184 Research Grants NZet-alNov-19

NEWS IN *NATURE*

20 NOVEMBER 2019

Science funders gamble on grant lotteries

A growing number of research agencies are assigning money randomly.

Albert Einstein famously insisted that God does not play dice. But the Health Research Council of New Zealand does. The agency is one of a growing number of funders that award grants partly through random selection. Earlier this year, for example, David Ackerley, a biologist at Victoria University of Wellington, received NZ$150,000 (US$96,000) to develop new ways to eliminate cells — after his number came up in the council's annual lottery.

"We didn't think the traditional process was appropriate," says Lucy Pomeroy, the senior research investment manager for the fund, which began its lottery in 2015. The council was launching a new type of grant, she says, which aimed to fund transformative research, so wanted to try something new to encourage fresh ideas.

Traditionalists beware: the forces of randomness in research are, if not quite on the march, then certainly plotting their next move. At a meeting at the University of Zurich in Switzerland on 19 November, supporters of the approach argued that blind chance should have a greater role in the scientific system. And they have more than just grant applications in their sights. They say lotteries could be used to help select which papers to publish — and even which candidates to appoint to academic jobs.

Luck of the draw

"Random chance will create more openness to ideas that are not in the mainstream," says Margit Osterloh, an economist at the University of Zurich who studies research governance and organized the meeting, which was intended to promote the idea among academics. She says that existing selection processes are inefficient. Scientists have to prepare lengthy applications, many of which are never funded, and assessment panels spend most of their time sorting out the specific order in which to place mid-ranking ideas. Low- and high-quality applications are easy to rank, she says. "But most applications are in the midfield, which is very big." Most importantly, she argues, standard assessments don't perform as well as policymakers, publishers and university officials assume. "Referees and all kinds of evaluation bodies do not have really good working criteria."

The Swiss National Science Foundation (SNSF) is the latest funder to experiment with random selection. Earlier this year, it asked assessment panels to draw lots to help decide which early-career scientists should receive postdoctoral fellowships. It is now evaluating the scheme and SNSF president Matthias Egger spoke about it at the Zurich meeting. Other programmes that rely on lottery systems to award some grant types include another New Zealand government fund called the Science for Technological Innovation National Science Challenge (SfTI), which introduced random selection in 2015. Germany's largest private funding agency, the Volkswagen Foundation in

Hannover, has also used lotteries to allocate some of its Experiment! grants since 2017.

'We actually do have a hat'

The process is not entirely random. Typically, funders screen applications to ensure they meet a minimum standard, then projects are given numbers and selected at random by a computer until all of the cash has been allocated.

"It just takes a lot of angst out of it," says Don Cleland, a process engineer at Massey University in Palmerston North, New Zealand, and a member of the team that oversees the SfTI fund. Given the money to fund 20 projects, an assessment panel doesn't need to agonize over which application ranks 20th and which comes 21st, he says. They can just agree that both are good enough to be funded and them put them into the hat. "We actually do have a hat," Cleland says.

The fund tells applicants how far they got in the process, and feedback from them has been positive, he says. "Those that got into the ballot and miss out don't feel as disappointed. They know they were good enough to get funded and take it as the luck of the draw."

The idea has some theoretical backing. A number of researchers have analysed various selection methods and suggested that incorporating randomness has advantages over the current system, such as reducing the bias that research routinely shows plagues grant-giving, and improving diversity among grantees[1].

The acceptance criteria for entering the lottery can be tweaked, for example, to give more weighting to scientists from minority ethnic backgrounds or to those who aren't backed by wealthy institutions. People from wealthy institutions or privileged backgrounds often have access to resources that help them to achieve success by standard metrics. And the conventional system tends to benefit them, says Cleland, because it focuses on candidates' track records rather than the strength of their ideas. "We want those with the best ideas to rise to the top."

Competitive arguments

Cleland argues that other funders should try it. But not everyone agrees. Despite benefitting from a grant lottery, Ackerley says he doesn't approve of them. "I spend a lot of time on grant-review panels and I like to think they do a reasonable job," he says. "I've done reasonably well out of competitive grants and I suppose the selfish reason is that I might not do so well out of a lottery system."

Because applications to funds that use lottery systems only need to satisfy basic criteria, they tend to be shorter. "I think there's a lot of value to writing a high-quality proposal," Ackerley says.

Osterloh, who recently triggered lively debate of her arguments in the pages of *Research Policy* after publishing them in the journal[2], says selection by random chance could have a wider benefit because those who benefit from lotteries do not feel so entitled. "If you know you have got a grant or a publication which is selected partly randomly then you will know very well you are not the king of the Universe, which makes you more humble," she says. "This is exactly what we need in science."

doi: 10.1038/d41586-019-03572-7

References

1. 1.

Fang, F. C. & Casadevall, A. *mBio* **7**, e00422-16 (2016).

    ○
    ▪ PubMed
    ▪ Google Scholar

2. 2.

Osterloh, M. & Frey, B. S. *Res. Policy* **49**, 103831 (2020).

    ○
    ▪ Article
    ▪ Google Scholar

Download references

https://www.nature.com/articles/d41586-019-03572-7?fbclid=IwAR2UnbAlOmV2SXLGvSRAMjAmFGaqWkT_SpwlMRq3qzOyww05aJyUu0m0Sqk

REF1 FENG

https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4959526/

mBio. 2016 Mar-Apr; 7(2): e00422-16.

Published online 2016 Apr 12. doi: 10.1128/mBio.00422-16

PMCID: PMC4959526

PMID: 27073093

Research Funding: the Case for a Modified Lottery

Ferric C. Fang, Editor in Chief, *Infection and Immunity*[a] and Arturo Casadevall, Founding Editor in Chief, *mBio*[b]

Author information Copyright and License information Disclaimer

This article has been corrected. See mBio. 2016 May 17; 7(3): e00694-16.

This article has been cited by other articles in PMC.

ABSTRACT

The lottery is in the business of selling people hope, and they do a great job of that.

—*John Oliver (1)*

Go to:

EDITORIAL

The American research establishment has been facing the most prolonged funding crisis in its history. After a doubling in funding at the turn of the 20th century, the budget of the National Institutes of Health (NIH) was flat from 2003 to 2015, translating into a 25% reduction in actual buying power after taking inflation and the increasing costs of research into account (2). Although the increased NIH support in the 2016 spending bill is welcome news (3), this does not alter long-term uncertainty regarding the federal commitment to scientific research. The research funding crisis has been paralleled by other problems in science, including concerns about the reliability of the scientific literature, demographic imbalances, and various antiscience campaigns that question evolutionary theory, the usefulness of vaccines, human impact on climate change, and even the occurrence of the moon landings. What is perhaps most remarkable in this time of crisis and change is how little scientific leaders and governmental officials have done to combat these trends. Although each of these problems merits its own essay, we focus here on the allocation of U.S. biomedical research funds by the NIH. Specifically, we provide a detailed justification for the proposal that the NIH distribute funding through a modified lottery system, as briefly described in an Op-Ed in the *Wall Street Journal*last year (4).

Go to:

BIOMEDICAL RESEARCH FUNDING ALLOCATION IN THE UNITED STATES

The primary source of biomedical research funds in the United States is the NIH, which has an annual budget of approximately 30 billion dollars. The NIH-supported research enterprise consists of two groups: intramural researchers housed in NIH facilities and extramural investigators who are mostly housed in universities, medical schools, institutes, and industry. The ratio of funds spent on the intramural and extramural programs is roughly 1:10. In both cases, the allocation of funds is made according to peer review, but the NIH uses two very different mechanisms for assessing investigators. Intramural investigators are usually evaluated through retrospective peer review, where their recent accomplishments are used to make funding decisions, a mechanism similar to that used by the Howard Hughes Medical Institute. In contrast, funding allocations to the extramural program, which comprises the overwhelming majority of the NIH budget, is allocated by a mechanism of prospective peer review in which scientists must write grant proposals detailing future work that are reviewed and criticized by a panel of experts known as a study section. The difference in funding mechanisms used by the intramural and extramural programs is significant because it shows that there is already some flexibility in the approach used by the NIH to distribute its research dollars. In this essay, we will focus on the prospective peer review mechanism used to allocate funds to extramural investigators. The fundamentals of NIH extramural peer review have not changed in a half-century. The process involves writing a proposal that is reviewed by a panel of "peers" and assigned a priority score that is converted to a percentile ranking. The NIH then funds proposals depending on the amount of money available, with the payline being that percentile ranking up to which funding is possible. At the time that the system was designed, paylines exceeded 50% of the grant applications received. However, in recent decades there has been a precipitous drop in the proportion of grants that are funded. Today's paylines and success rates are at historically low levels, hovering at around 10% in some institutes. Despite a drastic reduction in the likelihood of funding success, the essential features of NIH peer review and funding allocation have not changed.

Go to:

SHORTCOMINGS OF THE PRESENT SYSTEM

What is the desired product of scientific research? This question does not have a simple answer, but one measurable outcome is the generation of primary research publications, which are in turn cited by other publications. Remarkably, NIH study sections are unable to accurately predict which grant applications are likely to exhibit the highest publication productivity. Although a recent analysis of more than 130,000 NIH-funded grant applications reported a correlation between percentile scores and productivity (5), those findings contrast with several earlier studies showing poor predictive power for grant application peer review. Consequently, we reanalyzed the subset of the data for the grants awarded scores in the 20th percentile or better and found that the predictive ability of peer review was scarcely superior to what would be achieved by random chance and that differences in the median productivity exhibited by grants with high or low scores within this range were trivial (6). Our results corroborate earlier studies of more than 400 competing renewal R01 applications at the National Institute of General Medical Sciences (7) and 1,492 R01 applications at the National Heart, Lung, and Blood Institute (8). Hence, the available evidence makes a powerful case that the primary mechanism for biomedical research fund allocation in the United States is inadequate for prioritizing which applications to fund. The aforementioned analyses were preceded by studies suggesting that the NIH peer review process lacks statistical rigor. Only two to three reviewers in a typical study section carefully read an individual grant application and provide comments, and this reviewer sample size is too low to provide an acceptable level of precision (9). This criticism is not unique to the NIH, as studies from many countries have identified problems with the precision of grant peer review. In Canada, Mayo et al. found that the use of only two primary reviewers results in considerable randomness in funding decisions that could be improved by involving an entire 11-member review panel in the assessment of each application (10). Graves et al. examined variability in scores for the National Health and Medical Research Council of Australia and concluded that 59% of funded grants could miss funding simply on the basis of random variability in scoring (11). An analysis of applications to the Australian Research Council found interrater reliability for reviews to be poor (12), and researchers in Finland did not find that the reliability of grant peer review is improved by panel discussions (13). A French study observed that individual reviewers do not even tend to exhibit agreement on the weighting of criteria used for the grant review process (14).

A central weakness in the current system may be that experts are being asked to confidently predict the future of a scientific project, an inherently uncertain proposition. In this regard, the University of Pennsylvania psychologist Philip Tetlock showed that experts not only fared poorly in attempting to predict the future but also overrated their own abilities to do so (15). Another question is whether publication productivity is even the best metric on which to judge scientific success. Are study sections able to recognize potentially transformative research? Probably not, because intense competition for funding encourages both reviewers and applicants to be more cautious. The very structure of the NIH peer review system may encourage conformity and discourage innovation (16) of the type that could lead to scientific revolutions (17). As Nobel laureate Roger Kornberg has observed, “In the present climate especially, the funding decisions are ultraconservative. If the work that you propose to do isn’t virtually certain of success, then it won’t be funded. And of course, the kind of work that we would most like to see take place, which is groundbreaking and innovative, lies at the other extreme” (18). The NIH recognizes this problem and has created the Transformative Research Award Program, but of course, this does not solve the problem that transformative breakthroughs are often only evident as such after the fact (19).

There is also the critically important issue of bias. Sources of potential bias in peer review include cronyism and preference or disfavor for particular research areas, institutions, individual scientists, gender, or professional status. Reviewer bias can potentially have a major effect on the course of science and the career success of individual applicants. One meta-analysis of peer review studies found evidence of gender bias, such that women were approximately 7% less likely to obtain funding than men (20). Studies focusing specifically on the NIH have found comparable success in men and women submitting new R01 applications but lower success rates for women submitting renewal applications (21). There is also a continuing concern about racial bias in NIH peer review outcomes. Despite a number of initiatives following a study showing that black applicants were significantly less likely to be awarded NIH funding after controlling for educational background, country of origin, training, previous awards, publication record, and employer characteristics (22), as yet there is no evidence that the racial gap in funding success has improved (23). NIH peer reviewers tend to give better scores to applications closer to their area of expertise, and several studies have suggested that reviewers are influenced by direct or indirect personal relationships with an applicant (24).

The influence of grant reviewers in determining the fate of an application is directly proportional to the payline. This is an essential criticism of the current system, for it makes single individuals disproportionately powerful in their ability to influence the outcome of peer review. When generous paylines are available, applicants are likely to succeed even if there are scientific disagreements between applicants and/or reviewers. However, with shrinking paylines, the negative assessment by a single individual is often sufficient to derail a proposal. In this environment, a few individuals can profoundly influence the direction of research in an entire field. Reviewers are typically appointed for 4-year terms, allowing them to influence their fields for protracted periods of time. A Bayesian hierarchical statistical model applied to 18,959 R01 proposals scored by 14,041 reviewers found substantial evidence of reviewer bias that was estimated to impact approximately 25% of funding decisions (25). Day performed a computer simulation of peer review and found that very small amounts of bias can skew funding rates (25). This is not a new problem—in 1981, Cole et al. found that the odds of a proposal submitted to the National Science Foundation (NSF) getting funded were largely based on chance—the chance that specific reviewers would be chosen (26). "Targeting" research on the basis of program priorities can exacerbate the problem of bias and perversely lead to missed opportunities in basic research. The history of science is filled with stories of landmark discoveries by scientists who were looking for something else entirely—a third of anticancer drugs have been found by serendipity rather than by targeted cancer drug discovery research (27). Yet, funding agencies continue to attempt to target research funding to perceived priority areas, while support for undirected investigator-initiated projects has declined sharply (28).

Both applicants and reviewers have adapted to the funding crisis in ways that may be counterproductive to science. Applicants have responded by writing more grant applications, which takes time away from their research. As most applications are not funded, this largely represents futile effort. Some scientists estimate that half or more of their professional time is spent in seeking funding (29). In contrast, reviewers are asked to decide between seemingly equally meritorious applications and may respond by prioritizing them on the basis of "grantsmanship" (30), which has never been shown to correlate with research productivity or innovation. One of the most controversial aspects of NIH grant policy was the decision to limit applicants to two submissions of a research proposal (31). Under this policy, at a time when paylines were as low as 6%, many projects deemed meritorious by study sections were not only rejected but prohibited from resubmission for 37 months. With the current pace of science, this led to the death of many perfectly good ideas. Although this policy has now been modified to allow investigators to resubmit their projects as new grants (32), substantial damage has been done.

Peer review is used in both the ranking of grant applications and the evaluation of scientific papers. However, there are significant differences in how peer review of grant applications and papers operates. For grant applications, reviewers are chosen by an administrator who may or may not have in-depth knowledge of the relevant field, and review panels do not necessarily include the expertise necessary to review all proposals. For papers, reviewers are chosen by an editor who usually has expertise in the subject matter and can select reviewers with specific expertise in the subject area. Hence, a major difference between study section and manuscript peer review is that the latter is more likely to achieve a close match between subject matter and expertise. Accordingly, grant review is a more capricious process than manuscript review, and a single rogue reviewer can sink an application by assigning low scores without even needing to provide a convincing rationale for those scores. Publication decisions are made by editors, who can directly discuss areas of disagreement with authors and overrule single negative reviews at their discretion. Furthermore, authors have the option to appeal rejection decisions or submit their work to another journal. In contrast, there is no process for negotiation with scientific review administrators and little or no alternative to NIH funding. Another major difference is that the negative consequences of peer review differ for manuscript and grant applications, since the former usually find another publishing venue, whereas a denied grant application means that the proposed work cannot be done. Therefore, peer review of grant applications is of much greater importance for science than peer review of scientific manuscripts.

A critical aspect of the current crisis is that success rates for grant applications have fallen by more than two-thirds since the 1960s (33), and yet the system for fund allocation has essentially remained the same. A recent survey of researchers submitting proposals to the National Aeronautics and Space Administration (NASA), the NIH, and the NSF showed that even highly productive researchers are facing a 50% likelihood of not obtaining funding in the current cycle, resulting in the defunding of one-eighth of active programs following three such cycles (34). The authors of this survey estimated that at current funding rates, 78% of applicants will be unable to obtain federal funding for their research. This raises two obvious questions: (i) why has the system remained the same and (ii) why do scientists persist in this low-yield activity? Although we are not privy to discussions and decisions that have occurred among government leaders, it seems likely that the system has remained the same in the hope that national funding allocations will improve and because of the inertia involved in changing a mechanism that had worked relatively well for decades. As to why scientists persist in trying, the literature on the psychology of gambling behavior may provide some clues. People feeling desperate about their prospects will purchase lottery tickets as a surrogate for hope (35). Desperation is certainly prevalent in today's scientific community (36). Entrapment in a system due to a previous investment of time and resources is also commonly invoked as an explanation for gambling (37), and many scientists have difficulty envisaging an alternative career path. In fact, current trends in science demand so much specialization (38) that most scientists are unable to shift into fields where funding may be more plentiful. Intelligence and a high level of executive function, as seen in most scientists, are correlated with susceptibility to maladaptive decision-making and the "gambler's fallacy" (39). Risk-taking behavior may even have a neurological basis. Optimism has been described as a *sine qua non* for scientists (40), and irrational optimism correlates with reduced tracking of estimation errors by the right inferior prefrontal gyrus (41).

Go to:

PROPOSALS FOR REFORM

Recent systematic studies show that NIH grant peer review fails in its primary goal of stratifying meritorious applications when it comes to predicting the primary research outcome of citation metrics (6,–8). Despite data to the contrary, the CSR (NIH Center for Scientific Review) continues to defend its methods (42). Recent reforms in NIH peer review have failed to address the inherent unfairness of the system (43). The NIH spends a lot of money on grant peer review. The annual budget of the CSR is $110 million, which pays for more than 24,000 scientists reviewing approximately 75,000 applications and attending approximately 2,500 panel meetings (42). The costs are not only economic. Writing and reviewing grants are extremely time-consuming and divert the efforts of scientists away from doing science itself. Specifically, the NIH is asking scientists who perform peer review to perform the impossible, e.g., discriminate among the best proposals, which results in arbitrary decisions, leads to psychological stress on both reviewers and applicants, and may not be funding the most important science. Recognizing the flaws in the current grant funding process, some scientists have suggested alternative approaches that would represent a radical departure from the present peer review system. Johan Bollen has suggested having scientists vote on who deserves funding (44). Michele Pagano recommends basing funding for established scientists on track record and a one-page summary of their plans (45). This approach has some empirical support, as prior publication productivity has been shown to correlate with future productivity of R01 grant recipients (46). John Ioannidis has proposed a number of options ranging from awarding small amounts of funding to all applicants to assigning grants randomly or basing awards on an applicant's publication record (47). Recently, we proposed that the NIH adopt a hybrid approach based on a modified lottery system (4).

Go to:

LESSONS FROM THE WORLD OF FINANCE

The debate over the optimal strategy for allocating funds for scientific research has interesting parallels with the decisions involved in making financial investments. In 1973, the economist Burton Malkiel published his now-classic book, *A Random Walk down Wall Street* (48). Malkiel argued that investors cannot consistently outperform stock market averages, and therefore, a passive investment strategy can be just as effective as an active one. In fact, very few professional investors consistently outperform the market. A study called "Does Past Performance Matter?" by S&P Dow Jones found that only 2 out of 2,862 funds were able to remain in the top quarter over five successive years, worse than might be predicted by random chance alone—"If all of the managers of these mutual funds hadn't bothered to try to pick stocks at all—if they had merely flipped coins—they would, as a group, probably have produced better numbers" (49). Even Warren Buffett has instructed in his will to "Put 10% in short-term government bonds and 90% in a very low-cost index fund … I believe the long-term results from this policy will be superior to those attained by most investors—whether pension funds, institutions, or individuals—who employ high-fee managers" (50). In 2007, the statistician Nassim Nicholas Taleb published the acclaimed book *The Black Swan* (51), which argued that the most influential events were both highly improbable and unpredictable. According to Taleb, investors should not attempt to predict such events but instead should construct a system that is sufficiently robust to withstand negative events and maximize the opportunity to benefit from positive ones. Applied to science, this suggests that it may be futile for reviewers to attempt to predict which grant applications will produce unanticipated transformational discoveries. In this regard, our recent review of revolutionary science suggests that historical scientific revolutions lack a common structure, with transformative discoveries emerging from puzzle solving, serendipity, inspiration, or a convergence of disparate observations (19). Consequently, a random strategy that distributes funding as broadly as possible may maximize the likelihood that such discoveries will occur. Taleb underscores the limits of human knowledge and cautions against relying on the authority of experts, emphasizing that explanations for phenomena are often possible only with hindsight, whereas people consistently fail in their attempts to accurately predict the future.

Four European economists have raised the question "Given incomplete knowledge of the market, is a random strategy as good as a targeted one?" (52, 53). A computer simulation was performed using data from British, Italian, German, and American stock indices. The authors compared four different conventional investment strategies with a random approach. Over the long run, each strategy performed similarly, but the random strategy turned out to be the least volatile, i.e., the least risky strategy with little compromise in performance. Given that assigning funds for investment or research allocation each involves a wager on future success with incomplete information, these lessons from the world of finance have relevance to science funding. Among the advantages of index funds are that randomization of the investment process can reduce “herding behavior” and financial “bubbles” (which raises the question of whether we are heading for microbiome and precision medicine “bubbles”—but that is a discussion for another time). An indexed strategy for picking stocks reduces the administrative costs associated with fund management, just as a modified lottery system for grant allocation could reduce the administrative costs of review.

Go to:

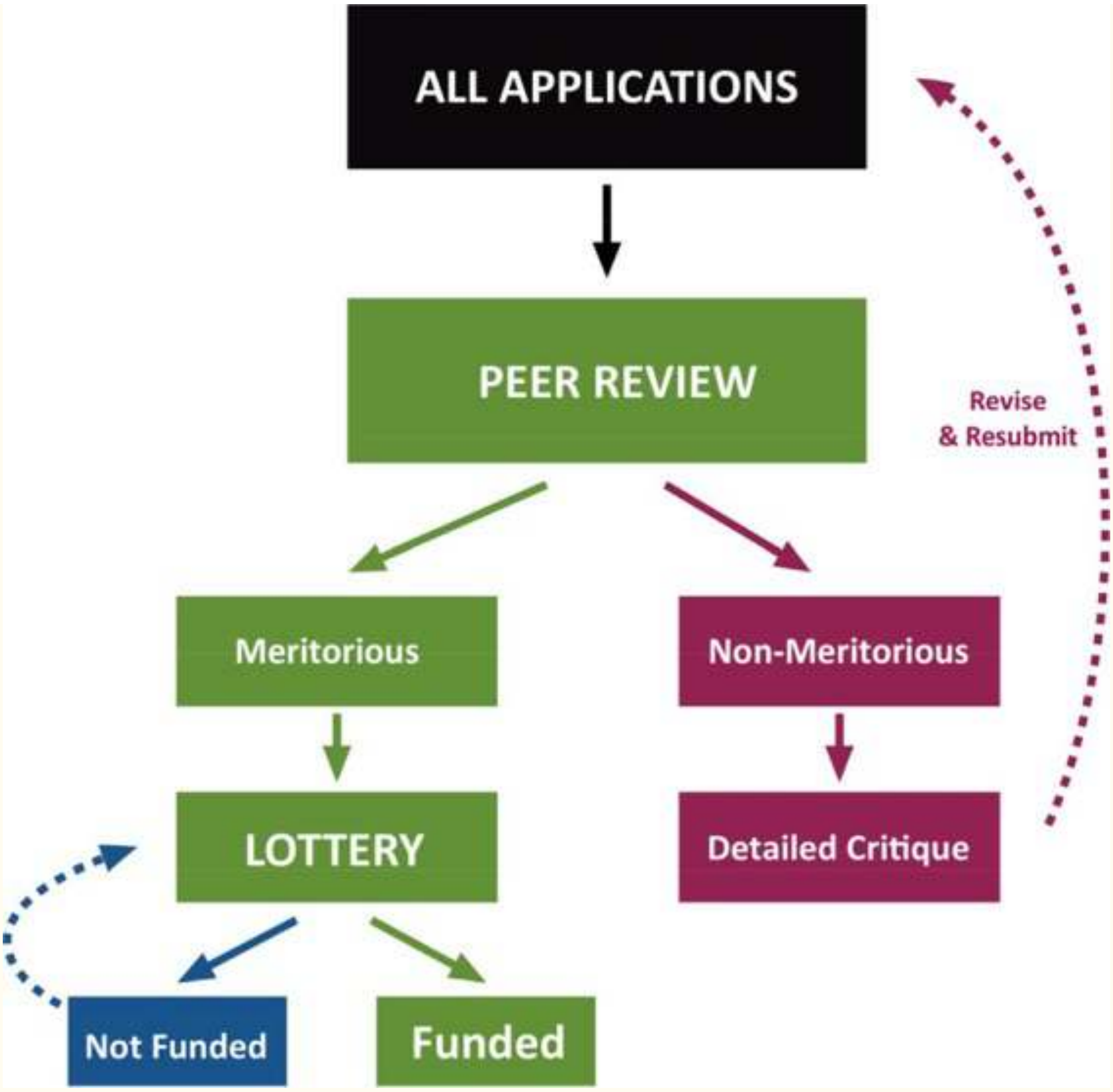
GOALS OF A FUNDING ALLOCATION SYSTEM

As we consider reform proposals for grant peer review, it is important to state some basic principles that we believe are likely to be accepted by the majority of scientists. First, we recognize that there are qualitative and quantitative differences among research proposals. Clearly, not all scientific projects are equally meritorious. We currently rely on the assessment of experts in the form of peer review to determine those differences. An ideal system would be a meritocracy that identified and funded the best science, but the available evidence suggests that the current process fails in this regard, and the goal might in fact be impossible. Second, we argue that some form of peer review will be required for funding allocation. Although we have catalogued many problems with the current peer review system, it is essential to have grant proposals evaluated by panels of scientists who have expertise in the area. Although experts may not be able to discriminate between meritorious proposals, they are still generally able to weed out proposals that are simply infeasible, are badly conceived, or fail to sufficiently advance science. Third, scarce research funds should be distributed in a fair and transparent manner. While fairness is likely to be partly in the eye of the beholder, there are mechanisms that are generally acknowledged to be fair. Specifically, there is a need to neutralize biases in funding decisions. Otherwise, the enormous power of reviewers at a time of unfavorable paylines will distort the course of science in certain fields. In this regard, there is evidence for increasing inefficiency in the translation of basic discovery into medical goods (54, 55). Although the causes for this phenomenon are undoubtedly complex, any bias in funding decisions affects the type of research done, which in turn influences potential downstream benefits for society. Should the review process favor new investigators? A case can certainly be made for the importance of providing support to new investigators, as they represent the future of science (56). This should not be taken to suggest that older investigators are unimportant. In fact, higher publication productivity has been seen for competing renewals than for new grants, and for projects directed by senior investigators (57). Nevertheless, we recognize that established investigators have significant advantages relative to new investigators with regard to experience, prior productivity, reputation in the field, and laboratories that are already established and productive. In a world of plentiful research funds, new investigators are able to compete successfully for funding with established laboratories. However, in times of funding scarcity, differences between established and new investigators can become magnified to favor established investigators over new ones. Established investigators benefit from the so-called "Matthew effect," whereby those with resources and prestige are more likely to receive further rewards (58). Consequently, steps should be taken to improve the opportunities for new investigators as a matter of science planning policy. A modified lottery system could immediately benefit young investigators by creating a more level playing field.

Go to:

POTENTIAL BENEFITS OF A FUNDING LOTTERY

Given overwhelming evidence that the current process of grant selection is neither fair nor efficient, we instead suggest a two-stage system in which (i) meritorious applications are identified by peer review and (ii) funding decisions are made on the basis of a computer-generated lottery (Fig. 1). The size of the meritorious pool could be adjusted according to the payline. For example, if the payline is 10%, then the size of the meritorious pool might be expected to include the top 20 to 30% of applications identified by peer review. This would eliminate or at least alleviate certain negative aspects of the current system, in particular, bias. Critiques would be issued only for grants that are considered nonmeritorious, eliminating the need for face-to-face study section meetings to argue over rankings, which would bring about immediate cost savings. Remote review would allow more reviewers with relevant expertise to participate in the process, and greater numbers of reviewers would improve precision. Funding would be awarded to as many computer-selected meritorious applications as the research budget allows. Applications that are not chosen would become eligible for the next drawing in 4 months, but individual researchers would be permitted to enter only one application per drawing, which would reduce the need to revise currently meritorious applications that are not funded and free scientists to do more research instead of rewriting grant applications. New investigators could compete in a separate lottery with a higher payline to ensure that a specific portion of funding is dedicated to this group or could be given increased representation in the regular lottery to improve their chances of funding. Although the proposed system could bring some cost savings, we emphasize that the primary advantage of a modified lottery would be to make the system fairer by eliminating sources of bias. The proposed system should improve research workforce diversity, as any female or underrepresented minority applicant who submits a meritorious application will have an equal chance of being awarded funding. There would also be benefits for research institutions. A modified lottery would allow research institutions to make more reliable financial forecasts, since the likelihood of future funding could be estimated from the percentage of their investigators whose applications qualify for the lottery. In the current system, administrators must deal with greater uncertainty, as funding decisions can be highly unpredictable. Furthermore, we note that program officers could still use selective pay mechanisms to fund individuals who consistently make the lottery but fail to receive funding or in the unlikely instance that important fields become underfunded due to the vagaries of luck.



Open in a separate window

FIG 1

Proposed scheme for a modified funding lottery. In stage 1, applications are determined to be meritorious or nonmeritorious on the basis of conventional peer review. Nonmeritorious applications may be revised and resubmitted. In stage 2, meritorious applications are randomized by computer and funding is awarded to as many applications as funds permit on the basis of randomly generated priority scores.

The proposed system would treat new and competing renewal applications in the same manner. Historically, competing applications have enjoyed higher success rates than new applications, for reasons including that these applications are from established investigators with a track record of productivity. However, we find no compelling reason to justify supporting established programs over new programs.

Although we recognize that some scientists will cringe at the thought of allocating funds by lottery, the available evidence suggests that the system is already in essence a lottery without the benefits of being random (6). Furthermore, we note that lotteries are already used by society to make difficult decisions. Historically, a lottery was used in the draft for service in the armed forces. Today, lotteries are used to select students for charter schools (59), to determine the order of selection in the National Basketball Association draft, to issue green cards for permanent residency, and even to allocate scarce medical resources (60). Modified lotteries have been advocated as the fairest way in which to allocate scarce medical resources such as vaccines and organs for transplantation (61, 62). If lotteries could be used to select those who served in Vietnam, they can certainly be used to choose proposals for funding. We note that we are not the first to arrive at this idea (63). In fact, the New Zealand Health Research Council has already adopted a lottery system to select recipients of investigator-initiated Explorer grants (64).

The institution of a funding lottery would have many immediate advantages. First, it will maintain an important role for peer review at the front end, to decide which applications are technically sound enough to merit inclusion in the lottery. Second, it will convert the current system with its biases and arbitrariness into a more transparent process. Third, it will lessen the blow of grant rejection, since it is easier to rationalize bad luck than to feel that one failed to make the cut due to a lack of merit. Fourth, it will relieve reviewers from having to stratify the top applications, since it is increasingly obvious that this is not possible. Fifth, meritorious but unfunded proposals could continue to have a shot at receiving funding in the future instead of being relegated to the dustbin. Sixth, it will be less expensive to administer, and some of the funds currently used for the futile exercise of ranking proposals could be devoted instead to supporting actual scientific research. Seventh, it should decrease cronyism and bias against women, racial minorities, and new investigators. Eighth, it would give administrators in research institutions a greater capacity to make financial projections based on the percentage of their investigators who qualify for the lottery. Ninth, the system will be less noisy, will be fairer, and may promote new areas of investigation by removing favoritism for established fields that are better represented in review panels. Tenth, the realization that many meritorious projects remain unfunded may promote more serious efforts to improve research funding and study alternative approaches to peer review. In fact, the success rate of the lottery would provide a clear number for society and politicians to understand the degree to which meritorious research proposals remain unfunded, and this would hopefully lead to an increased budgetary allocation for research and development. Under the current system, the underfunding of science is hidden by the fallacious mantra that the most worthy science continues to be funded, which provides an excuse for inaction. A recent NSF report indicated that 68% of applications were rated as meritorious but only a third of these are funded (65).

Go to:

CONCLUDING REMARKS

The biologist E. O. Wilson has compared scientists to prospectors searching for gold (66): "In the 17th, 18th and 19th centuries, making scientific discoveries was like picking nuggets off the ground." But, prospecting today is more challenging. The rewards are still great, but the big finds are more elusive. Targeted initiatives would direct all scientists to look for new lodes in the same place, while "transformative research" initiatives aim to fund only those who strike it rich. Neither strategy is optimal. Society must accept that science, as John Ioannidis has astutely observed, is an inherently "low-yield endeavor" (67). However, this low-yield endeavor has consistently improved the lot of humanity since the scientific revolution of the 17th century and remains humanity's best bet for finding solutions to deal with such challenges as climate change, pandemics and disease, a faltering green revolution, and the need for new energy sources (68, 69). To continue to reap the maximal benefits of scientific exploration, researchers must be encouraged to search as far and wide as possible, leaving no stone unturned, even though only some will be successful in their quests. As Nassim Nicholas Taleb has written, "The reason markets work is because they allow people to be lucky, thanks to aggressive trial and error, not by giving rewards or incentives for skill" (51). We must provide our scientists with an opportunity to get lucky.

Go to:

ACKNOWLEDGMENTS

Some of the ideas presented in this article emerged from discussions with individuals interested in problems with peer review and funding allocation. We thank these individuals, whose critical comments have made us think further.

Furthermore, we disclose that both of us have been recipients of NIH funding and as active investigators continue to seek funding in the current system. Otherwise, the authors have no funding sources or competing interests relating to this article to declare.

Go to:

FUNDING STATEMENT

The funders had no role in study design, data collection and interpretation, or the decision to submit the work for publication.

Go to:

NOTES

The views expressed in this Editorial do not necessarily reflect the views of this journal or of ASM.

Go to:

FOOTNOTES

Citation Fang FC, Casadevall A. 2016. Research funding: the case for a modified lottery. mBio 7(2):e00422-16. doi:10.1128/mBio.00422-16.

Go to:

REFERENCES

1. Oliver J. 2014. Last Week Tonight. Episode 24, 9 November 2014https://www.youtube.com/watch?v=9PK-netuhHA.

2. Szabo L. 23 April 2014. NIH director: budget cuts put U.S. science at risk. USA Today, McLean,

VA: http://www.usatoday.com/story/news/nation/2014/04/23/nih-budget-cuts/8056113/. [Google Scholar]

3. Mervis J. 18 December 2015. Budget agreement boosts science. Science Insider. American Association for the Advancement of Science, Washington, DC: http://www.sciencemag.org/news/2015/12/updated-budget-agreement-boosts-us-science. [Google Scholar]

4. Fang FC, Casadevall A. 14 April 2014. Taking the Powerball approach to funding medical research. Wall Street Journal. http://www.wsj.com/articles/SB10001424052702303532704579477530153771424.

5. Li D, Agha L. 2015. Research funding. Big names or big ideas: do peer-review panels select the best science proposals? Science 348:434–438. doi:10.1126/science.aaa0185. [PubMed] [CrossRef] [Google Scholar]

6. Fang FC, Bowen A, Casadevall A. 2016. NIH peer review percentile scores are poorly predictive of grant productivity. Elife 5:e13323. doi:10.7554/eLife.13323. [PMC free article] [PubMed] [CrossRef] [Google Scholar]

7. Danthi N, Wu CO, Shi P, Lauer M. 2014. Percentile ranking and citation impact of a large cohort of National Heart, Lung, and Blood Institute-funded cardiovascular R01 grants. Circ Res 114:600–606. doi:10.1161/CIRCRESAHA.114.302656. [PMC free article] [PubMed] [CrossRef] [Google Scholar]

8. Berg J. 2013. On deck chairs and lifeboats. ASBMB Today 12(2):3–6. [Google Scholar]

9. Kaplan D, Lacetera N, Kaplan C. 2008. Sample size and precision in NIH peer review. PLoS One 3:e2761. doi:10.1371/journal.pone.0002761. [PMC free article][PubMed] [CrossRef] [Google Scholar]

10. Mayo NE, Brophy J, Goldberg MS, Klein MB, Miller S, Platt RW, Ritchie J. 2006. Peering at peer review revealed high degree of chance associated with funding of grant applications. J Clin Epidemiol 59:842–848. doi:10.1016/j.jclinepi.2005.12.007. [PubMed] [CrossRef] [Google Scholar]
11. Graves N, Barnett AG, Clarke P. 2011. Funding grant proposals for scientific research: retrospective analysis of scores by members of grant review panel. BMJ343:d4797. doi:10.1136/bmj.d4797. [PMC free article] [PubMed] [CrossRef] [Google Scholar]
12. Marsh HW, Jayasinghe UW, Bond NW. 2008. Improving the peer-review process for grant applications: reliability, validity, bias, and generalizability. Am Psychol 63:160–168. doi:10.1037/0003-066X.63.3.160. [PubMed] [CrossRef] [Google Scholar]
13. Fogelholm M, Leppinen S, Auvinen A, Raitanen J, Nuutinen A, Väänänen K. 2012. Panel discussion does not improve reliability of peer review for medical research grant proposals. J Clin Epidemiol 65:47–52. doi:10.1016/j.jclinepi.2011.05.001. [PubMed] [CrossRef] [Google Scholar]
14. Abdoul H, Perrey C, Amiel P, Tubach F, Gottot S, Durand-Zaleski I, Alberti C. 2012. Peer review of grant applications: criteria used and qualitative study of reviewer practices. PLoS One 7:e46054. doi:10.1371/journal.pone.0046054. [PMC free article] [PubMed] [CrossRef] [Google Scholar]
15. Tetlock P. 2005. Expert political judgment: how good is it? How can we know?Princeton University Press, Princeton, NJ. [Google Scholar]
16. Nicholson JM, Ioannidis JP. 2012. Research grants: conform and be funded. Nature 492:34–36. doi:10.1038/492034a. [PubMed] [CrossRef] [Google Scholar]

17. Kuhn TS. 1962. The Structure of Scientific Revolutions. University of Chicago Press, Chicago, IL. [Google Scholar]
18. Lee C. 28 May 2007. Slump in NIH funding is taking toll on research. Washington Post, Washington, DC: http://www.washingtonpost.com/wp-dyn/content/article/2007/05/27/AR2007052700794.html. [Google Scholar]
19. Casadevall A, Fang FC. 2016. Revolutionary science. mBio 7:e00158-16. doi:10.1128/mBio.00158-16. [PMC free article] [PubMed] [CrossRef] [Google Scholar]
20. Bornmann L, Mutz R, Daniel H. 2007. Gender differences in grant peer review: a meta-analysis. J Informetr 1:226–238. doi:10.1016/j.joi.2007.03.001. [CrossRef] [Google Scholar]
21. Pohlhaus JR, Jiang H, Wagner RM, Schaffer WT, Pinn VW. 2011. Sex differences in application, success, and funding rates for NIH extramural programs. Acad Med 86:759–767. doi:10.1097/ACM.0b013e31821836ff. [PMC free article][PubMed] [CrossRef] [Google Scholar]
22. Ginther DK, Schaffer WT, Schnell J, Masimore B, Liu F, Haak LL, Kington R. 2011. Race, ethnicity, and NIH research awards. Science 333:1015–1019. doi:10.1126/science.1196783. [PMC free article] [PubMed] [CrossRef] [Google Scholar]
23. Hayden EC. 2015. Racial bias continues to haunt NIH grants. Nature 527:286–287. doi:10.1038/527286a. [PubMed] [CrossRef] [Google Scholar]
24. Wessely S. 1998. Peer review of grant applications: what do we know? Lancet352:301–305. doi:10.1016/S0140-6736(97)11129-1. [PubMed] [CrossRef] [Google Scholar]

25. Day TE. 2015. The big consequences of small biases: a simulation of peer review. Res Policy 44:1266–1270. doi:10.1016/j.respol.2015.01.006. [CrossRef] [Google Scholar]

26. Cole S, Cole JR, Simon GA. 1981. Chance and consensus in peer review. Science 214:881–886. doi:10.1126/science.7302566. [PubMed] [CrossRef] [Google Scholar]

27. Hargrave-Thomas E, Yu B, Reynisson J. 2012. Serendipity in anticancer drug discovery. World J Clin Oncol 3:1–6. doi:10.5306/wjco.v3.i1.1. [PMC free article][PubMed] [CrossRef] [Google Scholar]

28. Mandel HG, Vesell ES. 2008. Declines in NIH R01 research grant funding. Science 322:189a. doi:10.1126/science.322.5899.189a. [PubMed] [CrossRef] [Google Scholar]

29. Siliciano R. 19 May 2007. Testimony before the Senate Committee on Appropriations: Subcommittee on Labor, Health and Human Services, Education, and Related Agencies. U.S. Senate Committee on Appropriations, Washington, DC. [Google Scholar]

30. Price M. 4 June 2013. A shortcut to better grantsmanship. Science Careers. American Association for the Advancement of Science, Washington, DC: http://sciencecareers.sciencemag.org/career_magazine/previous_issues/articles/2013_06_04/caredit.a1300119. [Google Scholar]

31. Benezra R. 2013. Grant applications: undo NIH policy to ease effect of cuts. Nature 493:480. doi:10.1038/493480e. [PubMed] [CrossRef] [Google Scholar]

32. Kaiser J. 2014. Biomedical funding. At NIH, two strikes policy is out. Science344:350. doi:10.1126/science.344.6182.350. [PubMed] [CrossRef] [Google Scholar]

33. Atkinson RD, Ezell SJ, Giddings LV, Stewart LA, Andes SM. 2012. Leadership in decline: assessing U.S. international competitiveness in biomedical research. Information Technology and Innovation Foundation and United for Medical; Research, Washington, DC: http://www2.itif.org/2012-leadership-in-decline.pdf. [Google Scholar]
34. Von Hippel T, von Hippel C. 2015. To apply or not to apply: a survey analysis of grant writing costs and benefits. PLoS One 10:e0118494. doi:10.1371/journal.pone.0118494. [PMC free article] [PubMed] [CrossRef] [Google Scholar]
35. Clotfelter CT, Cook PJ. 1989. Selling hope: state lotteries in America. Harvard University Press, Cambridge, MA. [Google Scholar]
36. Basken P, Voosen P. 24 February 2014. Strapped scientists abandon research and students. Chronicle of Higher Education, Washington, DC: http://chronicle.com/article/Strapped-Scientists-Abandon/144921/. [Google Scholar]
37. Staw BM. 1976. Knee-deep in the big muddy: a study of escalating commitment to a chosen course of action. Organ Behav Hum Perform 16:27–44. doi:10.1016/0030-5073(76)90005-2. [CrossRef] [Google Scholar]
38. Casadevall A, Fang FC. 2014. Specialized science. Infect Immun 82:1355–1360. doi:10.1128/IAI.01530-13. [PMC free article] [PubMed] [CrossRef] [Google Scholar]
39. Xue G, He Q, Lei X, Chen C, Liu Y, Chen C, Lu ZL, Dong Q, Bechara A. 2012. The gambler's fallacy is associated with weak affective decision making but strong cognitive ability. PLoS One 7:e47019. doi:10.1371/journal.pone.0047019. [PMC free article] [PubMed] [CrossRef] [Google Scholar]

40. Foster R. 26 September 2013. Optimism is a sine qua non for scientists. Times Higher Education, London, United Kingdom: https://www.timeshighereducation.com/features/optimism-is-a-sine-qua-non-for-scientists/2007563.article. [Google Scholar]

41. Sharot T, Korn CW, Dolan RJ. 2011. How unrealistic optimism is maintained in the face of reality. Nat Neurosci 14:1475–1479. doi:10.1038/nn.2949. [PMC free article] [PubMed] [CrossRef] [Google Scholar]

42. Mervis J. 2014. Peering into peer review. Science 343:596–598. doi:10.1126/science.343.6171.596. [PubMed] [CrossRef] [Google Scholar]

43. Fang FC, Casadevall A. 2009. NIH peer review reform—change we need, or lipstick on a pig? Infect Immun 77:929–932. doi:10.1128/IAI.01567-08. [PMC free article] [PubMed] [CrossRef] [Google Scholar]

44. Bollen J, Crandall D, Junk D, Ding Y, Börner K. 2014. From funding agencies to scientific agency: collective allocation of science funding as an alternative to peer review. EMBO Rep 15:131–133. doi:10.1002/embr.201338068. [PMC free article] [PubMed] [CrossRef] [Google Scholar]

45. Pagano M. 2006. American Idol and NIH grant review. Cell 126:637–638. doi:10.1016/j.cell.2006.08.004. [PubMed] [CrossRef] [Google Scholar]

46. Kaltman JR, Evans FJ, Danthi NS, Wu CO, DiMichele DM, Lauer MS. 2014. Prior publication productivity, grant percentile ranking, and topic-normalized citation impact of NHLBI cardiovascular R01 grants. Circ Res 115:617–624. doi:10.1161/CIRCRESAHA.115.304766. [PMC free article] [PubMed] [CrossRef] [Google Scholar]

47. Ioannidis JPA. 2011. More time for research: fund people not projects. Nature477:529–531. doi:10.1038/477529a. [PubMed] [CrossRef] [Google Scholar]

48. Malkiel BG. 1973. A Random Walk down Wall Street. W. W. Norton, New York, NY. [Google Scholar]
49. Sommer J. 14 March 2015. How many mutual funds routinely rout the market? None. New York Times, New York, NY: http://www.nytimes.com/2015/03/15/your-money/how-many-mutual-funds-routinely-rout-the-market-zero.html?_r=0. [Google Scholar]
50. Buffett WE. 2014. Berkshire Hathaway 2013 annual report, p 20. Berkshire Hathaway, Omaha, NE. [Google Scholar]
51. Taleb NN. 2007. The Black Swan: the Impact of the Highly Improbable. Random House, New York, NY. [Google Scholar]
52. Biondo AE, Pluchino A, Rapisarda A, Helbing D. 2013. Are random trading strategies more successful than technical ones? PLoS One 8:e68344. doi:10.1371/journal.pone.0068344. [PMC free article] [PubMed] [CrossRef] [Google Scholar]
53. Knapp A. 22 March 2013. Computer simulation suggests that the best investment strategy is a random one. Forbes, Jersey City, NJ: http://www.forbes.com/sites/alexknapp/2013/03/22/computer-simulation-suggests-that-the-best-investment-strategy-is-a-random-one/#204abcd97e7c. [Google Scholar]
54. Scannell JW, Blanckley A, Boldon H, Warrington B. 2012. Diagnosing the decline in pharmaceutical R&D efficiency. Nat Rev Drug Discov 11:191–200. doi:10.1038/nrd3681. [PubMed] [CrossRef] [Google Scholar]
55. Bowen A, Casadevall A. 2015. Increasing disparities between resource inputs and outcomes, as measured by certain health deliverables, in biomedical research. Proc Natl Acad Sci U S A 112:11335–11340. doi:10.1073/pnas.1504955112. [PMC free article] [PubMed] [CrossRef] [Google Scholar]

56. National Institutes of Health 2016. New and early stage investigator policies. Office of Extramural Research, National Institutes of Health, Bethesda, MD.http://grants.nih.gov/grants/new_investigators/. [Google Scholar]
57. Druss BG, Marcus SC. 2005. Tracking publication outcomes of National Institutes of Health grants. Am J Med 118:658–663. doi:10.1016/j.amjmed.2005.02.015. [PubMed] [CrossRef] [Google Scholar]
58. Merton RK. 1968. The Matthew effect in science: the reward and communication systems of science are considered. Science 159:56–63. doi:10.1126/science.159.3810.56. [PubMed] [CrossRef] [Google Scholar]
59. O'Sullivan F. 21 June 2010. Charter school lotteries leave lots behind. Care2, Redwood City, CA: http://www.care2.com/causes/charter-lotteries-leave-lots-behind.html. [Google Scholar]
60. Persad G, Wertheimer A, Emanuel EJ. 2009. Principles for allocation of scarce medical interventions. Lancet 373:423–431. doi:10.1016/S0140-6736(09)60137-9. [PubMed] [CrossRef] [Google Scholar]
61. Annas GJ. 1985. The prostitute, the playboy, and the poet: rationing schemes for organ transplantation. Am J Public Health 75:187–189. doi:10.2105/AJPH.75.2.187. [PMC free article] [PubMed] [CrossRef] [Google Scholar]
62. McLachlan HV. 2012. A proposed non-consequentialist policy for the ethical distribution of scarce vaccination in the face of an influenza pandemic. J Med Ethics 38:317–318. doi:10.1136/medethics-2011-100031. [PubMed] [CrossRef] [Google Scholar]
63. Greenberg D. 11 February 2008. Peer review at NIH: a lottery would be better. Chronicle of Higher Education, Washington, DC: http://chronicle.com/blogs/brainstorm/peer-review-at-nih-a-lottery-would-be-better/5696. [Google Scholar]

64. Health Research Council of New Zealand 2015. Funding opportunities. Explorer grants. Health Research Council of New Zealand, Auckland, NZ: http://www.hrc.govt.nz/funding-opportunities/researcher-initiated-proposals/explorer-grants. [Google Scholar]

65. National Science Board 2013. Report to the National Science Board on the National Science Foundation's merit review process fiscal year 2012. NSB-13-33. National Science Board, Arlington, VA. [Google Scholar]

66. Wilson EO. 2014. The Meaning of Human Existence. Liveright, New York, NY. [Google Scholar]

67. Freedman DH. November 2010. Lies, damned lies, and medical science. The Atlantic, Washington, DC: http://www.theatlantic.com/magazine/archive/2010/11/lies-damned-lies-and-medical-science/308269/. [Google Scholar]

68. Fang FC, Casadevall A. 2012. Reforming science: structural reforms. Infect Immun 80:897–901. doi:10.1128/IAI.06184-11. [PMC free article] [PubMed] [CrossRef] [Google Scholar]

69. Casadevall A, Fang FC. 2012. Reforming science: methodological and cultural reforms. Infect Immun 80:891–896. doi:10.1128/IAI.06183-11. [PMC free article][PubMed] [CrossRef] [Google Scholar]

How to avoid borrowed plumes in academia

Author links open overlay panelMargitOsterlohBruno S.Frey

Show more

https://doi.org/10.1016/j.respol.2019.103831Get rights and content

Highlights

- Impact factors are still highly influential because a majority of authors benefit.
- Changes in performance management must occur at the institutional level.
- Evaluation has to take fundamental uncertainty in research into account.
- Focal randomization mitigates biased selection of articles.
- Scholarly diversity instead of one-dimensional rankings is supported.

Abstract

Publications in top journals today have a powerful influence on academic careers although there is much criticism of using journal rankings to evaluate individual articles. We ask why this practice of performance evaluation is still so influential. We suggest this is the case because a majority of authors benefit from the present system due to the extreme skewness of citation distributions. "Performance paradox" effects aggravate the problem. Three extant suggestions for reforming performance management are critically discussed. We advance a new proposal based on the insight that fundamental uncertainty

is symptomatic for scholarly work. It suggests focal randomization using a rationally founded and well-orchestrated procedure.

- **Previous** article in issue
- **Next** article in issue

Keywords

Journal rankings

Impact factor

Journal quality lists

Skewed citation distribution

Focal random selection

1. Introduction

Publication in peer-reviewed scholarly journals has today become the currency of performance for the evaluation of scholars, departments, faculties, and universities. Journals are ranked according to quality criteria, most importantly the journal impact factor. It is defined as the mean number of citations in a particular year of articles published in that journal in the previous two years or five years. Some journals are ranked according to journal quality lists, such as the Association of Business Schools (ABS) Guide in Great Britain (e.g. Mingers and Willmott, 2013) and the "Top Five" in economics (e.g. Hamermesh, 2018).[1] It has been empirically demonstrated that the "Top Five" have a powerful influence on tenure and promotion decisions and has even been denounced as the "tyranny of the top five" by a Nobel Prize laureate (Heckman and Moktan, 2018). Journal quality lists rely not only on journal metrics but also on qualitatively informed indicators of reputation. In both cases, the quality of a journal is widely believed to reflect the quality of any article published therein. Originally designed to evaluate scientific journals, today journal quality lists and impact factors are increasingly used to evaluate individual articles and authors. They strongly influence decisions on tenure, research funding, and the pursuit of career goals. For example, the British ABS Academic Journal Guide claims to give scholars “a recognized currency on which career progress can be based” (ABS The Association of Business Schools ABS, 2015: 5). In many academic institutions, scholars receive a financial bonus for a publication in one of the top journals (e.g. Fuyuno and Cyranoski, 2006; Macdonald and Kam, 2007; Shao and Shen, 2011).

However, this practice has been strongly criticized for several years (Seglen, 1997; Moed and Van Leeuwen, 1996; Laband and Tollison, 2003; Starbuck, 2005; Oswald, 2007; Singh et al., 2007; Adler and Harzing, 2009; Frey and Rost, 2010; Baum, 2011; Macdonald and Kam, 2011; Mingers and Willmott, 2013; Alberts, 2013; Osterloh and Frey, 2014; Wilsdon et al., 2015; Martin, 2016; Larivière et al., 2016; Berg, 2016; Callaway, 2016; Waltman, 2016; Wang et al., 2017), even by Eugene Garfield, the inventor of the impact factor (Garfield, 1973). The San Francisco Declaration on Research Assessment (DORA (San Francisco Declaration on Research Assessment) and DORA, 2012), which has been endorsed by many leading institutions, clearly states: “Do not use journal-based metrics, such as Journal Impact Factors, as a surrogate measure of the quality of individual research articles, to assess an individual scientist’s contributions, or in hiring, promotion, or funding decisions.” The recently released “Statement by three national academies (Académie des Sciences, Leopoldina and Royal Society) on good practice in the evaluation of researchers and research programmes”[2] also asserts that “[i]mpact factors of journals should not be considered in evaluating research outputs”.

Nevertheless, to date, these critiques have not diminished the impact of either impact factors or journal quality lists. Instead, journal rankings have become

more widespread and increasingly important for academic careers and research funding (e.g. Harzing, 2015; Martin, 2016; Vogel et al., 2017). Top-tier journals have become the ultimate fetish token (Willmott, 2011) for many scholars. According to a survey of the perceptions of young economists the pursuit of top journal publications “has become the obsession of the next generation” (Heckman and Moktan, 2018: 1).
This paper has two aims. The first is to understand why impact factors and journal lists are still so influential to evaluate individual papers even though they are strongly criticized by many influential scholars and institutions. This criticism is based on the heavily skewed distribution of citations in scholarly journals. Why are impact factors and journal lists not abolished as proxies for the quality of single articles? Second, while the criticisms of this practice are many, few suggestions have been made for changes at the institutional level to overcome the problem. We discuss such proposals and present a novel, radical proposition: purposeful focal randomization. To our knowledge, this is the first proposal for change using the insight that uncertainty is fundamental to research, translating it into performance management.

The second section of this paper complements the literature that questions the use of impact factors and journal quality lists to evaluate individual articles because of the strong skewness of citations in scholarly journals. We ask whether the citation rates of articles accumulated over five years are more useful in evaluating publications than yearly citation rates. We show empirically that this is not the case. There is still a substantial overlap in the distribution of citations between high-, middle- and low-ranked business journals. In the third section, we inquire why impact factors and journal quality lists have not been abolished even though they have attracted such strong criticism. We argue that this is mainly due to the fact that the majority of authors benefits from journal quality lists, which is aggravated by the "performance paradox" and lock-in effects. In the fourth section, we discuss proposals on how the present unsatisfactory situation can be overcome by changes at the institutional level. We present and discuss our own proposal.

2. Skewed distributions of citations

The use of journal lists to evaluate the quality of research – whether derived from metrics or qualitatively-informed indicators - takes for granted that publishing in a "good journal" is a signal of "good research". The most influential journal rankings today rely largely on the two-year journal impact factor (JIF) published by Clarivate Analytics (formerly Thomson Reuters), which owns and publishes the Journal Citation Reports (formerly known as the ISI Web of Knowledge).[3] The JIF was originally developed to help librarians identify the most important journals (see Archambault and Larivière, 2009) according to the numbers of citations of the articles published in those journals.

The use of citation counts as a performance indicator has its own problems (e.g. Starbuck, 2005; Adler and Harzing, 2009; Macdonald and Kam, 2010). To take citations as a proxy for quality is questionable. At best it can inform us whether an article can be considered interesting and influential since citations acknowledge the impact an author has on the work of others (e.g. Antonakis et al., 2014; Alvesson and Sandberg, 2013; Hamermesh, 2018). Nevertheless, citations are widely accepted as a performance indicator for articles and journals (e.g. Goodall, 2009; Vogel et al., 2017), though most scholars agree they should not be used as the only determinant.[4] However, those who use impact factors for an article or a journal – be it as a proxy for quality or for other reasons –must ex ante have accepted that citations matter, because impact factors are based on citations.

It is questionable using the impact factor as a quality indicator for a whole journal, but it is a clear misuse employing the impact factor of a journal as a quality indicator for a *single* article in that journal. This is due to the highly skewed distribution of citations.[5] Nevertheless, such misuse has not decreased (e.g. Heckman and Moktan, 2018), although an increasing number of studies argues that scholars should abolish it.

An impressive example of the misuse of impact factors was published recently in *Nature* (Callaway, 2016). This article refers to a study considering the natural sciences (Larivière et al., 2016), which reveals that 74.8 percent of the articles published in *Nature* (2015) were cited below the 2-year impact factor of 38.1, which reflects the average number of citations for articles in that journal. The most cited paper was referenced 905 times. Three quarters of authors benefit from the minority of authors with many citations. The equally renowned journal *Science*shows almost the same result: 75.5% of the papers published in 2015 garnered less than the impact factor of 34.7. The most successful paper was cited 694 times.

A similar pattern was demonstrated earlier in the field of organization and management by Baum (2011). He examined five journals[6] and collected the citations per year in 2008 of articles published from 1990 to 2007. He concludes that the impact factor has little credibility as a proxy for the quality of an article published in these journals. Using the JIF in such a way results in incorrect attribution of article quality more than half the time. Only a small correlation was found between the number of citations for an individual article and the impact factor of the publishing journal. Baum (2011) firmly recommends that we need to stop this misuse.

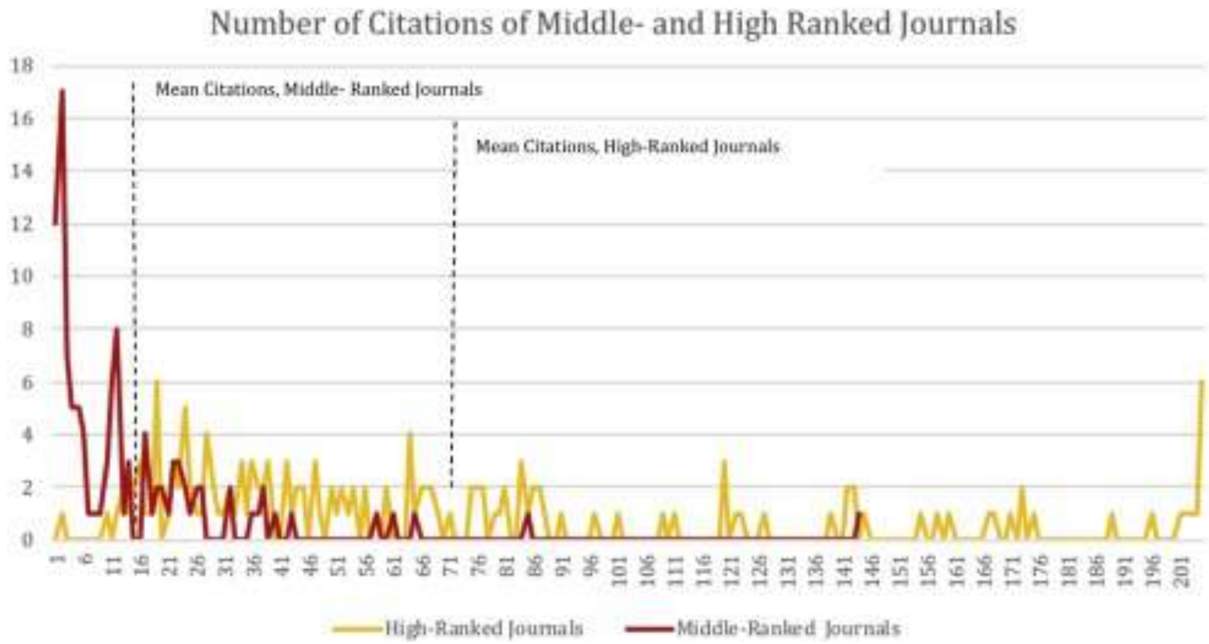
Many other influential scholars[7] and academic institutions have banned the use of JIFs as proxy for the quality of a single article, notably the International

Mathematical Union (2008), the San Francisco Declaration on Research Assessment (DORA (San Francisco Declaration on Research Assessment) and DORA, 2012), the Leiden Manifesto (Hicks et al., 2015), and the Metric Tide report (Wilsdon et al., 2015).

Yearly citation rates and short-term citation windows might be too narrow to evaluate the impact of articles measured by citations. Annual citation rates typically peak after three to five years (International Mathematical Union [IMU], 2008: 7; Mingers, 2008).[8] Perhaps the accumulation of citations across several years shows a less skewed distribution; this might justify evaluating individual articles by the journal in which they were published. Therefore, we undertake a citation analysis of individual articles and use cumulative citations per article over a five-year period, starting in the second year after publication. In contrast to the five-year Journal Impact Factor, we do not consider citations in the year immediately after publishing, because there is typically a citation lag. Instead, we take all articles published in 2010 in nine management journals and add all citations gained per article during the five years from 2012 to 2016. By doing so, we avoid the weakness of short citation windows (Martin, 2016) that favor "shooting stars" over "sleeping beauties" (Mingers, 2008). However, the period is short enough to avoid significant general changes in citation behavior.[9] We take into account three top-tier journals: *The Academy of Management Review* (*AMR*)*, The Journal of Management* (*JM*), and *The Academy of Management Journal* (*AMJ*), which take the first three positions out of 121 ranked by impact factor in the Business category in 2017.[10] As a comparison, we analyze three middle-ranked Journals (ranked 49 to 51): *Research-Technology Management (RTM), Small Business Economics (SBE)* and *Journal of Engineering and Technology Management (JET-M),*[11] and three low-tier journals (ranked 99 to 101): *The Asia Pacific Business*

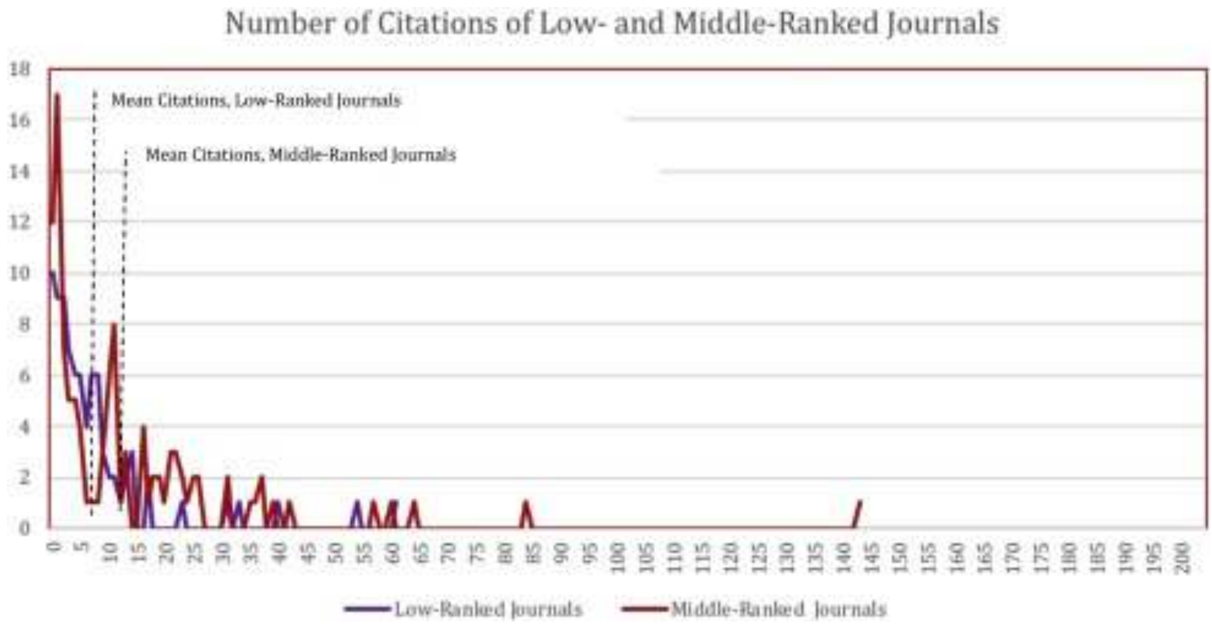
Review (*APBR*), *The Journal of Business Economics and Management*(*JBEM*), and *Organization Dynamics* (*OD*).[12] We count the citations of all 348 articles published in these journals in 2010 from 2012 up to 2016.

Fig. 1, Fig. 2 show the number of articles published in these journals in 2010, the number of citations over the five-year period 2012–2016, the citations per article, and the average number of citations per article. Table A1 in the appendix shows the statistics.



1. Download : Download high-res image (216KB)
2. Download : Download full-size image

Fig. 1. Distribution of Citations in Middle Ranked Journals (red) and in High-Ranked Journals (yellow).



1. Download : Download high-res image (185KB)
2. Download : Download full-size image

Fig. 2. Distribution of Citations in Low Ranked Journals (red) and in Middle-Ranked Journals (blue).

In Fig. 1 the yellow line indicates the citation patterns of the high-ranked journals *AMR*, *JM*, and *AMJ*, comprising 149 articles and 10,294 citations. They reveal that there is still a strong skewness and a long tail of the distribution, even when we consider cumulative citations across five years starting with the second year after publication. The most cited article draws 314 citations, more than four times the average citation rate of 69. A large majority of contributions—no less than 64.4%—are cited below average.

The red line indicates the citation pattern of the middle-ranked journals *RTM, SBE*, and *JET-M*. In total, in these journals 110 articles have been cited 1505 times. This distribution is also skewed due to the fact that 12 articles have not been cited at all, but one single article has been cited 144 times. The average number of citations is 13.7; 67.3% of the articles are cited less than the average.

In Fig. 2, the red line reproduces the citation patterns of the middle-ranked journals (as in Fig. 1). The blue line indicates the distribution of the 84 articles and 641 citations in the low-ranked journals *APBR*, *JBEM,* and *OD*. The citations are also strongly skewed and have a long tail. Of course, the number of citations is much lower than in the high- and middle- ranked journals; the average number of citations being 7.6. Five articles are cited more than 30 times, the maximum is 61. In this group, 65.5% of the articles are cited less than the average.

There is a considerable overlap in the citation distributions between the high-, middle- and low-ranked journals. The least cited article in *AMR* received 15 citations, in *AMJ* 12 citations, and in *JM* 1 citation. To attribute an article that receives 143 citations in a middle-ranked journal (or 61 citations in a low-ranked journal) to be less important than an article cited 1, 12 or 15 times in a high-ranked journal is questionable. One could even argue that being cited from a middle or low-ranked journal has to be valued more highly than being cited from a top journal, since it is harder to be noticed in a low-impact journal (Balaban, 2012).[13]

To sum up, many articles whose frequency of citation is high were published in less well-ranked journals, and vice versa. As we have demonstrated, this is not only true for short-window citations, but also with cumulative citations across five years starting with the second year after publication. Therefore, it is highly problematic to equate publication in “good” academic journals with “good” research and to consider publication in low-ranked journals automatically as signifying less good research.[14]

3. Why are journal rankings still so influential?

Despite the strong criticism, many scholars believe in journal rankings and have even internalized them as part of their identity (Alvesson and Sandberg, 2013). Publishing in a high-impact journal has become far more important than the content of research (e.g. Frey, 2009; Mingers and Willmott, 2013). This might be why the reward center in the brain of authors is activated when they expect a publication in a top journal (Paulus et al., 2015).

Could it be the case that impact factors and journal lists are still so influential because they possess positive qualities that outweigh their disadvantages? Advocates of the "paper quality theory" (Mingers and Xu, 2010) argue in this vein that top journals have more qualified reviewers and have editors who are better able to select promising articles than those of less highly ranked journals. This is certainly correct for journals on average. It is exactly what the JIF establishes, provided citations are taken as a proxy for the scholarly influence of a paper. Moreover, high journal rankings of management journals not only display some discriminatory power in interdisciplinarity, theoretical diversity, and (recombinant) innovativeness (Vogel et al., 2017; but see Wang et al., 2018), but also indicate a minimum threshold of quality. High impact factors also correlate with high rejection rates and thus stronger competition (e.g. Haensly et al., 2008). Further, the strongest driver of citations in management journals is the ranking of the journal itself (Mingers and Xu, 2010), which might be interpreted as a signal of the quality of high-ranked journals.

However, there are two arguments against the "paper quality theory" which assumes that high-ranked journals publish only the best papers (Mingers and Xu, 2010). First, although top journals on average publish more highly cited articles, there is a great deal of randomness in their editorial selections (Rothwell and Martyn, 2000; Bedeian, 2003; Starbuck, 2005; Siler et al., 2015). As discussed, the great majority of articles published in top-tier journals are cited far below the impact factor of the publishing journals. Most articles are cited little. This suggests that even the best referees and editors are able to assess the future impact of an article to only a limited degree. Reviewers' ratings of impact correlate only 0,14 with later citations for published articles (Gottfredson, 1978; Starbuck, 2015). The reason is not any lack of expertise or fairness, though biases may play a role (e.g. Bornmann, 2011). More importantly, it is a consequence of fundamental uncertainty in research (Bush, 1945; Dasgupta and David, 1994; Nelson, 1959, 2004; Stephan, 1996); that is, possible innovations are unknown, outcomes and alternatives are ambiguous,[15] serendipity is ubiquitous,[16] and individual ambiguity-aversion differs much (Krahnen et al., 2014). Such uncertainty is demonstrated by inconclusive reviews (Nightingale and Scott, 2007), low prognostic quality of reviews and low interrater reliability between the judgments of peers (Peters and Ceci, 1982; Starbuck, 2005, 2015; Bornmann, 2011; Nicolai et al., 2015). It is also indicated by empirical findings on the "luck of the reviewer draw" (Cole et al., 1981; Bornmann and Daniel, 2009), which in many cases is decisive for the acceptance or rejection of a grant proposal or paper. This phenomenon is illustrated by rejections of articles by authors who later won the Nobel Prize (Gans and Shepherd, 1994; Campanario, 1996; The Guardian, 2013[17]). This is not very often the case. However, Campanario (1995; 2009) discusses no less than nineteen Nobel class papers in the natural sciences that were rejected or had major difficulties during the review process.

Second, the journal effect theory (Mingers and Xu, 2010) argues that journal rankings activate strong Matthew effects, by which "success breeds success" (Merton, 1968; Starbuck, 2005; Espeland and Sauder, 2007). The high rank of a journal attracts more readers and thus more citations, which leads to a circular causality. This means that, in contrast to what Garfield (1973) intended, the impact factor of a journal has a considerable impact on the average citation rate. This consequence was shown in a natural experiment by Larivière and Gingras (2010). Duplicate articles published in high-ranked journals produced twice as many citations on average as their identical counterparts in lower-ranked journals.

Summing up the arguments, many influential scholars and institutions are justified in their assertion that - as the International Mathematical Union stated - classifying articles according to the ranking of the journals in which they were published is an "insidious misuse" (IMU, 2008: 9). Nevertheless, the role that impact factors and journal quality lists play in the evaluation of single articles has not diminished (e.g. Heckman and Moktan, 2018; Vogel et al., 2017). Baum (2011: 464) statement is still valid: "Typically, a measure found to be ill-conceived, unreliable, and invalid will fall into disrepute and disuse among the members of a scientific community. Remarkably, this has not been the case with the IF among organization theorists; indeed it is, if anything, gaining attention and being applied more frequently…." . Why is this the case? First, a majority of the authors whose papers are accepted for publication benefit from this measure. It is exactly the skewed distribution of citations that is beneficial for many authors. As argued, the quality of two thirds to three quarters of all articles is overestimated if they are evaluated according to the impact factor of the journal in which they were published. Thus, a majority of authors in a good journal can claim to have published well even if their work has been cited little. They are able to adorn themselves with borrowed plumes, while only a minority[18] would benefit from being accepted in a higher-ranked journal. It is not surprising that the majority of winners are not inclined to abolish the present system.

Second, performance indicators tend to establish a "performance paradox" (Gupta and Meyer, 1994; Frost and Brockmann, 2014).[19] Indicators not only cause reactivity (Espeland and Sauder, 2007) but may also cause perverse learning or lock-in effects (Osterloh, 2010). This is the case when people focus on performance indicators but not on the performance they are supposed to indicate. They tend to improve indicators ("playing to the test") without improving the performance characteristics the indicators are designed to measure. This practice may even worsen performance, for instance by goal displacement (Ordonez et al., 2009), gap-spotting research (Alvesson and Sandberg, 2013), and ranking games (Osterloh and Frey, 2014). Once a certain performance indicator has become established, people who have gained success with this indicator will make a strong effort to maintain its relevance, even if it has been proven to be misleading.

Such lock-in effects are reinforced by ever-growing bureaucracies. In many universities, report and reward systems are established that are aligned to journal rankings and impact factors. Research administrators increasingly allocate budgets and funds according to these criteria (e.g. Laudel, 2006; Bleiklie et al., 2015). Because funding inequality has increased strongly (Zhi and Meng, 2016; Katz and Matter, 2017), authors, deans, and research communities have "to play the game" (Macdonald and Kam, 2007; Frost and Brockmann, 2014). As a consequence, a ranking bureaucracy and even a ranking management industry have emerged (Mingers and Willmott, 2013).

Lock-in effects are also reinforced by adaptive expectations. Organizations' members are willing to adopt certain measurement criteria when they assume that others do so. If scholars expect influential scholars or committees to use impact factors as a proxy for quality, they adopt these criteria for their own work. They also direct their attention accordingly. A self-fulfilling prophecy may set in (Ferraro et al., 2005; Espeland and Sauder, 2007).

Lock-in effects might also be strengthened by the fact that the information about the acceptance of a paper is available earlier than that about citation counts. In contrast, citation counts as a proxy for quality need several years to make any sense. The impact factor of a journal provides scholars seemingly with a speedy quality indicator, in particular because impact factors are freely available.[20]

Lastly, it might be argued that no suitable alternatives exist to impact factors and journal lists, which are easy to handle.[21] Because time and resources are limited for assessing the huge amount of research we face, heuristics to select what to read are desirable. However, heuristics may be misleading. As we have demonstrated, this is the case when using quality indicators of journals (such as JIF or quality lists) to evaluate particular articles. We therefore focus on institutional changes inducing the use of more helpful heuristics.

4. Proposals for change

Although the use of journal rankings has been widely criticized, few proposals exist for changing the current practice of performance management in academia. Most concern the individual level. In particular, it has been suggested that the papers should be read instead of relying on journal rankings (e.g. Moed, 2007; DORA (San Francisco Declaration on Research Assessment) and DORA, 2012; Wilsdon et al., 2015; Alberts, 2013; Berg, 2016; Heckman and Moktan, 2018). This is certainly good advice, but hard to put into practice. We first discuss three extant proposals to reform performance evaluation. We then introduce our own suggestion based on the insight that research is characterized by fundamental uncertainty. All four proposals refer to the institutional level.

A first proposal intends to change the academic journal system as a whole. It suggests to evaluate scholarly work through “open post-publication peer review” (Kriegeskorte, 2012; Osterloh and Kieser, 2015). The internet allows manuscripts to be published as they are and to be evaluated ex post. This procedure starts with the publication of a paper in an online public repository. The author asks a senior scholar to try to find two to four reviewers willing to comment publicly on the paper. This creates transparency within the reviewing process and a plurality of perspectives. Some contributions will elicit inspiring debates; others will be ignored. The papers that have inspired the most interesting discussions might be presented to a broader audience as the state of art in special issues. However, unintended consequences may occur. First, the reputation of the senior scholar and of the reviewers will have a great impact on the attention that the paper receives. In contrast, today it is the reputation of a journal that has been acquired for a long time within a research community that counts for the attention for an article. Second, since comments and reviews are conducted publicly, junior scholars may be reluctant to critique the work of senior scholars. In addition, old boys´ networks might play an undesirable role, and cronyism could arise. Ultimately, the system of open post-publication peer review could lead to a ranking of publication outlets that produces similar problems as the evaluation of single articles according to the quality of a journal.

In contrast to the first proposal the following three accept the crucial role of journals to focus on topical and relevant issues. The second proposal suggests that every journal publishing its JIF should also publish the distribution of citations (Larivière et al., 2016). In the meantime, this proposal has been taken on board by Clarivate Analytics.[22] This proposal could apply to journal quality lists in general. For those who believe in citations as a signal of scholarly impact it can be used to reveal the extensive overlap between the citation distributions of different journals. It will broaden awareness of the spread of citations. It can also be used to measure how often an author's publications are cited above (or below) the impact factors of the journals he or she has appeared in. An alternative would be to provide parameters of distribution such as median or inter-quartile ranges, but a visual representation is more powerful. This suggestion meets the demands that editors and reviewers usually make on authors to make their data traceable.[23]

This suggestion has the advantage of being close to current practice and therefore of being accepted widely. It should, however, be taken into account that the time frame used by JIF is too narrow to evaluate a paper´s influence.[24] Moreover, the distribution of citations still relies on the questionable assumption that citations are a good measure of scholarly impact and that the present reviewing and acceptance procedures accurately reveal the "collective wisdom" (Laband, 2013) of the scientific community.

A third proposal is the publication of a manuscript on an "as is" basis (Tsang and Frey, 2007). A paper is reviewed double-blind as usual. The reviewers are given only two options when advising to the editor: to accept or reject the paper. The option to revise and resubmit is ruled out. The editor then decides whether the manuscript is published as it is or not. If the paper is accepted, then it is up to the authors to incorporate the comments of the reviewers into the paper. The editor also publishes a comment that addresses differences of view among reviewers and him- or herself. This suggestion would speed up the review process and the dissemination of new knowledge. It would unburden reviewers from evaluating revised and resubmitted papers. It also would avoid that authors feel as if they were coerced by the reviewers instead of being advised (Bedeian, 2003; Frey, 2003). Most importantly, this suggestion would make clear to both the authors and the readers that being accepted by a high-impact journal is not a universal quality indicator. The editors would be burdened with a higher responsibility than today to achieve and to demonstrate the state of "organized skepticism" (Merton, 1942) and “creative disagreement” (Harnad, 1979) that is at the heart of scholarly work. But it might encourage editors to publish more imaginative studies.

Our own – the fourth - proposal to overcome the performance paradox and the lock-in effect is based on the insight that uncertainty about future success is symptomatic of scholarly work (Bush, 1945; Nelson, 2004; Stephan, 1996). This insight can be liberating (Starbuck, 2015). Therefore, we translate it into the peer review system. Uncertainty can be used to the advantage of scholarship with the following procedure:

When reviewers agree on the excellent quality of a paper, it should be accepted, preferably on an “as is” basis (Tsang and Frey, 2007). Papers perceived unanimously as valueless are rejected immediately. Papers that are evaluated differently by the referees are randomized. Empirical research has found reviewers´ evaluations to be more congruent with poor contributions (Cicchetti, 1991; Bornmann, 2011; Moed, 2007; Siler et al., 2015) and fairly effective in identifying extremely strong contributions (Li and Agha, 2015). However, reviewers’ ability to predict the future impact of contributions has been shown to be particularly limited in the middle range in which reviewers´ judgements conform to a low degree (Fang et al., 2016).[25]Such papers could undergo a random draw.

Why should contributions to which the referees do not agree be randomized? This procedure reduces the “conservative bias”, that is the bias against unconventional ideas. Referees subjectively have more information on research projects that are close to existing knowledge. Moreover, information on those contributions is more consistent. With unorthodox contributions referees have less – and usually inconsistent - information. But such ideas yield may well high returns in the future. Under these circumstances a randomized choice among the unorthodox contributions is advantageous. Brezis (2007) shows in a numerical model that the optimal ranking mechanism is to accept contributions to which all referees have agreed and to reject those that all referees have put on the bottom and the variance is high. [26] It is the different level and different consistency of information between conventional and unorthodox contributions that is key to focal randomization among papers that referees disagree upon. Gilles (2008) and Engwall (2014) argue in a similar vein. They refer to the theory of statistical tests involving two types of error: type I errors ("reject errors") implying that a correct hypothesis is rejected, and type 2 errors implying that a false hypothesis is accepted ("accept errors"). The former matters more than the latter. "Reject errors" stop promising new ideas, sometimes for a long time, while "accept errors" lead to a waste of money, but may be detected soon once published. This is the reason why it is more difficult to identify "reject errors" than "accept errors".[27] To avoid the negative consequences of "reject errors", risks must be diversified. Fang and Casadevall (2016:158) support this argument by stating that "[*j*]ust as passively managed diversified stock portfolios that rely on random fluctuations of the stock market generally outperform active management based on expert predictions, a modified lottery-based funding strategy would maximize the return on society's investment". The suggestion of partly focal randomization of grants has already been put in practice by two big funding agencies.[28] Other research councils share such considerations.[29]

Our proposal applies these insights to the selection of journal articles. Disagreement among journal referee reports matters more than those among those on grant applications. In the latter case referees usually engage in extensive consultation and mutual adjustments before the final decision is made (Reinhart, 2010). Reducing the “conservative bias” by focal randomization of controversial papers not only diversifies risk of rejecting fruitful ideas, but in addition has an incentivizing effect. It encourages researchers to submit unorthodox ideas that otherwise have a hard time being published (e.g. Alvesson and Sandberg, 2013).

Rational scholars might feel uneasy with randomization mechanisms. However, with focal randomization scholars remain in power. They decide which papers are published or rejected immediately and which enter the randomization process. The purposeful use of random mechanisms in academia is not new. It played a role in the 18th century at the University of Basel. Vacant professorial chairs were filled by lot from a list of three candidates (Burckhardt, 1916; Stolz, 1986; Frey and Osterloh, 2015).[30] At that time the main purpose was to weaken old boys´ networks. Today the main purpose is to ensure diversity that is crucial for the progress of scholarly work (Starbuck, 2015). It also serves to encourage the submission of unorthodox yet promising ideas. The “tyranny of the top five” and their role in tenure and promotion decisions is de-emphasized, and the signaling function among a diversity of journals is redistributed. These goals are explicitly stated by Nobel Prize laureate Heckman (Heckman and Moktan, 2018: 54). Moreover, Matthew effects and lock-in effects are mitigated.

Our proposal moreover unburdens editors considerably from the problem of dealing with low interrater reliability and contradictory reviews. In contrast to the unintended randomness attributed to the peer review process (e.g. Peters and Ceci, 1982; Starbuck, 2005; Bornmann and Daniel, 2009; Rothwell and Martyn, 2000; Graves et al., 2011; Smith, 2015; Nicolai et al., 2015), which is sometimes close to an unintended lottery (Rothwell and Martyn, 2000; Bedeian, 2003; Siler et al., 2015), this suggestion applies randomness in a strictly controlled and rational way.

Such a system would also possess some disadvantages. First, random procedures do not differentiate between good and bad quality. This is the reason why they are preceded by a pre-selection based on quality. It is important to note that the better the pre-selection works, the less the quality of the remaining papers can be distinguished. In this case, the variance in quality is reduced. It becomes much harder to decide which is “the best” or the “second best” paper (March and March, 1977; Denrell et al., 2014). Through focal randomization, the seeming disadvantage becomes an advantage, since otherwise personal preferences and unintended randomness might be decisive (Brezis, 2007). Second, random decisions are considered by many people to be “irrational”. However, seemingly rational decisions are often marred by many biases (Kahnemann, 2011). An example is awarding prizes in some competitions, which turns out to be unintentionally random (Ginsburgh and Weyers, 2014). In such cases, the rationality of decision processes is a façade; an intentionally random decision based on mathematical probabilities would be much more rational. Third, more articles of low quality could be submitted if scholars knew that random selection played a role. But it could equally be the case that more unorthodox high-quality articles would be submitted because authors would feel more encouraged than with the present system.

5. Concluding remarks

The present practice of performance management in academia based on journal quality lists and impact factors needs reform. Publication in a “good” journal does not indicate that the article is “good”. Empirical research shows that about two-thirds to three-quarters of all published articles are overvalued by these criteria. In contrast, frequently cited articles which have had the misfortune to be published in low-ranked journals are undervalued. We show that this is true for both short citation windows and five-year spans.

We discuss why the present practice has gained so much influence. We suggest this is the case because a majority of authors benefits unduly from the present system. Moreover, performance paradox effects, lock-in effects, and ranking bureaucracies block reforms. Therefore, appealing to scholars individually is not sufficient to change the present practice of performance management. Instead, proposals are needed for changes at the institutional level that give incentives to mitigate the obsession of top journal publications. We discuss three suggestions made in the literature. The first is to inform scholars regularly about the skewed distribution of citations of articles and to show the overlap in the distributions for different-tier journals. The second, more far-reaching, proposal is “open post-publication peer review”, which abolishes ex-ante double-blind peer reviews. The third proposal is the publication of manuscripts on the basis of double-blind ex-ante reviews but “as-is”.

Our own proposal is the most radical. It is based on the insight that fundamental uncertainty is symptomatic for scholarly work. This is indicated by the low prognostic quality of reviews and the low interrater reliability revealed by many empirical analyses. Our suggestion takes this evidence into account. It suggests the introduction of a partly random mechanism. Focal randomisation takes place after a thorough preselection of articles by peer reviews. Such a rationally founded and well-orchestrated procedure promises to downplay the importance (or even “tyranny”) of top journals and to encourage more unorthodox research than today.

All four proposals could be initiated in an experimental way, preferably as field experiments. Their outcomes could be evaluated after some years. In any case, they serve to enrich the discussion about the inevitable uncertainty of quality indicators in science.

Declaration of Competing Interest

No conflict of interest.

Acknowledgements

We thank Katja Rost and Jolanta Lindenmann for helping us collecting the data and with the empirical analysis.

Appendix A

Table A1. Statistics of Citations in Low-, Middle- and High –Ranked Journals over five years.2012–2016.

Statistics			
Number of Citations			
low-ranked journals	**N**	Valid	83
		Missing	0
	Mean		7.5904
	Std. Error of Mean		1.17660
	Median		5.0000
	Std. Deviation		10.71930
	Variance		114.903
	Minimum		.00
	Maximum		61.00
	Percentiles	**25**	2.0000

Statistics

		50	5.0000
		75	8.0000
middle-ranked journals	**N**	Valid	110
		Missing	0
	Mean		13.6818
	Std. Error of Mean		1.85594
	Median		9.0000
	Std. Deviation		19.46530
	Variance		378.898
	Minimum		.00
	Maximum		143.00
	Percentiles	**25**	1.0000
		50	9.0000
		75	19.2500
high-ranked journals	**N**	Valid	154
		Missing	0
	Mean		71.5000
	Std. Error of Mean		4.53468
	Median		52.5000
	Std. Deviation		56.27387

Statistics

Variance		3166.748
Minimum		1.00
Maximum		205.00
	25	27.7500
Percentiles	**50**	52.5000
	75	87.7500

References

ABS The Association of Business Schools [ABS], 2015

ABS The Association of Business Schools [ABS]

Academic Journal Guide 2015 (2015)

Retrieved September 4, 2017, from

http://www.kfs.edu.eg/com/pdf/20820152253917.pdf

Adler and Harzing, 2009

N.J. Adler, A.-W. Harzing**When knowledge wins: transcending the sense and nonsense of academic rankings**

Acad. Manag. Learn. Educ., 8 (1) (2009), pp. 72-95

CrossRefView Record in ScopusGoogle Scholar

Aguinis et al., 2014

H. Aguinis, D.L. Shapiro, E.P. Antonacopoulou, T.G. Cummings**Scholarly impact: a pluralist conceptualization**

Acad. Manag. Learn. Educ., 13 (4) (2014), pp. 623-639

CrossRefView Record in ScopusGoogle Scholar

Alberts, 2013

B. AlbertsEditorial: impact factor distortions

Science, 340 (6134) (2013), p. 787

Retrieved August 31, 2017, from

http://science.sciencemag.org/content/340/6134/787

CrossRefView Record in ScopusGoogle Scholar

Alvesson and Sandberg, 2013

M. Alvesson, J. Sandberg**Has Management Studies lost its way? Ideas for more imaginative and innovative research**

J. Manag. Stud., 50 (1) (2013), pp. 128-152

CrossRefView Record in ScopusGoogle Scholar

Antonakis et al., 2014

J. Antonakis, N. Bastardoz, Y. Liu, C.A. Schriesheim**What makes papers highly cited?**

Leadersh. Q., 25 (2014), pp. 152-179

ArticleDownload PDFView Record in ScopusGoogle Scholar

Archambault and Larivière, 2009

É. Archambault, V. Larivière**History of the journal impact factor: contingencies and consequences**

Scientometrics, 79 (3) (2009), pp. 639-653

Google Scholar

Balaban, 2012

A.T. Balaban**Positive and negative aspects of citation indices and journal impact factors**

Scientometrics, 92 (2) (2012), pp. 241-247

CrossRefView Record in ScopusGoogle Scholar

Baum, 2011

J.A.C. Baum**Free-Riding on Power Laws: questioning the validity of the Impact Factor as a measure of research quality in organization studies**
Organization, 18 (4) (2011), pp. 449-466
CrossRefView Record in ScopusGoogle Scholar

Bedeian, 2003

A.G. Bedeian**The manuscript review process: The proper roles of authors, referees, and editors**
J. Manag. Inq., 12 (4) (2003), pp. 331-338
View Record in ScopusGoogle Scholar

Berg, 2016

J. Berg**JIFfy pop**
Science, 353 (6299) (2016), p. 523
CrossRefView Record in ScopusGoogle Scholar

Bleiklie et al., 2015

I. Bleiklie, J. Enders, B. Lepori**Organizations as penetrated hierarchies: environmental pressures and control in professional organizations**
Organ. Stud., 36 (7) (2015), pp. 873-896
CrossRefView Record in ScopusGoogle Scholar

Bornmann, 2011

L. Bornmann**Scientific peer review**

Annu. Rev. Inf. Sci. Technol., 45 (1) (2011), pp. 199-245

View Record in ScopusGoogle Scholar

Bornmann and Daniel, 2009

L. Bornmann, H.D. Daniel**The luck of the referee draw: the effect of exchanging reviews**

Learn. Publ., 22 (2) (2009), pp. 117-125

CrossRefView Record in ScopusGoogle Scholar

B
r
e
z
i
s
,
2
0
0
7

E.S. Brezis**Focal randomization: an optimal mechanism for the evaluation of R&D projects**

Sci. Public Policy, 34 (2007), pp. 691-698

CrossRefView Record in ScopusGoogle Scholar

A. Burckhardt**Ueber die Wahlart der Basler Professoren, besonders im 18. Jahrhundert**

Basler Zeitschrift für Geschichte und Altertumskunde, 15 (1916), pp. 28-46

View Record in ScopusGoogle Scholar

H. Buchstein**Demokratie Und Lotterie. Frankfurt Am Main**

Campus Verlag (2009)

Google Scholar

V. Bush**Science: the Endless Frontier. A Report to the President by Vannevar Bush, Director of the Office of Scientific Research and Development. July, 1945**

Retrieved September 4, 2017, from

United States Government Printing Office, Washington, DC (1945)

https://www.nsf.gov/od/lpa/nsf50/vbush1945.htm

Google Scholar

E. Callaway**Beat it, Impact Factor! Publishing elite turns against controversial metric**

Nature, 535 (7611) (2016), pp. 210-211

CrossRefView Record in ScopusGoogle Scholar

J.M. Campanario**Commentary on influential books and journal articles initially rejected because of negative referees evaluations**

Sci. Commun., 16 (1995), pp. 304-325

CrossRefView Record in ScopusGoogle Scholar

J.M. Campanario**Using citation classics to study the incidence of serendipity in scientific discovery**

Scientometrics, 37 (1) (1996), pp. 3-24

View Record in ScopusGoogle Scholar

J.M. Campanario**Rejecting and resisting Nobel class discoveries: accounts by Nobel Laureates**

Scientometrics, 81 (2) (2009), pp. 549-565

CrossRefView Record in ScopusGoogle Scholar

D.V. Cicchetti**The reliability of peer review for manuscript and grant submissions: a cross-disciplinary investigation**

Behav. Brain Sci., 14 (1) (1991), pp. 119-135

View Record in ScopusGoogle Scholar

S. Cole, J.R. Cole, G.A. Simon**Chance and consensus in peer review**

Science, 214 (4523) (1981), pp. 881-886

View Record in ScopusGoogle Scholar

P. Dasgupta, P.A. David**Toward a new economics of science**

Res. Policy, 23 (5) (1994), pp. 487-521

View Record in ScopusGoogle Scholar

J. Denrell, C. Fang, C. Liu**Perspective–Chance explanations in the management sciences**

Organ. Sci., 26 (3) (2014), pp. 923-940

Google Scholar

DORA (San Francisco Declaration on Research Assessment), DORA**(San Francisco Declaration on Research Assessment)**

Retrieved August 30, 2017, from

(2012)

http://www.ascb.org/dora/

Google Scholar

G. Dosi, F. Malerba, B. Ramello, F. Silva**Information, appropriability, and the generation of innovative knowledge four decades after Arrow and Nelson: an introduction**

Ind. Corp. Chang., 15 (6) (2006), pp. 891-901

CrossRefView Record in ScopusGoogle Scholar

L. Engwall**On the quality of quality assessments**

W. Blockmans, D. Engwall, D. Wearie (Eds.), Bibliometrics: Use and Abuse in the Review of Research Performance, Portland Press, London (2014), pp. 93-104

Google Scholar

W.N. Espeland, M. Sauder**Rankings and reactivity: how public measures recreate social worlds**

Am. J. Sociol., 113 (1) (2007), pp. 1-40

CrossRefView Record in ScopusGoogle Scholar

F.C. Fang, A. Bowen, A. Casadevall**NIH peer review percentile scores are poorly predictive of grant productivity**

eLife, 5 (2016), Article e13323, 10.7554/eLife.13323

Google Scholar

F.C. Fang, A. Casadevall**Grantfunding: playing the odds**

Science, 352 (6282) (2016), p. 158

CrossRefView Record in ScopusGoogle Scholar

F. Ferraro, J. Pfeffer, R.I. Sutton**Economics language and assumptions: how theories can become self-fulfilling**

Acad. Manag. Rev., 30 (1) (2005), pp. 8-24

View Record in ScopusGoogle Scholar

B.S. Frey**Publishing as prostitution? – choosing between one's own ideas and academic success**

Public Choice, 116 (1-2) (2003), pp. 205-223

View Record in ScopusGoogle Scholar

B.S. Frey**Economists in the PITS**

Int. Rev. Econ., 56 (4) (2009), pp. 335-346

CrossRefView Record in ScopusGoogle Scholar

B.S. Frey, M. Osterloh**Zufall in der wissenschaft**

konomenstimme (2015)

December 16, 2015. Retrieved from

http://www.oekonomenstimme.org/artikel/2015/12/zufallsauswahl-in-der-wissenschaft/

Google Scholar

B.S. Frey, K. Rost**Do rankings reflect research quality?**

J. Appl. Econ., 13 (1) (2010), pp. 1-38

ArticleDownload PDFCrossRefView Record in ScopusGoogle Scholar

J. Frost, J. Brockmann**When quality is equated with quantitative productivity – Scholars caught in a performance paradox**

Zeitschrift für Erziehungswissenschaft, 17 (6) (2014), pp. 25-45

CrossRefView Record in ScopusGoogle Scholar

I. Fuyuno, D. Cyranoski**Cash for papers: putting a premium on publication**

Nature, 441 (7095) (2006), p. 792

CrossRefView Record in ScopusGoogle Scholar

J.S. Gans, G.B. Shepherd**How are the mighty fallen: rejected classic articles by leading economists**

J. Econ. Perspect., 8 (1) (1994), pp. 165-179

CrossRefGoogle Scholar

E. Garfield**Citation impact depends upon the paper, not the journal!**

Don’t count on citation by association

Curr. Contents Clin. Med., 22 (1973), pp. 5-6

View Record in ScopusGoogle Scholar

V. Ginsburgh, S. Weyers**Nominees, winners, and losers**

J. Cult. Econ., 38 (4) (2014), pp. 291-313

CrossRefView Record in ScopusGoogle Scholar

M. Gittelman, B. Kogut**Does good science lead to valuable knowledge? Biotechnical firm and the evolutionary logic of citation patterns**

Manage. Sci., 49 (4) (2003), pp. 366-382

View Record in ScopusGoogle Scholar

A.H. Goodall**Highly cited leaders and the performance of research universities**

Res. Policy, 38 (7) (2009), pp. 1079-1092

ArticleDownload PDFView Record in ScopusGoogle Scholar

S.D. Gottfredson**Evaluating psychological research reports: dimensions, reliability, and correlates of quality judgments**

Am. Psychol., 33 (10) (1978), pp. 920-934

View Record in ScopusGoogle Scholar

N. Graves, A.G. Barnett, P. Clarke**Cutting random funding decisions**

Nature, 469 (7330) (2011), p. 299

CrossRefView Record in ScopusGoogle Scholar

V. Gupta, M.W. Meyer**The performance paradox**

Res. Organ. Behav., 16 (1994), pp. 309-369

CrossRefView Record in ScopusGoogle Scholar

P.J. Haensly, P.E. Hodges, S.A. Davenport**Acceptance rates and journal quality: an analysis of journals in economics and finance**

J. Bus. Financ. Librariansh., 14 (1) (2008), pp. 2-31

CrossRefGoogle Scholar

D.S. Hamermesh**Citations in economics: measurement, uses, and impacts**

J. Econ. Lit., 56 (1) (2018), pp. 115-156

View Record in ScopusGoogle Scholar

S. Harnad**Creative disagreement**

Sciences (New York), 19 (1979), pp. 18-20

CrossRefView Record in ScopusGoogle Scholar

A.-W. Harzing

Journal Quality List, 62th ed. (2015)

Retrieved from

https://harzing.com/resources/journal-quality-list

A.-W. Harzing, R. van der Wal**A Google Scholar h-index for Journals: an alternative metric to measure journal impact in Economics & Business?**

J. Am. Soc. Inf. Sci. Technol., 60 (1) (2009), pp. 41-46

CrossRefView Record in ScopusGoogle Scholar

J.J. Heckman, S. Moktan**Publishing and Promotion in Economics: the Tyranny of the Top Five**

© NBER Working Paper Series Working (2018)

25093

Google Scholar

D. Hicks, P. Wouters, L. Waltman, S. de Rijcke, I. Rafols**Bibliometrics: the leiden manifesto for research metrics**

Nature, 520 (2015), pp. 429-431

CrossRefView Record in ScopusGoogle Scholar

International Mathematical Union [IMU]**Citation Statistics. A report. Corrected version, 16/12/08**

Retrieved October 4, 2018, from

(2008)

https://www.mathunion.org/imu-net/archive/2008/imu-net-029b

Google Scholar

D. Kahnemann**Thinking, Fast and Slow**

Farrar, Straus and Giroux, New York (2011)

Google Scholar

Y. Katz, U. Matter**On the Biomedical Elite. Inequality and Stasis in Scientific Knowledge Production**

Unpubl. Manuscript, Harvard Medical School and Faculty of Economics and Business, University of Basel (2017)

Google Scholar

F. Knight**Risk, Uncertainty, and Profit**

Mifflin, Boston, New York (1921)

Google Scholar

J.P. Krahnen, P. Ockenfels, Wilde**Measuring Ambiguity Aversion: A Systematic Experimental Approach**

SSRN_ID2457026

(2014)

Google Scholar

N. Kriegeskorte**Open evaluation: a vision for entirely transparent post-publication peer review and rating for science**

Front. Comput. Neurosci., 6 (2012), pp. 1-18

Google Scholar

D.N. Laband**On the use and abuse of economics journal rankings**

Econ. J., 123 (2013), pp. F223-54

CrossRefView Record in ScopusGoogle Scholar

D.N. Laband, R.D. Tollison**Dry holes in economic research**

Kyklos, 56 (2003), pp. 161-174

View Record in ScopusGoogle Scholar

V. Larivière, Y. Gingras**The impact factor's matthew effect: a natural experiment in Bibliometrics**

J. Am. Soc. Inf. Sci. Technol., 61 (2) (2010), pp. 424-427

View Record in ScopusGoogle Scholar

V. Larivière, V. Kiermer, C.J. MacCallum, M. McNutt, M. Patterson, B.Pulverer, S. Swaminathan, S. Taylor, S. Curry**A Simple Proposal for the Publication of Journal Citation Distributions**

Retrieved September 11, 2018, from
(2016)
http://www.biorxiv.org/content/early/2016/09/11/062109
Google Scholar

G. Laudel**The 'quality myth': promoting and hindering conditions for acquiring research funds**
High. Educ., 52 (3) (2006), pp. 375-403
CrossRefView Record in ScopusGoogle Scholar

D. Li, L. Agha**Big names or big ideas: do peer-review panels select the best science proposals?**
Science, 348 (2015), pp. 434-438
CrossRefView Record in ScopusGoogle Scholar

S. Macdonald, J. Kam**Ring a ring o' roses: quality journals and gamesmanship in management studies**
J. Manag. Stud., 44 (4) (2007), pp. 640-655
CrossRefView Record in ScopusGoogle Scholar

S. Macdonald, J. Kam**Counting footnotes: citability in management studies**
Scand. J. Manag., 26 (2) (2010), pp. 189-203
ArticleDownload PDFView Record in ScopusGoogle Scholar

S. Macdonald, J. Kam**The skewed few: people and papers of quality in management studies**

Organization, 18 (4) (2011), pp. 467-475

CrossRefView Record in ScopusGoogle Scholar

B. Manin**The Principles of Representative Government**

Cambridge University Press, Cambridge (1997)

Google Scholar

J.C. March, J.G. March**Almost random careers: the Wisconsin school superintendency, 1940-1972**

Adm. Sci. Q., 22 (3) (1977), pp. 377-409

CrossRefView Record in ScopusGoogle Scholar

P. Martenson, U. Fors, W. Sven-Bertil, U. Zander, G.H. Nilsson**Evaluating research: a multidisciplinary approach to assessing research practice and quality**

Res. Policy, 45 (2016), pp. 593-603

Google Scholar

B.R. Martin**An new editorial team, and a new indicator of where RP stands**

Res. Policy, 44 (1) (2015), pp. 1-3

ArticleDownload PDFView Record in ScopusGoogle Scholar

B.R. Martin**Editorial: Editors´ JIF-boosting stratagems – Which are appropiate and which not?**

Res. Policy, 45 (1) (2016), pp. 1-7

ArticleDownload PDFView Record in ScopusGoogle Scholar

R.K. Merton**Science and technology in a democratic order**

J. Legal Polit. Sociol., 1 (1942), pp. 115-126

Google Scholar

R.K. Merton**The Matthew Effect in Science**

Science, 159 (3810) (1968), pp. 56-63

View Record in ScopusGoogle Scholar

J. Mingers**Exploring the dynamics of journal citations: modelling with Scurves**

J. Oper. Res. Soc., 59 (8) (2008), pp. 1013-1025

CrossRefView Record in ScopusGoogle Scholar

J. Mingers, F. Xu**The drivers of citations in management science journals**

Eur. J. Oper. Res., 205 (2) (2010), pp. 422-430

ArticleDownload PDFView Record in ScopusGoogle Scholar

J. Mingers, H. Willmott**Taylorizing business school research: on the 'one best way' performative effects of journal ranking lists**

Hum. Relat., 66 (8) (2013), pp. 1051-1073

CrossRefView Record in ScopusGoogle Scholar

H.F. Moed, T.N. Van Leeuwen**Impact factors can mislead**

Nature, 381 (1996), p. 186

View Record in ScopusGoogle Scholar

H.F. Moed**The future of research evaluation rests with an intelligent combination of advanced metrics and transparent peer review**

Sci. Public Policy, 34 (8) (2007), pp. 575-583

CrossRefView Record in ScopusGoogle Scholar

R.R. Nelson**The simple economics of basic scientific research**

J. Polit. Econ., 67 (3) (1959), pp. 297-306

CrossRefGoogle Scholar

R.R. Nelson**The market economy, and the scientific commons**

Res. Policy, 33 (3) (2004), pp. 455-471

ArticleDownload PDFView Record in ScopusGoogle Scholar

A.T. Nicolai, S. Schmal, Ch Schuster**Interrater reliability of the Peer review process in management journals**

I.M. Welpe, J. Wollersheim, S. Ringelhan, M. Osterloh (Eds.), Incentives and Performance. Governance of Research Organizations, Cham: Springer (2015), pp. 107-120

CrossRefView Record in ScopusGoogle Scholar

P. Nightingale, A. Scott**Peer review and the relevance gap: ten suggestions for policy-makers**

Sci. Public Policy, 34 (8) (2007), pp. 543-553

CrossRefView Record in ScopusGoogle Scholar

L.D. Ordonez, M.E. Schweitzer, A.D. Galinsky, M.H. Bazerman**Goals gone wild: the systematic side effects of overprescribing goal setting**

Acad. Manag. Perspect., 23 (1) (2009), pp. 6-16

View Record in ScopusGoogle Scholar

M. Osterloh**Governance by numbers. Does it really work in research?**

Analyse & Kritik: Zeitschrift für Sozialtheorie, 32 (2) (2010), pp. 267-283

View Record in ScopusGoogle Scholar

rey, 2014

M. Osterloh, B.S. Frey**Ranking games**

Eval. Rev., 39 (1) (2014), pp. 102-129

Google Scholar

and Kieser, 2015

M. Osterloh, A. Kieser**Double-blind Peer review: how to Slaughter a sacred cow**

I. Welpe, J. Wollersheim, S. Ringelhan, M. Osterloh (Eds.), Incentives and Performance. Governance of Research Organizations, Cham: Springer (2015), pp. 307-324

CrossRefView Record in ScopusGoogle Scholar

swald, 2007

A.J. Oswald**An examination of the reliability of prestigious scholarly journals: evidence and implications for decision-makers**

Economica, 74 (293) (2007), pp. 21-31

CrossRefView Record in ScopusGoogle Scholar

Paulus et al., 2015

F.M. Paulus, L. Rademacher, T.A.J. Schäfer, L. Müller-
Pinzler, S. Krach**Journal impact factor shapes scientists' reward signal in the prospect of publication**
PLoS One, 10 (11) (2015)
Retrieved September 11, 2018, from
http://journals.plos.org/plosone/article?id=10.1371/journal.pone.0142537
Google Scholar

Peters and Ceci, 1982
D. Peters, S.J. Ceci**Peer review practices of psychological journals: the fate of published articles, submitted again**
Behav. Brain Sci., 5 (2) (1982), pp. 187-195
View Record in ScopusGoogle Scholar

Reinhart, 2010
M. Reinhart**Peer review practices: epistemic and social aspects in the assesment of scientific work**
Res. Eval., 19 (5) (2010), pp. 317-331
CrossRefView Record in ScopusGoogle Scholar

Rossner et al., 2007
M. Rossner, H. Van Epps, E. Hill**Show me the data**
J. Cell Biol., 179 (6) (2007), pp. 1091-1092
CrossRefView Record in ScopusGoogle Scholar

Rossner et al., 2008
M. Rossner, H. Van Epps, E. Hill**Irreproducible results: a response to Thomson Scientific**
J. Cell Biol., 180 (2) (2008), pp. 254-255
CrossRefView Record in ScopusGoogle Scholar

Rost et al., 2017

K. Rost, T. Teichert, A. Pilkington**Social network analytics for advanced bibliometrics: referring to actor roles of management journals instead of journal rankings**

Scientometrics, 112 (3) (2017), pp. 1631-1657

CrossRefView Record in ScopusGoogle Scholar

Rothwell and Martyn, 2000

P.M. Rothwell, C.N. Martyn**Reproducibility of peer review in clinical neuroscience. Is agreement between reviewers any greater than would be expected by chance alone?**

Brain, 123 (9) (2000), pp. 1964-1969

View Record in ScopusGoogle Scholar

Seglen, 1997

P.O. Seglen**Why the Impact Factor of journals should not be used for evaluating research**

Br. Med. J., 314 (7079) (1997), pp. 498-502

View Record in ScopusGoogle Scholar

Shao and Shen, 2011

J. Shao, H. Shen**The outflow of academic papers from China: why is it happening and can it be stemmed?**

Learn. Publ., 24 (2) (2011), pp. 95-97

View Record in ScopusGoogle Scholar

Siler et al., 2015

K. Siler, K. Lee, L. Bero**Measuring the effectiveness of scientific gatekeeping**

PNAS, 112 (2) (2015), pp. 360-365

CrossRefView Record in ScopusGoogle Scholar

Singh et al., 2007

G. Singh, K.M. Haddad, S. Chow**Are articles in “top” management journals necessarily of higher quality?**

J. Manag. Inq., 16 (4) (2007), pp. 319-331

CrossRefView Record in ScopusGoogle Scholar

Smith, 2015

R. Smith**Ineffective at any dose? Why peer review simply doesn’t work**

Times Higher Education Supplement (2015)

May 28, 2015 (2205) available at

https://www.questia.com/magazine/1P3-3701977521/ineffective-at-any-dose-why-peer-review-simply-doesn-t

Google Scholar

Starbuck, 2005

W.H. Starbuck**How much better are the most prestigious journals? The statistics of academic publication**

Organ. Sci., 16 (2) (2005), pp. 180-200

CrossRefView Record in ScopusGoogle Scholar

Starbuck, 2015

W.H. Starbuck**Issues and trends in publishing behavioral science: a quarrelsome crew struggling with a disintegrating boat on a stormy sea**

I.M. Welpe, J. Wollersheim, S. Ringelhan, M. Osterloh (Eds.), Incentives and Performance. Governance of Research Organizations, Cham: Springer (2015), pp. 3-18

CrossRefView Record in ScopusGoogle Scholar

Stephan, 1996

P.E. Stephan**The economics of science**

J. Econ. Lit., 34 (3) (1996), pp. 1199-1235

View Record in ScopusGoogle Scholar

Stolz, 19

8

6

P. Stolz**Parteienwettbewerb, politisches Kartell und Tausch zwischen sozioökonomischen gruppen**

Schweizerische Zeitschrift für Volkswirtschaft und Statistik, 122 (1986), pp. 657-675

View Record in ScopusGoogle Scholar

J. Sydow, G. Schreyögg, J. Koch**Organizational path dependence: opening the black box**

Acad. Manag. Rev., 34 (4) (2009), pp. 689-709

View Record in ScopusGoogle Scholar

The Guardian**Dan Shechtman: “Linus Pauling Said I Was Talking Nonsense**

Jan. 6. Retrieved Januar 8, 2018, from

(2013)

https://www.theguardian.com/science/2013/jan/06/dan-shechtman-nobel-prize-chemistry-interview

Google Scholar

E.W.K. Tsang, B.S. Frey**The as-is journal review process: let authors own their ideas**

Acad. Manag. Learn. Educ., 6 (1) (2007), pp. 128-136

View Record in ScopusGoogle Scholar

D. Van Reybrouck**Against Elections. The Case for Democracy**

Penguin Random House, London (2016)

Google Scholar

R. Vogel, F. Hattke, J. Petersen**Journal rankings in management and business studies: What rules do we play by?**

Res. Policy, 46 (2017), pp. 1707-1722

ArticleDownload PDFView Record in ScopusGoogle Scholar

M.L. Wallace, V. Larivière, Y. Gingras**Modeling a century of citation distributions**

J. Informetr., 3 (4) (2009), pp. 296-303, 10.1016/j.joi.2009.03

ArticleDownload PDFView Record in ScopusGoogle Scholar

J.P. Walsh**2010 Presidential address: embracing the sacred in our secular scholarly world**

Acad. Manag. Rev., 36 (2) (2011), pp. 215-234

View Record in ScopusGoogle Scholar

L. Waltman**A review of the literature on citation impact indicators**

J. Informetr., 10 (2016) (2016), pp. 365-391

ArticleDownload PDFView Record in ScopusGoogle Scholar

J. Wang, R. Veugelers, Stephan P**Bias against novelty in science: A cautionary tale for users of bibliometric indicators**

Res Policy, 46 (2017), pp. 1416-1436

ArticleDownload PDFView Record in ScopusGoogle Scholar

H. Willmott**Journal list fetishism and the perversion of scholarship: reactivity and the ABS list**

Organization, 18 (4) (2011), pp. 429-442

CrossRefView Record in ScopusGoogle Scholar

J. Wilsdon, L. Allen, E. Belfiore, P. Campbell, S. Curry, S. Hill, R. Jones, R.Kain, S. Kerridge, M. Thelwall, J. Tinkler, I. Viney, P. Wouters, J. Hill, B. Johnson**The Metric Tide: Report of the Independent Review of the Role of Metrics in Research Assessment and Management**

Retrieved September 11, 2018, from
(2015), 10.13140/RG.2.1.4929.1363
http://blogs.lse.ac.uk/impactofsocialsciences/files/2015/07/2015_metrictide
Google Scholar

Q. Zhi, T. Meng**Funding allocation, inequality, and scientific research output: an empirical study based on the life science sector of Natural Science Foundation of China**
Scientometrics, 106 (2016), pp. 603-628
CrossRefView Record in ScopusGoogle Scholar

See also the *Handelsblatt* Ranking in Germany http://www.handelsblatt.com/politik/konjunktur/vwl-ranking/.

2

https://www.leopoldina.org/de/publikationen/detailansicht/publication/good-practice-in-the-evaluation-of-researchers-and-research-programmes-2017/

3

For a review of the literature on different citation impact indicators see Waltman (2016).

4

See e.g. the extensive model for evaluating research quality by Martenson et al. (2016).

5

In addition, many other criticisms have been leveled at the robustness of the journal impact factor, such as that JIFs are field specific, vary with the type of paper, include self-citations, can be manipulated, and are calculated from data that are neither transparent nor openly available to the public; see Martin (2015; 2016).

6

Academy of Management Journal, *Administrative Science Quarterly*, *Organization Science*, *Journal of Management Studies*, and *Organization Studies.*

7

See most prominently the panel discussion among five famous economists (Georges Akerlof, Angus Deaton, Drew Fudenberg, Lars Hansen, James Heckman), among them four Nobel Prize laureates, at the American Economic Association Annual Meeting January 7, 2017 in Chicago on "Publishing and Promotion in Economics: The Curse of the Top Five", https://www.youtube.com/watch?v=PqdKMQNXM2A.

8

Conversely, it has been shown that articles that are not cited within five years are unlikely to be remembered later (Gittelman and Kogut, 2003).

9

Citation practices have evolved over time. Citations per article approximately doubled between 1980 and 2004 (see Wallace et al., 2009). In management journals, impact factors have evolved accordingly, see e.g. Walsh (2011). This problem arises when considering Oswald's (2007) study, which analyzed the cumulative citations of articles in six journals in economics across 25 years. He found that five articles in two top journals had not been cited once during that time, whereas some articles in lower-ranked journals were cited 43 to 199 times. See also Antonakis et al. (2014). They found that 7 percent of all articles published in *The Leadership Quarterly* from 1990 to 2012 were never cited.

10

The two-year impact factors of these journals in 2017 are 9.4, 7.7, and 7.4, respectively.

11

The two-year impact factors of these journals in 2017 are 1,796, 2.857 and 2.686, respectively.

12

The two-year impact factors of these journals in 2017 are 1.0, 0.97, and 0.93, respectively.

13

This does not mean that we agree with the assumption that high citation rates are a measure of scholarly quality. Instead, we intend to demonstrate that if one adheres to impact factors one has agreed ex ante on citation as a proxy of quality.

14

We concentrate on journal rankings according to the JIF. Other kinds of journal list such as the British ABS list and the h-index for journals might lead to different journal rankings. In particular the h-index for journals provides a more accurate measure of journal quality than JIF (Harzing and van der Wal, 2009; Martin, 2015). However, the problem remains that evaluating single articles based on the quality of the publishing journal leads in the majority of cases to incorrect assessments, due to the skewed distribution of citations (e.g. Hamermesh, 2018).

15

in the sense of Knightian uncertainty (Knight, 1921), see e.g. Dosi et al. (2006).

16

that is, search might lead to results far from the expected ones.

17

In this article, Daniel Shechtman, the Nobel prize winner for chemistry in 2011, talks about the massive initial rejection of his research even by a former Nobel prize winner.

18

except the authors in the highest-ranked journal

19

The performance paradox literature argues similarly as the literature on organizational path dependencies, see e.g. Sydow et al. (2009). However, path dependencies usually start with a useful innovation. This is not the case with the JIF as a performance indicator for single articles.

20

Impact factors are readily available, but unfortunately, they are not easy to check. The data used by the providers of the JIF are not open to the public, see Martin (2016).

21

There are suggestions to use other indicators than impact factors, (e.g. Rost et al., 2017) or to apply a mix of different indicators (Aguinis et al., 2014). These suggestions are welcome; however, they are not easy to handle.

22

See https://clarivate.com/blog/science-research-connect/the-2018-jcr-release-is-here/

23

In earlier times the data that Thomson Reuter uses to produce the JIF were not openly available, and efforts to replicate individual impact values had failed (Rossner et al., 2007, 2008).

24

This is the reason why our own analysis presented above uses cumulative citations over a five year time span.

25

Li and Agha (2015) as well as Fang et al. (2016) refer to grant applications.

26

Brezis (2007) refers to R&D projects.

27

Engwall (2014) argues that „reject errors "will become larger the higher the percentage of desk rejections is. He presumes that due to „reject errors "the most innovative research will be found in low impact factors. See for empirical evidence Siler et al. (2016).

28

Volkswagen Foundation, https://www.volkswagenstiftung.de/en/funding/our-funding-portfolio-at-a-glance/experiment Health Research Council of New Zealand Explorer Grants. http://www.hrc.govt.nz/funding-opportunities/researcher-initiated-proposal/explorer-grants;

29

E.g. German Council of Science and Humanities https://www.wissenschaftsrat.de/index.php?id=1405&L=;

30

In political governance too, mixed procedures of random elements and voting were common, for instance in classical Athens and in medieval Venice and Florence (Manin, 1997; Buchstein, 2009; Van Reybrouck, 2016).

View Abstract