Article-level assessment of influence and translation in biomedical research

George M. Santangelo*
Office of Portfolio Analysis, Division of Program Coordination, Planning, and Strategic Initiatives, National Institutes of Health, Bethesda, MD 20892

ABSTRACT Given the vast scale of the modern scientific enterprise, it can be difficult for scientists to make judgments about the work of others through careful analysis of the entirety of the relevant literature. This has led to a reliance on metrics that are mathematically flawed and insufficiently diverse to account for the variety of ways in which investigators contribute to scientific progress. An urgent, critical first step in solving this problem is replacing the Journal Impact Factor with an article-level alternative. The Relative Citation Ratio (RCR), a metric that was designed to serve in that capacity, measures the influence of each publication on its respective area of research. RCR can serve as one component of a multifaceted metric that provides an effective data-driven supplement to expert opinion. Developing validated methods that quantify scientific progress can help to optimize the management of research investments and accelerate the acquisition of knowledge that improves human health.

The modern biomedical research enterprise is a complex ecosystem, encompassing topics as diverse as social psychology and physical chemistry. Optimizing the management of biomedical research projects, including the >70,000 awards made by the National Institutes of Health (NIH) each year, requires a deep understanding of this complexity on both a qualitative and quantitative level. Administrators in academia face a similar challenge as they pursue the most promising opportunities when hiring faculty and making capital investments. As decision makers have increasingly turned to metrics for assistance in pruning these large and elaborate decision trees, the ongoing use of inadequate metrics has fomented numerous protests (van Diest *et al.*, 2001; Colquhoun, 2003; Cherubini, 2008; Papatheodorou *et al.*, 2008; Bertuzzi and Drubin, 2013; Cagan, 2013; Eisen *et al.*, 2013; Schekman, 2013; Alberts *et al.*, 2014; Casadevall and Fang, 2014, 2015; Casadevall *et al.*, 2016; Collins and Tabak, 2014; Ioannidis and Khoury, 2014; Pierce, 2014;

Begley and Ioannidis, 2015; Bowen and Casadevall, 2015; Fang and Casadevall, 2015; Berg, 2016; Bohannon, 2016; Callaway, 2016; Lariviere *et al.*, 2016). Widespread concern about the use of flawed metrics derives not only from an awareness of their limitations but also from the reality that careers, and ultimately perhaps the scientific enterprise at large, are at stake.

GETTING BEYOND JOURNAL IMPACT FACTOR: VALIDATED METRICS CAN MAKE A POSITIVE CONTRIBUTION

The Journal Impact Factor (JIF) is one example of a statistically flawed measurement of citation activity that has seen broad adoption over the past few decades as a proxy for both the quality and impact of research publications. The technical shortcomings of JIF as a metric have been amply documented (Price, 1976; Seglen, 1997; Nature, 2005, 2013; Alberts, 2013; Cagan, 2013; Johnston, 2013; Misteli, 2013; Pulverer, 2013; Van Noorden, 2014; Berg, 2016; Bohannon, 2016; Callaway, 2016; Hutchins *et al.*, 2016; Lariviere *et al.*, 2016). Most importantly, JIF is mathematically invalid because it is calculated as the *average* number of times the entire collection of articles in a given journal is cited, but in reality, citations follow a *log-normal* rather than a Gaussian distribution (Stringer *et al.*, 2008). This means that two articles in the same journal—indeed, even in the same issue of that journal—can differ in citation activity by up to three logs. Field-specific differences, both in citation activity and in access to high-profile publication venues, further invalidate the use of JIF to assess the influence of a scientist's research at the article level. JIF is also of little or no value in comparing research outputs

of developed and developing nations or different types of institutions because meaningful apples-to-apples assessments require an article-level metric that includes a benchmarking step.

That JIF is an inappropriate tool for selecting winners in the intense competition for status and resources is highlighted by the observation that everything published in journals with a high JIF (≥28), taken together, accounts for only a small fraction of the most influential articles (Bertuzzi, 2015; Hutchins *et al.*, 2016). (Note that influence, not impact or quality, is what citation activity measures; more about that later.) Awareness of the flaws in JIF and the accompanying prisoner's dilemma (Axelrod and Hamilton, 1981; Erren *et al.*, 2016), in which scientists failing to pursue publication in high-JIF journals run the very real risk of being outcompeted by their peers, date at least as far back as 1997, and so do suggestions for combatting what was already by then a growing problem (Seglen, 1997). Despite this long-standing criticism, JIF has persisted in the absence of a quantitative approach that is a collectively agreed-upon alternative.

What method could replace JIF as a means of evaluating scientific productivity? Some have argued that the only acceptable solution is a return to the tradition of making career-determining judgments by relying exclusively on expert opinion, supplemented in the time-honored manner by peer recommendations, informal discussions, and other subjective criteria. Although this would seem to be an ideal solution, there are many circumstances in which it is unworkable in practice. Decision makers sometimes lack the bandwidth to review thoroughly the entire corpus of publications relevant to a given decisional task, whether they are considering a large cohort of applicants for an open faculty position, developing a funding opportunity announcement, or reviewing applications in a peer review meeting. Exponential expansion of the number of entries in PubMed over the past few decades, culminating in the current rate of more than one million publications per year (Figure 1), can represent a serious challenge to experts who wish to rely exclusively on

manual inspection when assessing contributions from a group of applicants to their chosen fields of research. This is especially true because many of these same experts, in addition to judging grant applications and vetting potential future colleagues, are regular reviewers of submissions to a variety of journals and conferences and are also occasionally called upon to review promotion and tenure dossiers. Indeed, the time and effort occupied by this portion of the typical workload of principal investigators, which makes a marginal contribution to their own career advancement and is done out of a sense of duty, may have helped to stimulate the widespread adoption of JIF as a proxy for quality.

Beyond these logistical realities, exclusive reliance on expert opinion has its own methodological drawbacks. Although metrics should never be used as a substitute for human judgment, the latter is far from perfect and can succumb to a wide variety of conscious and subconscious biases (for an excellent discussion of bias in peer review, see Lee *et al.*, 2013). In addition to gender bias—the more favorable treatment of men in peer review (Wenneras and Wold, 1997; Budden *et al.*, 2008; Larivière *et al.*, 2013; Urry, 2015; Lerback and Hanson, 2017; Overbaugh, 2017)—potential biases include prestige bias—the "benefit of the doubt" given to scientists who have gained a cumulative advantage by benefiting from greater and greater prestige as they disproportionately garner a larger and larger share of limited resources (sometimes called the "Matthew effect"; Merton, 1968; Peters and Ceci, 1982; Bornmann and Daniel, 2006); affiliation bias—favoritism based on semiformal or informal relationships between applicants and reviewers that are not addressed by conflict of interest rules (Wenneras and Wold, 1997; Sandström and Hällsten, 2008); content-based bias—preference for those who study similar problems and/or belong to a similar "school of thought" (Ferber, 1986; Travis and Collins, 1991; Lee and Schunn, 2011); and confirmation bias—the tendency to favor ideas and/or evidence that reinforces, rather than challenges, one's views (Ernst *et al.*, 1992; Nickerson, 1998). Well-designed and thoroughly tested metrics have the potential to expose bias where it exists and may therefore ultimately play an important role in assisting experts who might otherwise, however unintentionally, continue to apply those biases systematically when reviewing the work of others. Ironically, one of the most commonly held prestige biases might be the outsized esteem conferred on those who publish in high-JIF journals at the expense of scientists who publish work of equal or greater value elsewhere (Eisen *et al.*, 2013; Eyre-Walker and Stoletzki, 2013).

A MATHEMATICALLY SOUND, VALIDATED ALTERNATIVE: THE RELATIVE CITATION RATIO

As captured in the aphorism "you can't replace something with nothing," reliance on JIF as a proxy for quality might continue unless decision makers are convinced that there is a valid alternative. At the NIH, we therefore sought to develop such an alternative. Before any new metric can serve in this role, it must at a minimum be both statistically and bibliometrically sound and lend itself to careful interpretation within a broader context. The only valid way to assess the influence of publications is at the article level. Replacing journal-level with article-level assessment would place the many highly influential articles that appear in JIF < 28 journals on an equal footing with those in JIF ≥ 28 journals.

To judge fairly whether an individual article is well or poorly cited, it is necessary to identify other works in the same field with which it can be compared. Our metric, the Relative Citation Ratio (RCR), uses an article's co-citation network—that is, the other articles that appear alongside it in reference lists—to define its unique and customized field. An important advantage of this method of field
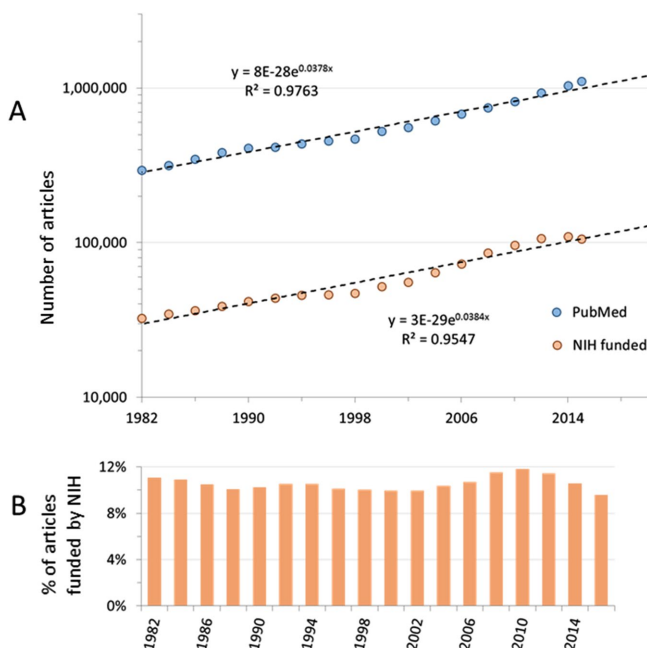

FIGURE 1: Exponential expansion of the number of articles in PubMed (1982–2015). (A) Blue-filled circles, all articles; pink-filled circles, NIH-funded articles. (B) Percentage of articles in PubMed authored by NIH awardees.

normalization is that it leverages the collective experience of publishing scientists, who are the best judges of what relevance one article has to another. RCR also time-normalizes by considering citation rates rather than raw citation counts and, unlike any other available metric, uses a customizable benchmark to allow comparisons between specific peer groups (Figure 2; Hutchins et al., 2016). Extensive testing and validation has demonstrated that the RCR method meets the requirements that should be demanded of any metric before it is adopted as a component in data-driven decision making: RCR values are calculated in a transparent fashion, benchmarked to peer performance, correlated with expert opinion, scalable from small to large collections of publications, and freely accessible (publicly available through our web-based iCite tool (icite.od.nih.gov; Hutchins et al., 2016), with open-source code (https://github.com/NIHOPA/Relative-Citation-Ratio-Manuscript).

Because even thoroughly tested and validated metrics can be misused, it is exceedingly important to define what they do and do not measure. As currently applied to scientific publications, it is a daunting challenge to measure, let alone to predict, the impact of specific advances in knowledge. For this reason, we do not say that RCR values measure the *impact*, but rather the *influence* of each article relative to what is expected, given the scope of its scientific topic. Redefining citations as a measure of influence is important because it addresses valid and often-expressed concerns about the variety of ways in which authors cite the work of others; examples include citing a report that the citer disagrees with or considers flawed, adding citations to please a journal editor (Wilhite and Fong, 2012), and citing more recent articles or reviews rather than the original report of a scientific advance. The concept of influence, broader and more neutral than the lens through which citations are now commonly viewed, also provides important context for understanding how RCR should be interpreted—as an indicator of the value that the relevant academic audience places on transmission of the information contained in one or more publications.

It is also important to keep in mind that citations follow a power law or log-normal distribution (Wang et al., 2013), meaning that as a researcher explores the literature, some of the articles that match his or her area of interest are far more visible than others. Consequently the choice of which article to cite is at least partly informed by the choices that other researchers previously made. The results of a landmark study on the relationship between quality and success in a competitive market illustrate the importance (and potential for bias) that is inherent in these early choices; winners are neither entirely random nor entirely meritocratic, and success is strongly influenced by intangible social variables (Salganik et al., 2006). This means that, although highly influential publications are as a rule important and of high quality, the converse is not necessarily true. RCR is therefore designed to contribute to decision making, not as a substitute for human judgment, but as one component of a multifaceted approach that places expert opinion at the center of assessments of scientific productivity. As we say at the NIH, bodies of work that have higher or lower RCR values than the benchmarked median value of 1.0 should be "flagged for inspection" and interpreted in the broader context of a diverse set of parameters. Note also that, given the inherent noise and intangible variables that affect the output of individuals, NIH decision makers are using RCR and other parameters only to measure the outputs of groups of researchers, not to assess individuals for the purpose of making funding decisions.

NEXTGEN PORTFOLIO ANALYSIS: USING RCR AS ONE COMPONENT OF A DIVERSE METRIC

Others have also highlighted the value of developing multifaceted assessments. For example, Ioannidis and Khoury (2014) incorporated just such a diverse set of parameters into a metric that they termed PQRST (productivity, quality, reproducibility, sharing of data and other resources, and translation. Of course, productivity goes beyond merely totting up the number of article in a portfolio. Therefore one way to adapt their innovative idea is to replace P with I (influence), yielding IQRST; RCR (or weighted RCR; Hutchins et al., 2016) can be used as the I component of the metric. Certain aspects of R and S might also be amenable to quantitation (Schimmack, 2014; Olfson et al., 2017). By its very definition, Q requires a qualitative (i.e., human) judgment. T, for translation, can also be measured by quantifying citations by clinical trials or guidelines, with two caveats. First, several years must elapse to obtain a sufficient signal. Second, this is a measure of direct translation; it does not detect the equally important occurrences of extended translation, defined as a series of citations that connect basic research with impact on human health. For this reason, Weber (2013) proposed the "triangle of biomedicine" to identify and visualize extended translation (Figure 3). We are working to adapt this framework into a tool to analyze a selected portfolio of articles and track its progress toward translation. Development of this and other metrics can contribute to NextGen data-driven decision making by funders of biomedical research.

A further advantage of effective article-based metrics is that they can be validly extended to compare collections of articles. Figure 4 shows a real-world example of the use of article-based metrics to
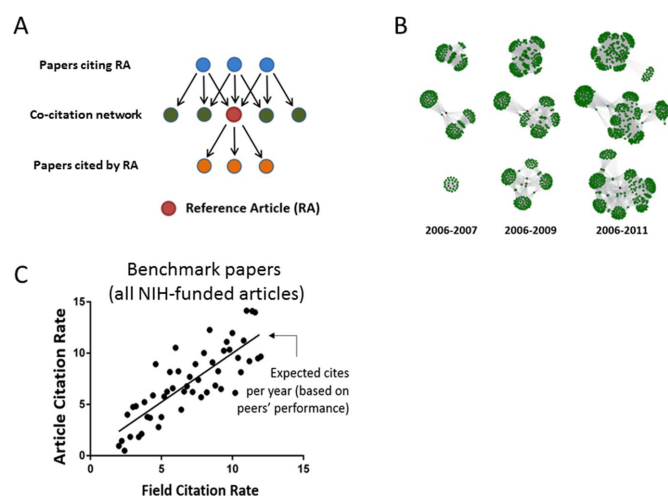

FIGURE 2: Using co-citation networks to calculate RCR. (A) Schematic of a co-citation network. The red circle represents a reference article (RA); blue circles represent articles citing the RA, and orange circles represent articles cited by the RA. The co-citation network of the RA is represented in green and is defined as the collection of articles cited by articles that also cite the RA. (B) Examples of co-citation network growth over time. The growth of the networks of articles co-cited (green circles) in articles that also cite representative RAs (red circles) published in 2006 (three examples from top to bottom). From left to right in each row, the co-citation networks expand between 2006 and 2011. (C) Normalizing article citation rates to calculate RCR. Expected citations rates are generated by benchmarking NIH-funded article citation rates to the citation rates for articles in their co-citation network (field citation rate). RCR is calculated by dividing the article citation rate of each article by its expected citation rate. This figure was adapted from Figures 1 and 3 of Hutchins et al. (2016).
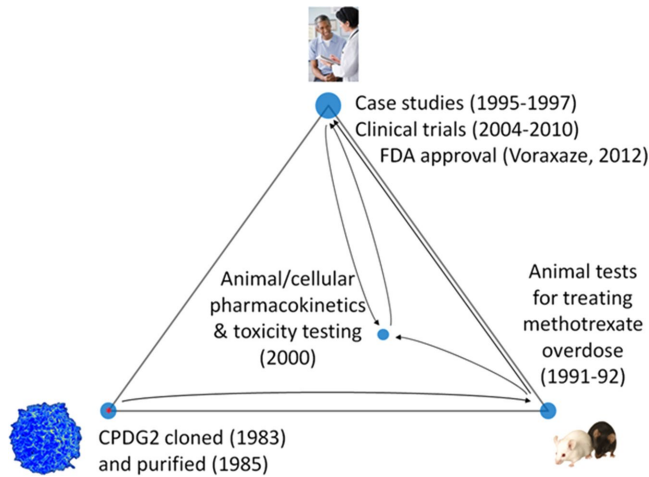
FIGURE 3: Capturing indirect translation with Griffin Weber's triangle of biomedicine. Sample visualization of indirect translation leading to the development of Voraxaze, a Food and Drug Administration–approved drug used to treat toxic plasma methotrexate concentrations in patients with impaired renal function. The vertices of the triangle correspond to cellular/molecular (bottom left), animal (bottom right), and human (top) research.
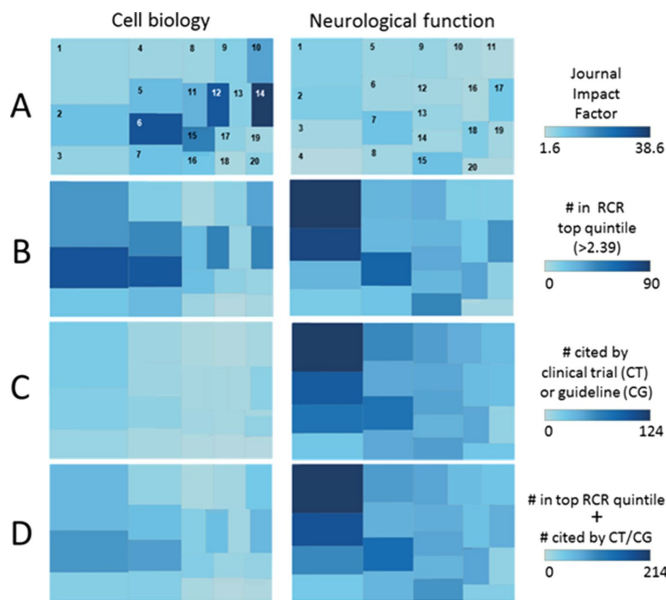

FIGURE 4: Comparison of I and T for NIH-funded articles in the areas of cell biology and neurological function from 2007 to 2012. These treemaps compare several metrics for the 20 journals in which NIH-funded cell biologists (left) and neurobiologists (right) publish their work most frequently. Each rectangle within the map represents one of the journals (numbered 1–20; see Tables 1 and 2 for lists of the journals), and its size is proportional to the number of NIH-funded publications in that journal. The rectangles are shaded from light blue to dark blue, based on increasing values of the corresponding journal for four different metrics: (A) JIF, (B) number of articles with an RCR in the top quintile, (C) number of articles cited by a clinical trial or guideline, and (D) the sum of articles with an RCR in the top quintile and articles cited by a clinical trial or guideline. Relative values can be assessed by comparing the darkness of the blue shading. For example, rectangle 1 under Neurological function (representing the journal *NeuroImage*) has a relatively low JIF (light blue) but a relatively high number of publications with RCR in the top quintile (dark blue).

assess a large number of publications in two distinct but related areas of biomedical research—basic cell biology and neurological function. We randomly sampled 2000 NIH-funded articles from the 20 journals in which NIH investigators from these fields most frequently published their work between 2007 and 2012. As we showed recently, although practitioners in those fields do not have equal access to high-JIF journals (Tables 1 and 2 and Figure 4A), the distribution of the respective article-level RCR values is statistically indistinguishable (Hutchins *et al.*, 2016). This comparison is illustrated in treemaps of the randomly selected articles from Tables 1 and 2. Although there are a comparable number of highly influential articles in both areas of science (those in the top quintile, RCR > 2.39; Figure 4B), there is a striking difference in the extent to which the research exhibits direct translation to clinical trials or guidelines (Figure 4C). Combining these measurements of I and T (Figure 4D) shows just how poorly JIF represents the broadly disseminated contributions that typify progress in biomedical research.

This analysis was repeated for all NIH-funded publications in 2012. Overall, as expected from our previous work (Hutchins *et al.*, 2016), only 8% of the most influential PubMed articles in 2012 (top quintile, RCR > 2.39) were published in high-profile journals (JIF ≥ 28). The treemaps in Figure 5 illustrate this for the journals ranked from 1st to 21st in terms of having published the most articles authored by NIH investigators in 2012 (Table 3). Again, the inadequacy of the JIF metric (Figure 5A) as a way to represent the research that is either most influential (Figure 5B) or directly translates into clinical work (Figure 5C) is apparent. The data for three journals that published similar numbers of NIH-funded articles in 2012—*Clinical Cancer Research*, *Cell*, and *Science*—provide a revealing example. The combined number of influential and clinically relevant articles is comparable for all three journals (Figure 5D), despite the approximately fourfold lower JIF of *Clinical Cancer Research* (Table 3). This modified approach therefore both balances the measurement of influence by doing so at the article level and appropriately credits the expected larger number of clinically relevant articles in the lower-JIF journal.

USING THE SCIENTIFIC METHOD TO PROMOTE THE ADVANCEMENT OF SCIENCE

The goal of replacing journal-level with article-level assessments is actively being pursued at the NIH. Recent analyses conducted by the NIH Office of Extramural Research and the National Institute of General Medical Sciences have used the RCR method to measure outcomes of awarded grants (Basson *et al.*, 2016; Dorsey, 2016; Lauer, 2016a,b). The NIH will continue to promote the shift to article-level assessments in partnership with the scientific community, including collaborators at other domestic and international funding agencies, private foundations, and academic institutions, as part of an ongoing effort to implement data-driven decision making that improves the shared stewardship of research investments. Indeed, use of RCR has already spread outside of the NIH; the Wellcome Trust in the United Kingdom and Fondazione Telethon in Italy have now adopted RCR as part of their suite of portfolio analysis tools (Naik, 2016). Although the scope of this effort traverses the boundaries of biomedical research, the wisdom of the Hippocratic Oath provides a guiding principle: first, do no harm. When comparing portfolios of research investments, it is critical to ensure that those comparisons are "apples to apples." For example, as shown in Figure 4, when measuring at the journal level, we get the wrong answer; although NIH-funded studies of neurological function and cell biology appear in two very different sets of journals, at the article level, one is at least as influential as the other. Similarly, the

Ranka	Journal	Number of publicationsb	2012 JIF	Percentage of publications with RCR > 2.39c	Percentage of publications cited by CT/CGd
1	J Biol Chem	341	4.7	11.1	5.6
2	Proc Natl Acad Sci USA	215	9.7	34.4	7.4
3	Mol Biol Cell	150	4.8	11.3	6.0
4	PLoS One	141	3.7	8.5	5.0
5	J Cell Biol	122	10.8	35.2	5.7
6	Cell	109	32.0	65.1	11.0
7	Curr Biol	106	9.5	21.7	3.8
8	Dev Biol	95	3.9	3.2	2.1
9	Development	95	6.2	9.5	1.1
10	Mol Cell	78	15.3	42.3	6.4
11	Dev Cell	70	12.9	31.4	5.7
12	Science	64	31.0	70.3	7.8
13	J Cell Sci	63	5.9	11.1	6.3
14	Nature	63	38.6	69.8	20.6
15	Nat Cell Biol	53	20.8	39.6	11.3
16	PLoS Genet	50	8.5	10.0	2.0
17	Mol Cell Biol	49	5.4	18.4	10.2
18	Dev Dyn	48	2.6	0.0	0.0
19	Invest Ophthalmol Vis Sci	47	3.4	14.9	23.4
20	Cell Cycle	41	5.3	7.3	0.0

[a]Rank indicates the relative frequency with which NIH-funded cell biologists published their work in the given journal.
[b]Two thousand articles were randomly selected from among the NIH-funded cell biology articles in the given 20 journals.
[c]Number of articles in the top quintile of RCR values.
[d]Number of articles cited by a clinical trial (CT) or clinical guideline (CG).

TABLE 1: Cell biology journals (2007–2012).

Ranka	Journal	Number of publicationsb	2012 JIF	Percentage of publications with RCR > 2.39c	Percentage of publications cited by CG/CTd
1	NeuroImage	228	6.3	39.5	54.4
2	J Neurosci	153	6.9	53.6	62.7
3	Neuropsychologia	133	3.5	19.5	57.1
4	J Acoust Soc Am	125	1.6	11.2	18.4
5	Psychopharmacology (Berl)	125	4.1	20.8	47.2
6	PLoS One	115	3.7	20.9	37.4
7	Biol Psychiatry	111	9.2	58.6	69.4
8	Cognition	99	3.5	18.2	43.4
9	J Cogn Neurosci	92	4.5	31.5	52.2
10	Brain Res	90	2.9	14.4	44.4
11	J Speech Lang Hear Res	89	2.0	13.5	34.8
12	J Neurophysiol	85	3.3	30.6	51.8
13	Psychol Sci	81	4.5	38.3	56.8
14	Dev Sci	74	3.6	33.8	51.4
15	Proc Natl Acad Sci U S A	74	9.7	59.5	64.9
16	Psychiatry Res	73	2.5	21.9	42.5
17	Neurology	71	8.2	54.9	40.8
18	Cereb Cortex	63	6.8	52.4	69.8
19	Behav Brain Res	62	3.3	14.5	17.7
20	Exp Brain Res	57	2.2	7.0	40.4

[a]Rank indicates the relative frequency with which NIH-funded neurobiologists published their work in the given journal.
[b]Two thousand articles were randomly selected from among the NIH-funded neurobiology articles in the given 20 journals.
[c]Number of articles in the top quintile of RCR values.
[d]Number of articles cited by a clinical trial (CT) or clinical guideline (CG).

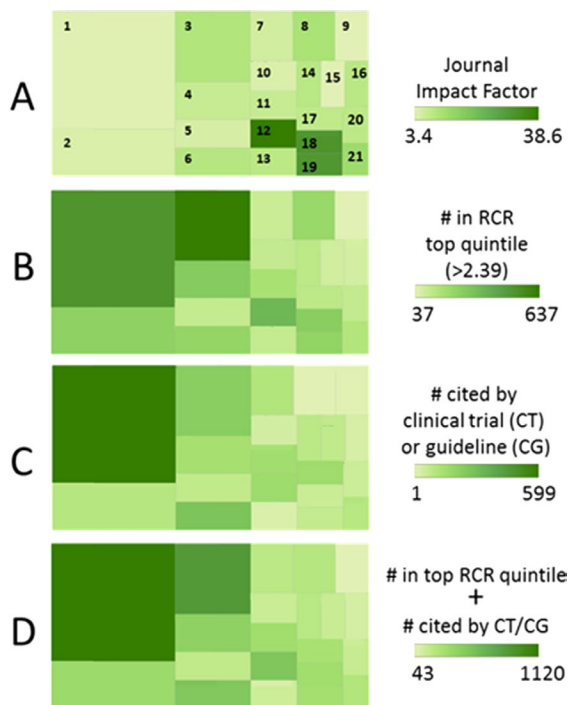TABLE 2: Neurobiology journals (2007–2012).

FIGURE 5: Measuring I and T for all NIH-funded articles in 2012. Rectangles are as described in the Figure 4 legend. The number in each rectangle corresponds to the list of journals in Table 3 (ranked by the relative frequency with which NIH-funded scientists published their work in that journal).

effectiveness of new funding initiatives cannot be properly analyzed without carefully determining the most appropriate methods and control groups. When samples sizes allow, propensity score matching should be used to eliminate confounding variables that can lead to erroneous conclusions. Another caveat is that, as is true for both clinical trials and preclinical research, these studies must be effectively powered by using sufficiently large sample sizes. Fortunately, improvements in computational methodology and database management now readily permit such large-scale analyses; calculation of RCR values for 24 million articles (Hutchins *et al.*, 2016) took less than 1 day to complete. In short, the hallmarks of the scientific method, including due diligence in selecting the appropriate questions, methods, controls, and standards of analysis, are just as essential when attempting to analyze research portfolios and/or track scientific advances.

FUTURE DIRECTIONS

This new area of research that can inform decision making, sometimes termed "science of science," is understandably of great interest to science funders and stakeholders alike. From the policy and program management perspective, it has the exciting potential to guide decisions by revealing overlapping investments, detecting emerging areas, and demarcating research gaps. In doing this work, it is essential not to lose sight of the fact that the most impactful advances in science, those that result in paradigm shifts (Kuhn, 1962), are by their very nature anecdotal, and the resulting ripple effects are difficult to track effectively. Indeed, effective tracking of what Kuhn called "normal science" has the potential to optimize the distribution of research investments in a way that

Rank[a]	Journal	Total number of publications	JIF	Percentage of publications with RCR > 2.39[b]	Percentage of publications cited by CG/CT[c]	Percentage I + T[d]
1	PLoS One	4792	3.7	10.9	12.5	23.4
2	J Biol Chem	1916	4.7	12.9	5.2	18.1
3	Proc Natl Acad Sci USA	1752	9.7	36.4	13.1	49.5
4	J Neurosci	928	6.9	30.4	15.1	45.5
5	J Virol	703	5.1	14.8	10.8	25.6
6	Blood	697	9.1	33.0	39.3	72.3
7	J Immunol.	695	5.5	13.1	15.3	28.4
8	J Am Chem Soc	685	10.7	31.5	0.1	31.6
9	Biochemistry	552	3.4	6.7	1.1	7.8
10	Biochim Biophys Acta	458	4.4	23.1	7.9	31.0
11	NeuroImage	432	6.9	38.9	35.2	74.1
12	Nature	430	38.6	83.5	37.4	120.9
13	Nucleic Acids Res	427	8.3	24.6	5.4	30.0
14	Cancer Res	369	8.7	32.0	23.3	55.3
15	Invest Ophthalmol Vis Sci	366	3.4	20.8	20.2	41.0
16	PLoS Genet	365	8.5	19.2	7.7	26.9
17	Clin Cancer Res	351	7.8	33.9	45.3	79.2
18	Cell	347	32.0	76.7	15.6	92.3
19	Science	341	31.0	69.2	19.9	89.1
20	PLoS Pathog	313	8.1	33.9	15.7	49.6
21	J Clin Invest	284	12.8	49.6	33.1	82.7

[a]Rank indicates the relative frequency with which NIH awardees published their work in the given journal.
[b]Number of articles in the top quintile of RCR values.
[c]Number of articles cited by a clinical trial (CT) or clinical guideline (CG).
[d]Number of articles of high influence (RCR > 2.39; I) plus the number that exhibited direct translation (cited by a clinical trial or guideline; T). Note that a given publication may have high influence and a citation from a clinical trial or guideline, so percentage I + T can be >100%.

TABLE 3: All journals (2012).

increases the likelihood that paradigm-challenging research can flourish.

As founders of a new field of research that can take full advantage of the rapid proliferation of ever more sophisticated computational resources and methodologies, science-of-science scholars are poised to make seminal discoveries that at a minimum can reveal features of normal science and how it progresses. It is increasingly straightforward to develop new methods of analysis, build powerful algorithms, and share them globally. The quality of a wide variety of data fields, including research awards, publications, citations, patents, drugs, and devices, to name but a few, continues to improve, as does the interoperability of the databases and systems that house them. Stewards of research resources have a duty to explore the resulting new opportunities in support of data-driven decision making whenever strong evidence indicates that the use of such methodologies provides an undistorted lens through which to view research investments and the resulting productivity and/or impact. That said, there must be strong evidence that any new methodology has undergone the most rigorous testing to validate its capacity to distinguish between "fool's gold" and real discoveries. It is also crucial to implement such new tools wisely, understanding that no single metric or approach tells the whole story and using the outputs of science-of-science research only to supplement, never to replace, human judgment. The summative result of these efforts will be a bright future for the scientific enterprise as we strive together to optimize the rate of scientific discovery and demonstrate the value of investments in research and the resulting impact on human health and beyond.

REFERENCES

Alberts B (2013). Impact factor distortions. Science 340, 787.

Alberts B, Kirschner MW, Tilghman S, Varmus H (2014). Rescuing US biomedical research from its systemic flaws. Proc Nat Acad Sci USA 111, 5773–5777.

Axelrod R, Hamilton WD (1981). The evolution of cooperation. Science 211, 1390–1396.

Basson J, Lorsch J, Dorsey T (2016). Revisiting the dependence of scientific productivity and impact on funding level. Available at https://loop.nigms.nih.gov/2016/07/revisiting-the-dependence-of-scientific-productivity-and-impact-on-funding-level/ (accessed 21 March 2017).

Begley CG, Ioannidis JP (2015). Reproducibility in science: improving the standard for basic and preclinical research. Circ Res 116, 116–126.

Berg J (2016). JIFfy Pop. Science 353, 523.

Bertuzzi S (2015). A new and stunning metric from NIH reveals the real nature of scientific impact. Available at www.ascb.org/2015/activation-energy/a-new-and-stunning-metric-from-nih-reveals-the-real-nature-of-scientific-impact/ (accessed 21 March 2017).

Bertuzzi S, Drubin DG (2013). No shortcuts for research assessment. Mol Biol Cell 24, 1505–1506.

Bohannon J (2016). Hate journal impact factors? New study gives you one more reason. Science, doi: 10.1126/science.aag0643.

Bornmann L, Daniel H-D (2006). Potential sources of bias in research fellowship assessments: effects of university prestige and field of study. Res Eval 15, 209–219.

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 USA 112, 11335–11340.

Budden AE, Tregenza T, Aarssen LW, Koricheva J, Leimu R, Lortie CJ (2008). Double-blind review favours increased representation of female authors. Trends Ecol Evol 23, 4–6.

Cagan R (2013). The San Francisco declaration on research assessment. Dis Model Mech 6, 869–870.

Callaway E (2016). Beat it, impact factor! Publishing elite turns against controversial metric. Nature 535, 210–211.

Casadevall A, Bertuzzi S, Buchmeier MJ, Davis RJ, Drake H, Fang FC, Gilbert J, Goldman BM, Imperiale MJ, Matsumura P, et al. (2016). ASM journals eliminate impact factor information from journal websites. mBio 7, e01150–16.

Casadevall A, Fang FC (2014). Causes for the persistence of impact factor mania. mBio 5, e00064–00014.

Casadevall A, Fang FC (2015). Impacted science: impact is not importance. mBio 6, e01593–01515.

Cherubini P (2008). Impact factor fever. Science 322, 191.

Collins FS, Tabak LA (2014). Policy: NIH plans to enhance reproducibility. Nature 505, 612–613.

Colquhoun D (2003). Challenging the tyranny of impact factors. Nature 423, 479; discussion, 480.

Dorsey T (2016). P01 outcomes analysis. Available at https://loop.nigms.nih.gov/2016/04/p01-outcomes-analysis/ (accessed 21 March 2017).

Eisen JA, Maccallum CJ, Neylon C (2013). Expert failure: re-evaluating research assessment. PLoS Biol 11, e1001677.

Ernst E, Resch KL, Uher EM (1992). Reviewer bias. Ann Int Med 116, 958.

Erren TC, Shaw DM, Morfeld P (2016). Analyzing the publish-or-perish paradigm with game theory: the prisoner's dilemma and a possible escape. Sci Eng Ethics 22, 1431–1446.

Eyre-Walker A, Stoletzki N (2013). The assessment of science: the relative merits of post-publication review, the impact factor, and the number of citations. PLoS Biol 11, e1001675.

Fang FC, Casadevall A (2015). Competitive science: is competition ruining science? Infect Immun 83, 1229–1233.

Ferber MA (1986). Citations: are they an objective measure of scholarly merit? Signs J Women Cult Soc 11, 381–389.

Hutchins BI, Yuan X, Anderson JM, Santangelo GM (2016). Relative Citation Ratio (RCR): a new metric that uses citation rates to measure influence at the article level. PLoS Biol 14, e1002541.

Ioannidis JP, Khoury MJ (2014). Assessing value in biomedical research: the PQRST of appraisal and reward. J Am Med Assoc 312, 483–484.

Johnston M (2013). We have met the enemy, and it is us. Genetics 194, 791–792.

Kuhn TS (1962). The Structure of Scientific Revolutions, Chicago: University of Chicago Press.

Lariviere V, Kiermer V, MacCallum CJ, McNutt M, Patterson M, Pulverer B, Swaminathan S, Taylor S, Curry S (2016). A simple proposal for the publication of journal citation distributions. bioRxiv doi: 10.1101/062109.

Larivière V, Ni C, Gingras Y, Cronin B, Sugimoto CR (2013). Bibliometrics: global gender disparities in science. Nature 504, 211–213.

Lauer M (2016a). Measuring impact of NIH-supported publications with a new metric: the relative citation ratio. Available at https://nexus.od.nih.gov/all/2016/09/08/nih-rcr/ (accessed 21 March 2017).

Lauer M (2016b). Applying the relative citation ratio as a measure of grant productivity. Available at https://nexus.od.nih.gov/all/2016/10/21/applying-the-relative-citation-ratio-as-a-measure-of-grant-productivity/ (accessed 21 March 2017).

Lee CJ, Schunn CD (2011). Social biases and solution for procedural objectivity. Hypatia 26, 352–373.

Lee CJ, Sugimoto CR, Zhang G, Cronin B (2013). Bias in peer review. J Am Soc Inf Sci Technol 64, 2–17.

Lerback J, Hanson B (2017). Journals invite too few women to referee. Nature 541, 455–457.

Merton RK (1968). The Matthew effect in science. The reward and communication systems of science are considered. Science 159, 56–63.

Misteli T (2013). Eliminating the impact of the impact factor. J Cell Biol 201, 651–652.

Naik G (2016). The quiet rise of the NIH's hot new metric. Nature 539, 150.

Nature (2005). Not-so-deep impact. Nature 435, 1003–1004.

Nature (2013). The maze of impact metrics. Nature 502, 271.

Nickerson RS (1998). Confirmation bias: a ubiquitous phenomenon in many guises. Rev Gen Psychol 2, 175–220.

Olfson M, Wall MM, Blanco C (2017). Incentivizing data sharing and collaboration in medical research—the s-index. J Am Med Assoc Psychiatry 74, 5–6.

Overbaugh J (2017). Gender bias: track revisions and appeals. Nature 543, 40.

Papatheodorou SI, Trikalinos TA, Ioannidis JP (2008). Inflated numbers of authors over time have not been just due to increasing research complexity. J Clin Epidemiol 61, 546–551.

Peters DP, Ceci SJ (1982). Peer-review research: objections and obligations. Behav Brain Sci 5, 246.

Pierce GN (2014). Is open-access publishing the wave of the future in science? Can J Physiol Pharmacol 92, iii.

Price DDS (1976). A general theory of bibliometric and other cumulative advantage processes. J Am Soc Inform Sci 27, 292–306.

Pulverer B (2013). Impact fact-or fiction? EMBO J 32, 1651–1652.

Salganik MJ, Dodds PS, Watts DJ (2006). Experimental study of inequality and unpredictability in an artificial cultural market. Science 311, 854–856.

Sandström U, Hällsten M (2008). Persistent nepotism in peer-review. Scientometrics 74, 175–189.

Schekman R (2013). How journals like Nature, Cell and Science are damaging science. Available at www.theguardian.com/commentisfree/2013/dec/09/how-journals-nature-science-cell-damage-science (accessed 21 March 2017).

Schimmack U (2014). The Replicability-Index (R-Index): quantifying research integrity. Available at https://replicationindex.wordpress.com/2014/12/01/quantifying-statistical-research-integrity-r-index/ (accessed 21 March 2017).

Seglen PO (1997). Why the impact factor of journals should not be used for evaluating research. BMJ 314, 498–502.

Stringer MJ, Sales-Pardo M, Nunes Amaral LA (2008). Effectiveness of journal ranking schemes as a tool for locating information. PLoS One 3, e1683.

Travis GDL, Collins HM (1991). New light on old boys: cognitive and institutional particularism in the peer review system. Sci Technol Hum Values 16, 322–341.

Urry M (2015). Science and gender: scientists must work harder on equality. Nature 528, 471–473.

van Diest PJ, Holzel H, Burnett D, Crocker J (2001). Impactitis: new cures for an old disease. J Clin Pathol 54, 817–819.

Van Noorden R (2014). Transparency promised for vilified impact factor. Nature News, doi:10.1038/nature.2014.15642.

Wang D, Song C, Barabasi AL (2013). Quantifying long-term scientific impact. Science 342, 127–132.

Weber GM (2013). Identifying translational science within the triangle of biomedicine. J Trans Med 11, 126.

Wenneras C, Wold A (1997). Nepotism and sexism in peer-review. Nature 387, 341–343.

Wilhite AW, Fong EA (2012). Scientific publications. Coercive citation in academic publishing. Science 335, 542–543.