980).

Figure E.3: Distribution of Z statistics in the full sample of estimates, excluding observations with outlying estimates or outlying standard errors.

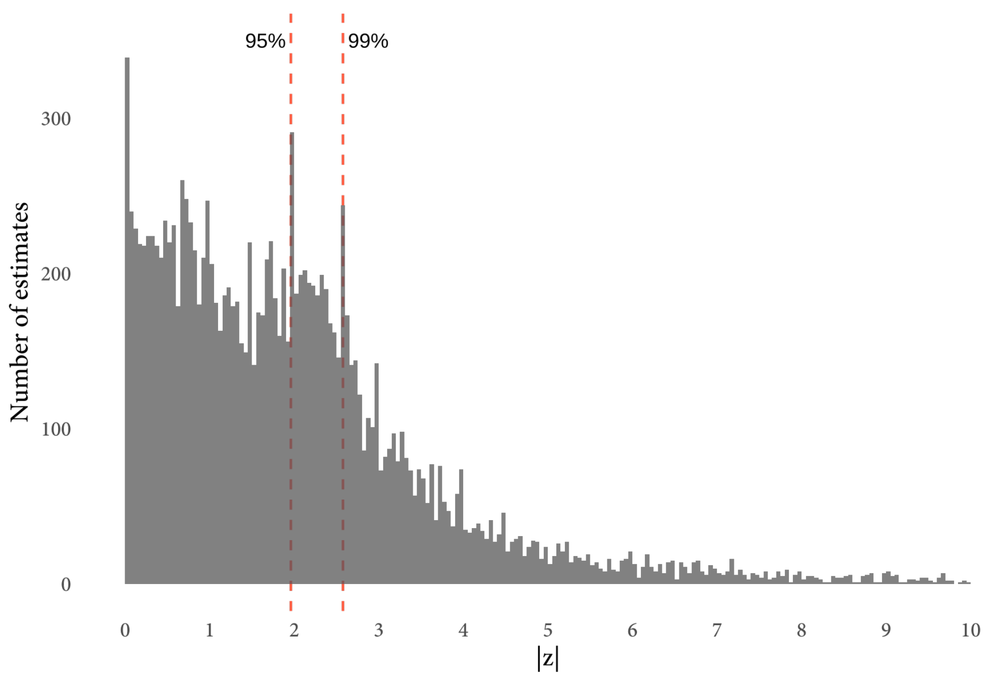
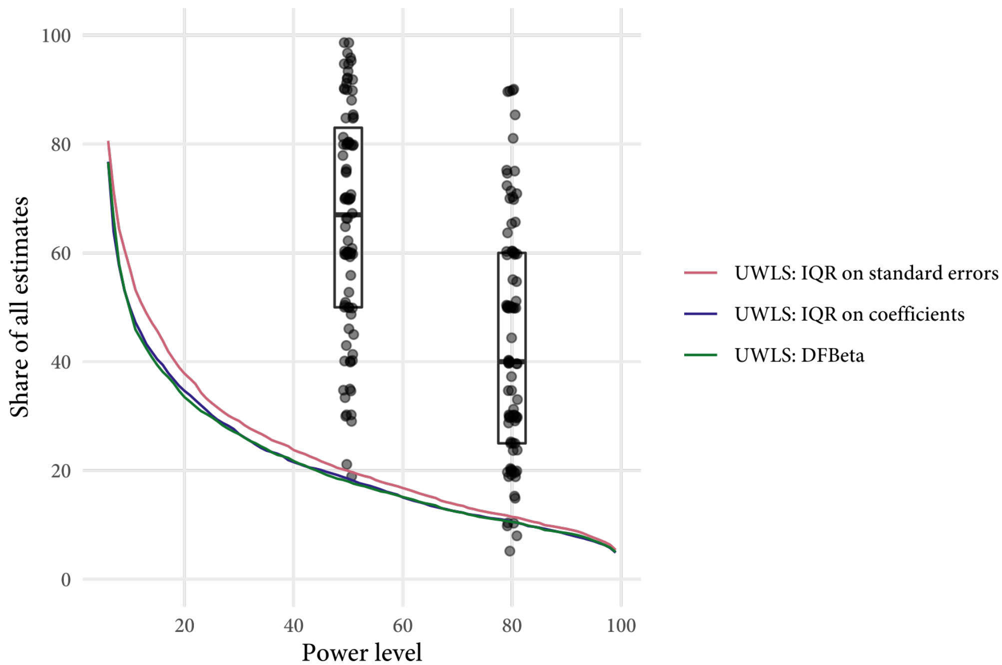


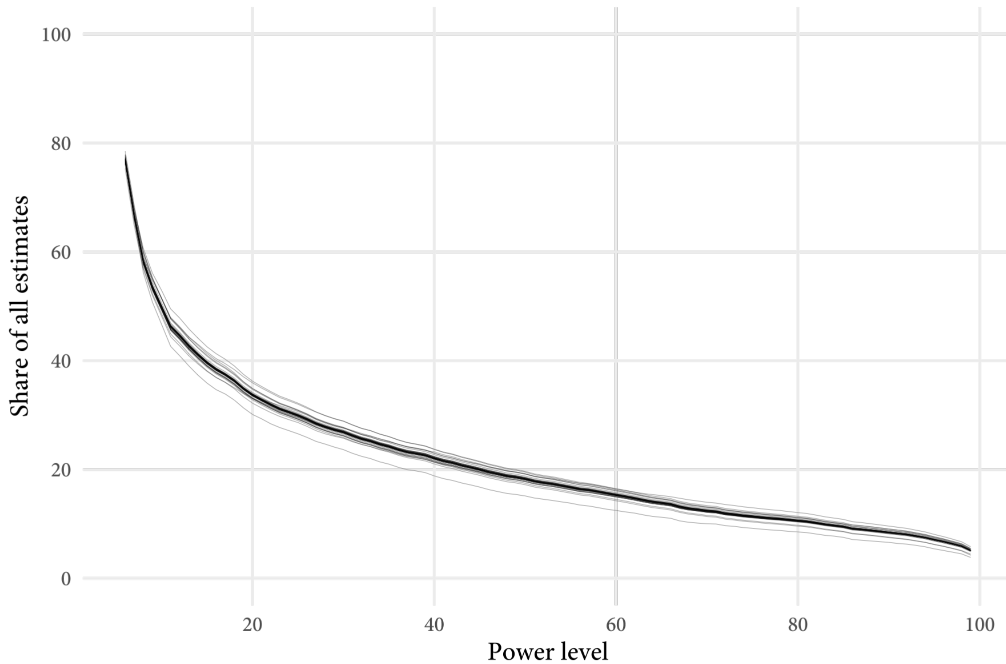
Figure E.4: Share of estimates by power level, excluding outliers.



F OUTLIERS: META-ANALYSES

To ensure that no single meta-analytic paper is driving the results, we replicate the UWLS estimates portion of Figure 2 while sequentially withholding the data from one paper at a time. The lines are plotted with high transparency, so the dark black line emerges only from overplotting. The results are quite stable.

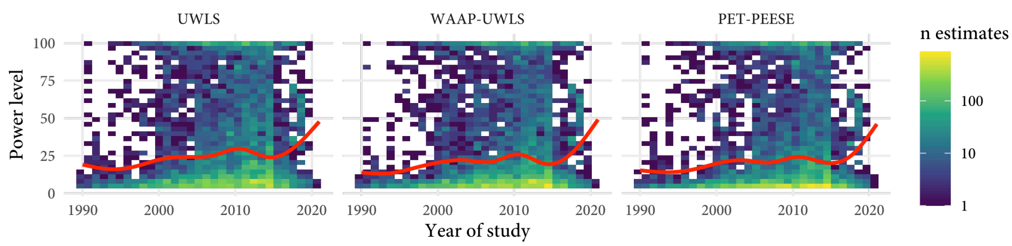
Figure F.5: Share of estimates by power level, excluding one full meta-analytic paper at a time. Unrestricted weighted least squares estimates.



G POWER OVER TIME

Mean power appears to be modestly rising over time. Our three methods for estimating population mean effects suggest that power may be rising more quickly in the past five years but data for those years are fairly sparse. We do not show trends for tests before 1990 as the data are very sparse. Median power also increases over this time but less dramatically.

Figure G.6: Relationship between the year of the test and its power.



H SOFTWARE BIBLIOGRAPHY

Arel-Bundock, Vincent. 2022. *modelsummary: Summary Tables and Plots for Statistical Models and Data: Beautiful, Customizable, and Publication-Ready*. https://vincentarelbundock.github.io/modelsummary/.

Barrett, Malcolm. 2021. *ggokabeito: Okabe-Ito Scales for ggplot2 and ggraph*. https://github.com/malcolmbarrett/ggokabeito

Blair, Graeme, Jasper Cooper, Alexander Coppock, and Macartan Humphreys. 2019. "Declaring and Diagnosing Research Designs." *American Political Science Review* 113 (3): 838–59. https://declaredesign.org/paper.pdf.

Cameron, Allan, and Teun van den Brand. 2022. *geomtextpath: Curved Text in ggplot2*. https://allancameron.github.io/geomtextpath/.

Dowle, Matt, and Arun Srinivasan. 2021. *data.table: Extension of data.frame*. https://rdatatable.gitlab.io/data.table/

Lang, Michel. 2017. "checkmate: Fast Argument Checks for Defensive R Programming." *The R Journal* 9 (1): 437–45. https://doi.org/10.32614/RJ-2017-028.

Müller, Kirill. 2020. *here: A Simpler Way to Find Your Files*. https://here.r-lib.org/

Murdoch, Duncan. 2020. *tables: Formula-Driven Table Generation*. https://r-forge.r-project.org/projects/tables/.

R Core Team. 2022. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. https://www.R-project.org/.

Timm, Andrew. 2019. *retrodesign: Tools for Type S (Sign) and Type M (Magnitude) Errors*. https://github.com/andytimm/retrodesign.

Vaughan Davis, Matt Dancho (2022). *furrr: Apply Mapping Functions in Parallel using Futures*. https://github.com/DavisVaughan/furrr.

Wickham, Hadley, Mara Averick, Jennifer Bryan, Winston Chang, Lucy D'Agostino McGowan, Romain François, Garrett Grolemund, et al. 2019. "Welcome to the tidyverse." *Journal of Open Source Software* 4 (43): 1686. https://doi.org/10.21105/joss.01686.

Wickham H, Jennifer Bryan (2022). *readxl: Read Excel Files*. https://readxl.tidyverse.org.

Wickham, Hadley, Evan Miller, and Danny Smith. 2022. *haven: Import and Export SPSS, Stata and SAS Files*. https://haven.tidyverse.org/

Wickham Hadley 2021. *conflicted: An Alternative Conflict Resolution Strategy*. https://conflicted.r-lib.org, https://github.com/r-lib/conflicted.

Xie, Yihui. 2022. *xaringan: Presentation Ninja*. https://github.com/yihui/xaringan.

Zeileis, Achim, Susanne Köll, and Nathaniel Graham. 2020. "Various Versatile Variances: An Object-Oriented Implementation of Clustered Covariances in R." *Journal of Statistical Software* 95 (1): 1–36. https://doi.org/10.18637/jss.v095.i01.

Zhu, Hao. 2021. *kableExtra: Construct Complex Table with knitr::kable() + %>%*. https://haozhu233.github.io/kableExtra/