2021

The Transparency of Quantitative Empirical Legal Research (2018–2020)

Jason Chin
University of Sydney

Kathryn Zeiler
Boston University School of Law

Natali Dilevski
University of Sydney

Alexander Holcombe
University of Sydney

Rosemary Gatfield- Jeffries
University of Sydney

See next page for additional authors

Authors

Jason Chin, Kathryn Zeiler, Natali Dilevski, Alexander Holcombe, Rosemary Gatfield- Jeffries, Ruby Bishop, Simine Vazire, and Sarah Schiavone

BU School of Law

THE TRANSPARENCY OF QUANTITATIVE EMPIRICAL LEGAL RESEARCH (2018-2020)

Boston University School of Law
Research Paper Series No. 22-4

February 28, 2022

Kathryn Zeiler
Boston University School of Law

Jason M. Chin
School of Law, University of Sydney

Natali Dilevski
School of Psychology, University of Sydney

Alex O. Holcombe
School of Psychology, University of Sydney

Rosemary Gatfield-Jeffries
University of Sydney

Ruby Bishop
School of Law, University of Sydney

Simine Vazire
School of Psychology, University of Melbourne

Sarah R. Schiavone
University of California, Davis

The Transparency of Quantitative Empirical Legal Research (2018–2020)

Jason M. Chin[1], Kathryn Zeiler[2], Natali Dilevski[3], Alex O. Holcombe[3], Rosemary Gatfield-Jeffries[4], Ruby Bishop[1], Simine Vazire[5], Sarah R. Schiavone[6]

December 31, 2021

[1] School of Law, University of Sydney, [2] School of Law, Boston University, [3] School of Psychology, University of Sydney, [4] University of Sydney, [5] School of Psychology, University of Melbourne, [6] University of California, Davis

Abstract

Scientists are increasingly concerned with making their work easy to verify and build upon. Associated practices include sharing data, materials, and analytic scripts, and preregistering protocols. This has been referred to as a "credibility revolution". The credibility of empirical legal research has been questioned in the past due to its distinctive peer review system and because the legal background of its researchers means that many often are not trained in study design or statistics. Still, there has been no systematic study of transparency and credibility-related characteristics of published empirical legal research. To fill this gap and provide an estimate of current practices that can be tracked as the field evolves, we assessed 300 empirical articles from highly ranked law journals including both faculty-edited journals and student-edited journals. We found high levels of article accessibility (86% could be accessed without a subscription, 95% CI = [82%, 90%]), especially among student-edited journals (100% accessibility). Few articles stated that a study's data are available, (19%, 95% CI = [15%, 23%]), and only about half of those datasets are reportedly available without contacting the author. Preregistration (3%, 95% CI = [1%, 5%]) and availability of analytic scripts (6%, 95% = [4%, 9%]) were very uncommon. We suggest that empirical legal researchers and the journals that publish their work cultivate norms and practices to encourage research credibility.

Introduction

> The credibility of scientific claims is rooted in the evidence supporting them, which includes the methodology applied, the data acquired, and the process of methodology implementation, data analysis and outcome interpretation. Claims become credible by the community reviewing, critiquing, extending and reproducing the supporting evidence. However, without transparency, claims only achieve credibility based on trust in the confidence or authority of the originator. Transparency is superior to trust.[1]

Increasing the transparency of research is a key component of the ongoing "credibility revolution" occurring in many fields.[2] This movement, as the above epigraph explains, seeks to improve research credibility by ensuring that claims can be tested and critiqued by other researchers. Further benefits of the credibility revolution are efficiency, in that transparent research is reusable by other researchers to explore new questions,[3] and that transparent research enhances public trust in science, comporting with lay expectations about how science ought to be conducted.[4] Despite its work being cited by courts and policymakers,[5] the field of empirical legal research has so far largely refrained from engaging in significant reforms.[6] In this article, we measure the transparency and other related characteristics of 300 empirical legal studies published between 2018 and 2020 in law journals rated highly by traditional metrics. For the

[1] Marcus R. Munafò et al., *A manifesto for reproducible science*, 1 NAT. HUM. BEHAV. 1 (2017) at 4-5.
[2] *Id*.; Joshua D. Angrist & Jörn-Steffen Pischke, *The credibility revolution in empirical economics: How better research design is taking the con out of econometrics*, 24(2) J. ECON. PERSPECT. 3 (2010); Simine Vazire, *Implications of the credibility revolution for productivity, creativity, and progress*, 13(4) PERSPECT. PSYCHOL. SCI. 411 (2018); Garret Christensen et al., *Open Science Practices are on the Rise: The State of Social Science (3S) Survey*, *MetaArXiv*, https://osf.io/preprints/metaarxiv/5rksu.
[3] Munafò et al., *supra* note 1 at 2-3.
[4] CARY FUNK ET AL., *Trust and Mistrust in Americans' Views of Scientific Experts*, Pew Research Center (2019) 24; Justin T. Pickett & Sean Patrick Roche, *Questionable, Objectionable or Criminal? Public Opinion on Data Fraud and Selective Reporting in Science*, 24 SCI. ENG. ETHICS 151 (2018).
[5] Kathryn Zeiler, *The Future of Empirical Legal Scholarship: Where Might We Go from Here?* 66 J. LEGAL EDUC. 78 (2016); Jason M. Chin, Malgorzata Lagisz & Shinichi Nakagawa, *Where is the evidence in evidence-based law reform?* U.N.S.W.L.J (2021).
[6] Zeiler, *id.*; Jason M. Chin & Kathryn Zeiler, *Replicability in Empirical Legal Research*, 17 ANNU. REV. LAW SOC. SCI. 239 (2021).

purposes of this article, we define empirical research as research that performs quantitative analysis on data.[7]

The credibility revolution and the role of research transparency

The "credibility revolution"[8] responded to a "crisis"[9] reported in many fields, in which researchers were unable to replicate the findings of published studies.[10] This crisis has been well-documented in the context of several social scientific fields that often run adjacent to legal research, such as economics,[11] criminology,[12] and psychology.[13] Failures to replicate and other controversies were especially well-publicized in psychology.[14] These included: a leading journal's publication of nine studies finding evidence for extrasensory perception,[15] failures to replicate the "power pose" phenomenon (including that standing in powerful positions inhibits the release of stress hormones),[16] the discovery that several leading researchers' work was

[7] This generally tracks the definition provided in Michael Heise, *The past, present, and future of empirical legal scholarship: judicial decision making and the new empiricism*, UNIV. ILL. LAW REV. 819 (2002); we acknowledge that many definitions of empirical legal research have been offered, see Shari Seidman Diamond & Pam Mueller, *Empirical Legal Scholarship in Law Reviews*, 6 ANNU. REV. LAW SOC. SCI. 581 (2010) at 582-583. As we detail below, our definition is useful for the present study, which measures transparent practices.
[8] Vazire, *supra* note 2; Munafò, *supra* note 1.
[9] Monya Baker, *1,500 scientists lift the lid on reproducibility*, 533 Nature 452 (2016).
[10] Open Science Collaboration (OSC), *Estimating the Reproducibility of Psychological Science*, 349 SCIENCE 3451 (2015); Richard A. Klein et al., *Investigating variation in replicability: A 'many labs' replication project*, 45(3) SOC. PSYCHOL. 142 (2014); Richard A. Klein et al., Many Labs 2: Investigating Variation in Replicability Across Samples and Settings, 1 ADV. METH. & PRACT. PSYCHOL. SCI. 443 (2018); Charles Ebersole et al., *Many Labs 3: Evaluating participant pool quality across the academic semester via replication*, 67 J. EXP. SOC. PSYCHOL. 68 (2016); Richard A. Klein et al., *Many Labs 4: Failure to Replicate Mortality Salience Effect With and Without Original Author Involvement*, https://psyarxiv.com/vef2c (accessed 2020); Colin F. Camerer et al., *Evaluating replicability of laboratory experiments in economics*, 351 Science 1433 (2016); Colin F. Camerer et al., *Evaluating the replicability of social science experiments in Nature and Science between 2010 and 2015*, 2 NAT. HUM. BEHAV. 637 (2018).
[11] Angrist & Pischke, *supra* note 2; Sarah Necker, *Scientific misbehavior in economics*, 43 RES. POL. 1747 (2014).
[12] Jason M. Chin et al., Questionable Research Practices and Open Science in Quantitative Criminology, J. QUANT. CRIM. (2021).
[13] Leslie K. John et al., *Measuring the Prevalence of Questionable Research Practices With Incentives for Truth Telling*, 23(5) PSYCHOL. SCI. 524 (2012).
[14] See Leif D. Nelson, Joseph Simmons, Uri Simonsohn, *Psychology's Renaissance*, 69 ANNU. REV. PSYCHOL. 511 (2018).
[15] E. J. Wagenmakers et al., *Why psychologists must change the way they analyze their data: The case of psi: Comment on Bem*, 100(3) J. PERS. SOC. PSYCHOL. 426 (2011).
[16] Eva Ranehill et al., *Assessing the Robustness of Power Posing: No Effect on Hormones and Risk Tolerance in a Large Sample of Men and Women*, 26(5) PSYCHOL. SCI. 653 (2015).

fraudulent, and, ultimately, the results of several large multi-lab replication efforts finding that studies published in eminent journals could only be replicated at the original level of statistical significance about 50% of the time.[17]

The revolution involves a host of changes to the research process, such as improved transparency, higher standards of evidence, and more replication research.[18] Transparency-focused reforms make the data and process underlying results more accessible, which make it easier for other researchers to verify, correct, and build upon existing research.[19] Transparency can also be advanced through preregistration (or registration or a pre-analysis plan as it is called in medical research and economics respectively), which is a time-stamped statement of the research protocols and hypotheses that is posted prior to data collection. Preregistration helps reduce and detect selective reporting, also known as "publication bias" or the "file drawer effect".[20] When some research goes unpublished, it can be difficult or impossible for researchers and the public to assess the strength of the evidence for a hypothesis, an empirical phenomenon, or a theory. Preregistration has long been standard, and even legally required in certain cases (e.g., publication of clinical medical trials .[21]

Fuller reporting, in the form of data sharing as well as providing more details of methods and statistical analyses performed, allows other researchers to better scrutinize findings and detect errors in research.[22] For instance, researchers recently discovered a case of data fraud in a

[17] See Nelson, Simmons & Simonsohn et al., *supra* note 14 at 513.
[18] Vazire, *supra* note 2; E. Miguel, *Promoting Transparency in Social Science Research*, 343 SCIENCE 30 (2014).
[19] B. A. Nosek et al., *Promoting an open research culture*, 348 SCIENCE 1422 (2015).
[20] Brian A. Nosek et al., *The Preregistration Revolution*, 115(11) PNAS 2600 (2018).
[21] Kay Dickersin & Iain Chalmers, *Recognizing, investigating and dealing with incomplete and biased reporting of clinical research: from Francis Bacon to the WHO*, 104 J. R. SOC. MED. 104 (2011) 532. One common concern with preregistrations is that they open up the research to be "scooped" by others. However, this can be addressed by embargoing the preregistration for the duration of the data collection and analysis.
[22] Simine Vazire & Alex O. Holcombe, *Where are the Self-Correcting Mechanisms in Science?* REV. GEN. PSYCHOL. (2021).

study purporting to find that signing one's name before versus after providing information in a document reduces dishonesty.[23] This study has been cited often for its legal and policy consequences,[24] including by the UK Behavioural Insights Team (i.e., Nudge Unit).[25] Beyond availability of the raw data, which helped other researchers to uncover the fraud, replication also played a role. Failures to replicate other studies in the paper led to increased scrutiny of the entire set of results, which eventually led researchers to take a closer look at the data. One of the authors of the problematic paper, who had worked on the non-fraudulent studies reported within the same article, wrote in response to the discovery of the fraud:[26]

> Though very painful, this experience has reinforced my strong commitment to the Open Science movement. **As it clearly shows, posting data publicly, pre-registering studies, and conducting replications of prior research is key to scientific progress**.

Transparency can also make research more efficient because other researchers can leverage open data and materials to test new questions, and to synthesize existing data in meta-analyses.[27] On the other hand, research efforts can be wasted in the absence of open data in the sense that those data cannot be obtained by subsequent researchers seeking reuse them. This is because researchers change emails and institutions or leave academic research behind altogether, make them unavailable to share data upon request.[28] Moreover, many researchers who are

---

[23] Uri Simonsohn, Joseph Simmons & Leif D. Nelson, [98] Evidence of Fraud in an Influential Field Experiment About Dishonesty, http://datacolada.org/98 (accessed 2021).
[24] See: Brigitte C. Madrian, *Applying Insights from Behavioral Economics to Policy Design*, 6 ANNU. REV. ECONOM. 663 (2014).
[25] Cabinet Office Behavioural Insights Team, Applying behavioural insights to reduce fraud, error and debt (2012), https://vng.nl/sites/default/files/knowledge_base_compliance/Rapport_201608_Applying_behavioural_insights.pdf (accessed 2021)
[26] Francesca Gino, Gino-memo-data-colada-August16.pdf, http://datacolada.org/storage_strong/Gino-memo-data-colada-August16.pdf (accessed 2021) [**emphasis added**].
[27] Iain Chalmers & Paul Glasziou, *Avoidable waste in the production and reporting of research evidence*, 374 LANCET 86 (2009).
[28] Timothy H. Vines et al., *The Availability of Research Data Declines Rapidly with Article Age*, 24 CURR. BIOL. 94 (2014); Jelte M. Wicherts et al., *Willingness to Share Research Data is Related to the Strength of the Evidence and the Quality of Reporting of Statistical Results*, 6(11) 1 PLOS ONE (2011).

reachable decline to share data and materials when they are contacted, or promise to deliver the data but never follow through.[29]

It is also worth noting that transparent and open research inspires public trust and comports with the public's expectations. For instance, a recent Pew survey found that the public was more likely to trust research when it was openly shared than when it was not.[30] Other researchers have found that the public looks unfavorably upon selectively reporting data that confirms the researcher's hypothesis.[31]

Recent reforms to the peer review and publication process seek to encourage researchers to adopt transparency and credibility-related reforms.[32] For example, over 1,100 journals have integrated the Transparency and Openness (TOP) Guidelines into their guidelines,[33] which, under some implementations, require that researchers provide data, code, methods, and/or preregistration statements.[34] Many journals are also offering authors the option of conducting their research as a registered report. In a registered report, researchers submit a "stage one" manuscript for peer review prior to data collection. The manuscript contains only necessary background and a research plan, and if it passes peer review, the journal commits to publishing the manuscript as long as the authors follow the agreed-upon plan. Recent studies suggest that registered reports are both higher quality and less subject to publication bias.[35] For example, the

---

[29] *Id*.
[30] Funk et al., *supra* note 4; see also Hyunjin Song, David M. Markowitz & Samuel H. Taylor, *Trusting on the Shoulders of Open Giants? Open Science Increases Trust in Science for the Public and Academics*, https://osf.io/g328c (accessed 2021).
[31] Pickett & Roche, *supra* note 4.
[32] See Vazire, *supra* note 2.
[33] Center for Open Science, The TOP Guidelines were created by journals, funders, and societies to align scientific ideals with practices, https://www.cos.io/initiatives/top-guidelines (accessed 2021).
[34] Nosek et al., 2015 *supra* note 19.
[35] Courtney K Soderberg et al. *Initial evidence of research quality of registered reports compared with the standard publishing model*, 5(8) NAT. HUM. BEHAV. 990 (2021); Anne M. Scheel, Mitchell Schijen & Daniël Lakens, *An excess of positive results: Comparing the standard Psychology literature with Registered Reports*, 4(2) AMPPS (2021).

proportion of studies that reject the null hypothesis in favor of the alternative hypothesis is approximately 50% in registered reports, compared to 95% in traditional articles.[36]

Measuring transparency and credibility-related features of published research

Several metascientific studies, across a variety of fields, have conducted "state-of-the-science" audits, in which recent published studies are randomly sampled and coded for various transparency and credibility-related features.[37] These metascientific studies have generally found very low levels of transparency. One study examined psychology articles published from 2014-2017.[38] Only about 2% of the studies sampled had available data, approximately 17% had available materials, and 3% were preregistered.[39] Note, however, that studies published during this timeframe were likely conducted in the early days of the reported crisis in psychology.[40]

While these findings are worrisome, recent reforms may have led to an increase in transparency related practices in recent years. For instance, journals that implemented open data policies (e.g., requiring open data under some circumstances) show substantial increases in the proportion of studies with open data, albeit with imperfect compliance.[41] Similarly, policies that encourage transparency related practices through optional badges (e.g., preregistered articles

[36] Sheel et al., *id*.
[37] Tom E. Hardwicke et al., *An empirical assessment of transparency and reproducibility-related research practices in the social sciences (2014–2017)*, 7(2) R. SOC. OPEN SCI. 190806 (2020); Tom E. Hardwicke et al., *Estimating the Prevalence of Transparency and Reproducibility-Related Research Practices in Psychology (2014–2017)*, PERSPECT. PSYCHOL. SCI. (2021); Austin L. Johnson et al., *An assessment of transparency and reproducibility-related research practices in otolaryngology*, 130(8) THE LARYNGOSCOPE 1894 (2020); Mopileola Tomi Adewumi et al., *An evaluation of the practice of transparency and reproducibility in addiction medicine literature*, 112 ADDICTIVE BEHAVIORS 106560 (2021); Elizabeth R. Tenney et al., *Open Science and Reform Practices in Organizational Behavior Research over Time (2011 to 2019)*, https://psyarxiv.com/vr7f9/ (accessed 2021).
[38] Hardwicke et al., 2021, *Id*.
[39] *Id*.
[40] Nelson et al., *supra* note 14.
[41] Tom E. Hardwicke et al., *Data availability, reusability, and analytic reproducibility: evaluating the impact of a mandatory open data policy at the journal Cognition*, 5 R. SOC. OPEN SCI. 180448 (2018); Anisa Rowhani-Farid & Adrian G. Barnett, *Has open data arrived at the British Medical Journal (BMJ)? An observational study*, 6 BMJ OPEN e011784 (2016); Antica Culina, *Low availability of code in ecology: A call for urgent action*, 18(7) PLOS BIOL. e3000763 (2020).

receive a visible signifier on their first page showing the study was preregistered, see https://osf.io/tvyxz/wiki/home/) are correlated with uptake of those practices.[42] For example, the journal *Psychological Science* showed an uptick in articles reporting data availability (to nearly 40%) after it began awarding badges. However, psychology, at this time, was undergoing a widely discussed crisis and a randomized trial of badges in a different field found no effect, suggesting that awarding badges is not sufficient in all contexts to change research behavior.[43] Finally a survey across many fields directly asking researchers about when they first engaged in a transparency-related practice (open data, open materials, open code, and preregistration) found that uptake has increased in recent years, suggesting that recent reforms and initiatives are moving the needle.[44]

Empirical legal research

Numerous researchers have questioned the credibility of empirical legal research. In a relatively early critique, Epstein and King reviewed all law journal articles published over a ten-year period that contain the word "empirical" in the title.[45] They found numerous errors, generally centering around poor transparency and reproducibility. For instance, many authors had not fully described how they gathered data and then reasoned from that data to their conclusion. Similar critiques have been levied since then, such as reports that empirical legal

[42] Mallory C. Kidwell et al., *Badges to Acknowledge Open Practices: A Simple, Low-Cost, Effective Method for Increasing Transparency*, 14(5) PLoS BIOL. e1002456 (2016); Anisa Rowhani-Farid & Adrian G. Barnett, *Badges for sharing data and code at Biostatistics: an observational study*, 7 F1000RESEARCH 90 (2018).
[43] Anisa Rowhani-Farid, Adrian Aldcroft & Adrian G. Barnett, *Did awarding badges increase data sharing in BMJ Open? A randomized controlled trial*, 7 R. SOC. OPEN SCI. 7191818 (2020). Another reason badges may not have been successful in this study is that the field was less aware of them, with the associated journal not taking as active a role in explaining to authors what each badge signifies.
[44] Christensen et al., *supra* note 2.
[45] Lee Epstein & Gary King, *The Rules of Inference*, 69 U. CHI. L. REV. 1 (2002).

studies misinterpret statistical results (e.g., p-values), misapply statistical methods, and fail to verify that the assumptions underlying their methods were met.[46]

While these issues are not specific to empirical legal research, several other characteristics of the field heighten existing concerns and may make the field harder to improve than other cognate fields. For instance, as individuals formally trained in the law rather than in empirical science, many authors of empirical legal work have less methodological expertise than researchers in other sciences. Moreover, many legal journals are not peer-reviewed, with submissions examined solely by student editors. Many of these editors likely lack the knowledge to appropriately review empirical work and may make publication decisions based on other features. Moreover, author guidelines at law journals, especially those that are student-reviewed and -edited, include almost no transparency requirements.[47]

While these features of empirical legal scholarship are worrying, we know little about the state of the research today. The last large study that assessed factors related to credibility was conducted 20 years ago. It included only articles with "empirical" in the title[48] and the results were not quantified in a way that makes them easy to update and revisit. Beyond these limits, we now have more specific questions to ask about empirical legal research: are authors making data available and if so, how are they doing that? Are empirical legal researchers taking up the more

[46] Shari Seidman Diamond, *Empirical Legal Scholarship: Observations on Moving Forward*, 113 NW. U. L. REV. 1229 (2019).; Zeiler, *supra* note 5; Gregory Mitchell, *Empirical legal scholarship as scientific dialogue*, 83 N.C. L. REV. 167 (2004). In other metaresearch in empirical legal research, Diamond and Mueller (*supra* note 7) tracked the amount of empirical research in law journals, finding that only about 10% of articles in highly ranked U.S. law journals contained original empirical work – see also Michael Heise, *An Empirical Analysis of Empirical Legal Scholarship Production*, 1990-2009, 2011 U. ILL. L. REV. 1739 (2011). And Hall and Wright examined trends in the use of one particular empirical legal research methodology—systematic analysis of judicial decisions. They found that papers in this area rarely cited methodological articles and seemed to reinvent the wheel, methodologically, in each iteration: Mark A. Hall & Ronald F. Wright, *Systematic Content Analysis of Judicial Opinions*, 96 CALIF. L. REV. 63 (2008).

[47] Chin & Zeiler, *supra* note 6.

[48] Epstein & King, *supra* note 45 at 15-16.

recent (in social science, at least) practice of preregistration? This study seeks to address these questions and others.

Our results are mixed, but most suggest that transparency reforms are needed. While most empirical legal research is freely available, a substantial portion sits behind paywalls. Most of the 300 articles we examined contain no statements regarding funding or conflicts of interest, although just over half of the articles published in faculty-edited journals do contain statements related to funding sources. We find that both student-edited and faculty-edited journals do poorly when it comes to availability of data and analysis scripts. Materials availability is more common—about half of the articles in both types of journals mention that materials are available. Very few of the studies were preregistered, and very few attempt to replicate original work. After presenting our results, we suggest steps both authors and journal editors can take to enhance the transparency of empirical legal research.

Methods

Overview and design

To estimate the transparency of credibility-related features of recent empirical legal research, we examined a sample of 300 law journal articles published between 2018 and 2020. We chose this sample size because it is consistent with many previous transparency studies.[49] Based on those authors' reports[50] of how long it took them to extract the relevant features of each article, we judged that 300 articles was a practical target number to code given our available

[49] N = 250 in Hardwicke et al., 2020, *supra* note 37; N = 250 in Hardwicke et al., 2021, *supra* note 37; N = 286 in Johnson et al., *supra* note 37; N = 244 in Adewumi et al., *supra* note 37; N = 2234 in Tenney et al., *supra* note 37; N = 300 also provides a margin of sampling error of about 6%, although we did not include this in our *a priori* justification: American Association for Public Opinion Research, Margin of Sampling Error/Credibility Interval, https://www.aapor.org/Education-Resources/Election-Polling-Resources/Margin-of-Sampling-Error-Credibility-Interval.aspx (accessed 2021).

[50] E-mail from Tom E. Hardwicke to Jason M. Chin (Jan. 26, 2020).

resources. To provide a comparison between the student-edited journals (which tend to not use peer review, but rather the judgment of student editors to make acceptance decisions) and faculty-edited journals (which tend to rely on peer review), we chose 150 articles from each. We classified articles as empirical if they included original quantitative analyses using descriptive or inferential statistics of original or pre-existing data (e.g., survey studies, content analyses of judicial decisions, meta-analyses).

As described below, we coded features such as statements about the availability of data, preregistration, and declarations of conflicts of interest. This is the first study of its kind in empirical legal research, and we are not testing hypotheses; thus, the results should be considered descriptive and exploratory. This study is preregistered and provides open data, code, and materials (https://osf.io/msjqf/).

We deviated from previous studies measuring transparency in two main ways. First, previous studies using this type of protocol focused on fields whose journals contain a high proportion of empirical research (e.g., psychology, organizational behavior research, otolaryngology, addiction medicine),[51] so they randomly sampled studies without screening out studies that did not use empirical methods. This approach would have been inappropriate for the current study because it would have led us to include a large number of non-empirical studies (~90% of the published work, according to prior estimates).[52] As a result, we developed an approach for early screening of non-empirical research (see literature search string below). We also deviated from some previous studies by sampling only from highly-ranked journals. This may have biased our results towards finding higher research transparency than the field generally has, because higher rank typically translates to greater selectivity, and thus should in principle

[51] See the sources at *supra* note 37.
[52] Diamond & Mueller, *supra* note 7.

enable higher standards. Note also that given the perceived importance of the journals in our sample, low levels of transparency would be especially concerning.

Identifying empirical articles: Search string used to generate sample

To develop a search string to more efficiently identify and sample articles that met our specifications, we conducted a preliminary examination of the literature. We coded 2019-2020 articles from 10 law journals that Washington and Lee ranks in the top 25 (1,024 total articles).[53] Through reading those articles, we identified 92 (or 9% of the sample) meeting our definition of empirical within this dataset (https://osf.io/hyk8c/, with the code producing these descriptive statistics at https://osf.io/9q47g/). It is worth noting that Diamond and Mueller found that, using a somewhat similar definition of "empirical" we have adopted,[54] 8.5% of articles in top ranked journals contained empirical work.[55]

Using the knowledge from that preliminary examination, we first considered two different ways of more quickly identifying empirical articles without reviewing the full text. First, we considered selecting only articles with the word "empirical" in the title as Epstein and King had done in their landmark study. However, only 10% of the empirical articles in the preliminary examination sample had the word "empirical" in their title. This strategy, therefore, would miss a great deal of empirical work, raising concerns about the representativeness of the sample and making it more difficult to find our target of 300 recent empirical studies. We also

[53] We used the 2019 list, which was the latest available when we started coding. To get a broad range of journals, we chose the top 5 on the list (Yale Law Journal, Harvard Law Review, Stanford Law Review, Columbia Law Review, and University of Pennsylvania Law Review) and the bottom 5 (Fordham Law Review, Boston College Law Review, Boston University Law Review, Cornell Law Review, and Northwestern University Law Review). We began coding in January 2021, so any issues released after that date are not included (sometimes, a year's issue is not released until the following year); see Washington & Lee Law, W&L Journal Rankings, https://managementtools4.wlu.edu/LawJournals/ (accessed 2021).

[54] Diamond & Mueller, *supra* note 7 at 589-90.

[55] They found this in the top 10 journals of the W&L rankings, whereas we looked at journals in the top 1-5 and top 20-25.

considered selecting only articles with "empirical" in their abstract; however, that strategy would have missed approximately 50% of the articles identified by the more intensive method used in our preliminary examination.

Ultimately, we decided to use the words in the abstracts and titles of the 92 empirical articles we identified in our preliminary examination, and to write a search string based on those words. That search string is:

> ABS ("content analysis" OR data* OR behavioral OR behavioural OR empirical OR experiment OR meta-ana* OR multidimensional OR multivariate OR quantitative OR statistical OR study OR studies OR survey OR systematic)

One limitation of this strategy is that, in our preliminary examination, about 8% of the empirical articles we identified did not have an abstract. As a result, any search strategy that uses abstract searches is bound to miss a small proportion of empirical articles. However, this method balances efficiency (i.e., full text searches would yield too many false positives for our team to review) and breadth.

Sample

Figure 1 details our sampling process and exclusions. We used the search string described above to search Scopus for articles published from January 1, 2018 to the date of our search, January 29, 2021. We populated our overall sample of 300 articles with 150 articles from the top 25 student-edited journals from the Washington and Lee rankings (based on its "combined score" in 2019, https://managementtools4.wlu.edu/LawJournals/) and 150 faculty-edited journals from the 25 journals (by impact 2019 factor) in the Web of Science's "law" database.[56] That is, we applied our search string to both of those journal lists, which can be found in our

[56] Using the same method of selecting student-edited and faculty-edited journals as Chin & Zeiler, *supra* note 6.

supplementary materials (https://osf.io/ravhe/). The Washington and Lee search returned 596 articles and the Web of Science search returned 859 articles.

[**Figure 1 about here**]

Because searches returned several of what we classified as non-empirical articles (e.g., the abstract contained the word "data" to describe data regulation laws), one author (JC) randomly sorted both lists and then screened out articles that did not meet our criteria until we reached the pre-specified sample of 150 articles for each group (Figure 1). Of the 596 articles in the W&L sample, we needed to review 510 to obtain our sample of 150 (i.e., 31% of those reviewed were selected, the rest were excluded). For the Web of Science sample, we needed to review 383 to find 150 empirical articles (i.e., 40.1% of those reviewed were selected, the rest were excluded). The relatively high rate of exclusions suggests that our search string was overly inclusive, adding more work for us but reducing the chances that we missed a large proportion of empirical articles. The articles screened out and the reasons for their exclusion are described in our supplementary materials (W&L: https://osf.io/qf7sc; Web of Science: https://osf.io/vbu63). After we initiated coding of these articles with the protocol below, we found that 8 were incorrectly categorized as empirical, so we selected the next 8 from the list as replacements. These are the numbers that are reflected in Figure 1 and above.

Coding procedure

Articles were coded using the structured form developed by Hardwicke and colleagues (see supplemental materials – https://osf.io/4btxv/).[57] Following the Hardwicke et al. protocol (as well as other transparency coding projects for systematic reviews, see O'Dea et al.),[58] each

[57] Hardwicke et al., 2021, *supra* note 37.
[58] Rose E. O'Dea et al., *Preferred reporting items for systematic reviews and meta-analyses in ecology and evolutionary biology: a PRISMA extension*, 96 BIOL. REV. 1695 (2021).

article was coded by two of the authors, with disagreements resolved through discussion between those coders and a third author if the coders could not agree (for disagreements, see https://osf.io/7q32m/). The coders were all trained on five articles and did not begin coding the target sample of articles until they reached consensus on the five training articles. As we discuss below, two items were difficult to code, and so we discontinued coding them and do not present the result for them. For multiple-study articles (we defined studies as distinct data collection activities), we coded only the first-reported study. Coding one article in the student-edited sample took about 30-45 minutes. Coding an article in the faculty-edited sample took about 10-20 minutes. This reflects the longer length of the articles in the student-edited sample and that their methods and data were frequently difficult to locate due to the lack of a standard article format.

[Table 1 about here]

The features of the articles that we coded are detailed in the coding sheet (https://osf.io/4btxv/) and in Table 1 (and further detailed in our preregistration, https://osf.io/msjqf/registrations). Some of these features are relevant background information on the studies, such as the statistics used by the researchers, the nature of the data, and data sources. Others are relevant to the transparency and credibility of the research, such as whether authors stated that data and analysis scripts were available, and whether the study was preregistered.

With respect to data availability, Hardwicke et al. attempted to code whether authors provided a clear reference to where the data could be found (without vouching for its availability or taking steps to make it available).[59] Due to difficulty coding this item, they did not report this and instead collapsed these types of data references into "no – there was no data availability

[59] Hardwicke et al., 2021, *supra* note 37.

statement". Because we expected the current study to include several cases of authors analyzing pre-existing data and datasets, we initially attempted to preserve this as a distinct item in our coding form. However, our coders also encountered difficulty with it (e.g., sometimes articles would provide a vague reference to another article, and, when we accessed that article, it referenced yet other articles). So, our results also collapse these types of data references into the "no data availability statements" category. We did, however, include a separate item for secondary data studies (Table 1) in which we coded whether authors provided an index of the secondary data items (e.g., references to the judicial decisions included).[60]

Deviations from preregistration

Our study deviated from our preregistration in two ways. First, we originally planned to code sample size, but did not complete this coding because studies did not provide a single sample size. Second, as noted above, we originally planned to code whether the authors provided a source of the data that could be verified, but we did not complete this because it was impractical.

Results

Overall, we found a low level of transparency on the characteristics we measured. Only 19% of articles stated that their data are available, and we were able to access that data in only about half of those cases.[61] Preregistration and availability of analytic scripts were also very

[60] For an example of this approach, see Bijal Shah, *Executive (Agency) Administration*, 72 STANFORD LAW REV. 641 (2020). Although, raw data can be provided in many cases. For instance, see Oona A. Hathaway, Curtis A. Bradley & Jack L. Goldsmith, *The Failed Transparency Regime for Executive Agreements: An Empirical and Normative Analysis*, 134(2) HARV. L. REV. 629 (2020) in which the authors digitized the data they relied on and made them available on the Harvard Dataverse.

[61] Recall that some of the variables we measured are on the level of the article (i.e., article accessibility and if the article is accessible, where it is accessible; conflict of interest statement; funding statement) with all other pertaining to the first reported study within an article. For simplicity, we will refer to the units described below as "articles". We acknowledge that there may be some bias in coding only the first reported study in that first reported studies may be different in some ways that subsequent studies in an article. However, we judged it to be unlikely that the variables we were interested in (e.g., data availability statements, preregistrations) would differ in any meaningful

uncommon, and, in fact, almost nonexistent in the empirical legal research examined here. However, we found several positive aspects of the literature to build on. For instance, about 50% of studies employing original data stated that at least some materials were available. And, article accessibility was high among the empirical legal research examined here, especially among articles in student-edited journals (100% of those articles were available without library access). These findings are detailed below.

Sample characteristics

General characteristics of our sample are reported in Table 2, specifically the proportion of articles that: analyzed original or secondary data; used human participants; reported an experiment; were a synthesis (which we operationalized as studies that self-identified as a systematic review or meta-analysis); and reported descriptive or descriptive and inferential statistics. Secondary data analysis was more common (65% of studies, 95% CI = [59%, 70%]) than analysis of original data. Secondary data were also more frequently employed in the student-edited journals (79%, 95% CI = [73%, 85%]) than in the faculty-edited journals (51%, 95% CI = [43%, 59%]). Furthermore, 40% (95% CI = [35%, 46%]) of studies relied on human participants. This figure was 21% (95% CI = [15%, 27%]) among the student-edited journals and 60% (95% CI = [53%, 69%]) among the faculty-edited journals.

[Table 2 about here]

Turning to methodology, experiments (which require random assignment according to our definition) were rare in our sample (18% of studies, 95% CI = [14%, 22%]). Syntheses were very uncommon, with only six in the sample (all six in the faculty-edited sample). Most articles (68% (95% CI = [62%, 73%])) contained descriptive *and* inferential statistics (the remaining

way across studies, and we would expect authors to adopt the same transparency approach across studies in an article.

32% reported only descriptive statistics). 78% (95% CI = [72%, 85%]) of the faculty-edited articles used inferential statistics versus 57% (95% CI = [49%, 65%]) in the student-edited sample.

Among the 194 articles that used secondary data, 53 (27%) articles analyzed judicial decisions, 11 (6%) analyzed company documents, and a further 11 analyzed statutes or legislation (see the breakdown at https://osf.io/usfy4/). Human participants were recruited from a variety of groups, with 12 of the 121 articles (10%) sampling from university students, 35 (29%) sampling from the general population, and 75 (61%) sampling from special populations. Those special populations included difficult-to-reach groups such as judges, young offenders, and government employees (see the list at https://osf.io/m589c/).[62]

Article accessibility

The articles in our sample were generally easy to access as compared to estimates from previous metascientific studies in criminology and psychology (Table 3, Figure 2).[63] 86% (95% CI = [82%, 90%]) of articles had publicly available versions – 100% of the student-edited journal articles and 71% (95% CI = [65%, 79%]) of the faculty-edited group. 70% of articles (95% CI = [65%, 76%]) were gold open access, meaning they were accessible on journals' websites. This was the case for 100% of the articles in student-edited journals, whereas 41% (95% CI = [33%, 49%]) of the faculty-edited articles were gold open access. Empirical legal researchers also regularly use pre- and post-print services to provide open access versions of their work. 42%

[62] We were interested in special populations because law, as an applied field, has a special interest in certain groups and stakeholders.

[63] Matthew P. J. Ashby, *The Open-Access Availability of Criminological Research to Practitioners and Policy Makers*, 32(1) J. CRIM. JUS. EDUC. 1 (2021); Hardwicke et al., 2021, *supra* note 37 at 5: "Among the 237 English-language articles, we obtained a publicly available version for 154 (65%, 95% CI = [59%, 71%]".

(95% CI = [36%, 48%]) of articles in the overall sample were downloadable on SSRN and 22% (95% CI = [18%, 27%]) were downloadable on ResearchGate.

[Table 3 about here]

[Figure 2 about here]

Conflicts of interest and funding statements

Turning to conflicts of interest and funding statements, we found that most articles did not provide any such declaration. In fact, 89% (95% CI = [86%, 93%]) of articles did not include a conflicts of interest statement. Conflicts of interest statements were more common in the faculty-edited journals with only one article in the student-edited sample containing such a statement. As to statements of funding sources, 60% (95% CI = [55%, 66%]) of articles contained no statement. Again, such statements appear to be rarer in the student-edited sample (see Table 3).

Data availability

The availability of the data, analysis scripts, and materials in our sample was generally low (Table 3, Figure 3). Just 19% (95% CI = [15%, 23%]) of articles provided a statement that data are available. Of articles with data availability statements, the most common means for sharing data were via a third-party repository (39%, 95% CI = [26%, 53%]), by contacting the author (28%, 95% CI = [16%, 42%], and via a personal or institutional website (21%, 95% CI = [9%, 35%]) (see https://osf.io/67t9y/ for a full breakdown). We checked whether the data referenced in the statements were readily available (i.e., whether we could access them without further steps, such as contacting the author). Only about half (53%, 95% CI = [40%, 66%]) were readily available, making the effective data availability rate about 10%.

[Table 3 about here]

In the social sciences, much of the move towards providing data availability statements has occurred in the context of psychological research, where original data are often collected. As a result, it may be useful to drill down on articles reporting on original data. Limiting our analysis to these articles (N = 106), we found 29% (95% CI = [22%, 39%]) included a data availability statement, whereas only 13% (95% CI = [9%, 18%]) of articles reporting on secondary data did so (N = 194).

For secondary data, as noted above, we coded the steps authors took to provide information about the dataset. In most cases, authors did not provide any details about the dataset (https://osf.io/xczpy/). In 26 of the 194 (13%, 95% CI = [8%, 20%]) articles reporting on secondary data, the authors provided an index of the secondary data (e.g., a list of judicial decisions relied on). Several others linked to sources, such as external websites, that were no longer accessible (https://osf.io/xczpy/).

Analysis-script availability

Very few studies provided a statement about the availability of their analysis scripts (6%, 95% CI = [4%, 9%]). Providing analysis code is especially important when reporting inferential statistics (e.g., to determine the exact statistical test and assumptions the authors used), but of the 203 studies that relied on inferential statistics, only 8% (95% CI = [5%, 12%]) made their code available. Even these figures are somewhat inflated, however, because only for approximately half of the articles with script availability statements could we access the scripts without taking further steps (again, due to dead links and statements indicating that the scripts were available on request).

Materials availability

The materials availability results presented in Table 3 and Figure 3 are limited to studies with original data. We presented them this way because sharing of study materials (e.g., survey instruments, vignettes) is arguably less applicable to analysis of existing data. However, some studies analyzing secondary data do involve useful materials that could be shared, such as the coding sheets used by researchers who tally different sorts of judicial decisions. Of studies that reported on original data, about 44% (95% CI = [35%, 54%]) stated that materials were available. Recall that this figure does not mean that all materials were made available, but rather that authors stated that at least some materials were available. Moreover, we were able to access materials for only 39 of the 47 (83%, 95% CI = [74%, 94%]) studies that stated that materials were available, making the effective material available rate about 37% among studies that report on original data.

Preregistration and replication

Preregistration and replication were extremely uncommon, at 3% (95% CI = [1%, 5%]) and 4% (95% CI = [2%, 6%]) of studies, respectively. Of the 8 preregistered studies, we could not access the preregistrations for 2 of them. The purported locations of the 8 preregistrations were: the Open Science Framework (5 studies), the AsPredicted.org registry (1 study), the PROSPERO registry (for syntheses; 1 study), and the Evidence in Governance and Politics (EGAP) registry (which is hosted by the Open Science Framework; 1 study).

Discussion

Our results suggest that there is ample room to substantially improve empirical legal research transparency. Our hope is that our results encourage researchers in the field of empirical legal research to move forward in making their work verifiable and reusable. In particular, we

note that although articles in our sample generally had low levels of transparency and credibility-related characteristics we measured, these results are not much different than many other fields (Table 4).[64] On a more positive note, with respect to article accessibility, empirical legal research performs very well, especially for articles published in student-edited journals.

Comparing student-edited and faculty-edited journals on other transparency and credibility-related characteristics, we generally did not find large differences. However, student-edited journals did seem to have a smaller proportion of articles with conflicts of interest and funding statements. The latter may be due to law professors relying largely on internal funding. However, many authors may have affiliations that should be disclosed, such as governmental appointments, affiliations with think tanks, and company directorships or board memberships.

While we urge caution in comparing our results to those from transparency and credibility studies in other fields, such a comparison may be instructive (see Table 4). In particular, we did not observe large differences (other than in materials availability, see below) between empirical legal research and other fields. However, the two comparison studies in Table 4 (sampling from social science generally and otolaryngology) did not restrict their samples to highly ranked journals.[65] On the other hand, our study sampled only from what many would describe as the top journals in the field. It would be reasonable to expect that these journals should be leading the field in producing verifiable and reusable work. Moreover, the other studies focus on articles published in the mid-2010s, and so we might expect stronger adoption of transparency and credibility reforms in our sample. In other words, the results of our study likely provide an optimistic comparison with other fields of research.

[64] See the sources at *supra* note 37.
[65] See the sources at *supra* note 37.

Table 4 also contains two comparisons with studies that have sampled only from journals that have implemented transparency and openness guidelines. In particular, Culina and colleagues sampled only from ecology journals that had implemented data and analysis script availability policies (both mandatory guidelines and encouragements).[66] And, Hardwicke and colleagues examined data availability of studies published by the journal *Cognition*, which had implemented a mandatory data availability policy.[67] As can be seen in Table 4, recent articles in those journals show markedly higher levels of data and script availability than our study found in empirical legal research. We cannot say what caused the relatively high levels of data and script availability in these journals, but these results suggest journal guidelines may play an important role in reform efforts.

Our results are limited in other respects. First, empirical legal research is a multi-disciplinary field, which uses a panoply of methods from several research traditions.[68] As a result, some forms of transparency may be less applicable for some methods than for others. We attempted to take this into account by reporting results for some of these practices separately for different types of studies (e.g., reporting materials transparency for studies reporting on original data; reporting analysis script transparency for studies reporting inferential statistics). In this respect, our results may overestimate transparency levels by restricting analyses to only one subset of studies, when in fact the practice would be beneficial for a broader range of studies. For example, many studies reporting on secondary data would nevertheless be more reproducible if they shared materials such as coding sheets used by research assistants who coded legislation or judicial decisions.

[66] Culina et al., *supra* note 41.
[67] Hardwicke et al., *supra* note 41.
[68] Hall & Wright, *supra* note 46 at 63.

We also did not take the additional step of contacting authors to see whether statements that data, materials, or analysis scripts were available upon request would be honored. Note, however, that such options become impossible over time and some research finds that, even in the short term, researchers do not often respond productively to such requests.[69] In addition, this method of transparency presents a significant obstacle for third parties who wish to access these artifacts for purposes that the authors may view as adversarial (e.g., because the requesters suspect an error in the original article).

Our results are also limited in that we did not attempt to take into account data sharing limits such as privacy and proprietary datasets.[70] However, we did code whether any statement was made about data availability, which would have included statements about barriers to sharing data, and we did not find any studies that explained their lack of data sharing in such terms, so this may not have been prevalent. Alternatively, authors simply might not have reported their inability to share the data. Moreover, we attempted to code other means of transparency for secondary data analysis (e.g., indexes of cases relied on) and found that few papers took up any such options. Future meta-research projects may wish to take a more focused approach, targeting specific empirical legal research methods to better understand their norms and limits related to transparency practices adoption.[71]

We also highlight that the mere presence of data, analysis scripts, and preregistration does not mean that the findings will be reproducible. Systematic research has found that data is

[69] Vines et al., *supra* note 28; Wicherts et al., *supra* note 28.
[70] Michelle N. Meyer, *Practical Tips for Ethical Data Sharing*, 1(1) ADV. METH. & PRACT. PSYCHOL. SCI. 131 (2018).
[71] As discussed above, we saw approaches to more transparent handling of secondary data ranging from providing a detailed index of the secondary data (Shah, *supra* note 60) to digitizing the data and making it publicly available (Hathaway, Bradley & Goldsmith, *supra* note 60). Best practices documents ought to be created that explain the scenarios in which such methods are possible and desirable.

often not well documented, making it difficult to reproduce findings.[72] Future projects should consider focusing on a smaller number of studies for which some data are available to determine if the results are fully reproducible.[73]

Where do we go from here? As we reviewed above, transparency has proven vital in uncovering flaws, limitations, and fraud in published work. We call on journals to adopt policies to increase the transparency of published studies. This may be especially important for journals that are not commonly peer-reviewed, such as student-edited journals, because peer review detects some flaws and errors. Even then, however, studies that have deliberately added errors to studies have found that peer reviewers detect just a minority of those errors.[74] Only with a high level of transparency can we hope that errors in important studies are likely to be caught, as transparency enables post-publication peer review.

If law journal editors are concerned about finding appropriately experienced reviewers, they may wish to make use of the new "Peer Community In" (PCI) initiative, whose members provide free evaluations of and recommendations on preprints submitted to PCI's journal-agnostic review process. Several journals (e.g., *Royal Society Open Science*, *PeerJ*) have agreed to accept registered reports approved by Peer Community In with no further review. For law journals that do not regularly receive empirical submissions, the PCI initiative, or something like it, may be an attractive way forward.

The fact that at least some datasets employed in empirical legal research studies are proprietary and cannot be made publicly available should not cause the field to shy away from

---

[72] Tom E. Hardwicke et al., *Analytic reproducibility in articles receiving open data badges at the journal Psychological Science: an observational study*, 8 R. SOC. OPEN SCI. 201494 (2018); Riana Minocher et al., *Estimating the reproducibility of social learning research published between 1955 and 2018*, 8 R. SOC. OPEN SCI. 210450 (2018); Hardwicke et al., 2018, *supra* note 37.
[73] *Id*.
[74] Sara Schroter, et al., What errors do peer reviewers detect, and does training improve their ability to detect them? 101(10) J. R. SOC. MED. 507 (2008).

general data availability requirements. For example, in psychology it is common for privacy issues to preclude data sharing. Journal guidelines in this field balance privacy and other ethical constraints on data sharing with data availability by asking authors to explain any restrictions in the manuscript and requiring data sharing if such an explanation cannot be provided.[75] An example of such a statement is: "The conditions of our ethics approval do not permit public archiving of anonymised study data. Readers seeking access to the data should contact the lead author X or the local ethics committee at the Department of Y, University of Z. Access will be granted to named individuals in accordance with ethical procedures governing the reuse of sensitive data. Specifically, requestors must meet the following conditions to obtain the data [insert any conditions, e.g. completion of a formal data sharing agreement, or state explicitly if there are no conditions]."[76]

Finally, empirical legal research can take advantage of the larger movement in the social sciences, medicine, and many other fields, by leveraging the technology, training, and ideas flowing from those credibility revolutions. Free technologies like the Open Science Framework provide a place not just to store data, but to collaborate, establish version control, preregister, and even store video stimuli. Other examples include tools like github (a data and code repository), AsPredicted (a general study registry), Declare Design (a tool for creating a preregistration), and the American Economic Association's registry for randomized controlled trials. Straightforward guides to data staring, preregistering, and many other transparency and credibility-related activities are now available.[77] At least one guide specific to some empirical legal research

[75] Cortex, *Transparency and Openness Promotion (TOP) guidelines*, https://www.elsevier.com/__data/promis_misc/Cortex-TOP-author-guidelines.pdf (accessed 2021).
[76] *Id*.
[77] Olivier Klein, *A Practical Guide for Transparency in Psychological Science*, 4 COLLABRA: PSYCHOLOGY 4 (2018) 20; (2019). Sophia Crüwell et al., *Seven easy steps to open science: An annotated reading list*, 227 ZEITSCHRIFT FÜR PSYCHOLOGIE, 237 (2019).

methodologies is also available, and we hope more are on the way.[78] With these tools at their fingertips – and as a field whose data and results are often of great public importance – there is little reason researchers in the field of empirical legal research should not become leaders in the move towards transparency and credibility.

[78] Jason M. Chin et al., 3(1) *Improving the Credibility of Empirical Legal Research: Practical Suggestions for Researchers, Journals and Law Schools*, LAW, TECHNOLOGY AND HUMANS (2021).

Table 1

Variable	Further details
Article accessibility	• Was the article available through the journal's website (without university library access, i.e., gold open access)? • Was the article available through another service (e.g., ResearchGate, SSRN)?
Conflict of interest	Does the article include a statement indicating whether there were any conflicts of interest?
Funding	Does the article include a statement indicating whether there were funding sources?
Experimental design	Is it an experiment? For our purposes, experiments are studies in which some variable is manipulated by the researcher (e.g., some participants are randomly assigned to a condition).
Synthesis	Is it a synthesis (e.g., meta-analysis, systematic review)? For our purposes, a synthesis is a quantitative analysis of other studies/articles.
Replication	Does the article claim to report a replication study?
Human subjects	Were there human subjects? For our purposes, this means measuring and/or aggregating responses from individuals or groups. This does not include judicial decisions written by judges and analogous data.
Original or secondary data	For our purposes, original data are data the authors collected or generated that did not exist before. Secondary data are data that already existed (e.g., analyses of judicial decisions or contracts).
Data availability	• Does the article state whether or not data are available? • How does the statement indicate the data are available? • Can you access, download, and open the data files?
Analysis script availability	• Does the article state whether or not analysis scripts are available? • How does the statement indicate the analysis scripts are available? • Can you access, download, and open the analysis files?
Materials availability	• Does the article state whether or not materials are available? • How does the statement indicate the materials are available? • Can you access, download, and open the materials files?
Preregistration	• Does the article state whether or not the study (or some aspect of the study) was preregistered? • Where does the article indicate the preregistration is located? • Can you access and open the preregistration?

Note. The primary measured variables in our analysis. The full set of variables can be found in the full structured coding form (https://osf.io/4btxv/).

Table 2

All			
Variable	Response	N	% [95% CI]
Original or secondary data	Original	106	35% [30%, 41%]
	Secondary	194	65% [59%, 70%]
Human subjects	No	179	60% [54%, 65%]
	Yes	121	40% [35%, 46%]
Experimental design?	No	247	82% [78%, 87%]
	Yes	53	18% [14%, 22%]
Synthesis	No	294	98% [97%, 99%]
	Yes	6	2% [1%, 3%]
Statistics used	Descriptive	97	32% [27%, 38%]
	Descriptive & Inferential	203	68% [62%, 73%]
Studies in student-edited journals			
Original or secondary data	Original	32	21% [15%, 28%]
	Secondary	118	79% [73%, 85%]
Human subjects	No	119	79% [73%, 86%]
	Yes	31	21% [15%, 27%]
Experimental design?	No	129	86% [81%, 92%]
	Yes	21	14% [9%, 20%]
Synthesis	No	150	100% [100%, 100%]
	Yes	0	0% [0%, 1%]
Statistics used	Descriptive	64	43% [35%, 51%]
	Descriptive & Inferential	86	57% [49%, 65%]
Studies in faculty-edited journals			
Original or secondary data	Original	74	49% [41%, 58%]
	Secondary	76	51% [43%, 59%]
Human subjects	No	60	40% [33%, 49%]
	Yes	90	60% [53%, 69%]
Experimental design?	No	118	79% [73%, 85%]
	Yes	32	21% [15%, 28%]
Synthesis	No	144	96% [93%, 99%]
	Yes	6	4% [1%, 7%]
Statistics used	Descriptive	33	22% [16%, 29%]
	Descriptive & Inferential	117	78% [72%, 85%]

Note. Overview of the samples of empirical legal studies. The variables are: original or secondary data, whether there were human subjects, whether the study was an experiment, whether it was a synthesis (i.e., systematic review or meta-analysis), and whether it used descriptive statistics or descriptive statistics along with inferential statistics.

Table 3

		All		Student-edited		Faculty-edited	
Variable	Response	N	% [95% CI]	N	% [95% CI]	N	% [95% CI]
Article accessibility	Paywall only	43	14% [11%, 18%]	0	0% [0%, 1%]	43	29% [22%, 36%]
	Available	257	86% [82%, 90%]	150	100% [100%, 100%]	107	71% [65%, 79%]
Conflicts of interest	No statement	267	89% [86%, 93%]	149	99% [99%, 100%]	118	79% [73%, 85%]
	Conflicts	2	1% [0%, 4%]	1	1% [0%, 2%]	1	1% [0%, 7%]
	No conflicts	31	10% [7%, 14%]	0	0% [0%, 1%]	31	21% [15%, 27%]
Funding	No statement	180	60% [55%, 66%]	112	75% [69%, 82%]	68	45% [37%, 54%]
	No funding	2	1% [0%, 6%]	0	0% [0%, 7%]	2	1% [0%, 10%]
	Private	41	14% [8%, 19%]	27	18% [12%, 25%]	14	9% [1%, 18%]
	Public	56	19% [13%, 24%]	6	4% [0%, 11%]	50	33% [25%, 42%]
	Public & private	21	7% [2%, 13%]	5	3% [0%, 11%]	16	11% [3%, 19%]
Data availability	No statement	242	81% [76%, 85%]	124	83% [77%, 89%]	118	79% [73%, 85%]
	Says available	57	19% [15%, 23%]	25	17% [11%, 23%]	32	21% [15%, 28%]
	Not available	1	0% [0%, 5%]	1	1% [0%, 7%]	0	0% [0%, 7%]
Analysis script availability	No statement	281	94% [91%, 96%]	142	94% [91%, 96%]	139	93% [89%, 97%]
	Says available	19	6% [4%, 9%]	8	6% [4%, 9%]	11	7% [4%, 12%]
Materials availability	No statement	59	56% [46%, 65%]	17	53% [38%, 71%]	42	57% [46%, 68%]
	Says available	47	44% [35%, 54%]	15	47% [31%, 65%]	32	43% [32%, 55%]
Preregistration	No statement	292	97% [96%, 99%]	147	98% [97%, 100%]	145	97% [95%, 100%]
	Says preregistered	8	3% [1%, 5%]	3	2% [1%, 4%]	5	3% [1%, 6%]
Replication	No	289	96% [95%, 98%]	147	98% [97%, 100%]	142	95% [92%, 98%]
	Yes	11	4% [2%, 6%]	3	2% [1%, 4%]	8	5% [3%, 9%]

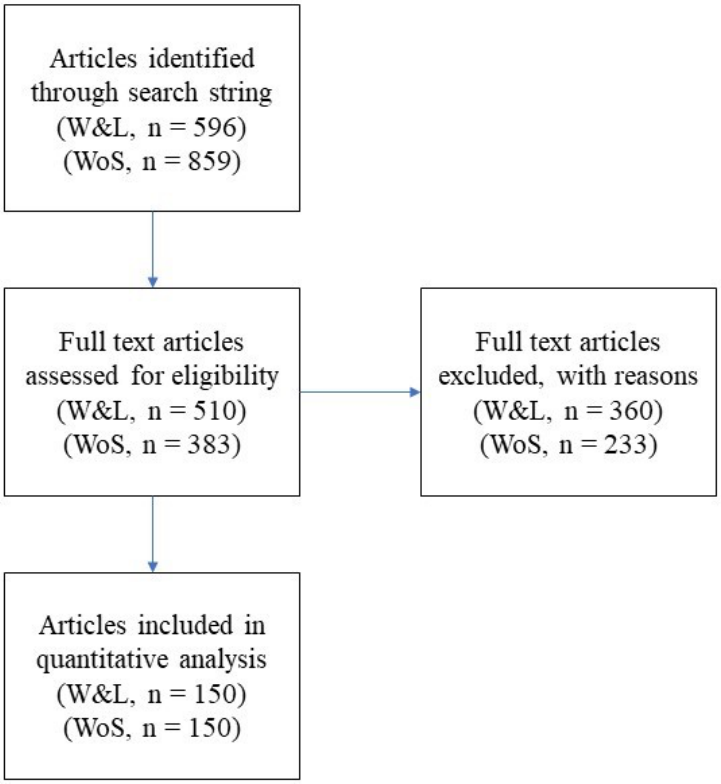
Note. Transparency and credibility-related features of empirical legal research. The variables are: article accessibility, the presence and content (if applicable) of statements about funding, conflicts of interest, data availability, materials availability, and analysis script availability. We further coded whether there was a statement that the study was preregistered and whether the authors described the study as a replication. The figures for materials availability include only the articles that collected original data. Note that this figure reflects availability statements. As discussed in text, actual accessibility was considerably lower.

Table 4

		Current study	Hardwicke et al., 2020	Johnson et al., 2019	Culina et al., 2020	Hardwicke et al., 2018
Field(s)		Empirical Legal	Social Sciences	Otolaryngology	Ecology	Psychology
Reform(s)?		--	--	--	Journal guidelines	Journal guidelines
Articles analyzed	*N*	300	250	300	346	174
	Publication years	2018 - 2021	2014 - 2017	2014 - 2018	2015 - 2019	2015 - 2017
Article availability	Paywall only	14%	54.0%	77.7%	--	--
	Publicly available	86%	40.4%	22.3%	--	--
Data availability	No statement	81%	80.8%	96.7%	--	22%
	Says available	19%	7.0%	2.0%	79%	78%
	Not available	0%	0.6%	1.3%	--	0%
	N	300	156	151	346	174
Analysis script availability	No statement	94%	98.7%	99.4%	--	--
	Says available	6%	1.3%	0.7%	27%	--
	N	300	156	151	346	--
Materials availability	No statement	56%	89.4%	94.5%	--	--
	Says available	44%	10.6%	4.8%	--	--
	N	106	151	145	--	--
Preregistration	No statement	97%	100%	95.4%	--	--
	Says preregistered	3%	0%	4.0%	--	--
	N	300	156	151	--	--
Replication	No	96%	98.7%	100%	--	--
	Yes	4%	1.3%	0%	--	--
	N	300	156	151	--	--

Note. A comparison of studies measuring transparency-related factors. Culina et al. (2020) and Hardwicke (2018) focused on journals that had recently implemented transparency guidelines (Culina et al. studied such journals in Ecology; Hardwicke et al. focused on the journal, *Cognition*). A fuller description of the methodological differences between these studies and an expanded table is available (https://osf.io/z6tx3/).

Figure 1

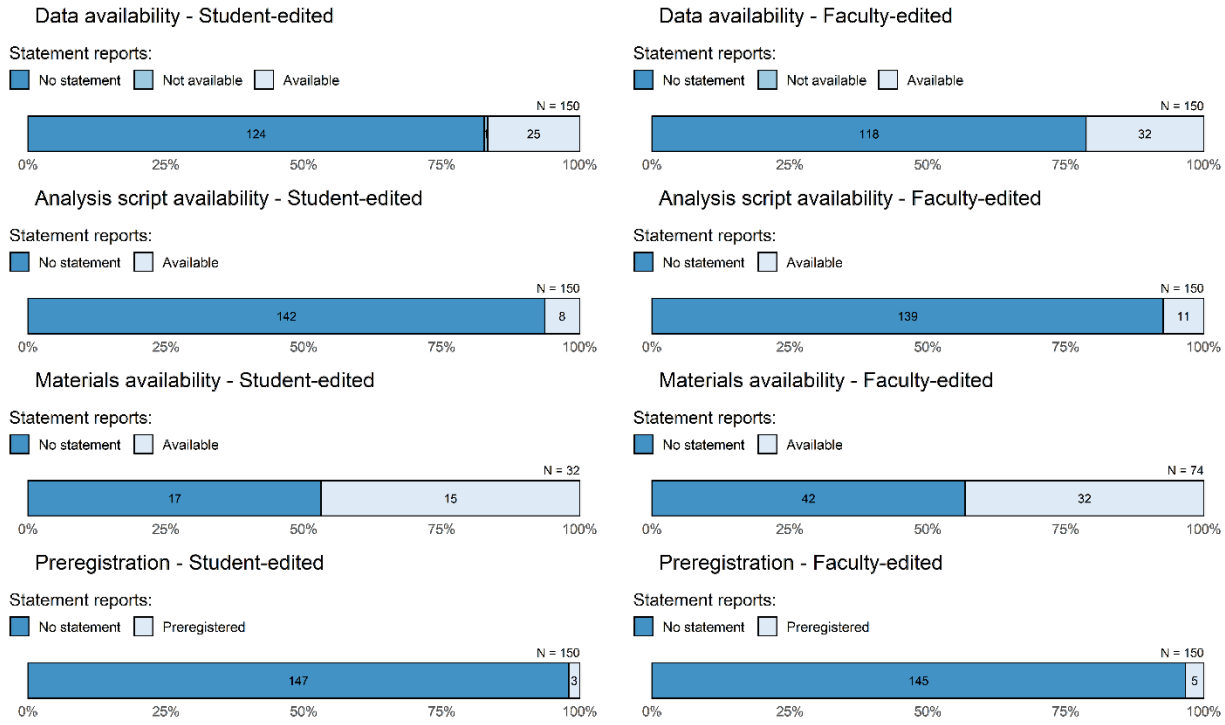


Note. The screening procedure for building the student-edited (W&L) and faculty-edited samples (WoS). Articles were first identified through the Scopus search string described in the methods. They were then screened for eligibility in random order until the samples were complete. The excluded articles and the reasons for their exclusion are at https://osf.io/qf7sc for the student-edited journals, and at https://osf.io/vbu63 for the faculty-edited journals.

Figure 2

Note. Article availability, funding statements, and conflict of interest statements in empirical legal research. The left column includes articles from the student-edited sample and the right column is from the faculty-edited sample. Numbers within bars refer to the number of articles that meet the given standard.

Figure 3



Note. Assessment of transparency and credibility-related characteristics of empirical legal research. The student-edited sample is reported in the left column and the right column is the faculty-edited sample. Numbers within bars refer to the number of articles that meet the given standard. Data availability, analysis script availability, and preregistration bars include the full sample (150 per group), whereas the bars for materials availability include only the articles that collected original data. Note that this figure reflects availability statements. As discussed in text, actual accessibility was considerably lower.

CRediT Statement

Conceptualization: Jason M. Chin.
Data curation: Jason M. Chin and Alex O. Holcombe.
Formal analysis: Jason M. Chin.
Funding acquisition: Jason M. Chin and Alex O. Holcombe.
Investigation: Jason M. Chin, Natali Dilevski, Alex O. Holcombe, Rosemary Gatfield-Jeffries, Simine Vazire, and Ruby Bishop.
Methodology: Jason M. Chin, Kathryn Zeiler, Natali Dilevski, Alex O. Holcombe, Rosemary Gatfield-Jeffries, and Simine Vazire.
Project administration: Jason M. Chin.
Supervision: Jason M. Chin and Alex O. Holcombe.
Validation: Jason M. Chin.
Visualization: Jason M. Chin.
Writing - original draft: Jason M. Chin.
Writing - review & editing: Jason M. Chin, Kathryn Zeiler, Natali Dilevski, Alex O. Holcombe, Rosemary Gatfield-Jeffries, Ruby Bishop, Simine Vazire, and Sarah R. Schiavone.

Created using the tenzing app.[79]

[79] Alex O. Holcombe et al., *Documenting contributions to scholarly articles using CRediT and tenzing*, 15(12) PLOS ONE e0244611 (2020).