

Measuring the effectiveness of scientific gatekeeping

Kyle Siler^{a,1}, Kirby Lee^b, and Lisa Bero^c

^aDepartment of Strategic Management, Rotman School of Management, University of Toronto, Toronto, ON, Canada M5S 3E6; ^bDepartment of Clinical Pharmacy, University of California, San Francisco, CA 94143; and ^cFaculty of Pharmacy and Charles Perkins Centre, University of Sydney, Sydney, NSW 2006, Australia

Edited by Susan T. Fiske, Princeton University, Princeton, NJ, and approved November 18, 2014 (received for review September 21, 2014)

Peer review is the main institution responsible for the evaluation and gestation of scientific research. Although peer review is widely seen as vital to scientific evaluation, anecdotal evidence abounds of gatekeeping mistakes in leading journals, such as rejecting seminal contributions or accepting mediocre submissions. Systematic evidence regarding the effectiveness—or lack thereof—of scientific gatekeeping is scant, largely because access to rejected manuscripts from journals is rarely available. Using a dataset of 1,008 manuscripts submitted to three elite medical journals, we show differences in citation outcomes for articles that received different appraisals from editors and peer reviewers. Among rejected articles, desk-rejected manuscripts, deemed as unworthy of peer review by editors, received fewer citations than those sent for peer review. Among both rejected and accepted articles, manuscripts with lower scores from peer reviewers received relatively fewer citations when they were eventually published. However, hindsight reveals numerous questionable gatekeeping decisions. Of the 808 eventually published articles in our dataset, our three focal journals rejected many highly cited manuscripts, including the 14 most popular; roughly the top 2 percent. Of those 14 articles, 12 were desk-rejected. This finding raises concerns regarding whether peer review is ill-suited to recognize and gestate the most impactful ideas and research. Despite this finding, results show that in our case studies, on the whole, there was value added in peer review. Editors and peer reviewers generally—but not always—made good decisions regarding the identification and promotion of quality in scientific manuscripts.

peer review | innovation | decision making | publishing | creativity

Peer review alters science via the filtering out of rejected manuscripts and the revision of eventually published articles. Publication in leading journals is linked to professional rewards in science, which influences the choices scientists make with their work (1). Although peer review is widely cited as central to academic evaluation (2, 3), numerous scholars have expressed concern about the effectiveness of peer review, particularly regarding the tendency to protect the scientific status quo and suppress innovative findings (4, 5). Others have focused on errors of omission in peer review, offering anecdotes of seminal scientific innovations that faced emphatic rejections from high-status gatekeepers and journals before eventually achieving publication and positive regard (6–8). Unfortunately, systematic study of peer review is difficult, largely because of the sensitive and confidential nature of the subject matter. Based on a dataset of 1,008 manuscripts submitted to three leading medical journals—*Annals of Internal Medicine*, *British Medical Journal*, and *The Lancet*—we analyzed the effectiveness of peer review. In our dataset, 946 submissions were rejected and 62 were accepted. Among the rejections, we identified 757 manuscripts eventually published in another venue. The main focus of our research is to examine the degree to which editors and peer reviewers made decisions and appraisals that promoted manuscripts that would receive the most citations over time, regardless of where they were published.

Materials and Methods

To analyze the effectiveness of peer review, we compared the fates of accepted and rejected—but eventually published—manuscripts initially submitted to three leading medical journals in 2003 and 2004, all ranked in the top 10 journals in the Institute for Scientific Information Science Citation Index. These journals are *Annals of Internal Medicine*, *British Medical Journal*, and *The Lancet*. In particular, we examined how many citations published articles eventually garnered, whether they were published in one of our three focal journals or rejected and eventually published in another journal. To gauge postpublication impact and scientific influence, citation counts as of April 2014 were culled from Google Scholar (see Figs. S1 and S2 and SI Appendix for citation comparisons of major scholarly databases). We also examined the logarithms of those counts because citations in academia tend to be distributed exponentially, with a few articles garnering a disproportionate number of citations (9, 10). In our sample, 62 manuscripts were accepted of 1,008 submitted, yielding an overall 6.2% acceptance rate over that time period. Among rejected manuscripts, we identified 757 articles that were eventually published elsewhere after their rejection from our three focal journals. The remaining 189 rejected manuscripts (18.8%) were either altered beyond recognition when published or “file-drawer” by their authors. Eleven accepted manuscripts had missing or incomplete data, leaving a sample of 51 accepted manuscripts in our dataset.

Our study was approved by the Committee on Human Research at the University of California, San Francisco. As it is not possible to completely remove all identifiers from the raw data and protect the confidentiality of all the participating authors, reviewers, and editors, all the data (archival and taped) are stored securely and can be accessed only by the research team at the University of California, San Francisco. One journal required that the authors give permission to be part of the study. All authors from that journal subsequently granted permission. We wish to express our gratitude to those authors, as well as to the journal editors for sharing their data with us.

Citation analysis has long been used to analyze intellectual history and social behavior in science. It is important to consider the strengths and limitations of citation data in the context of our research. Scientists cite work

Significance

Peer review is an institution of enormous importance for the careers of scientists and the content of published science. The decisions of gatekeepers—editors and peer reviewers—legitimize scientific findings, distribute professional rewards, and influence future research. However, appropriate data to gauge the quality of gatekeeper decision-making in science has rarely been made publicly available. Our research tracks the popularity of rejected and accepted manuscripts at three elite medical journals. We found that editors and reviewers generally made good decisions regarding which manuscripts to promote and reject. However, many highly cited articles were surprisingly rejected. Our research suggests that evaluative strategies that increase the mean quality of published science may also increase the risk of rejecting unconventional or outstanding work.

Author contributions: K.S., K.L., and L.B. designed research; K.S., K.L., and L.B. performed research; K.S. analyzed data; and K.S. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

Data deposition: Our data are confidential and securely stored at the University of California, San Francisco.

¹To whom correspondence should be addressed. Email: ksiler@gmail.com.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1418218112/-DCSupplemental.

for a myriad of reasons (11, 12). However, the vast majority of citations are either positive or neutral in nature (13). We worked with the assumption that scientists prefer to build upon other quality research with their own work. As Latour and Woolgar (14) suggested, citation is an act of deference, as well as the means by which intellectual credit and content flows in science. Relatedly, we also assumed that most scientists want to produce quality work and will seldom attempt to garner credit and attention by blatantly doing bad work. Thus, on the whole, the attention and impact associated with citations provides a reasonable measure of quality. Citations provide an objective and quantitative measure of credit and attention flows in science.

Results

Our results suggest that gatekeepers were at least somewhat effective at recognizing and promoting quality. The main gatekeeping filters we identified were (i) editorial decisions regarding which manuscripts to desk-reject and (ii) reviewer scores for manuscripts sent for peer review. Desk-rejections are manuscripts that an editor decides not to send for peer review after an initial evaluation. This choice entails no further journal or personal resources being expended on gestating or considering the article, although the lack of peer review means that the authors will be free to submit their article elsewhere relatively quickly. An article sent for peer review can still be rejected as well. Reviewers, usually anonymous scholars with relevant expertise, provide feedback to authors regarding their manuscripts, which journal editors use to decide whether to reject, recommend revisions (seldom with a guarantee of eventual publication), or accept an article for publication.

Merton (1) posited that science tends to reward high-status academics merely by virtue of their previously attained status, dubbing this self-fulfilling prophecy the “Matthew Effect.” Examining rejected and accepted manuscripts separately helps rule out potential Matthew Effects affecting citation outcomes, because any citation discrepancies cannot be explained by the halo or reputational effects of being published in one of our three elite focal journals.

Rejected Manuscripts. Generally, the journal editors in our study made good appraisals regarding which articles to desk-reject. Desk-rejected articles eventually published in other journals received fewer citations than those that went through at least one round of peer review before rejection. Of 1,008 articles in our dataset, 772 were desk-rejected by at least one of our focal-journals. Eventually published desk-rejected articles ($n = 571$) received on average 69.80 citations, compared with 94.65 for articles sent for peer review before rejection ($n = 187$; $P < 0.01$).

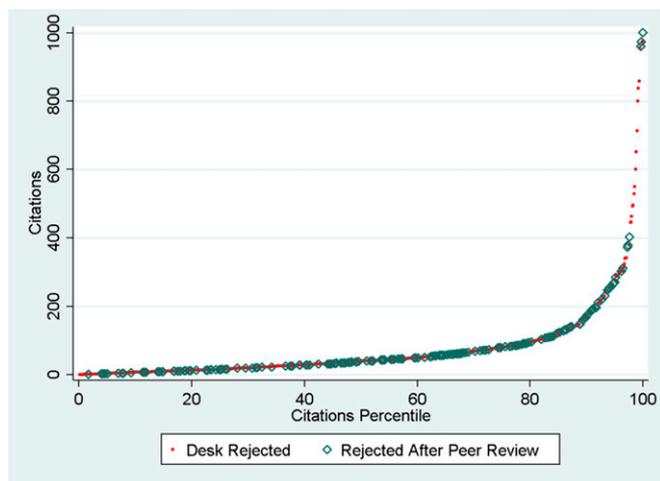


Fig. 1. Citation distribution of rejected articles (peer reviewed vs. desk-rejected).

Because citations are often distributed exponentially, with a few articles garnering disproportionate attention, we also used the logarithm of citation counts as a dependent variable to diminish the potential influence of a few highly cited outlier articles. Logging citations yields similar results, with desk-rejections receiving a mean of 3.44 logged citations, compared with 3.92 for peer-reviewed rejections ($P < 0.001$). Fig. 1 shows a graphical illustration of desk-rejections and peer review rejections. In general, articles chosen for peer review tended to receive significantly more citations than desk-rejected manuscripts. However, a number of highly cited articles were desk-rejected, including 12 of the 15 most-cited cases. Of the other 993 initially submitted manuscripts, 760 (76.5%) were desk-rejected. This finding suggests that in our case study, articles that would eventually become highly cited were roughly equally likely to be desk-rejected as a random submission. In turn, although desk-rejections were effective with identifying impactful research in general, they were not effective in regards to identifying highly cited articles.

Articles sent for peer review may have benefited from receiving feedback from attentive reviewers. However, the magnitude of the difference between the desk-rejected and nondesk-rejected articles, as well as the fact that 12 of the most highly cited articles were desk-rejected and received little feedback, suggests that innate quality of initial submissions explains at least some of the citation gap. Further, if a future highly cited manuscript was aided by critical feedback attached to a “reject” decision from a journal, it renders the decision not to at least grant an opportunity at revision an even more egregious mistake. Despite the importance of peer review, its scope of influence with regard to changing and gestating articles may be limited. The core innovations and research content of a scientific manuscript are rarely altered substantially through peer review. Instead, peer review tends to focus on narrower, more technical details of a manuscript (15). However, there is some evidence that peer review improves scientific articles, particularly in the medical sciences (16).

Peer reviewers also appeared to add value to peer review with regards to the promotion and identification of quality. We assigned reviewer scores to quantify the perceived initial quality of each submitted manuscript. For each “accept” recommendation, three points were assigned, two for “minor changes,” one for “major changes,” and zero for reject. From these values, each manuscript received a mean score. Whereas some initially lauded manuscripts may have improved little over the peer review process, and other marginal initial submissions may have improved greatly, initial perceived quality was related to citation outcomes. For manuscripts with two or more reviewers, there were weak but positive correlations between initial scores and eventual citations received. Reviewer scores were correlated 0.28 ($P < 0.01$; $n = 89$) with citations and 0.21 with logged citations ($P < 0.05$). Although the effects of the peer-reviewer scores on citation outcomes appeared weaker than those of editors making desk-rejections, it is worth noting the survivor bias of manuscripts that are not desk-rejected. It is generally easier to distinguish scientific work of very low quality than it is to recognize finer gradations distinguishing the best contributions in advance of publication (17, 18).

Related to the Matthew Effect (1) and underscoring the importance of social status in science, evaluators have been found to judge equivalent work from high-status sources more favorably than work from lower-status contributors (11, 19, 20). Consequently, placement in a prestigious journal can bolster the visibility and perceived quality of work of many scientists. Evaluating complex work is difficult, so scientists often rely on heuristics to judge quality; status of scholars, institutions, and journals are common means of doing so (21, 22). Unsurprisingly, citations received by manuscripts were positively correlated with the impact factor of the journal in which it was eventually published. Journal impact factor was correlated 0.54

with citations ($P < 0.001$; $n = 757$) and 0.42 with logged citations ($P < 0.001$). Of course, it is difficult to disentangle exactly how much these positive correlations were a result of (i) higher-quality articles being published in competitive high impact journals and (ii) visibility or halo effects associated with publishing in more prestigious journals.

Accepted Manuscripts. Next, we examined manuscripts that survived the peer review process and were eventually published in one of our three target journals. Among the 40 articles that were scored by at least two peer reviewers, there were weak positive correlations between reviewer scores and citations received (0.21; $n = 40$) and citations logged (0.26). Both of these correlations fell short of statistical significance, in part because of the small sample. As another way of gauging the ability of peer reviewers to identify quality manuscripts, we compared submissions that had at least one “reject” recommendation from a reviewer with those that never received such a recommendation. Many have bemoaned the practice at most highly competitive academic journals of rejecting an article that does not achieve a positive consensus among all reviewers (23). After all, some rejection recommendations are unwise. Among accepted articles, rejection recommendations from reviewers were associated with substantial differences in citation outcomes. Published articles that had not received a rejection recommendation from a peer reviewer ($n = 30$) received a mean of 162.80 citations, compared with 115.24 for articles that received at least one rejection recommendation ($n = 21$; $P < 0.10$). Similarly, manuscripts without rejection recommendations received a mean of 4.72 logged citations, compared with 4.33 logged citations for articles for which a peer reviewer had recommended rejection ($P < 0.10$). Results approached significance, suggesting that among accepted manuscripts, rejection recommendations—or lack thereof—were at best weakly predictive of future popularity.

If anything, the differences in citation outcomes for rejected and accepted articles with more initially favorable peer reviews likely underestimates the effectiveness of scientific gatekeeping. Assuming eventually published manuscripts were at least not made worse in subsequent submissions after initial rejection, and in some cases improved, this should mitigate quality discrepancies between articles that received more and less favorable initial assessments.

Fates of Rejected vs. Accepted Articles. Although our results suggest that gatekeeping at our three focal journals was effective, it was far from perfect. When examining the entire population of 808 eventually published manuscripts, our three focal journals rejected the 14 most highly cited articles. This entails 15 total cases because one article was rejected by two of our focal journals before eventual publication. Most of the 14 most-cited articles were published in lower-impact journals than the initial target journal. In some cases, the impact factor of the publishing journal was substantially lower than the initial target journal from which the manuscript was rejected. The “best” acceptance decision one of our focal journals made was publishing the 16th most-cited case, which placed in the 98th percentile of submitted manuscripts. Despite the 15 glaring omissions at the top, on the whole gatekeepers appeared to make good decisions. Citation percentiles for accepted articles ranged from a minimum of 17.06 to a maximum of 98.15. The median percentile was 79.36, with quartiles at 53.28 (25th percentile) and 91.22 (75th percentile). Fig. 2 provides a graphical illustration of the distribution of citations for accepted and rejected articles. As is generally the case in science, citations are distributed exponentially. Near the 85th percentile, citations increase sharply. Whether this does—or should—influence the evaluative strategies of gatekeepers is an open question. These results also raise the question of to what degree these positive outcomes for accepted submissions, particularly

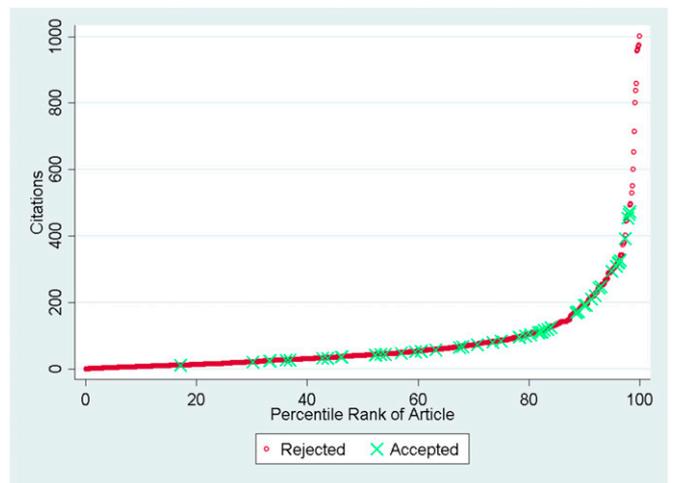


Fig. 2. Citation distribution of accepted and rejected articles.

the dearth of barely or never-cited articles, are because the prestige of the journal and/or the innate quality of the manuscripts.

To ensure that these findings were not excessively influenced by the majority of articles published in less-eminent journals, we repeated the previous analysis, restricting articles to those published in journals with at least an 8.00 impact factor. The correlation between impact factor and citations received remained positive, but was roughly halved to 0.28. Interestingly, after running t tests comparing rejected and accepted manuscripts for articles published in high-impact journals, rejected articles received more citations. Specifically, rejected articles averaged 212.77 citations and accepted manuscripts averaged 143.22 ($P < 0.05$). The relationship is similar with logged citations (4.77 for rejections, 4.53 for acceptances), although it falls short of statistical significance, suggesting that it is a few highly cited articles underpinning the significant difference. Multiple regression analysis, allowing for impact factor to be used as a control variable, yielded similar results (Tables S1 and S2). Controlling for impact factor shows that surprisingly, rejected manuscripts were more cited than accepted manuscripts when published in prestigious journals.

By restricting analysis to journals with impact factors greater than 8.00, we are more likely to be cherry-picking gatekeeping mistakes and ignoring the vast cache of articles that were “rightfully” sent down the journal hierarchy. In addition to a status symbol, journal impact factors represent positive, but often noisy signals of manuscript quality. Authorial ambition may also be relevant; there was likely some self-selection in our manuscript pool. We looked solely at articles initially submitted to elite journals, as opposed to manuscripts that the authors targeted for a less eminent journal in the first place. In some cases, authors know their articles better than peer reviewers, who possess disparate sources of expertise. Further, ambition can portend success. For example, Dale and Krueger (24) found that future earnings were better predicted by the highest-status college to which a student applied, as opposed to attended. Low acceptance rates also likely create a substantial pool of high-quality articles that would fit well in other high-quality journals. Alberts et al. expressed concern that very high rejection rates tend to squeeze out innovative research (25). High-status journals with very low acceptance rates tend to emphasize avoiding errors of commission (publishing an unworthy article) over avoiding errors of omission (rejecting a worthy article).

Explaining and Justifying Rejection. Most articles received similar boilerplate text in their rejection letters. In the anomalous case of the #3 ranked article, which received uniformly favorable

feedback, the standard rejection letter was modified slightly to acknowledge that the article had received positive reviews but was being rejected anyway. The standard editorial response sent to many rejected authors emphasizes a need for “strong implications” and a need to “change practice.” These correspondences reveal the importance of what Gieryn (26) dubbed “boundary work” in demarcating worthy versus unworthy science. Some rejected submissions may have been excellent articles that did not fit with the perceived mission or image of the journal. In most cases, gatekeepers acknowledged that the article they were rejecting had at least some merit, but belonged in a different, usually specialist, journal. Normal science and replication are more likely to occur in such journals. In contrast, generalist journals tend to assume, or at least perceive, a different role in science, particularly when the publication is high-status. Since professional power and status tend to be linked to control over abstract knowledge (27, 28), the emphasis on general implications is not surprising. Although accruing citations is valuable for authors and journals, image is also important. High-status actors and institutions often strategically forego valuable, but lower-status, downmarket niches to preserve exclusive reputations and identities (29).

Based on the feedback from editors and peer reviewers reported in *SI Appendix*, Table 1 summarizes the most commonly stated justifications for rejection.

Gatekeepers prioritized novelty—at least as they perceived it—in their adjudications. High-status journals distinguish themselves by publishing cutting-edge science of theoretical importance (2, 30), so the emphasis on novelty is unsurprising. Further, because highly cited articles are expected to be novel and paradigm-shifting (5, 31), it is surprising that almost half of the top 15 cases were criticized for their lack of novelty. However, perceived novelty may only be relative. George Akerlof received the 2001 Nobel Prize in economics for his advances in behavioral economics. This work is well-known as a source of one of the most (in)famous rejection mistakes in contemporary scientific publishing. Akerlof’s seminal article, “The Market for ‘Lemons,’” initially faced stiff resistance—if not outright hostility—in peer review at elite economics journals. The first two journals where the article was submitted rejected it on the basis of triviality. The third journal rejected the article because it was “too” novel, and if the article was correct, it would mean that “economics would be different” (32). Relatedly, former *Administrative Science Quarterly* editor William Starbuck (33) observed that gatekeepers are often territorial and dogmatic in their criticisms, criticizing methods and data attached to theories they dislike and lauding methods and data attached to theories they prefer.

Our results suggest that Akerlof’s experience facing rejections with an influential article may not be abnormal. Mark Granovetter’s “The Strength of Weak Ties,” the most-cited article in contemporary sociology, was emphatically rejected by the first journal where he submitted the manuscript (34). Rosalyn Yalow received the Nobel Prize in medicine for work that was initially rejected by *Science* and only published in a subsequent journal after substantial compromise (6). Outside of science, at the

start of her career, famed author J. K. Rowling experienced 12 rejections of her first *Harry Potter* book before eventually finding a publisher willing to take a chance on her book for a small monetary advance (35). Highly cited articles and innovations experiencing rejection may be more the rule than the exception.

Decisions to reject or accept manuscripts can be complex. Multiple characteristics of articles and authors—and not just novelty—are associated with publication. Lee et al. (22) listed numerous potential sources of bias in peer review, including social characteristics of authors, as well as the intellectual content of their scientific work. The tendency of gatekeepers to prefer work closer to their own and the scientific status quo is a source of intellectual conservatism in science (5). Previous analysis of this dataset found that submitted manuscripts were more likely to be published if they have high methodological quality, larger sample sizes, a randomized, controlled design, disclosed funding sources, and if the corresponding author lives in the same country as that of the publishing journal (36). Scientists have expressed concern that this publication bias selectively distorts the corpus of published science and contributes to the publication of dubious results (37, 38).

Impact Factors of Submitted vs. Published Journals. Among the 14 most highly cited articles, most manuscripts were eventually published in a lower impact factor journal than the initial focal journal to which they were submitted. Table 2 reports the ratios of impact factors of initially submitted journal to published journal. Impact factor ratios are reported, as opposed to exact impact factors, as a precaution to preserve the anonymity of authors and their manuscripts. A higher ratio denotes a decrease in journal status in publication source, because the numerator (initially submitted journal) is greater than the denominator (published journal), whereas a lower ratio denotes an increase in status. Most articles ended up moving down the status hierarchy, because there are few journals in academia with impact factors higher than our three focal journals and authors tend to submit manuscripts to less-selective journals following rejection (39).

Time to Publication and Received Citations. We also examined the relationship between time to publication (specifically, time since initial submission to our three focal journals) and citations received. The correlation between time to publications and citations received was only -0.18 . However, Fig. 3 shows that there appears to be a “wall” after about 500 d for highly cited articles. After roughly 500 d, no manuscript received more than 200 citations. In fact, this weak negative correlation is driven by a small number of highly cited articles published in the first 500 d since initial submission. This finding suggests that if editors and peer reviewers made a mistake in overlooking or rejecting a future seminal contribution, the peer review system as a whole appears to be generally able to rectify the mistake within 2 y. If one of our three focal journals happened to reject an excellent article—out of error or perhaps an unlucky choice of peer reviewers—such articles appear to be published elsewhere relatively quickly. Of course, it is impossible to observe counterfactuals for meritorious rejected articles that were scooped, severely revised, never published, or consigned to an obscure publication outlet. Alternatively, in a fast-moving field such as medicine, results not published quickly are prone to becoming scooped or stale, and thus less likely to become highly cited. Repeated time-consuming rejections at other journals are also likely an indicator of the true quality of a manuscript, as authors tend to send an article further down the journal status hierarchy after each rejection (39).

Among accepted articles that we could track from submission to publication, the number of rounds of revision a manuscript endured before publication had no major effect on citation outcomes among accepted manuscripts. Increased rounds of

Table 1. Most common justifications for article rejection among top 15-cited cases

Justification	<i>n</i>
Lacking novelty	7
Methodological problems	4
Magnitude of results too small	4
No reason given	3
Insufficient data/evidence	2
Speculative results/questionable validity	2

Table 2. Impact factor comparisons of submitted versus published journals, 14 most highly cited articles

Citation rank	Impact factor ratio of submitted to published journal	Publication journal outcome
1	0.85	Higher
2 (first rejection)	0.18	Higher
2 (second rejection)	0.46	Higher
3	0.85	Higher
4	1.41	Lower
5	1.39	Lower
6	3.02	Lower
7	1.11	Lower
8	0.96	Higher
9	2.92	Lower
10	12.14	Lower
11	1.49	Lower
12	6.96	Lower
13	2.57	Lower
14	6.79	Lower

revision were weakly negatively correlated with citations (-0.13) and citations logged (-0.12) ($n = 40$). This suggests that there were diminishing returns to multiple rounds of revision. The function of numerous revisions in scientific journals may be more evaluative and bureaucratic than gestational.

Discussion

Our research suggests that in our case studies, both editors and peer reviewers made generally, but not always, good decisions with regards to identifying which articles would receive more citations in the future. These results are inured to potential Matthew Effects, as comparisons among rejected manuscripts showed that better-rated rejected articles still fared better than those that were evaluated more negatively. The rejection of the 14 most-cited articles in our dataset also suggests that scientific gatekeeping may have problems with dealing with exceptional or unconventional submissions. The fact that 12 of the 14 most-cited articles were desk-rejected is also significant. Rapid decisions are more likely to be informed by heuristic thinking (40). In turn, research that is categorizable into existing research frames is more likely to appeal to risk-averse gatekeepers with time and resource constraints, because people generally find uncertainty to be an aversive state (41). This may serve as an explanation for why exceptional and uncommon work may be particularly vulnerable to rejection.

Many of the problems editors and reviewers identified with the rejected articles were potentially fixable. However, journals have finite time and resources. There is no guarantee authors will be able to execute a successful revision. Some rejected manuscripts might have been casualties of a 6% acceptance rate, as opposed to inherently flawed or deficient. On the other hand, some rejected articles may not have fit well with the scope and mission of the initial focal journal, or the manuscript may have been improved since the initial submission. A focus on elite, cutting-edge research germane to high-status scientists may also open niches for popular contributions outside of those particular missions and constituencies, particularly dealing with practice and application.

It is also important to consider that citations are not necessarily a perfect indicator of scientific merit. Contagion processes are often decoupled from quality. Since the utility of information is often dependent on widespread adoption by others (42), diffusion patterns often have arbitrary influences and suboptimal innovations can gain precedence (43). Extreme performance can be more indicative of volatility than merit (44). Further, some

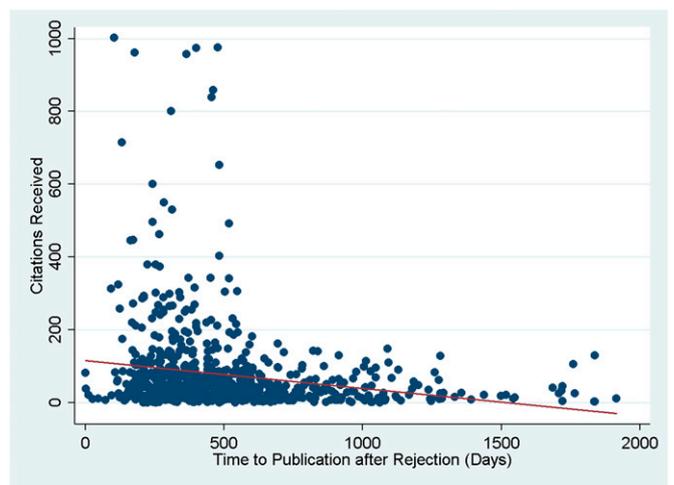


Fig. 3. Citation distribution of rejected articles by time to publication.

innovations with mass appeal may have greater diffusion potential appealing to lowest common denominators, while more complex or esoteric contributions can be harder for many to understand or cite. Status signals and other social factors can profoundly affect diffusion processes, as well as people's opinions and perceptions of quality (45). This is a key mechanism underpinning the Matthew Effect in science. After a scientific paper achieves widespread adoption, it is usually cited—often vaguely or inaccurately—as an exemplar of a general idea, as opposed to for the quality of its research (12). Impact factors, citations, intellectual credit, and visibility are the main means by which scholars and journals accrue status (14), so it is in the interest of editors to generally select highly cited articles for acceptance. Market forces also shape scientific incentives and journal status hierarchies, as journal revenues are linked to impact factors, even though using quantitative metrics to gauge the quality of a journal is controversial (46).

Prediction is a difficult task in most complex systems (47). This is particularly true in science, where cutting-edge research is inherently uncertain (30). Like with professional sports (48), financial investing (49), fine art (50), venture capitalism (51), books, movies, and music (52), identifying quality and predicting future performance with science and knowledge appears to be rife with noise and randomness. Former *American Economic Review* editor Preston McAfee (53) argued that gatekeeping errors in science are not only inevitable, they can also be efficient. Journals that do not occasionally wrongly reject unconventional submissions will publish articles of lower average quality on the whole. Gans and Shepherd (7) chronicled numerous gatekeeping mistakes in economics, resulting in widespread concerns among journal editors of rejecting another future classic article (53). However, changing evaluative cultures in an attempt to accommodate potentially exceptional articles may be costly. Dilemmas often exist between strategies that yield lower means but higher extreme values, and those with higher means and less variance (54). Because most new ideas tend to be bad ideas (55), resisting unconventional contributions may be a reasonable and efficient default instinct for evaluators. However, this is potentially problematic because unconventional work is often the source of major scientific breakthroughs (5). In turn, it worth considering whether the high mean and median citation rates for articles accepted in our three focal journals could have come at the expense of rejecting the some of the most highly cited articles. More generally, this finding raises the normative and strategic question of whether editors should try to maximize the mean or median quality of publications, even if such gatekeeping strategies or evaluative cultures entail a higher risk of rejecting highly cited articles. The

medical context in our study is relevant to the question of whether and to what degree science should be risk-averse. Much decision-making in medicine is influenced by the Hippocratic Oath, emphasizing the imperative to avoid harm above all else. This underscores the need for follow-up studies of peer review in different journals and disciplines to enable replication and comparison with different epistemic and scientific cultures.

Although editors and reviewers appear to be able to improve predictions of which articles will receive the most citations—if not while also improving scientific output via selection and gestation—both errors of omission and commission were prominent in our case studies. Lindsey (17) posited that among the highest quality, most competitive manuscripts at elite high-rejection journals, peer review is “little better than a dice roll.” Cole et al. (56) argued that roughly one-half of a competitive National Science Foundation granting process was a matter of luck. Even with effective peer

review, occasional acceptances of mediocre contributions and rejections of excellent manuscripts may be inevitable (33, 57). Given the effectiveness of desk-rejections as shown in our results, it seems that distinguishing poor scientific work is easier than distinguishing excellent contributions. However, the high rate of desk-rejections among extremely highly cited articles in our case studies suggests that although peer review was effective at predicting “good” articles, it simultaneously had difficulties in identifying outstanding or breakthrough work. In turn, the complexity of science appears to limit the predictive abilities of even the best peer reviewers and editors.

ACKNOWLEDGMENTS. We thank Aaron Panofsky for helping initiate this project. Our research was supported by Grant 5R01NS044500-02 from the Research Integrity Program, Office of Research Integrity/National Institutes of Health collaboration, as well as a Banting Postdoctoral Fellowship and Sloan Foundation Economics of Knowledge Contribution Project fellowship (to K.S.).

- Merton RK (1968) The Matthew Effect in science. *Science* 159(3810):56–63.
- Lamont M (2009) *How Professors Think: Inside the Curious World of Academic Judgment* (Harvard Univ Press, Cambridge, MA).
- Kassirer JP, Campion EW (1994) Peer review. Crude and understudied, but indispensable. *JAMA* 272(2):96–97.
- Mahoney MJ (1977) Publication prejudices: An experimental study of confirmatory bias in the peer review system. *Cognit Ther Res* 1(2):161–175.
- Horrobin DF (1990) The philosophical basis of peer review and the suppression of innovation. *JAMA* 263(10):1438–1441.
- Yalow RS (1977) Nobel Lecture. *Radioimmunoassay: A Probe For Fine Structure Of Biologic Systems*. December 8, 1977 at Stockholm Concert Hall, Stockholm, Sweden.
- Gans JS, Shepherd GB (1994) How are the mighty fallen: Rejected classic articles by leading economists. *J Econ Perspect* 8(1):165–179.
- Nature (2003) Coping with peer rejection. *Nature* 425(6959):645.
- Lotka AJ (1926) The frequency distribution of scientific productivity. *J Wash Acad Sci* 16:317–323.
- Barabási AL, Albert R (1999) Emergence of scaling in random networks. *Science* 286(5439):509–512.
- Lynn FB (2014) Diffusing through disciplines: Insiders, outsiders and socially influenced citation behavior. *Soc Forces* 93(1):355–382.
- Hargens LL (2000) Using the literature: Reference networks, reference contexts, and the social structure of scholarship. *Am Sociol Rev* 65(6):846–865.
- Bormmann L, Daniel H-D (2008) What do citation counts measure: A review of studies on citing behavior. *J Doc* 64(1):45–80.
- Latour B, Woolgar S (1979) *Laboratory Life: The Construction of Scientific Facts* (Sage, Los Angeles).
- Ellison G (2002) Evolving standards for academic publishing: A q-r theory. *J Polit Econ* 110(5):994–1034.
- Goodman SN, Berlin J, Fletcher SW (1994) Manuscript quality before and after peer review and editing at *Annals of Internal Medicine*. *Annals of Internal Medicine* 121(1):11–21.
- Lindsey D (1988) Assessing precision in the manuscript review process: A little better than a dice roll. *Scientometrics* 14(1-2):75–82.
- Cicchetti DV (1991) The reliability of peer review for manuscript and grant submissions: A cross-disciplinary investigation. *Behav Brain Sci* 14(1):119–135.
- Peters DP, Ceci SJ (1982) Peer review practices of psychological journals: The fate of published articles, submitted again. *Behav Brain Sci* 5(2):187–195.
- Leahy E (2004) The role of status in evaluating research: The case of data editing. *Soc Sci Res* 33(3):521–537.
- Long JS, Fox MF (1995) Scientific careers: Universalism and particularism. *Annu Rev Sociol* 24:45–71.
- Lee CJ, Sugimoto CR, Zhang G, Cronin B (2013) Bias in peer review. *JASIST* 64(1):2–17.
- Armstrong JS, Hubbard R (1991) Does the need for agreement among reviewers inhibit the publication of controversial findings? *Behav Brain Sci* 14(1):136–137.
- Dale SB, Krueger AB (2002) Estimating the payoff to attending a more selective college: An application of selection on observables and unobservables. *Q J Econ* 117(4):1491–1527.
- Alberts B, Kirschner MW, Tilghman S, Varmus H (2014) Rescuing US biomedical research from its systemic flaws. *Proc Natl Acad Sci USA* 111(16):5773–5777.
- Gieryn TF (1983) Boundary-work and the demarcation of science. *Am Sociol Rev* 48(6):781–795.
- Abbott AD (1981) Status and status strain in the professions. *Am J Sociol* 86(4):819–835.
- Smith PK, Trope Y (2006) You focus on the forest when you're in charge of the trees: Power priming and abstract information processing. *J Pers Soc Psychol* 90(4):578–596.
- Podolny JM (2005) *Status Signals: A Sociological Study of Market Competition* (Princeton Univ Press, Princeton, NJ).
- Cole S (1983) The hierarchy of the sciences? *Am J Sociol* 89(1):111–139.
- Kuhn TS (1962) *The Structure of Scientific Revolutions* (Univ of Chicago Press, Chicago).
- Akerlof GA (2003) Writing “The Market for ‘Lemons’”: A Personal and Interpretive Essay. Available at www.nobelprize.org/nobel_prizes/economic-sciences/laureates/2001/akerlof-article.html. Accessed June 2, 2014.
- Starbuck WH (2003) Turning lemons into lemonade: Where is the value in peer reviews? *J Manage Inq* 12(4):344–351.
- Available at scatter.files.wordpress.com/2014/10/granovetter-rejection.pdf. Accessed October 23, 2014.
- McGinty S (June 16, 2003) The J.K. Rowling Story. *The Scotsman*. Available at www.scotsman.com/lifestyle/books/the-jk-rowling-story-1-652114. Accessed December 9, 2014.
- Lee KP, Boyd EA, Holroyd-Leduc JM, Bacchetti P, Bero LA (2006) Predictors of publication: Characteristics of submitted manuscripts associated with acceptance at major biomedical journals. *Med J Aust* 184(12):621–626.
- Dickersin K (1990) The existence of publication bias and risk factors for its occurrence. *JAMA* 263(10):1385–1389.
- Ioannidis JPA (2005) Why most published research findings are false. *PLoS Medicine* 2(8):0696–0701.
- Calcagno VE, et al. (2012) Flows of research manuscripts among scientific journals reveal hidden submission patterns. *Science* 388(6110):1065–1069.
- Kahneman D (2011) *Thinking Fast and Slow* (Farrar, Straus and Giroux, New York).
- Mueller JS, Melwani S, Goncalo JA (2012) The bias against creativity: Why people desire but reject creative ideas. *Psychol Sci* 23(1):13–17.
- Granovetter M, Soong R (1986) Threshold models of interpersonal effects in consumer demand. *J Econ Behav Organ* 7:83–99.
- Arthur WB (1989) Competing technologies, increasing returns and lock-in by historical events. *The Economic Journal* 99:116–131.
- Denrell J, Liu C (2012) Top performers are not the most impressive when extreme performance indicates unreliability. *Proc Natl Acad Sci USA* 109(24):9331–9336.
- Salganik MJ, Dodds PS, Watts DJ (2006) Experimental study of inequality and unpredictability in an artificial cultural market. *Science* 311(5762):854–856.
- Seglen PO (1997) Why the impact factor of journals should not be used for evaluating research. *BMJ* 314(7079):498–502.
- Silver N (2012) *The Signal and the Noise* (Penguin, New York).
- Massey C, Thaler RH (2013) The loser's curse: Decision making and market efficiency in the National Football League draft. *Manage Sci* 59(7):1479–1495.
- Mauboussin MJ (2012) *The Success Equation* (Harvard Business Review Press, Boston).
- Dempster AM, ed (2014) *Risk and Uncertainty in the Art World* (Bloomsbury, London).
- Gompers P, Lerner J (2006) *The Venture Capital Cycle* (MIT Press, Cambridge, MA).
- Hirsch PM (1972) Processing fads and fashions: An organization-set analysis of cultural industry systems. *Am J Sociol* 77(4):639–659.
- McAfee RP (2014) Edifying editing. *Secrets of Economics Editors*, eds Szenberg M, Ramrattan L (MIT Press, Cambridge, MA), pp 33–44.
- Denrell J, March JG (2001) Adaptation as information restriction: The hot stove effect. *Organ Sci* 12(5):523–538.
- Levinthal DA, March JG (1993) The myopia of learning. *Strateg Manage J* 14(52):95–112.
- Cole S, Cole JR, Simon GA (1981) Chance and consensus in peer review. *Science* 214(4523):881–886.
- Stinchcombe AL, Ofshe R (1969) On journal editing as a probabilistic process. *Am Sociol* 4(2):116–117.