Northwestern | INSTITUTE FOR POLICY RESEARCH

RESEARCH EXCELLENCE • POLICY IMPACT

IPR Working Paper Series

WP-23-41

Knowledge Production Processes: An Analysis of Research Perseverance and the File Drawer Bias in Social Science Survey Experiments

Philip Moniz
University of Texas at Austin

James Druckman
Northwestern University and IPR

Jeremy Freese
Stanford University

Version: December 8, 2023

Abstract

The scientific process is difficult to evaluate because many of its stages typically evade observation. This includes whether one has success in obtaining funding for data collection, whether one perseveres should their funding application fail, whether one writes up the results of data analyses and submits a manuscript to a journal, and of those submissions whether publication occurs. Using data from applicants to a unique grant program to fund probability-sample survey experiments in the U.S. (Time-sharing Experiments for the Social Sciences), Moniz, Druckman, and Freese identify factors that influence each step. They find that research time, and not resources, plays a substantial role in determining whether the grant is funded, and, if not, whether the applicant proceeds with the project. The latter result likely reflects the availability of cheaper non-probability sample data sources that still require time to collect. Additionally, they document the substantial influence of obtaining statistically significant results in determining whether a scholar writes up and submits a paper (a variation of file drawer bias). Once a manuscript is submitted, however, statistical significance does not influence publication likelihood at all. Thus, file drawer bias emerges from researcher rather than editorial choices. The bias also is substantially smaller than it was a decade ago (Franco et al. 2014), suggesting increased recognition of the importance of null results. Overall, the researchers' findings identify how research time and statistical significance shape science, at least in the broad domain of survey experimental research, providing guidance for potential interventions in the scientific process.

This research was supported by National Science Foundation Grant SES-1628057. The survey was designed and implemented by J.N.D. and J.F. P.M. organized and analyzed the data. J.N.D. wrote the initial draft. J.F. and P.M. edited the paper.

Science offers an approach to accumulating knowledge. Its systematization along with its norms make it a particularly vaunted endeavor (Dietz 2013, Oreskes 2019). Evaluating the success of science though is difficult. Geering et al. (2020: 2) explain that "today's social scientists may be only slightly better equipped to vanquish error and construct an edifice of truth than their forbears – who conducted analyses with slide rulers and wrote up results with typewriters." The authors emphasize the need to assess science, and social science particularly, at the systemic level, that is, at the intersection of many individuals, organizations, and institutions. This aligns with metascience scholarship, perhaps the most notable of which involves publication bias, where a study enters the published record for reasons orthogonal to the study's quality (Franco et al. 2014). A large literature demonstrates these and other types of biases in the scientific process (e.g., Fanelli et al. 2018, Malhotra 2021). This has led to widespread discussions of reforms such as an emphasis on pre-registration and replication (Nosek et al. 2015, Christensen et al. 2019, Druckman 2022).

Our goal here is not to directly discuss these reforms.[1] Rather, we offer insight into scientific processes that are rarely observed to diagnose their prevalence and correlates (within one domain of the social sciences). The scientific method, heuristically, entails asking a question, developing a theory, deriving testable hypotheses, collecting data, and then analyzing and interpreting the findings. Ideally, any quality theory/hypothesis has data collected to evaluate it with the results becoming part of the scientific record (i.e., a peer-reviewed publication). As mentioned, scholars have accumulated substantial evidence that moving from a hypothesis to the collection and analysis of data to becoming part of the scientific peer-reviewed record is fraught

[1] There are various other metascience questions that we do not explore including demographic biases in the publication process (e.g., Teele and Thelen 2017), in the career process (e.g., Spoon et al. 2023), and in other domains of science (e.g., Ceci et al. 2023).

with biases, the best known of which is the file-drawer bias that favors statistically significant results, regardless of quality (Malhotra 2021). One of the more influential studies come from Franco et al. (2014) who use a grant program for survey experiments (described below) where they can identify the full set of data collected and then assess whether publication depends on statistical significance. They find that studies with significant results are 40 percentage points more likely to be published than those with null results. Much of the bias stems from authors discontinuing the paper writing process upon discovering null results; when it comes to the set of papers that have been written, they find a roughly 5 percentage point bias toward significant results in the publication process (Franco et al. 2014, appendix, Table S2).

Other work, mostly from the physical and medical sciences, explores what factors correlate with research productivity (e.g., publications, citation counts), showing that work environment (as opposed to individual attributes) plays a crucial role (Way et al. 2019, Zhang et al. 2022). And additional literature studies which factors shape who obtains grants (e.g., Working Group on Diversity in the Biomedical Research Workforce 2012, Erosheva et al. 2020). All of these research agendas are important. Yet a downside is they rarely have access to a large range of decision points. We next describe the precise subset of steps in the process that we explore.

Empirical Processes

We investigate the empirical portion of the scientific method, from the starting point of having designed a study to test a hypothesis. While this does not mean that the hypothesis is high quality, it does indicate that effort has been invested to develop the prediction and identify a way to evaluate it. The costs of getting to that stage are non-trivial and thus likely a signal of some developed knowledge. Our interest lies in the subsequent empirical processes.

In Figure 1, we present a simple model from the point of view of a scholar who has designed a study, with the ultimate question being what needs to occur for those ideas to enter the published record.[2] How this process unfolds of course will vary across the questions being explored and the methods employed. While we believe the figure captures most approaches, we emphasize that our explicit inquiry is narrower, involving survey experiments in the social sciences.

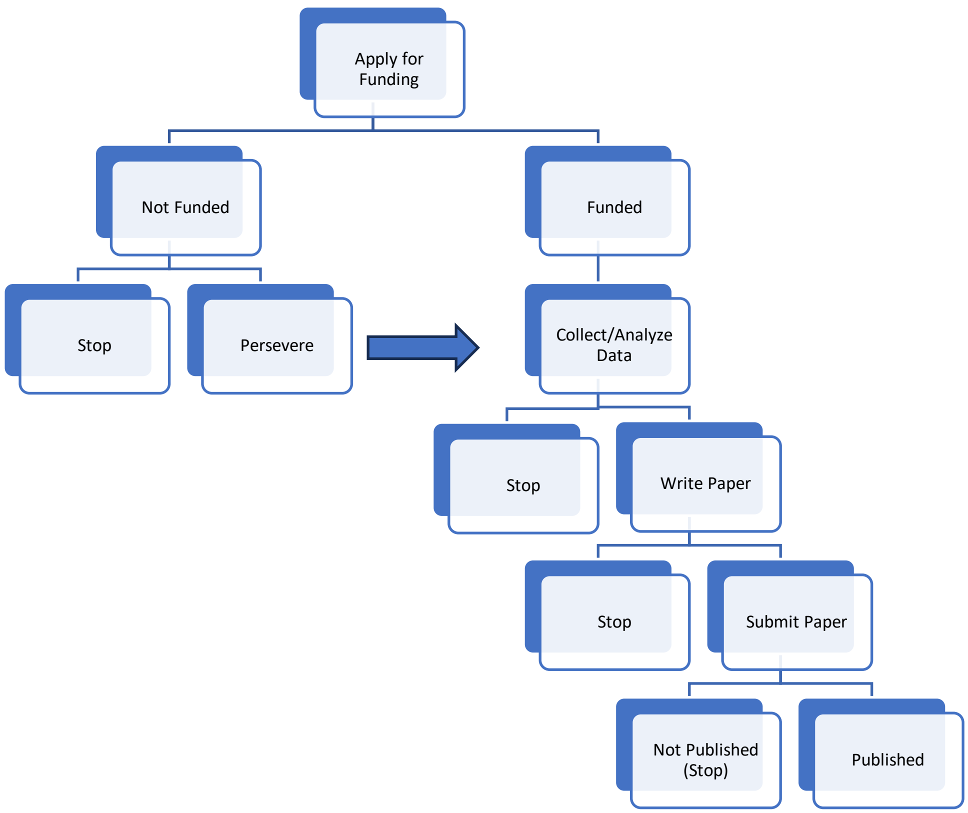
The first step is that the scholar needs to obtain resources for data collection, which for survey experiments means money to pay respondents and/or a vendor.[3] We characterize this step as *applying* for a grant to fund data collection with the outcome either being funded or not. If the grant is declined, the scholar can stop the project, meaning the idea will not be evaluated or published. Alternatively, the scholar can *persevere* by applying for a different grant (i.e., starting over), or finding other resources to pay for data collection, often meaning less expensive (and possibly lower quality) data.

If the grant is accepted or the researcher perseveres, we assume the data are collected and analyzed (we thus bundle the collection and analysis steps for now). At that point, which is the starting point in the aforementioned work by Franco et al. (2014), the scholar decides whether to *write* a paper with the results. If not, the process ends. If they do write a paper, then the scholar either *submits* it to a peer-reviewed outlet or fails to do so (ending the process). Finally, once submitted, the outlet (or outlets if multiple submissions are pursued) either *publishes* or rejects the paper with the latter being yet another hurdle to entering the scientific record. We realize the

[2] We recognize there are alternative ways to influence knowledge such as entering the gray literature (e.g., dissertations, conference papers); to the extent these are peer-reviewed processes, they could be substituted for the "published" outcome in our framework.
[3] The type/amount of resources vary not only for other types of work but also across survey experiments depending on the data source, a point to which we will return.

figure simplifies aspects and may not suitably represent other types of work (than survey experiments). That said, an upside is that it differs from virtually all prior metascience work by, at least somewhat, differentiating research process steps. It also provides a blueprint to investigate what factors correlate with challenges at these different steps and thus what could be done to lower hurdles.[4]

Figure 1: Empirical Knowledge Production Process



The challenge in understanding dynamics at each of these stages is typically an empirical one. We next explain our approach for obtaining relevant data to study each stage.

Time-sharing Experiments for the Social Sciences (TESS)

[4] We realize not all work should enter the published record and presumably some published work is less reliable than unpublished work. We are assuming, however, that in an ideal situation, the developed hypotheses move forward, undergo quality empirical evaluation, and become part of the scientific corpus. This comes with some assumptions that may not always be accurate.

Time-sharing Experiments for the Social Sciences (TESS) is a platform for conducting social science survey experiments fielded on probability-based samples of United States adults (https://www.tessexperiments.org/) (Mutz 2011). Established in 2001 with support from the National Science Foundation, TESS has provided, on a competitive basis, nearly 1,000 social scientists with the opportunity to test a broad range of innovative hypotheses. Investigators seeking to conduct an experiment using TESS submit a 5-page proposal (plus additional pages that detail the exact experimental treatments and measures) that is sent out for review, typically by two other scholars. The principal investigators use these reviews to decide which proposals to fund/implement; historically (over the course of the entire project), the acceptance rate for submitted proposals has been roughly 14% (although variable over-time). A funded project has data collected via a TESS contracted vendor who draws a probability sample.

TESS is, essentially, a proxy grant program for survey experimental studies to test hypotheses. This means one can assess which projects receive funding, which funded projects produce statistically significant effects, and which projects are published. We exploit this infrastructure to study each of the knowledge production steps in Figure 1. One the one hand, this has the advantage of not only investigating the multiple steps of the scientific process but also providing us with a set of studies that offer empirical tests of explicit hypotheses (as this characterizes virtually all TESS proposals), matching our theoretical presumption behind Figure 1. On the other hand, we are cognizant that this narrows the nature of our conclusions as one cannot generalize beyond social science survey experiments, which nonetheless constitute a central method for large parts of the social sciences (e.g., Sniderman 2011, 2018).[5]

[5] In their study of publication bias, Franco et al. (2018) point out that using TESS biases against finding publication bias since they are high quality studies. We agree, however, this does not apply to several of steps we explore (other than the publication step).

Our approach follows the lead of Franco et al.'s (2014) study; however, whereas Franco et al. focused (given their goals) on proposals accepted by TESS, we invited all *applicants*, regardless of acceptance, to complete a survey that asked about the trajectory of their proposed project. As we discuss next, this enabled us to identify the typically unobservable outcomes of whether investigators of declined proposals persevere, whether they collected data, whether any respondent submitted a paper to a journal, and whether their paper was accepted or rejected. Our approach has the additional advantage of allowing us to investigate over-time changes since we have analogous data and analyses to Franco et al. (2014) but from a later time period.

Survey of TESS Applicants

We identified the population of 794 TESS applicant-proposals (on which final decisions had been made) from October 2012 to January 2018.[6] We collected data from July 22, 2020, to September 23, 2020, sending multiple reminders and offering $10 to $25 for completion of the survey.[7] We chose the specific proposal dates because they encompassed a change in TESS's principal investigators (in late 2012) to a point when it seemed reasonable to assess whether the respondent had pursued a declined project and/or gone through the publication process (see Franco et al. 2014: 1503 for discussion). Franco et al.'s (2014) analysis included studies conducted between 2002 and 2012, and thus there is virtually no overlap, allowing us to compare our publication bias results to theirs to assess any over time changes.

We received a total of 544 responses for a 68.51% response rate.[8] Of those who responded, 107 (19.67%) had their proposals accepted and 437 (80.33%) had their proposals

---

[6] This population includes only those for whom we had or could locate a current e-mail address. Thus, it does not include all TESS applicants from this period.
[7] Compensation was higher for those who had multiple proposals; we also increased compensation offers with later reminders.
[8] If a respondents applied to TESS multiple times, they were asked about each application separately. Thus, our unit of analysis is applicant-proposal.

declined.[9] Thus, we see a slightly higher response rate for those with accepted proposals given the approximate actual acceptance rate is about 5-6% lower but the difference is not substantial. In the appendix, we compare the gender identification, race/ethnicity, and position (professional rank) of our sample to the sampling frame, showing that it matches very well.[10]

Our interest lies in in the five distinct decision points: 1) funding or not; 2) if not funded, persevere or not; 3) if funded/not funded and persevered, write paper or not; 4) if write paper, submit or not; and 5) if submit, publish or not. In Table 1, we provide an overview of our main measures. The exact question wordings are in the appendix. As should be evident, each item is fairly straightforward: funding success comes directly from TESS administrative data; a survey question asked whether the individual, if not funded, persevered on the project; a survey question asked TESS researchers/perseverers (who had collected data) whether they had analyzed their data; a survey question for those funded or persevering asked whether they wrote a draft of a paper; a survey question asked those who had a draft paper whether they had submitted it to a publication outlet; and a survey question asked those who submitted about the outcome. The survey included additional items we will use in supplemental analyses, including data and funding source for those who persevered, and, for published respondents, whether they perceive the outlet to be a top one.

The survey also included our key explanatory variables. Most important for the publication bias analysis is whether, for those who had analyzed data, they found statistically significant results regarding their main hypothesis (by whatever standard they used to interpret

[9] Franco et al. (2014) identify 249 TESS funded studies; their larger sample of TESS studies likely reflects that their time period was roughly twice as long as ours.

[10] We took a small random sample of TESS studies whose authors did not respond and used the study's data and searches for publication to assess file drawer bias. The results match the findings we report below. This is some evidence that survey non-response (although low) did not bias our findings.

statistical significance). In asking respondents for self-reported statistical significance instead of analyzing the data ourselves, we follow Franco et al. (2014) who explain that it is otherwise difficult to ascertain the exact analyses intended by the researchers. Even with access to the proposals, the specific analysis plan is not always present (the proposals are not equivalent to pre-analysis plans). More importantly, it matters most what the authors themselves consider their results (and thus how they present them) to be in terms of their decisions to write up and submit their findings.

Table 1: Main Outcomes Variables and Statistical Significance

Decision Point	**Variable**
Funding Decision	Funded or not funded (data from TESS on proposal disposition).
Perseverance	If not funded, asked whether project pursued or not.
Perseverance Data Collection/Analysis	If not funded, ask whether (if pursued project) data were collected and analyzed.
Funded Data Analysis (*all data had been collected)*	If funded, asked whether data were analyzed.
Write Paper	If funded and analyzed data or persevered and analyzed data, asked whether had completed a draft or final version of a paper.
Submitted	If paper written, asked whether had submitted to a peer reviewed outlet.
Published	If paper submitted, asked whether had been accepted or published.
Publication outlet	If accepted or published paper, asked whether outlet was considered "top."
Statistical significance	If funded and analyzed data or persevered and analyzed data, asked whether main hypothesis was statistically significant (by whatever standard the respondent used to interpret statistical significance).

While statistical significance is the key variable for steps after data analysis, other variables are relevant for whether TESS funded the proposal and whether those not funded persevered. We selected measures by drawing on literature that looks at success in obtaining grants and research productivity (Way et al. 2019, Zhang et al. 2022). This included four types of variables. First, we included various *demographics* and social characteristics including race/ethnicity, gender identification, age, whether the respondent had children, highest parental education level, and risk orientation (i.e., extent of risk-proneness) that research in the health context suggest can shape one's trajectory (e.g., Working Group on Diversity in the Biomedical Research Workforce 2012, Erosheva et al. 2020). Second, we added the size of one's annual discretionary research budget (on a 1-7 scale with higher values indicating a more sizeable budget) and the extent of opportunities to apply for internal institutional grants (on a 1-5 scale with higher scores indicating more opportunities) as either could offer *resources* in the absence of a successful grant application. Third, we gauged *time* available to pursue research by asking respondents the percentage of time they devoted to research (versus administration, teaching/advising, or other), and the number of Ph.D. students advised.[11] Finally, we included a host of *professional context* variables: the number of articles published in the last three years, the number of books published in the last three years, whether the person's department offered a Ph.D., whether the proposal included a co-author(s), and whether the respondent had tenure (e.g., an Associate or Full Professor).

Results

We start with the first step – what applicant characteristics correlate with being funded by TESS? The first column of Table 2 presents results of a regression of the applicant outcome on the previously discussed variables (recall 19.67% of proposals were funded).[12] It shows that the odds of being funded increase with having relatively more research time and decrease with the

[11] The number of Ph.D. students advised could work in either direction. The team-based nature of many sciences would lead one to expect that more Ph.D. students, in those disciplines, facilitates the pursuit of research ideas. But, even though team-based models have notably grown, they remain relatively rare in most of the social sciences, with students pursuing independent research. In this situation, advising more students would take time away from an advisor's own work.
[12] The N reduces due to missing values, mostly on age and race. The results are robust if we exclude those variables.

number of students advised – both indicators of time, which presumably facilitates the development and possible piloting of a project. For instance, moving from the 25th percentile of research time (which is 40%) to the 75% percentile (which is 80%) increases the probability of funding by 7-percentage points (.13 to .20). Age also increases the probability of acceptance, likely reflecting experience (age is extremely highly correlated with years since Ph.D., $r = .93$).[13] Having more published articles increases funding likelihood as well. We additionally find being a Black applicant and "other race" substantially decreases the probability of acceptance. The direct result is misleading as there were very few Black and "other race" respondents in our sample – just 1% and 2% of the sample, respectively. Thus, the number of negative decisions was negligible. Instead, the result reflects a notably poor record by TESS of attracting Black and "other race" applicants (see the appendix).[14] Otherwise, none of the other demographic variables (e.g., gender, and parental education) register significance.

Next, we turn to the decisions of those who did not receive funding. A total of 74.14% (324/437) of declined applicants report persevering, pursuing the project despite the rejection. Thus, for the majority of failed applicants, the TESS rejection did not deter their research. The second column of Table 2 presents regressions results with the outcome being whether a non-funded applicant reported pursuing the project upon being declined. Here, we do not differentiate how far the individual has gotten in terms of their perseverance (i.e., it does not require that they had already collected and analyzed their data). Interestingly, the results somewhat echo the funding findings. Time again clearly matters: those with more research time and those who

[13] For age, the difference between the 25th and 75th percentile – a 32-year-old vs. a 41-year-old, corresponds to a 4.8 percentage point increase (12.5% to 17.3%).
[14] As noted in the appendix, after the period of the data covered here, TESS launched a targeted population special competition that led to an apparent increase in minority applicants. Even so, clearly the program had done an inadequate job of attracting demographically diverse applicants.

advise fewer Ph.D. students (which, as explained, likely reflects having more time) were significantly more likely to persevere. We also found that having children falls just short of significance ($p = .12$) in a negative direction. The predicted probability of persevering for the average researcher in the 25th percentile of research time is 71.1% while that for a researcher in the 75th percentile is 81.8%, an 11-percnetage point difference. Additionally, we again see the benefit of experience with older respondents being significantly more likely to persevere.[15]

We find no evidence that access to resources (i.e., size of one's discretionary research budget, access to internal grant opportunities) influences perseverance.[16] Resources and time also are not substitutes as there is virtually no correlation between the resource and time measures. The result reflects the availability of much cheaper non-probability samples; indeed, 64% of those who persevere report obtaining data from a crowdsource labor market (e.g., Mechanical Turk), 27% from a national non-probability vendor (e.g., Dynata), and 6% from students or community members. All of these sources are dramatically cheaper than a national probability sample that TESS provides (3% report obtaining such a sample).

The researcher's demographic characteristics have scant influence. The only other significant relationships are a positive one with risk-proneness (reflecting the reality that pursuing any project after a decline is risky) and a negative one with the researcher's number of published articles (possibly reflecting investment in more projects and thus lower opportunity costs for dropping one given project). The exact mechanism underlying the time effect is an important question for future inquiry as it could reflect the literal time it takes to implement data

[15] In terms of predicted probabilities, the difference between the 25th and 75th percentile of age corresponds to a 13.7 percentage point increase (66.5% vs. 80.2%).
[16] Of those who collected data, 45% used their personal funds, 45% used other internal funding, and the remaining 9% acquired external funding.

collection (which the TESS program does) and/or the time a researcher perceives would be necessary to revise a project before data collection, given the unsuccessful grant proposal.

Table 2: Predicting Funding and Perseverance

	Funded	**Persevered**
Variable	Estimate (SE)	Estimate (SE)
Demographics		
Age	0.04 (0.02)*	0.08 (0.03)**
Female	0.26 (0.25)	0.10 (0.26)
Asian-American/ Pacific Islander (reference = White)	0.32 (0.37)	-0.05 (0.55)
Black (reference = White)	-14.82 (0.72)***	0.68 (1.22)
Hispanic (reference = White)	-1.16 (1.18)	-0.40 (0.42)
Other race (reference = White)	-14.57 (0.55)***	0.61 (1.16)
Parents' educ. BA (reference = no BA)	0.12 (0.41)	0.30 (0.37)
Parents' educ. MA or PhD (reference = no BA)	-0.06 (0.33)	0.03 (0.28)
Any children	-0.65 (0.26)	-0.43 (0.27)
Risk prone	-0.28 (0.17)	0.48 (0.18)**
Resources		
Research budget	0.12 (0.07)	0.01 (0.09)
Internal grant	0.11 (0.12)	0.09 (0.13)
Time		
Research time	0.01 (0.01)*	0.02 (0.01)**
Students advised	-0.02 (0.01)+	-0.03 (0.01)*
Professional Context		
Dept. confers Ph.D.	0.44 (0.31)	0.18 (0.30)
Coauthors	1.11 (0.31)	0.28 (0.26)
Tenure	0.22 (0.47)	0.45 (0.58)
Published articles	0.02 (0.01)*	-0.04 (0.02)*
Published books	-0.05 (0.17)	-0.07 (0.18)
Intercept	-4.91 (1.09)***	-4.17 (1.26)**
Num. obs.	500	398

***p<0.001, **p<0.01, *p<0.05, +p<0.1 for two-tailed tests. Estimates are logit coefficients. Standard errors clustered by researcher are in parentheses.

File Drawer Bias

We next turn to the processes that occur after data-collection. By definition, these analyses are relevant only to researchers who report having analyzed their data. This includes all 107 TESS recipients. Of those who persevered, 82.72% (268/324) had collected and analyzed their data (in the at least two-year period that had passed since their TESS application, the minimum passage of time).[17] While a high percentage, it is substantially lower than the 100% of TESS researchers who analyzed their data. This parallels the challenges of time insofar as for TESS-accepted scholars, experiments are virtually ready to launch and the programming, data cleaning, weighting, and so on, are all done via the TESS infrastructure.

The key variable for these analyses is whether the researcher reports finding statistical significance. We therefore present the results, for each outcome, separated by whether they were a TESS or persevering researcher, and whether they obtained statistical significance. These appear in Table 3. The main categories echo Figure 1: writing, submitting, and publishing.[18] We present regressions (with the explanatory variables used in the other analyses) for each outcome in the appendix; none of the substantive findings change.

[17] A total of 113 did not pursue the project (as noted). Otherwise, 34 reported pursuing the project but not yet collecting their data, and 16 had collected but not yet analyzed their data. Another 6 respondents report analyzing their data but then did not respond to questions about the results and/or subsequent activities (e.g., writing up, submitting) and thus we remove them from these analyses.

[18] This table matches the one from Franco et al. (2014) with a few caveats. First, as mentioned, our data are from a subsequent time period. Second, we include persevering researchers whereas Franco et al. (2014) included TESS researchers only. Third, Franco et al. (2014) included a category of mixed significance whereas we instead asked if the main hypothesis was significant or not (by whatever standard the respondent used, as explained). Fourth, we added a stage of submitted but not published since it is possible for a scholar to write up a paper but never submit it. We thought this was an important addition since it allows us to pinpoint any bias strictly attributable to the editorial and peer review processes, as opposed to authors' decisions.

Table 3: Outcomes for Researchers Who Analyzed Data

	TESS Researchers (N = 107)		Persevering Researchers (N = 268)	
	Statistically Significant (N = 61)	Not Statistically Significant (N = 46)	Statistically Significant N = 171)	Not Statistically Significant (N = 97)
Not Written	8.20%	28.26%	13.45%	41.24%
Written but not submitted	0.00%	15.22%	7.02%	13.40%
Submitted but not published	16.39%	10.87%	22.22%	9.28%
Published	75.41%	45.65%	57.31%	36.08%
Total	100.00%	100.00%	100.00%	100.00%

The results reveal several dynamics about the knowledge production process. First, focusing on the “published” row, we find clear evidence of a classic file drawer publication bias, for both TESS researchers and persevering researchers. In both cases, those with statistically significant results are substantially more likely to publish ($p < .01$ for both, for two-tailed tests). The bias looks larger for TESS researchers (roughly 30 percentage points, 75.41% - 45.65%, compared to 21 percentage points, 57.31% - 36.08%). Second, the “submitted but not published” row – that is, papers that were submitted to journals but not accepted shows that there is no apparent file drawer bias in the actual publication process at journals. In fact, the results suggest that journals themselves may have a preference for non-significant results. For TESS researchers, there is a roughly 5-percentage point advantage for non-significant results – only 10.87% of non-significant results submit but are not published versus 16.39% of significant results (although this is not a significant difference, $p = .42$). The respective statistics for persevering researchers are 9.28% and 22.22%, a significant 13-percentage point gap with statistically significant results being more likely to be submitted but not published ($p < .01$). Thus, if anything, journals with non-probability sample survey experiments give preferential treatment to null results. These results echo Brodeur et al.'s (2023) analysis of the journal submission process at a top applied microeconomics journal that finds the peer-review process is not the source of publication bias (instead they point to author behavior).[19] Further, while not exactly the same, it is similar to Berinsky et al.'s (2021) "gotcha bias" where null (non-significant) replication results are preferred.

Third, the endline file drawer publication bias emerges – not from a bias against significant findings in submissions – but rather because researchers either do not write up non-significant results or do not submit those that they have written up.[20] No TESS researcher who wrote up significant results failed to subsequently submit, while nearly 15.22% of those with non-significant results had not submitted ($p < .01$). The respective percentages for persevering researchers are 7.02% and 13.40%, a difference just short of conventional statistical significance ($p = .08$). Even more dramatic is the failure to write up findings. Among TESS researchers, we find a 20-percentage point gap in writing up the paper between those with significant findings and those without (8.20% versus 28.26%; $p < .01$). Among persevering researchers, there is an astounding roughly 27-percentage point gap (13.45% versus 41.24%; $p < .01$). In short, the file drawer bias stems entirely from author choices and not publication processes.

[19] They focus on p-hacking or the bunching of p-values around statistical significance.

[20] We recognize that it is possible some of those written may still be submitted but even if they were all submitted and published, we would still see an endline file drawer bias due to the failure to write up insignificant results. Moreover, the timing of the survey means that all of these researchers had a good amount of time to have written up and submitted their results. (Also, note that a paper submitted, declined and then not yet re-submitted would fall into the "submitted but not published" category.)

This result also explains, somewhat, the larger ostensible endline publication bias for TESS researchers relative to persevering researchers. It is a function of writing-up, not the peer-review process. The result can be most directly seen by simply computing the percentages of published papers *of all those submitted* (see the appendix for a full set of statistics). For TESS researchers, 82.14% of the statistically significant submissions were published versus 80.77% of insignificant submissions being published. For persevering researchers, 72.06% of statistically significant submissions were published versus 79.55% of insignificant submissions being published (again revealing a marginal preference for non-significant result). The root of the bias appears to be from persevering researchers being much less likely to write up non-significant results than TESS researchers (58.76% versus 71.74%; $p = .13$; this falls short of significance partially to the small number of non-significant TESS results).[21] Overall, the file drawer bias reflects researcher rather than editorial decisions.[22]

Over-time Changes and TESS versus Persevering Researchers

As mentioned, we can compare our results from the end of 2012 to 2018 against Franco et al. (2014) who used analogous TESS studies from 2002 to 2012. Our findings match theirs in terms of evidence of clear publication bias, with most of it stemming from the authors either choosing not to write up results or not submitting. As mentioned, Franco et al. report a roughly 5-percentage point bias toward significant results in the publication process (Franco et al. 2014:

[21] Also, slightly more TESS researchers (91.80%) with significant results write up results compared to preserving researchers (86.55%), although this is far from statistically significant ($p = .28$).

[22] Recall we followed Franco et al. (2014) in operationalizing statistical significance based on the authors' interpretations. We did, however, also ask respondents what level of statistical significance they used in interpreting their results. Of TESS researchers who reported they had statistically significant results, 72.9% used a p-value of .05 or lower with another 22.9% using .10. Of preserving researchers who reported that had statistically significant results, 92.6% used a p-value of .05 or lower with another 5.0% using .10. Thus, TESS researchers seemed more comfortable in using a lower standard p-value, perhaps anticipating leverage given their high-quality samples. Even so, given we find no endline publication bias differences, the varying use of significance levels is not a cause of file drawer bias per se, at least the peer review stage.

appendix, Table S2) whereas we find no significant bias and in fact the reverse among persevering researchers. Perhaps more important is that the file drawer problem is much less present in our data than in the Franco et al. data. They find about 21% of non-significant papers are published (in either non-top or top-tier journals) while 62% of strongly significant ones are published, for a 41-percentage point difference. Among our TESS researchers – the sample comparable to Franco et al. – the number of null papers published increased to nearly 46%; the number of significant papers published also increased to about 75%, but the overall gap dropped to 29 percentage points. The persevering researchers have lower overall publication rates, but their gap is even smaller at 21 percentage points. The decline in the file drawer gap stems from more authors writing up their papers: 65% of those with non-significant results did not write up their papers in the Franco et al. data whereas in our data, among TESS researchers, only 28% did not. In that sense, the Franco et al. paper itself – given its very high profile – likely had an influence in encouraging authors to write up null results. This coincided with the open science movement that also surely played a role (Christensen et al. 2019). Even so, the results make clear there is still room to go in terms of encouraging writing up *and* submitting, given we find no evidence of a bias against non-significant results in the publication process. In sum, the file drawer bias seems to have the same main source of failing to write up non-significant results (and not submitting those written) but the bias is smaller, thus reducing the overall file drawer bias.

Finally, since we have data from TESS and non-TESS researchers, we can assess whether researchers privilege TESS studies. We do this by exploring the different research trajectories that influence outcomes at each stage (see the appendix for the full set of statistics). At the writing stage, as mentioned, TESS researchers are more likely to write up non-significant results by 13 percentage points (71.74% versus 58.76%), although this falls short of conventional significance (p = .13). At the submission stage, the main difference is with significant results, with all (100%) TESS researchers submitting if they had written and 91.89% of persevering researchers submitting (still a high number) ($p < .05$). At the publication stage, it is again statistically significant results from TESS that are privileged with 82.14% of submitted TESS researcher papers being published versus 72.06% of submitted papers from persevering researchers being published (p = .14).[23] These are all suggestive results given the uneven statistical significance. Yet, they point toward peer-review and editorial processes advantaging data from TESS and researchers anticipating this in their submission decisions. That said, investigators perhaps too quickly choose not to write up non-significant results (that could ultimately be published). Further, we asked respondents whose work was published whether they considered the outlet in which it was published to be a "top" outlet. We find a sizeable difference between TESS and persevering authors, with nearly 77% of TESS researchers publishing in top outlets versus about 57% of persevering researchers ($p < .01$).

Researchers' tendency to privilege funded projects could stem from TESS scholars designing higher-quality projects, as they were accepted by TESS in the peer-review process (while the perseverers were not), or from the use of nationally representative probability samples by TESS as opposed to nearly all non-probability samples (as noted above) by persevering researchers.

Conclusion

The last decade has seen significant changes in the conduct of science, as epitomized by open science initiatives (Nosek et al. 2015, Christensen et al. 2019, Druckman 2022). This

[23] It could be that non-TESS researchers will end up publishing additional studies, insofar as it may take longer for them to collect and analyze their data. Thus, some of the differences we document could have shrunk with time.

includes increased attention to long standing questions of publication biases as well as discussion about the drivers of research productivity. A persistent challenge with documenting the research and publication processes is that much of it remains unobserved. We focused on the empirical stages of knowledge production when it comes to one widely used approach – social science survey experiments. While we recognize limits of the extent to which we can generalize our findings to other research approaches, the focus allowed us to leverage a unique data set from TESS, matching and expanding on the approach used by Franco et al.'s (2014) seminal article.

When it comes to gaining acceptance by TESS or persevering on a declined project, having time is essential. This likely reflects a general phenomenon that more time allows one to develop their ideas and pursue them when presented with hurdles. The lack of an effect for resources may be more idiosyncratic insofar as there are several relatively low-cost substitute data options for survey experimental work (and indeed, we find these are what most persevering researchers used). An implication is that, particularly, for early-career scholars who frequently employ this method, institutions should be cognizant of providing time protections (e.g., course releases, service reductions) and offering mentoring support with regard to time management. In some cases, this may be more valuable than resources. Additionally, the widespread use of lower-cost non-probability sample data sources highlights the importance of assessing their reliability and being transparent in reporting information about them (Jamieson et al. 2023).

Additionally, we replicate decades of work in documenting a file drawer bias in the publication process. Yet, to the extent this leads to a skewed distribution of knowledge, there is reason for optimism. The bias is smaller than that from earlier years (as documented by Franco et al. 2014). Further, it stems not from editorial or peer-review processes but rather researcher choices to not write up or submit null results. While we cannot dismiss the possibility that those decisions may correlate with other parts of the studies – meaning that they would not have been accepted if written/submitted – it seems reasonable to presume that at least a sizeable share of the file drawer bias reflects researcher miscalculation of publication processes. This aligns with Brodeur et al.'s (2023) findings on publication bias. It also coheres with Squazzoni et al. (2021) who explore gender bias in publication, showing much of it stems from authors' expectations rather than peer review or editorial decisions. The lesson is that it is difficult to alter beliefs about long-standing biases in academic publishing.[24] Squazzoni et al. (2021) emphasize the need for explicit signaling from publications. The same can be said when it comes to openness to null results. Several journals now explicitly invite replication studies (regardless of results) and/or pre-registered reports, and this surely has mattered, and its expansion could serve as a further antidote to the file drawer bias. Journals could more explicitly make clear that null results, generally, are welcome (contingent on quality).[25] This is all easier said than done insofar as publications also need to be careful to avoid a "gotcha" bias where null results of replications are favored (Berinsky et al. 2021) – indeed, we found some evidence of this with persevering researchers. Overall, the key point is additional communication from publication outlets could further contribute to addressing the file drawer bias, in addition to pre-registration reports and replications.

We also find some advantages for TESS studies relative to non-TESS studies. As mentioned, this could reflect the data sources used or the studies' general quality (given TESS studies survived the program's peer-review process). Identifying when biases reflect aspects of studies uncorrelated to quality versus those related to quality is extremely difficult. It is entirely

[24] As Brodeur et al. (2023: 1300) explain "a large set of economists (falsely) believe that editors and reviewers have strong preferences for significant results, leading them to engage in… withholding their nonsignificant results from journal submission."
[25] Granting entities too, including TESS, could consider imposing writing requirements (Franco et al. 2014: 1504).

possible that, given the number of errors that could occur in the data collection phase, that statistical significance correlates with quality (Malhotra 2021, Druckman 2022: 141). It remains an open question.

Our hope is that our work stimulates further investigations of the research lifecycle, even with a larger purview than we have used. This would include looking at both career and project trajectories. The research process is defined by challenges (Oliver 2004, Druckman 2022: 118-126), and identifying and addressing such hurdles contributes to advances in knowledge and improved career decisions.

References

Brodeur, Abel, Scott Carrell, David Figlio, and Lester Lusher. 2023. "Unpacking P-Hacking and Publication Bias." *American Economic Review* 113 (11): 2974–3002.

Ceci, Stephen J., Shulamit Kahn, and Wendy M. Williams. 2023. "Exploring Gender Bias in Six Key Domains of Academic Science: An Adversarial Collaboration." *Psychological Science in the Public Interest* 24 (1): 15–73.

Christensen, Garret, Jeremy Freese, and Edward Miguel. 2019. *Transparent and Reproducible Social Science Research: How to Do Open Science*. Berkeley, CA: University of California Press.

Dietz, Thomas. 2013. "Bringing Values and Deliberation to Science Communication." *Proceedings of the National Academy of Sciences* 110: 14081–87.

Druckman, James N. 2022. *Experimental Thinking: A Primer on Social Science Experiments*. New York: Cambridge University Press.

Erosheva, Elena A., Sheridan Grant, Mei-Ching Chen, Mark D. Lindner, Richard K. Nakamura, and Carole J. Lee. 2020. "NIH Peer Review: Criterion Scores Completely Account for Racial Disparities in Overall Impact Scores." *Science Advances* 6 (23): eaaz4868.

Fanelli, Daniele. 2018. "Is Science Really Facing a Reproducibility Crisis, and Do We Need It To?" *Proceedings of the National Academy of Sciences* 115 (11): 2628–31.

Franco, Annie, Neil Malhotra, and Gabor Simonovits. 2014. "Publication Bias in the Social Sciences: Unlocking the File Drawer." *Science* 345 (6203): 1502–5.

Gerring, John, James Mahoney, and Colin Elman. 2020. "Introduction." In Colin Elman, John Gerring, and James Mahoney, eds., *The Production of Knowledge: Enhancing Progress in Social Science*. Cambridge, UK: Cambridge University Press.

Jamieson, Kathleen Hall, Arthur Lupia, Ashley Amaya, Henry E. Brady, René Bautista, Joshua D. Clinton, Jill A. Dever, David Dutwin, Daniel L. Goroff, D. Sunshine Hillygus, Courtney Kennedy, Gary Langer, John S. Lapinski, Michael Link, Tasha Philpot, Ken Prewitt, Doug Rivers, Lynn Vavreck, David C. Wilson, and Marcia K. McNutt. 2023. "Protecting the Integrity of Survey Research." *PNAS Nexus* 2, (3): pgad049.

Malhotra, Neil. 2021. "Threats to the Scientific Credibility of Experiments." In James N. Druckman and Donald P. Green, eds., *Advances in Experimental Political Science*. New York: Cambridge University Press.

Mutz, Diana C. 2011. *Population-Based Survey Experiments*. Princeton, NJ: Princeton University Press.

Nosek, Brian A., George Alter, George C Banks, Denny Borsboom, Sara D Bowman, Steven J. Breckler, Stuart Buck, et al. 2015. "Promoting an Open Research Culture." *Science* 348 (6242): 1422–25.

Oliver, Jack E. 2004. *The Incomplete Guide to the Art of Discovery*. Ithaca, NY: Internet-First University Press.

Oreskes, Naomi. 2019. *Why Trust Science?* Princeton, NJ: Princeton University Press.

Sniderman, Paul M. 2011. "The Logic and Design of the Survey Experiment." In James N. Druckman, Donald P. Green, James H. Kuklinski, and Arthur Lupia, eds., *Cambridge Handbook in Experimental Political Science*. New York: Cambridge University Press.

Sniderman, Paul M. 2018. "Some Advances in the Design of Survey Experiments." *Annual Review of Political Science* 21: 259–75.

Spoon, Katie, Nicholas LaBerge, K. Hunter Wapman, Sam Zhang, Allison C. Morgan, Mirta Galesic, Bailey K. Fosdick, Daniel B. Larremore, and Aaron Clauset. 2023. "Gender and Retention Patterns among U.S. Faculty." *Science Advances* 9 (42): eadi2205.

Squazzoni, Flaminio, Giangiacomo Bravo, Mike Farjam, Ana Marusic, Bahar Mehmani, Michael Willis, Aliaksandr Birukou, Pierpaolo Dondio, and Francisco Grimaldo. 2021. "Peer Review and Gender Bias: A Study on 145 Scholarly Journals." *Science Advances* 7 (2): eabd0299.

Teele, Dawn Langan, and Kathleen Thelen. 2017. "Gender in the Journals: Publication Patterns in Political Science." *PS: Political Science & Politics* 50 (2): 433–47.

Way, Samuel F., Allison C. Morgan, Daniel B. Larremore, and Aaron Clauset. 2019. "Productivity, Prominence, and the Effects of Academic Environment." *Proceedings of the National Academy of Sciences* 116 (22): 10729–33.

Working Group on Diversity in the Biomedical Research Workforce. 2012. "Draft Report of the Advisory Committee to the Director Working Group on Diversity in the Biomedical Research Workforce." Bethesda, MD: National Institutes of Health.

Zhang, Sam, K. Hunter Wapman, Daniel B. Larremore, and Aaron Clauset. 2022. "Labor Advantages Drive the Greater Productivity of Faculty at Elite Universities." *Science Advances* 8 (46): eabq7056.

Appendix

Sample Demographics

In our sample, 86% of respondents identified as White, 7% as Asian-American, 4% as Hispanic, 1% as Black, and 2% as other/unknown. In terms of position, 49% were graduate students, 26% Assistant Professors, 7% Associate Professors, 8% Full Professors, and 10% other (e.g., post-doctoral fellow, lecturer). And, 59% were male, 40% female, and 1% other. To benchmark these statistics, we independently searched on each person in our sampling frame to identify their apparent race/ethnicity, gender identification, and position at the time of applying to TESS. It matches the sample very well. For race/ethnicity, the frame consisted of 78% White, 9% Asian-American, 5% Hispanic, 3% Black, and 5% other/unknown. (As noted in the text, TESS could dramatically improve in soliciting applications from minority applicants, particularly Black and Hispanic applicants; along these lines, after this data collection, it launched a targeted population special competition that led to an apparent increase in applicants from these groups. Clearly, more steps are needed given the very poor track record.) For position, we find that 47% were graduate students, 28% were Assistant Professors, 9% were Associate Professors, 8% were Full Professors, and 8% were other. Finally, 57% were male, 41% were female, and 2% were unknown.

Survey Questions

Over the past *three* years, roughly how many papers have you published in journals or edited books? (Include any forthcoming papers.) ________

Over the past *three* years, how many books have you published (as an author or editor)? (Include any forthcoming books.)__________

Over the course of your career, how many, if any, Ph.D. students have you advised (as the main advisor) (include current ABD students):_________

Do you have many opportunities to obtain internal funding from your department or college or place of work?
a. Not at all
b. A little
c. Somewhat
d. A fair amount
e. A lot

What is the size of your discretionary research budget, on an annual basis?
a. $0
b. $1-$999
c. $1000-$4,999
d. $5000-$9,999
e. $10,000-$19,999

f. $20,000-$49,999
e. More than $50,000

What is your age?

What is the highest level of education completed by one of your parents? (Think about the parent who has received the highest level of education.) You can check multiple entries when it comes to advanced degrees.
a. Less than high school
b. High school
c. Some college
d. 4 year college degree
e. M.A./M.S.
f. J.D.
g. M.D.
h. Ph.D.

Do you have any children?
a. No
b. Yes

How do you see yourself: are you generally a person who is fully prepared to take risks or do you try to avoid taking risks?
a. Not at all willing to take risks
b. Rarely willing to take risks
c. Occasionally willing to take risks
d. Frequently willing to take risks
e. Always willing to take risks

What is your gender?
a. Male
b. Female
c. Other

Which of the following do you consider to be your primary racial or ethnic group (*you may check more than one on this question*)?
a. White
b. Black/African American
c. Hispanic/Latino
d. Asian/Pacific Islander
e. Middle Eastern/North African
f. Native American
g. Other

What is the highest terminal degree *granted by your current department*?
a. Ph.D.

b. Master's Degree
c. Bachelor's Degree
d. I do not work at a school.

Which of the following best describes your position *when you submitted this TESS application*?
a. Emeritus Professor
b. Full Professor
c. Associate Professor
d. Assistant Professor
e. Continuing non-tenure track lecturer
f. Adjunct Professor
g. Postdoctoral Researcher
h. Graduate Student
i. Other

When you submitted this TESS proposal, roughly what percentage of your work time did you spend on teaching/advising, administrative, research, and "other"? Please fill in the appropriate percentages, which should sum to 100%.
a. Teaching/advising________
b. Administration__________
c. Research_________
d. Other____________

The TESS submission system provided information on the presence of co-authors.

FOR TESS-FUNDED RESEARCHERS

Many people take some time to analyze their data from TESS to ensure careful analyses (and due to other responsibilities that take up time). What about you? Have you analyzed the data from your TESS study?
a. No
c. Yes

Have you written up a draft or final version of your results?
a. No and I do not plan to do so
b. No and I have not started but I plan to do so
c. I am working on it but I'm not sure if I will ever finish
d. I am working on it and I hope to finish at some point
e. Yes

Have you submitted the paper to a peer reviewed outlet (e.g., journal, publisher, edited book)?
a. No
b. Yes

What is the current status of the paper?
a. It is not under review (e.g., it was declined and I have not re-submitted)
b. It is under initial review at a peer reviewed outlet (e.g. journal, book publisher)
c. It received a revise-and-re-submit and I am revising or have re-submitted
d. It has been accepted and is forthcoming
e. It is in print (on-line or in paper copy)

IF ACCEPTED OR IN PRINT:
At what outlet?

Do you consider this to be a "top" peer reviewed outlet?

Did you find statistically significant results when it comes to your main hypothesis or hypotheses (by whatever standard you use to interpret statistical significance)?
a. No
b. Yes

To the extent that you recall/know, at what level of statistical significance were your main results (using two-tailed tests)? Check all that apply.
a. $\leq .001$
b. $\leq .005$
c. $\leq .01$
d. $\leq .05$
e. $\leq .10$
f. Not significant, at least the .10 level.
g. other

FOR TESS NON-FUNDED RESEARCHERS

Please check which of the following best describes what happened.
a. I did not pursue the project in any form
b. I pursued the project, it remained similar, and I have collected and analyzed the data
c. I pursued the project, it remained similar, and I have collected but not analyzed the data
d. I pursued the project, it remained similar, but I have not yet collected data
e. I pursued the project but it changed somewhat, and I have collected and analyzed the data
f. I pursued the project but it changed somewhat, and I have collected but not analyzed the data
g. I pursued the project but it changed somewhat, but I have not yet collected data

IF COLLECTED DATA:
What was your data source? Check all that apply.
a. Mechanical Turk
b. A student subject pool

c. Community members / college staff
d. Lucid
e. Polimetrics/YouGov
f. Knowledge Networks/GfK/Ipsos
g. NORC
h. SSI / ResearchNow/ Dynata
i. Crowdsourcing data other than Mechanical Turk.
j. A vendor not listed above.

How did you pay for the data collection? Check all that apply.
a. My personal research funds
b. My advisor's research funds or grant (e.g., if you are a graduate student or post-doc)
c. Funds from the university but not my personal research funds (e.g., an internal grant)
d. An NSF grant
e. An external grant not from the NSF
f. My collaborator is paying for it
g. My personal money (i.e., not meant for research)
h. Other

Have you written up a draft or final version of your results?
a. No and I do not plan to do so
b. No and I have not started but I plan to do so
c. I am working on it but I'm not sure if I will ever finish
d. I am working on it and I hope to finish at some point
e. Yes

Have you submitted the paper to a peer reviewed outlet (e.g., journal, publisher, edited book)?
a. No
b. Yes

What is the current status of the paper?
a. It is not under review (e.g., it was declined and I have not re-submitted)
b. It is under initial review at a peer reviewed outlet (e.g. journal, book publisher)
c. It received a revise-and-re-submit and I am revising or have re-submitted
d. It has been accepted and is forthcoming
e. It is in print (on-line or in paper copy)

IF ACCEPTED OR IN PRINT:
At what outlet?

Do you consider this to be a "top" peer reviewed outlet?

Did you find statistically significant results when it comes to your main hypothesis or hypotheses (by whatever standard you use to interpret statistical significance)?
a. No
b. Yes

To the extent that you recall/know, at what level of statistical significance were your main results (using two-tailed tests)? Check all that apply.
a. ≤ .001
b. ≤ .005
c. ≤ .01
d. ≤ .05
e. ≤ .10
f. Not significant, at least the .10 level.
g. other

Supplementary Analyses

Research Trajectory Percentages

The below table uses the same data as in Table 3; however, it presents them differently, showing the percentage who act at each stage, given they moved on from the prior stage.

Table A-1: Percentages at Research Stages

TESS Researchers (N = 107)		Persevering Researchers (N = 268)	
Statistically Significant	Not Statistically Significant	Statistically Significant	Not Statistically Significant
Write Up			
91.80% (56/61)	71.74% (33/46)	86.55% (148/171)	58.76% (57/97)
Submit			
100% (56/56)	78.79% (26/33)	91.89% (136/148)	77.19% (44/57)
Publish			
82.14% (46/56)	80.77% (21/26)	72.06% (98/136)	79.55% (35/44)

Regressions

The regressions that follow parallel the results presented in Table 3, by adding control variables. Note that demographic variables are missing for some of the TESS researcher regressions due to insufficient numbers (see discussion in the appendix sample demographics section). Table A.2 shows that statistically significant results influence the decision to write up a draft of the findings. While the magnitude is larger for TESS researchers, it is not significantly so (i.e., the interaction is not significant). TESS researchers are, all else constant, more likely to write up their results. The other consistent finding, across groups, is that an increase in students advised decreases the likelihood of writing up results.

Table A-3 shows statistical significance increases the likelihood of submitting, with the TESS researcher coefficient being enormous (and the interaction significant) due to all TESS researchers with significant results who wrote up a paper submitting (see Table 3). Here, across groups, students advised actually increases the likelihood of submitting while age decreases it (perhaps reflecting less pressure concerning tenure).

Table A-4 shows statistical significance does not influence publication results, reflecting the findings in the paper that the file draw bias reflects the authors' decisions and not the peer review process. Age increases likely publication success (across groups), presumably reflecting experience.

Table A-2: Logit Model of Writing Up a Draft

	TESS Researchers	**Persevering Researchers**	**All Researchers (with interaction)**
Variable	Estimate (SE)	Estimate (SE)	Estimate (SE)
Significant results	2.88 (1.56) +	1.86 (0.40) ***	1.80 (0.38) ***
TESS researcher (reference = persevering researchers)			0.90 (0.45) *
Significant results x TESS researcher			-0.09 (0.76)
Demographics			
Age	0.07 (0.05)	0.05 (0.04)	0.04 (0.03)
Female	0.90 (0.92)	-0.04 (0.38)	0.05 (0.34)
Asian-American/ Pacific Islander (reference = White)	-1.05 (2.06)	0.38 (0.66)	0.20 (0.63)
Black (reference = White)		-0.88 (1.04)	-0.59 (0.97)
Hispanic (reference = White)	-19.72 (2.01) ***	15.16 (0.50) ***	0.90 (1.11)
Other race (reference = White)		-0.78 (1.23)	-0.60 (1.30)
Parents' educ. BA (reference = no BA)	0.41 (1.54)	0.00 (0.54)	0.28 (0.49)
Parents' educ. MA or PhD (reference = no BA)	-2.78 (1.29) *	-0.36 (0.43)	-0.38 (0.38)
Any children	-3.42 (1.31) **	0.51 (0.38)	0.11 (0.32)
Risk prone	0.15 (0.56)	-0.05 (0.24)	0.03 (0.22)
Resources			
Research budget	0.34 (0.25)	0.01 (0.12)	0.07 (0.11)
Internal grant	-0.70 (0.85)	-0.23 (0.19)	-0.29 (0.17) .
Time			
Research time	-0.00 (0.02)	0.01 (0.01)	0.00 (0.01)
Students advised	-0.18 (0.06) **	-0.05 (0.02) *	-0.06 (0.02) **
Professional Context			
Dept. confers Ph.D.	0.48 (1.26)	0.28 (0.46)	0.27 (0.42)
Coauthors	-1.24 (1.51)	-0.67 (0.39) +	-0.51 (0.35)
Tenure	0.25 (1.18)	-0.68 (0.80)	-0.37 (0.66)
Published articles	0.41 (0.20) *	0.01 (0.03)	0.04 (0.03)
Published books	2.83 (0.92) **	0.35 (0.26)	0.54 (0.26) *
Intercept	0.54 (3.24)	-0.81 (1.85)	-0.69 (1.64)
Num. obs.	102	247	349

***p<0.001, **p<0.01, *p<0.05, +p<0.1 for two-tailed tests.
Estimates are logit coefficients. Standard errors clustered by researcher are in parentheses.

Table A-3: Logit Model of Submitting to a Journal

	TESS Researchers	Persevering Researchers	All Researchers (with interaction)
Variable	Estimate (SE)	Estimate (SE)	Estimate (SE)
Significant results	376.12 (6.12) ***	1.07 (0.55) *	0.95 (0.53) +
TESS researcher (reference = persevering researchers)			-0.56 (0.71)
Significant results x TESS researcher			17.72 (0.89) ***
Demographics			
Age	-11.19 (0.18) ***	-0.12 (0.05) *	-0.13 (0.04) **
Female	0.02 (1.18)	-0.77 (0.53)	-0.56 (0.43)
Asian-American/ Pacific Islander (reference = White)	-138.96 (3.45) ***	-1.00 (0.94)	-1.23 (0.86)
Black (reference = White)		-2.19 (0.98) *	-2.04 (0.94) *
Hispanic (reference = White)		-2.15 (1.43)	-2.02 (1.35)
Other race (reference = White)		15.31 (1.01) ***	17.03 (0.99) ***
Parents' educ. BA (reference = no BA)	-9.03 (1.81) ***	0.26 (0.82)	0.02 (0.70)
Parents' educ. MA or PhD (reference = no BA)	-33.68 (1.12) ***	0.54 (0.67)	0.53 (0.57)
Any children	-135.50 (2.65) ***	0.95 (0.54) .	0.67 (0.45)
Risk prone	-31.92 (0.68) ***	0.15 (0.34)	0.08 (0.32)
Resources			
Research budget	-5.65 (0.32) ***	-0.20 (0.17)	-0.12 (0.15)
Internal grant	-32.54 (0.60) ***	0.42 (0.37)	0.27 (0.31)
Time			
Research time	-0.97 (0.04) ***	-0.01 (0.01)	-0.01 (0.01)
Students advised	10.23 (0.17) ***	0.29 (0.16) .	0.23 (0.10) *
Professional Context			
Dept. confers Ph.D.	-154.49 (2.74) ***	0.27 (0.71)	-0.07 (0.61)
Coauthors	-26.09 (1.18) ***	1.42 (0.57) *	1.63 (0.50) **
Tenure	17.43 (1.31) ***	-0.69 (1.18)	-0.13 (0.95)
Published articles	3.08 (0.11) ***	0.08 (0.06)	0.05 (0.05)
Published books	80.79 (1.36) ***	0.45 (0.53)	0.60 (0.50)
Intercept	869.21 (14.26) ***	3.01 (2.27)	4.22 (2.02) *
Num. obs.	89	192	281
***p<0.001, **p<0.01, *p<0.05, +p<0.1 for two-tailed tests. Estimates are logit coefficients. Standard errors clustered by researcher are in parentheses.			

Table A-4: Logit Model of Publishing

	TESS Researchers	**Persevering Researchers**	**All Researchers (with interaction)**
Variable	Estimate (SE)	Estimate (SE)	Estimate (SE)
Significant results	-9.17 (5.89)	-0.49 (0.48)	-0.50 (0.49)
TESS researcher (reference = persevering researchers)			-0.23 (0.71)
Significant results x TESS researcher			0.26 (0.80)
Demographics			
Age	0.77 (0.52)	0.10 (0.04) *	0.11 (0.04) **
Female	4.41 (3.64)	-0.27 (0.45)	0.14 (0.40)
Asian-American/ Pacific Islander (reference = White)	31.66 (9.85) **	0.23 (0.88)	1.12 (0.80)
Black (reference = White)		15.11 (1.03) ***	15.14 (1.02) ***
Hispanic (reference = White)		1.37 (0.95)	1.58 (0.96) .
Other race (reference = White)		1.06 (1.10)	0.87 (1.03)
Parents' educ. BA (reference = no BA)	16.39 (11.70)	0.75 (0.74)	1.25 (0.61) *
Parents' educ. MA or PhD (reference = no BA)	5.54 (4.69)	0.03 (0.61)	0.35 (0.49)
Any children	-4.10 (3.08)	-0.37 (0.40)	-0.16 (0.34)
Risk prone	2.10 (1.46)	0.12 (0.28)	0.19 (0.23)
Resources			
Research budget	-0.46 (0.37)	0.02 (0.14)	-0.12 (0.12)
Internal grant	2.78 (1.43) .	-0.34 (0.22)	-0.21 (0.18)
Time			
Research time	-0.08 (0.10)	0.01 (0.01)	0.01 (0.01)
Students advised	-0.92 (0.79)	-0.01 (0.04)	-0.00 (0.06)
Professional Context			
Dept. confers Ph.D.	12.24 (10.86)	0.59 (0.59)	0.90 (0.46) .
Coauthors	-8.14 (5.40)	0.37 (0.41)	-0.04 (0.38)
Tenure	42.39 (30.49)	-0.95 (0.86)	-0.18 (0.66)
Published articles	0.16 (0.16)	0.15 (0.06) *	0.13 (0.04) **
Published books	0.45 (1.78)	-0.20 (0.31)	-0.19 (0.29)
Intercept		3.01 (2.27)	4.22 (2.02) *
Num. obs.	82	167	249
***p<0.001, **p<0.01, *p<0.05, +p<0.1 for two-tailed tests. Estimates are logit coefficients. Standard errors clustered by researcher are in parentheses.			