scholcommlab

Use of the Journal Impact Factor in academic review, promotion, and tenure evaluations

Erin C. McKiernan[1,*], Lesley A. Schimanski[2], Carol Muñoz Nieves[2], Lisa Matthias[3], Meredith T. Niles[4], and Juan Pablo Alperin[2,5,**]

[1]Departamento de Física, Facultad de Ciencias, Universidad Nacional Autónoma de México
[2]Scholarly Communications Lab, Simon Fraser University
[3]John F. Kennedy Institute, Freie Universität Berlin
[4]Department of Nutrition and Food Sciences, Food Systems Program, University of Vermont
[5]School of Publishing, Simon Fraser University
[*]Corresponding author: emckiernan@ciencias.unam.mx
[**]Corresponding author: juan@alperin.ca

Abstract

The Journal Impact Factor (JIF) was originally designed to aid libraries in deciding which journals to index and purchase for their collections. Over the past few decades, however, it has become a relied upon metric used to evaluate research articles based on journal rank. Surveyed faculty often report feeling pressure to publish in journals with high JIFs and mention reliance on the JIF as one problem with current academic evaluation systems. While faculty reports are useful, information is lacking on how often and in what ways the JIF is currently used for review, promotion, and tenure (RPT). We therefore collected and analyzed RPT documents from a representative sample of 129 universities from the United States and Canada and 381 of their academic units. We found that 40% of doctoral, research-intensive (R-type) institutions and 18% of master's, or comprehensive (M-type) institutions explicitly mentioned the JIF, or closely related terms, in their RPT documents. Undergraduate, or baccalaureate (B-type) institutions did not mention it at all. A detailed reading of these documents suggests that institutions may also be using a variety of terms to indirectly refer to the JIF. Our qualitative analysis shows that 87% of the institutions that mentioned the JIF supported the metric's use in at least one of their RPT documents, while 13% of institutions expressed caution about the JIF's use in evaluations. None of the RPT documents we analyzed heavily criticized the JIF or prohibited its use in evaluations. Of the institutions that mentioned the JIF, 63% associated it with quality, 40% with impact, importance, or significance, and 20% with prestige, reputation, or status. In sum, our results show that the use of the JIF is encouraged in RPT evaluations, especially at research-intensive universities, and indicates there is work to be done to improve evaluation processes to avoid the potential misuse of metrics like the JIF.

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.27638v2 | CC BY 4.0 Open Access | rec: 9 Apr 2019, publ: 9 Apr 2019

Introduction

Originally developed to help libraries make indexing and purchasing decisions for their journal collections (Archambault & Larivière, 2009; Garfield, 2006; Haustein & Larivière, 2015), the Journal Impact Factor (JIF) has moved beyond libraries and into the realm of research evaluation, despite the wide criticisms and well-documented limitations of the metric (e.g., Brembs et al., 2013; Haustein & Larivière, 2015; Kurmis, 2003; Moustafa, 2015; PLoS Medicine Editors, 2006; Seglen, 1997; Sugimoto & Larivière, 2018; The Analogue University, 2019). Even the metric's own creator, Eugene Garfield, made it clear that the JIF is not appropriate for evaluating individuals or for assessing the importance and significance of individual works (Garfield, 1963). Yet, substantial increases in publication rates and the number of academics competing for grants, jobs, and promotions over the past few decades (i.e., 'hypercompetition') have in part led academics to rely on the JIF as a proxy measure to quickly rank journals and, by extension, the articles published in these journals and the individuals authoring them (Casadevall & Fang, 2014). The association between the JIF, journal prestige, and selectivity is strong, and has led academics to covet publications in journals with high JIFs (Harley et al., 2010). Publishers, in turn, promote their JIF to attract academic authors (Hecht et al., 1998; SpringerNature, 2018; Sugimoto & Larivière, 2018).

In some academic disciplines, it is considered necessary to have publications in journals with high JIFs to succeed, especially for those on the tenure track (for review see Schimanski & Alperin, 2018). There are even institutions in some countries that financially reward their faculty for publishing in journals with high JIFs (Fuyuno & Cyranoski, 2006; Quan et al., 2017), demonstrating an extreme but important example of how reliance on this metric may be distorting academic incentives. Even when the incentives are not so clear-cut, faculty still often report intense pressure to publish in these venues (Harley et al., 2010; Tijdink et al., 2016; Walker et al., 2010). Faculty also report that concerns about the JIF and journals' perceived prestige are limiting factors in their adoption of open access publishing (of California Libraries; Schroter et al., 2005; Swan & Brown, 2004), indicating how the effects of the JIF permeate to the broader scholarly publishing ecosystem.

This use — and potential misuse — of the JIF to evaluate research and researchers is often raised in broader discussions about the many problems with current academic evaluation systems (Moher et al., 2018). However, while anecdotal information or even formal surveys of faculty are useful in gauging its effect on the academic system, there is still a lot we do not know about the extent to which the JIF is used in formal academic evaluations. To our knowledge, there have been no studies analyzing the content of university review, promotion, and tenure (RPT) guidelines to determine the extent to which the JIF is being used to evaluate faculty, or in what ways. We therefore sought to answer the following questions: (1) How often is the JIF, and closely related terms, mentioned in RPT documents? (2) Are the JIF mentions supportive or cautionary? and (3) What do RPT documents assume the JIF measures? In the process of answering these questions, our study offered an opportunity to explore the context surrounding mentions of the JIF to qualitatively assess its use in the documents that guide formal evaluation.

Methods

Document collection

This paper reports a set of findings from a larger study (Alperin et al., 2019) for which we collected documents related to the RPT process from a representative sample of universities in the United States and Canada and many of their academic units. A detailed description of the methods for selecting institutions to include in our sample, how we classified them, how we collected documents, and the analysis approach is included in Alperin et al. (2019) and in the methodological note accompanying the public dataset Alperin et al. (2018). Briefly, we used the 2015 edition of the Carnegie Classification of Institutions of Higher Education (Carnegie Foundation for the Advancement of Teaching, 2015) and the 2016 edition of the Maclean's University Rankings (Rogers Digital Media, 2016), which respectively group U.S. and Canadian universities into those focused on doctoral programs (i.e., research intensive; R-type), those that predominantly grant master's degrees (M-type), and those that focus on undergraduate programs (i.e., baccalaureate; B-type). We classified academic units (e.g., department, school, or faculty) within an institution by discipline using the National Academies Taxonomy (The National Academies of Sciences, Engineering, and Medicine, 2006) into three major areas: Life Sciences (LS); Physical Sciences and Mathematics (PSM); and Social Sciences and Humanities (SSH). Additional units that could not be classified as belonging to a single area (e.g., a College of Arts & Sciences) were designated as multidisciplinary. We then used a combination of web searches, crowdsourcing, and targeted emailing to request documents related to the RPT process, including but not limited to collective agreements, faculty handbooks, guidelines, and forms. Some of these documents applied to the institution as a whole, while others applied only to specific academic units.

In the end, we obtained 864 documents related to the RPT process of 129 universities and of 381 academic units. These included documents from 57 R-type, 39 M-type, and 33 B-type institutions. The documents from the 381 academic units came from 60 of the 129 universities in the sample and included documents from 98 (25.7%) LS units, 69 (18.1%) PSM units, 187 (49.1%) SSH units, and 27 (7.1%) multidisciplinary units. However, to avoid pooling academic units from different institution types, and based on sample size considerations, we limited our disciplinary analysis to academic units from R-type institutions: 33 (28%) LS units, 21 (18%) PSM units, 39 (34%) SSH units, and 23 (20%) multidisciplinary units.

Document analysis and coding terminology

The RPT documents were loaded into QSR International's NVivo 12 qualitative data analysis software, where text queries were used to identify documents that mention specific terms. Because the language in RPT documents varies, we first searched all the documents for the words "impact" and "journal", and read each mention to identify terms that may be referencing the JIF. We classified these terms into three groups: (1) direct references to the JIF as a metric; (2) those that reference

journal impact in some way; and (3) indirect but possible references to the JIF. In the first group, we included the terms "impact factor", "impact score", "impact metric", and "impact index". In the second group, we included the terms "high-impact journal", "impact of the journal", and "journal('s) impact". The third group contains a larger number and variety of terms, such as "high-ranking journal", "top-tier journal", and "prestigious journal". For all terms, we considered both singular and plural equivalents. A map of the terms we found and their grouping into the three categories can be seen in Fig. 1. In our analysis, we looked at only the first two groups of terms, as we considered them to be unambiguously about the JIF (group 1) or sufficiently close to the notion of JIF (group 2). The terms in the third group, however, may or may not refer to the JIF. So while these terms could represent examples of ways in which the idea of the JIF is invoked without begin explicit, their mentions were not analyzed further for this study.

The results of each text query for the terms in groups 1 and 2 were placed in an NVivo "node" that contained the text surrounding each of the mentions. We then performed a "matrix coding query" to produce a table with institutions and academic units as rows, terms of interests as columns, and a 1 or a 0 indicating whether the institution or academic unit made mention of the term or not, with the ability to distinguish if the mention appeared in documents that pertain to the whole institution, to one or more academic units, or both. We considered an institution as making mention of a term if the term was present in at least one document from that institution or any of its academic units. More details on this process can be found in Alperin et al. (2019).

Qualitative analysis

We also exported the content of each node for a qualitative analysis of the JIF mentions. In some cases, the software extracted complete sentences, while in other cases it pulled only fragments and we retrieved the rest of the text manually to provide better context. Based on a detailed reading of the text, we classified each of the JIF mentions along two dimensions. First, we classified each mention as either: (1) *supportive* of the JIF's use in evaluations; (2) *cautious*, meaning the document expresses some reservations about the use of the JIF in evaluations; or (3) *neutral*, meaning the mention was neither supportive nor cautious, or not enough information was present in the document to make a judgement. In addition, we read each mention to determine what aspects of research were being measured with the JIF, if specified. Using categories we arrived at inductively, we classified each mention of the JIF as associating the metric with one or more of the following: (i) quality of the research and/or journal; (ii) impact, importance, or significance of the research or publication; (iii) prestige, reputation, or status of the journal or publication; or (iv) left unspecified, meaning the document mentions the JIF, but does not state what the metric is intended to measure. If an institution contained multiple mentions (for example, in two different academic units), it was counted under all the relevant categories.

To arrive at the classification, each mention was independently coded by two of the authors (EM and LM) using the definitions above. After an initial pass, the two coders agreed on all of the

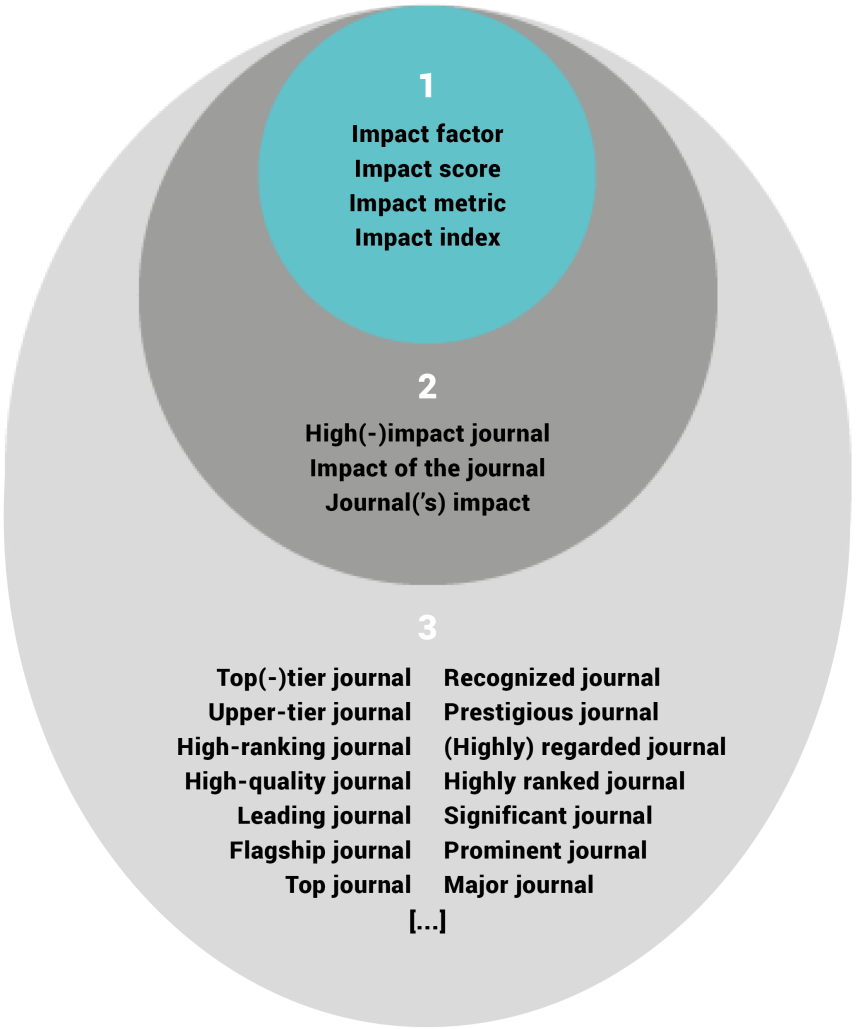

Figure 1: Grouping of terms related to the JIF. Terms found in RPT documents were classified as either: (1) referring directly to the JIF (inner ring); (2) referring in some way to journal impact (middle ring); or (3) indirect but probable references to the JIF. For simplicity, singular versions of each term are shown, but searches included their plural equivalents. Our analysis is based only on those terms found in groups 1 and 2 (the two innermost rings).

classifications for 86% of all mentions. The remaining mentions were independently coded by a third author (LS). In all instances, the third coder agreed with one of the previous two, and this agreement was taken as the final code.

Data availability

We have shared the data on which this paper is based in two different formats: (1) a spreadsheet with all the JIF-related mentions (including repetitions) extracted from the RPT documents, available as part of the larger public dataset (Alperin et al., 2018), and (2) a text document containing the mentions (minus repetitions), with terms of interest color coded and a qualitative assessment of each quote, available as supplemental information. We are not able to share the original RPT documents collected for this study, since the copyrights are held by the universities and academic units that created them. However, for publicly available documents, we included Wayback Machine web archive links to them in the shared spreadsheet.

Results

How often is the JIF mentioned in RPT documents?

While metrics in general are mentioned in RPT documents from 50% of institutions in our sample (Alperin et al., 2019), only 23% (30 of 129) of the institutions mentioned the JIF explicitly or used one of the JIF-related terms (see groups 1 and 2 in Fig. 1) in their RPT documents. The percentage was higher for R-type institutions (23 of 57; 40%) than for either M-type (7 of 39: 18%) or B-type (0 of 33; 0%) institutions (Table 1). Some mentions were found in the institutional-level documents, while others were found at the level of the academic unit (e.g., college, school, or department). Many of the mentions were from different academic units within the same university. Within the R-type institutions, the percentage of academic units that mention JIF-related terms was higher for LS (11 of 33; 33%) and PSM (6 of 21; 29%) than for SSH (8 of 39; 21%) or multidisciplinary units (4 of 23; 17%).

Are the JIF mentions supportive or cautionary?

The majority of mentions of the JIF were supportive of the metric's use in evaluations. Overall, 87% (26 of 30) of institutions that mentioned the JIF did so supportively in at least one of their RPT documents from our sample. Breaking down by institution type, 83% (19 of 23) of R-type and 100% (7 of 7) of M-type institutions had supportive mentions (Table 1). In contrast, just 13% (4 of 30) of institutions overall had at least one mention which expressed caution about using the JIF in evaluations (13% R-type; 14% M-type). Two institutions (University of Central Florida and University of Guelph) had both supportive and cautious mentions of the JIF, but originating from different academic units. Overall, 17% (5 of 30) of institutions had at least one neutral mention (17% R-type; 14% M-type). Examples of supportive and cautious mentions can be found in the following two sections. Examples of neutral mentions are in the supplemental information.

Table 1: Mentions of the JIF in RPT documents, overall and by institution type

		All	R-type	M-type	B-type
How many institutions mention the JIF?					
	n	129	57	39	33
	JIF mentioned	30 (23%)	23 (40%)	7 (18%)	0 (0%)
Are the JIF mentions supportive or cautionary?					
	n	30	23	7	0
	supportive	26 (87%)	19 (83%)	7 (100%)	-
	cautious	4 (13%)	3 (13%)	1 (14%)	-
	neutral	5 (17%)	4 (17%)	1 (14%)	-
What do institutions measure with the JIF?					
	n	30	23	7	0
	quality	19 (63%)	14 (61%)	5 (71%)	-
	impact/importance/significance	12 (40%)	8 (35%)	4 (57%)	-
	prestige/reputation/status	6 (20%)	5 (22%)	1 (14%)	-
	unspecified	23 (77%)	17 (74%)	6 (86%)	-

*Note: Percentages do not sum to one hundred in any given column, since many institutions had more than one JIF mention that could be classified differently. For example, an institution was marked as having a supportive mention if at least one RPT document from that institution, or any of its academic units, had a supportive mention. The same institution could also be counted under 'cautious' if a different academic unit within that institution had such a mention.

What do RPT documents assume the JIF measures?

Associating the JIF with quality

The most common specified association we observed in these RPT documents was between the JIF and quality. Overall, 61% (14 of 23) of R-type and 71% (5 of 7) of M-type institutions that mention the JIF in our sample associate the metric with quality (Table 1). This association can be seen clearly in the guidelines from the Faculty of Science at the University of Alberta (University of Alberta, 2012) that state:

> Of all the criteria listed, the one used most extensively, and generally the most reliable, is the quality and quantity of published work in refereed venues of international stature. Impact factors and/or acceptance rates of refereed venues are useful measures of venue quality...

While some RPT documents recommend using the JIF to determine the quality of a journal, others suggest that this metric can be used to indicate the quality of individual publications. An example of the latter comes from the College of Health Sciences and Professions at Ohio University (Ohio University, 2014):

> Markers of quality of publications may include impact factors of journals, number of citations of published work, and audience of journal.

Other guidelines create their own metrics using the JIF in their calculations and suggest this will incentivize high quality research, as seen in the following example from the Institute of Environmental Sustainability at Loyola University (Loyola University Chicago, 2015):

> For promotion to Professor, the candidate must have an average publication rate of at least one article per year published in peer-reviewed journals in the five-year period preceding the application for promotion. These articles should be regularly cited by other researchers in the field. We will consider both the quality of the journal (as measured by the journal's impact factor, or JIF) as well as the number of citations of each publication. We will employ the metric: Article Impact Factor (AIF) = (JIF * citations) where "citations" represents the number of citations for the particular publication. Employing this metric, faculty have incentive to publish in the highest quality journals (which will increase the JIF) and simultaneously produce the highest quality research manuscripts, potentially increasing the number of citations, and increasing the AIF.

In sum, there are repeated links made in the sampled RPT documents between the JIF, and research, publication, or journal quality.

Associating the JIF with impact, importance, or significance

The second most common specified association we observed in these RPT documents was between the JIF and the impact, importance, or significance of faculty research or publications,

found in 40% (12 of 30) of institutions in our sample. By institution type, 35% (8 of 23) of R-type and 57% (4 of 7) of M-type institutions made this association (Table 1). For example, guidelines from the Department of Psychology at Simon Fraser University (Simon Fraser University, 2015) link the JIF with impact:

> "The TPC [Tenure and Promotion Committee] may additionally consider metrics such as citation figures, impact factors, or other such measures of the reach and impact of the candidate's scholarship."

Promotion and tenure criteria from the University of Windsor (University of Windsor, 2016) link the JIF to publication importance:

> "Candidates will be encouraged to submit a statement that explains the importance of their publications, which may include factors such as journal impact factors, citation rates, publication in journals with low acceptance rates, high levels of readership, demonstrated importance to their field."

Guidelines from the Institute of Environmental Sustainability at Loyola University (Loyola University Chicago, 2015) associate the JIF with scientific significance:

> "Candidates should have at least four manuscripts in peer-reviewed journals published or in-press in the five years preceding application for tenure and promotion to Associate Professor. The length of articles and scientific significance, as measured by citations and journal impact factor, will also be considered, as will authorship on contributions to other scholarly works (e.g., reference and text books)."

In all of the above cases, the value of faculty research or individual publications is being evaluated, at least in part, based on the JIF.

Associating the JIF with prestige, reputation, or status

A third set of mentions of the JIF associated the metric with prestige, reputation, or status, typically referring to the publication venue. Overall, 20% (6 of 30) of institutions in our sample that mentioned the JIF made such an association. As with other concepts, there was variability by institution type, with 22% (5 of 23) of the R-type and 14% (1 of 7) of the M-type having at least one instance of this

association (Table 1). For example, guidelines from the Department of Sociology at the University of Central Florida (University of Central Florida, 2015) link the JIF with prestige:

> It is also true that some refereed journal outlets count for more than others. Publication in respected, highly cited journals, that is, counts for more than publication in unranked journals. The top journals in sociology and all other social sciences are ranked in the Thompson/ISI citation data base (which generates the well-known Impact Factors), in the Scopus data base, and in certain other citation data bases. In general, it behooves faculty to be aware of the prestige rankings of the field's journals and to publish in the highest-ranked journals possible. It is also advisable to include in one's tenure and promotion file information about the Impact Factors or related metrics for the journals where one's papers appear.

An evaluation rubric from the University of Windsor (University of Windsor, 2016) links the JIF with journal reputation:

> a) Publishes in journals or with publishing houses with a strong academic reputation[2]
>
> [2]Departments may wish to provide quantitative metrics such as journal impact factors as an element of their standards. Factors such as low acceptance rates, high levels of readership, importance to the field are also suggestive indicators in assessing quality and reputation.

Similarly, promotion and tenure forms from the University of Vermont (University of Vermont, 2016) associate the JIF with journal status:

> List all works reviewed prior to publication by peers / editorial boards in the field, such as journal articles in refereed journals, juried presentations, books, etc. Indicate up to five of the most important contributions with a double asterisk and briefly explain why these choices have been made. Include a description of the stature of journals and other scholarly venues and how this is known (e.g., impact factors, percentage of submitted work that is accepted, together with an explanation of the interpretation of these measures).

Overall, these documents show a focus on publication venue and use the JIF as a proxy measure for determining how much individual publications should count in evaluations based on where they

are published.

Many mentions do not specify what is measured with the JIF

Lastly, we were left with many instances where the JIF was mentioned without additional information on what it is intended to measure. Such unspecified mentions were found in the RPT documents of 77% (23 of 30) of institutions that mentioned the JIF. These correspond to 74% (17 of 23) of the R-type institutions and 86% (6 of 7) of the M-type institutions with mentions (Table 1). These mentions were often found in research and scholarship sections that ask faculty to list their publications and accompanying information about the publication venues, such as the JIF or journal rank. Some of these documents simply suggest the JIF be included, while others make it a requirement. For example, guidelines from the Russ College of Engineering and Technology at Ohio University (Ohio University, 2015) request the JIF in the following way:

> "List relevant peer-reviewed journal and conference papers published over the last five years (or since last promotion or initial appointment, whichever is less) related to pedagogy or other relevant areas of education. Include the journal's impact factor (or equivalent journal ranking data) and the number of citations of the article(s)."

Not all mentions of the JIF support its use

While the majority of the mentions found in our sample of RPT documents were either neutral or supportive of the JIF, we find that 13% of institutions had at least one mention which cautioned against or discouraged use of the JIF in evaluations. We observed varying levels of caution in these mentions. Some do not critique use of the JIF in general, but rather express concern that JIF data are not as relevant for their discipline as for others. For example, criteria for promotion and tenure from the School of Social Work at the University of Central Florida (University of Central Florida, 2014) state:

> "Journal impact factors will not be a primary criteria for the measurement of scholarly activity and prominence as the academic depth and breadth of the profession requires publication in a multitude of journals that may not have high impact factors, especially when compared to the stem [sic] disciplines."

Similarly, guidelines from the Department of Human Health and Nutritional Sciences at the University of Guelph (University of Guelph, 2008) call the JIF a 'problematic' index and discourage its use while again highlighting disciplinary differences:

> Discussion of journal quality (by those familiar with the field) may be included in the assessment in addition to consideration of the quality of individual research contributions. However, citation analyses and impact factors are problematic indices, particularly in comparisons across fields, and their use in the review process is not encouraged.

Other guidelines, such as those from the Faculty of Veterinary Medicine at the University of Calgary (University of Calgary, 2008), caution against relying solely on the JIF as a measure of quality, but still allow it to be considered:

> Special consideration is to be given to the quality of the publication and the nature of the authorship. Contributions of the applicant must be clearly documented. The reputation and impact of the journal or other publication format will be considered, but takes secondary consideration to the quality of the publication and the nature of the contributions. Impact factors of journals should not be used as the sole or deciding criteria in assessing quality.

Some RPT documents even seem to show disagreement within evaluation committees on the use of the JIF. For example, a document from the Committee on Academic Personnel at the University of California, San Diego (University of California, San Diego, 2015-2016) reads:

> CAP [Committee on Academic Personnel] welcomes data on journal acceptance rates and impact factors, citation rates and H-index, but some CAP members (as do senior staff of scholarly societies) retain various degrees of skepticism about such measures.

None of the RPT documents we analyzed heavily criticize the JIF or prohibit its use in evaluations.

Discussion

To our knowledge, this is the first large-scale study of RPT documents from a representative sample of U.S. and Canadian universities to analyze the use of the JIF in academic evaluations. We found that 23% of institutions in our sample mentioned the JIF or related terms in their RPT documents. The percentage was highest for R-type institutions at 40%, versus either M-type (18%) of B-type (0%) institutions. Mentions were largely supportive of JIF use, with 87% of institutions having at least one supportive mention. In contrast, just 13% of institutions had mentions which expressed

caution about use of the JIF in evaluations. None of the RPT documents we analyzed prohibit its use. With respect to what is being measured with the JIF, the most common positive association we observed was between the JIF and quality, with 63% of institutions making this link. Less common though still observed were associations made between the JIF and impact, importance, or significance (40% of institutions), and prestige, reputation, or status (20%).

How prevalent is the use of the JIF in evaluations?

Mentions of the JIF and related terms in RPT documents are not as ubiquitous as the amount of discussion of current evaluation systems would suggest – 23% of institutions in our sample used these terms explicitly. However, the results differ depending on institution type, which might suggest that the experiences at R-type universities (where mentions of the JIF were most prevalent) play an outsized role in discussions about evaluation. Furthermore, the analysis we present on the terms in groups 1 and 2 of our coding terminology (see Fig. 1) may represent only the tip of the iceberg. That is, while we analyzed only those terms that were very closely related to the JIF, we also observed (but did not analyze) terms such as 'major', 'prestigious', 'prominent', 'highly respected', 'highly ranked', and 'top tier' that may be associated with high JIFs in the minds of evaluators. It is impossible to know how RPT committee members interpret such phrases on the basis of the documents alone, but we suspect that some of these additional terms serve to invoke the JIF without explictly naming it. Take the following examples that leave open for interpretation what measure is used for determining a journal's status (emphasis added):

From the Department of Health Management & Informatics at the University of Central Florida (University of Central Florida, 2014):

> "Both quality and quantity of publications are important. Conventional evidence for quality includes publications in **high-ranking journals** and citation by other scholars."

From the College of Arts and Sciences, University of Vermont (University of Vermont, 2015):

> "Excellence in scholarly research is often demonstrated by the presence of works published in **top tier journals** and academic presses."

Both of these examples do not explicitly mention the JIF (and thus are not counted in our analysis), but do imply the need for some measure for ranking journals. It seems likely, given the ubiquity of the JIF, that some committee members will rely on this metric, at least in part, for such a ranking. In short, counting mentions of a restricted set of terms, as we have done here, is likely

an underestimate of the extent of the use of the JIF in RPT processes. However, we believe the in-depth analysis presented herein provides a glimpse into the current use of the JIF and may indicate how faculty are considering the metric in evaluations, particularly with respect to assessments of quality.

The JIF does not measure quality

The association between the JIF and quality was found in 63% of institutions in our sample. This raises the question, is there evidence that the JIF is a good indicator of quality? Although quality is hard to define, and even harder to measure, there are some aspects of methodological rigor which could be considered indicative of quality, such as sample sizes, experimental design, and reproducibility (Brembs, 2018). What is the relationship between these aspects of a study and the JIF?

Evidence suggests that methodological indicators of quality are not always found in journals with high JIFs. For example, Fraley & Vazire (2014) found that social and personality psychology journals with the highest JIFs tend to publish studies with smaller sample sizes and lower statistical power. Similarly, Munafò et al. (2009) report that higher-ranked journals tend to publish gene-association studies with lower sample sizes and overestimate effect sizes. Analyses of neuroscience and/or psychology studies show either no correlation (Brembs et al., 2013) or a negative correlation (Szucs & Ioannidis, 2017) between statistical power and the JIF. Charles et al. (2009) found that two thirds of a sample of clinical trial studies published in medical journals with high JIFs did not report all the parameters necessary to justify sample size calculations, or had problems with their calculations.

Several studies have also looked at different aspects of experimental design to assess methodological rigor and quality of a study. Chess & Gagnier (2013) analyzed clinical trial studies for ten different indicators of quality, including randomization and blinding, and found that less than 1% of studies met all ten quality criteria, while the JIF of the journals did not significantly predict whether a larger number of quality criteria were met. Barbui et al. (2006) also looked at clinical trial studies and used three different scales that take into account experimental design, bias, randomization, and more to assess quality. The authors found no clear relationship between the JIF and study quality (Barbui et al., 2006).

Others have suggested that reproducibility be used as a measure of quality, since it requires work to provide sufficient methodological care and detail. For example, Bustin et al. (2013) analyzed molecular biology studies and found key methodological details lacking, reporting a negative correlation between the JIF of the journal where the study was published and the amount of information provided in the work. Vasilevsky et al. (2013) analyzed articles from multiple disciplines and found that many resources (e.g., antibodies, cell lines) were not 'uniquely identifiable', reporting no relationship between the JIF and resource identifiability. Mobley et al. (2013) found that around

half of biomedical researchers surveyed reported they had been unable to reproduce a published finding, some from journals with a JIF over 20. Prinz et al. (2011) found, “that the reproducibility of published data did not significantly correlate with journal impact factors” (pg. 2).

Thus, at least as viewed through the aspects above, there is little to no evidence to justify a relationship between the JIF and research quality. A more comprensive review of these issues can be found in Brembs (2018).

Improving academic evaluation

The lack of evidence for linking the JIF with quality, along with the clearly prevalent association that the academic community makes between the two, has given rise to a number of proposals and initiatives to challenge the use of the JIF, promote the responsible use of metrics, and otherwise improve academic evaluations. These include the Leiden Manifesto (Hicks et al., 2015), the Metric Tide report (Wilsdon et al., 2015), the Next-Generation Metrics report (Wildson et al., 2017), and HuMetricsHSS (humetricshss.org), among others (for a review, see Moher et al. (2018)). Inasmuch as this project can be said to be contributing to these efforts by answering questions about the use of the JIF, we provide a brief description of a few of these projects and efforts.

Declaration on Research Assessment

Probably the most well-known such project is the Declaration on Research Assessment (DORA; sfdora.org). DORA outlines some of the limitations of the JIF, and puts forward a general recommendation that those evaluating academics and their research not use it, especially as a “surrogate measure of the quality of individual research articles" (sfdora.org/read). Particularly relevant to our current research is the DORA recommendation that asks institutions to:

> “Be explicit about the criteria used to reach hiring, tenure, and promotion decisions, clearly highlighting, especially for early-stage investigators, that the scientific content of a paper is much more important than publication metrics or the identity of the journal in which it was published.”

In June of 2018, DORA released its two-year strategic plan to expand its work towards improving academic evaluations (DORA Steering Committee, 2018). This work includes spreading awareness of alternatives to the JIF and collecting examples of good evaluation practices from funders, academic societies, and institutions (sfdora.org/good-practices).

To date, DORA has been signed by over 1,200 organizations and nearly 14,000 individuals worldwide. None of the institutions in our sample are DORA signatories, so we were unable to do

any analysis on this, but it would be interesting to study if and how commitment to DORA might be reflected in changes to an institution's RPT documents and evaluation processes.

Libraries taking the lead on responsible metrics

Libraries are at the forefront of promoting the responsible use of metrics. Academic libraries have developed online guides to help faculty learn about the correct uses of different metrics, including the JIF (e.g., Duke University Medical Center Library & Archives; University of Illinois at Urbana Champaign Library; University of Surrey Library; University of York Library). Libraries are also providing in-person advising and training for faculty in publishing and bibliometrics.

There are also several larger-scale library-led efforts. For example, the Association of College & Research Libraries (ACRL) has developed a Scholarly Communication Toolkit on evaluating journals (Association of College & Research Libraries), which outlines several ways to assess journal quality that go beyond metrics like the JIF. LIBER (Ligue des Bibliothèques Européennes de Recherche) has established a Working Group on Metrics, and recently recommended increased training in metrics and their responsible uses (Coombs & Peters, 2017). The Measuring your Research Impact (MyRI) project (http://myri.conul.ie/) is a joint effort by three Irish academic libraries to provide open educational resources on bibliometrics. The Metrics Toolkit is a collaborative project by librarians and information professionals to provide educational information on a variety of metrics, both traditional and alternative, that can be used to evaluate different aspects of research (www.metrics-toolkit.org). In particular, their guide on the JIF outlines the metric's limitations, along with appropriate and inappropriate use cases (http://www.metrics-toolkit.org/journal-impact-factor/).

Conclusions

Overall, our results support the claims of faculty that the JIF features in evaluations of their research, though perhaps less prominently than previously thought, at least with respect to formal RPT guidelines. Importantly, our analysis does not estimate use of the JIF beyond what is found in formal RPT documents, e.g., faculty members who serve on review committees and pay attention to this metric despite it not being explicitly mentioned in guidelines. Future work will include surveying faculty members, particularly those who have served on RPT committees, to learn more about how they interpret and apply RPT guidelines in evaluations and investigate some of the more subjective issues not addressed in this study.

Our results also raise specific concerns that the JIF is being used to evaluate the quality and significance of research, despite the numerous warnings against such use (Brembs, 2018; Brembs et al., 2013; Haustein & Larivière, 2015; Kurmis, 2003; Moustafa, 2015; Seglen, 1997; Sugimoto & Larivière, 2018; The Analogue University, 2019). We hope our work will draw attention to this

issue, and that increased educational and outreach efforts, like DORA and the library-led initiatives mentioned above, will help academics make better decisions regarding the use of metrics like the JIF.

Funding

This study was supported by the Open Society Foundations [OR2016-29841].

Acknowledgements

We are grateful to SPARC, the OpenCon community, the DORA Steering Committee (especially Catriona MacCallum and Anna Hatch), Chealsye Bowley, and Abigail Goben for discussions that shaped and improved this work. We also thank Elizabeth Gadd and Erika Mias, who suggested library guides and projects on responsible metrics to highlight in our Discussion.

Competing interests and disclosures

Erin McKiernan is a member of the DORA Steering Committee and an advisor for the Metrics Toolkit, both volunteer positions. The authors declare they have no other competing interests.

References

[1] Alperin, J., Muñoz Nieves, C., Schimanski, L., McKiernan, E., and Niles, M. Terms and Concepts found in Tenure and Promotion Guidelines from the US and Canada. Harvard Dataverse, V3, 2018. https://doi.org/10.7910/DVN/VY4TJE, UNF:6:PQC7QoilolhDrokzDPxxyQ== [fileUNF].

[2] Alperin, J., Muñoz Nieves, C., Schimanski, L., Fischman, G., Niles, M., and McKiernan, E. How significant are the public dimensions of faculty work in review, promotion, and tenure documents? *eLife*, 2019;8:e42254, 2019. https://doi.org/10.7554/eLife.42254.

[3] Archambault, E. and Larivière, V. History of the journal impact factor: Contingencies and consequences. *Scientometrics*, 79(3):635–649, 2009. https://doi.org/10.1007/s11192-007-2036-x.

[4] Association of College & Research Libraries. Scholarly Communication Toolkit: Evaluating Journals. Accessed October 2018 https://acrl.libguides.com/scholcomm/toolkit/evaluating.

[5] Barbui, C., Cipriani, A., Malvini, L., and Tansella, M. Validity of the impact factor of journals as a measure of randomized controlled trial quality. *Journal of Clinical Psychiatry*, 67(1):37–40, 2006. https://doi.org/10.4088/JCP.v67n0106.

[6] Brembs, B. Prestigious science journals struggle to reach even average reliability. *Frontiers in Human Neuroscience*, 12:37, 2018. https://doi.org/10.3389/fnhum.2018.00037.

[7] Brembs, B., Button, K., and Munafò, M. Deep impact: unintended consequences of journal rank. *Frontiers in Human Neuroscience*, 7:291, 2013. https://doi.org/10.3389/fnhum.2013.00291.

[8] Bustin, S., Benes, V., Garson, J., Hellemans, J., Huggett, J., Kubista, M., Mueller, R., Nolan, T., Pfaffl, M., Shipley, G., et al. The need for transparency and good practices in the qPCR literature. *Nature Methods*, 10(11):1063, 2013. https://doi.org/10.1038/nmeth.2697.

[9] Carnegie Foundation for the Advancement of Teaching. The Carnegie Classifications of Institutions of Higher Education, 2015. http://carnegieclassifications.iu.edu/.

[10] Casadevall, A. and Fang, F. Causes for the persistence of impact factor mania. *mBio*, 5(2): e00064–14, 2014. https://doi.org/10.1128/mBio.00064-14.

[11] Charles, P., Giraudeau, B., Dechartres, A., Baron, G., and Ravaud, P. Reporting of sample size calculation in randomised controlled trials. *BMJ*, 338:b1732, 2009. https://doi.org/10.1136/bmj.b1732.

[12] Chess, L. and Gagnier, J. Risk of bias of randomized controlled trials published in orthopaedic journals. *BMC Medical Research Methodology*, 13(1):76, 2013. https://doi.org/10.1186/1471-2288-13-76.

[13] Coombs, S. and Peters, I. The Leiden Manifesto under review: what libraries can learn from it. *Digital Library Perspectives*, 33(4):324–338, 2017. https://doi.org/10.1108/DLP-01-2017-0004.

[14] DORA Steering Committee. DORA Roadmap: A two-year strategic plan for advancing global research assessment reform at the institutional, national, and funder level, 2018. https://sfdora.org/2018/06/27/dora-roadmap-a-two-year-strategic-plan-for-advancing-global-research-assessment-reform-at-the-institutional-national-and-funder-level/.

[15] Duke University Medical Center Library & Archives. Publication Metrics. Accessed October 2018 https://guides.mclibrary.duke.edu/researchimpact/.

[16] Fraley, R. and Vazire, S. The n-pact factor: Evaluating the quality of empirical journals with respect to sample size and statistical power. *PloS ONE*, 9(10):e109019, 2014. https://doi.org/10.1371/journal.pone.0109019.

[17] Fuyuno, I. and Cyranoski, D. Cash for papers: putting a premium on publication. *Nature*, 441 (792), 2006. https://doi.org/10.1038/441792b.

[18] Garfield, E. Citation indexes in sociological and historical research. *American Documentation*, 14(4):289–291, 1963. https://doi.org/10.1002/asi.5090140405.

[19] Garfield, E. The history and meaning of the journal impact factor. *JAMA*, 295(1):90–93, 2006. https://doi.org/10.1001/jama.295.1.90.

[20] Harley, D., Acord, S., Earl-Novell, S., Lawrence, S., and King, C. Assessing the future landscape of scholarly communication: An exploration of faculty values and needs in seven disciplines. Center for Studies in Higher Education, UC Berkeley, 2010. http://escholarship.org/uc/cshe_fsc.

[21] Haustein, S. and Larivière, V. The use of bibliometrics for assessing research: Possibilities, limitations and adverse effects. In *Incentives and Performance*, pages 121–139. 2015. Available at https://ost.openum.ca/files/sites/132/2017/06/HausteinLariviereIncentives.pdf.

[22] Hecht, F., Hecht, B., and Sandberg, A. The journal “impact factor": a misnamed, misleading, misused measure. *Cancer Genetics and Cytogenetics*, 104(2):77–81, 1998. https://doi.org/10.1016/S0165-4608(97)00459-7.

[23] Hicks, D., Wouters, P., Waltman, L., de Rijcke, S., and Rafols, I. The Leiden Manifesto for research metrics. *Nature*, 520:429–431, 2015. https://doi.org/10.1038/520429a.

[24] Kurmis, A. Understanding the limitations of the journal impact factor. *Journal of Bone & Joint Surgery*, 85(12):2449–2454, 2003. https://bit.ly/2NrWxam.

[25] Loyola University Chicago. The Institute of Environmental Sustainability: Tenure and Promotion Guidelines, 2015.

[26] Mobley, A., Linder, S., Braeuer, R., Ellis, L., and Zwelling, L. A survey on data reproducibility in cancer research provides insights into our limited ability to translate findings from the laboratory to the clinic. *PloS ONE*, 8(5):e63221, 2013. https://doi.org/10.1371/journal.pone.0063221.

[27] Moher, D., Naudet, F., Cristea, I., Miedema, F., Ioannidis, J., and Goodman, S. Assessing scientists for hiring, promotion, and tenure. *PLoS Biology*, 16(3):e2004089, 2018. https://doi.org/10.1371/journal.pbio.2004089.

[28] Moustafa, K. The disaster of the impact factor. *Science and Engineering Ethics*, 21(1): 139–142, 2015. https://doi.org/10.1007/s11948-014-9517-0.

[29] Munafò, M., Stothart, G., and Flint, J. Bias in genetic association studies and impact factor. *Molecular Psychiatry*, 14(2):119, 2009. https://doi.org/10.1038/mp.2008.77.

[30] of California Libraries, U. Pay It Forward: Investigating a Sustainable Model of Open Access Article Processing Charges for Large North American Research Institutions.

[31] Ohio University. College of Health Sciences and Professions: Promotion and Tenure Policy, 2014.

[32] Ohio University. Russ College of Engineering and Technology: Minimal Criteria for Promotion and Tenure, 2015.

[33] PLoS Medicine Editors. The impact factor game. *PLoS Medicine*, 3(6):e291, 2006. https://doi.org/10.1371/journal.pmed.0030291.

[34] Prinz, F., Schlange, T., and Asadullah, K. Believe it or not: how much can we rely on published data on potential drug targets? *Nature Reviews Drug Discovery*, 10(9):712, 2011. https://doi.org/10.1038/nrd3439-c1.

[35] Quan, W., Chen, B., and Shu, F. Publish or impoverish: An investigation of the monetary reward system of science in China (1999-2016). *Aslib Journal of Information Management*, 69(5):486–502, 2017. https://doi.org/10.1108/AJIM-01-2017-0014.

[36] Rogers Digital Media. Maclean's University Rankings, 2016. https://www.macleans.ca/education/unirankings/.

[37] Schimanski, L. and Alperin, J. The evaluation of scholarship in academic promotion and tenure processes: Past, present, and future [version 1; referees: 2 approved]. *F1000Research*, 7: 1605, 2018. https://doi.org/10.12688/f1000research.16493.1.

[38] Schroter, S., Tite, L., and Smith, R. Perceptions of open access publishing: interviews with journal authors. *BMJ: British Medical Journal*, 330(7494):756, 2005. https://doi.org/10.1136/bmj.38359.695220.82 .

[39] Seglen, P. Why the impact factor of journals should not be used for evaluating research. *BMJ: British Medical Journal*, 314(7079):498, 1997. https://doi.org/10.1136/bmj.314.7079.497.

[40] Simon Fraser University. Department of Psychology: Criteria for Promotion, 2015.

[41] SpringerNature. Prospectus for the public offering, 2018. web.archive.org/web/20180507134223/http://proxy.dbagproject.de/mediacenter/ressourcen/pdf/emissionen/springernature_prospectus.pdf.

[42] Sugimoto, C. and Larivière, V. *Measuring Research: What Everyone Needs to Know*. Oxford University Press, 2018.

[43] Swan, A. and Brown, S. JISC/OSI journal authors survey report, 2004. Accessed November 2018 https://eprints.soton.ac.uk/261002/1/JISCOAreport1.pdf.

[44] Szucs, D. and Ioannidis, J. Empirical assessment of published effect sizes and power in the recent cognitive neuroscience and psychology literature. *PLoS Biology*, 15(3):e2000797, 2017. https://doi.org/10.1371/journal.pbio.2000797.

[45] The Analogue University. Calling all journal editors: Bury the metrics pages! *Political Geography*, 2019. https://doi.org/10.1016/j.polgeo.2018.09.002.

[46] The National Academies of Sciences, Engineering, and Medicine. Taxonomy of fields and their subfields, 2006. http://sites.nationalacademies.org/pga/resdoc/pga_044522.

[47] Tijdink, J., Schipper, K., Bouter, L., Pont, P., De Jonge, J., and Smulders, Y. How do scientists perceive the current publication culture? a qualitative focus group interview study among dutch biomedical researchers. *BMJ Open*, 6(2):e008681, 2016. http://dx.doi.org/10.1136/bmjopen-2015-008681.

[48] University of Alberta. Faculty of Science: Criteria for Merit Increments, Tenure and Promotion, 2012.

[49] University of Calgary. Faculty of Veterinary Medicine: Guidelines for Appointment, Promotion, and Tenure of Academic Staff, 2008.

[50] University of California, San Diego. Where CAP Stood, 2015-16, 2015-2016.

[51] University of Central Florida. Department of Health Management & Informatics, College of Health & Public Affairs: Criteria for Promotion and Tenure, 2014.

[52] University of Central Florida. School of Social Work: Criteria for Promotion and Tenure, 2014.

[53] University of Central Florida. Department of Sociology: Department Criteria for Tenure and Promotion, 2015.

[54] University of Guelph. Department Human Health and Nutritiona Sciences: Guidelines for Tenure, Promotion and Performance Assessment, 2008.

[55] University of Illinois at Urbana Champaign Library. Understanding Impact Factor and Other Bibliometrics. Accessed October 2018 http://guides.library.illinois.edu/c.php?g=621441&p=4328606.

[56] University of Surrey Library. Responsible use of metrics. Accessed October 2018 http://www.surrey.ac.uk/library/research/openresearch/understandingmetrics/responsible_use_of_metrics.htm.

[57] University of Vermont. College of Arts and Sciences: Statement Regarding the Appointment, Tenure, and Promotion of Tenure-Track Faculty, 2015.

[58] University of Vermont. Reappointment, promotion, and tenure (RPT) guidelines and forms – Green sheet form, 2016.

[59] University of Windsor. Sample Research Evaluation Rubrich, 2016.

[60] University of Windsor. Working Session on Developing Promotion and Tenure Criteria for Research, 2016.

[61] University of York Library. Bibliometrics: a Practical Guide. Accessed October 2018 https://subjectguides.york.ac.uk/bibliometrics.

[62] Vasilevsky, N., Brush, M., Paddock, H., Ponting, L., Tripathy, S., LaRocca, G., and Haendel, M. On the reproducibility of science: unique identification of research resources in the biomedical literature. *PeerJ*, 1:e148, 2013. https://doi.org/10.7717/peerj.148.

[63] Walker, R., Sykes, L., Hemmelgarn, B., and Quan, H. Authors' opinions on publication in relation to annual performance assessment. *BMC Medical Education*, 10(1):21, 2010. https://doi.org/10.1186/1472-6920-10-21.

[64] Wildson, J., Bar-Ilan, J., Frodeman, R., Lex, E., Peters, I., and Wouters, P. Next-generation metrics: Responsible metrics and evaluation for open science. 2017. https://ec.europa.eu/research/openscience/pdf/report.pdf.

[65] Wilsdon, J., Allen, L., Belfiore, E., Campbell, P., Curry, S., Hill, S., Jones, R., Kain, R., Kerridge, S., Thelwall, M., Tinkler, J., Viney, I., Wouters, P., Hill, J., and Johnson, B. The metric tide: Report of the independent review of the role of metrics in research assessment and management, 2015. https://doi.org/10.13140/RG.2.1.4929.1363.