

BMJ Open Comprehensive Researcher Achievement Model (CRAM): a framework for measuring researcher achievement, impact and influence derived from a systematic literature review of metrics and models

Jeffrey Braithwaite,¹ Jessica Herkes,¹ Kate Churruca,¹ Janet C Long,¹ Chiara Pomare,¹ Claire Boyling,¹ Mia Bierbaum,¹ Robyn Clay-Williams,¹ Frances Rapport,¹ Patti Shih,¹ Anne Hogden,¹ Louise A Ellis,¹ Kristiana Ludlow,¹ Elizabeth Austin,¹ Rebecca Seah,¹ Elise McPherson,¹ Peter D Hibbert,^{1,2} Johanna Westbrook¹

To cite: Braithwaite J, Herkes J, Churruca K, *et al.* Comprehensive Researcher Achievement Model (CRAM): a framework for measuring researcher achievement, impact and influence derived from a systematic literature review of metrics and models. *BMJ Open* 2019;9:e025320. doi:10.1136/bmjopen-2018-025320

► Prepublication history and additional material for this paper are available online. To view these files, please visit the journal online (<http://dx.doi.org/10.1136/bmjopen-2018-025320>).

Received 10 July 2018

Revised 4 February 2019

Accepted 6 February 2019



© Author(s) (or their employer(s)) 2019. Re-use permitted under CC BY-NC. No commercial re-use. See rights and permissions. Published by BMJ.

For numbered affiliations see end of article.

Correspondence to

Professor Jeffrey Braithwaite; jeffrey.braithwaite@mq.edu.au

ABSTRACT

Objectives Effective researcher assessment is key to decisions about funding allocations, promotion and tenure. We aimed to identify what is known about methods for assessing researcher achievements, leading to a new composite assessment model.

Design We systematically reviewed the literature via the Preferred Reporting Items for Systematic Review and Meta-Analysis Protocols framework.

Data sources All Web of Science databases (including Core Collection, MEDLINE and BIOSIS Citation Index) to the end of 2017.

Eligibility criteria (1) English language, (2) published in the last 10 years (2007–2017), (3) full text was available and (4) the article discussed an approach to the assessment of an individual researcher's achievements.

Data extraction and synthesis Articles were allocated among four pairs of reviewers for screening, with each pair randomly assigned 5% of their allocation to review concurrently against inclusion criteria. Inter-rater reliability was assessed using Cohen's Kappa (κ). The κ statistic showed agreement ranging from moderate to almost perfect (0.4848–0.9039). Following screening, selected articles underwent full-text review and bias was assessed.

Results Four hundred and seventy-eight articles were included in the final review. Established approaches developed prior to our inclusion period (eg, citations and outputs, h-index and journal impact factor) remained dominant in the literature and in practice. New bibliometric methods and models emerged in the last 10 years including: measures based on PageRank algorithms or 'altmetric' data, methods to apply peer judgement and techniques to assign values to publication quantity and quality. Each assessment method tended to prioritise certain aspects of achievement over others.

Conclusions All metrics and models focus on an element or elements at the expense of others. A new composite design, the Comprehensive Researcher Achievement

Strengths and limitations of this study

- A large, diverse dataset of over 478 articles, containing many ideas for assessing researcher performance, was analysed.
- Strengths of the review include executing a wide-ranging search strategy, and the consequent high number of included articles for review; the results are limited by the literature itself, for example, new metrics were not mentioned in the articles, and therefore not captured in the results.
- A new model combining multiple factors to assess researcher performance is now available.
- Its strengths include combining quantitative and qualitative components in the one model.
- The Comprehensive Researcher Achievement Model, despite being evidence oriented, is a generic one and now needs to be applied in the field.

Model (CRAM), is presented, which supersedes past anachronistic models. The CRAM is modifiable to a range of applications.

INTRODUCTION

Judging researchers' achievements and academic impact continues to be an important means of allocating scarce research funds and assessing candidates for promotion or tenure. It has historically been carried out through some form of expert peer judgement to assess the number and quality of outputs and, in more recent decades, citations to them. This approach requires judgements regarding the weight that should be assigned to the

number of publications, their quality, where they were published and their downstream influence or impact. There are significant questions about the extent to which human judgement based on these criteria is an effective mechanism for making these complex assessments in a consistent and unbiased way.^{1–3} Criticisms of peer assessment, even when underpinned by relatively impartial productivity data, include the propensity for bias, inconsistency among reviewers, nepotism, group-think and subjectivity.^{4–7}

To compensate for these limitations, approaches have been proposed that rely less on subjective judgement and more on objective indicators.^{3 8–10} Indicators of achievement focus on one or a combination of four aspects: quantity of researcher outputs (*productivity*); value of outputs (*quality*); outcomes of research outputs (*impact*); and relations between publications or authors and the wider world (*influence*).^{11–15} Online publishing of journal articles has provided the opportunity to easily track citations and user interactions (eg, number of article downloads) and thus has provided a new set of indices against which individual researchers, journals and articles can be compared and the relative worth of contributions assessed and valued.¹⁴ These relatively new metrics have been collectively termed *bibliometrics*¹⁶ when based on citations and numbers of publications, or *altmetrics*¹⁷ when calculated by alternative online measures of impact such as number of downloads or social media mentions.¹⁶

The most established metrics for inferring researcher achievement are the h-index and the journal impact factor (JIF). The JIF measures the average number of citations of an article in the journal over the previous year, and hence is a good indication of journal quality but is increasingly regarded as a primitive measure of quality for individual researchers.¹⁸ The h-index, proposed by Hirsch in 2005,¹⁹ attempts to portray a researcher's productivity and impact in one data point. The h-index is defined as the number (*h*) of articles published by a researcher that have received a citation count of at least *h*. Use of the h-index has become widespread, reflected in its inclusion in author profiles on online databases such as Google Scholar and Scopus.

Also influenced by the advent of online databases, there has been a proliferation of other assessment models and metrics,¹⁶ many of which purport to improve on existing approaches.^{20 21} These include methods that assess the impact of articles measured by: downloads or online views received, practice change related to specific research, take-up by the scientific community or mentions in social media.

Against the backdrop of growth in metrics and models for assessing researchers' achievements, there is a lack of guidance on the relative strengths and limitations of these different approaches. Understanding them is of fundamental importance to funding bodies that drive the future of research, tenure and promotion committees and more broadly for providing insights into how we recognise and value the work of science and scientists,

particularly those researching in medicine and health-care. This review aimed to identify approaches to assessing researchers' achievements published in the academic literature over the last 10 years, considering their relative strengths and limitations and drawing on this to propose a new composite assessment model.

METHOD

Search strategy

All Web of Science databases (eight in total, including the Web of Science Core Collection, MEDLINE and BIOSIS Citation Index) were searched using terms related to researcher achievement (*researcher excellence, track record, researcher funding, researcher perform*, relative to opportunity, researcher potential, research* career pathway, academic career pathway, funding system, funding body, researcher impact, scientific*productivity, academic productivity, top researcher, researcher ranking, grant application, researcher output, h*index, i*index, impact factor, individual researcher*) and approaches to its assessment (*model, framework, assess*, evaluat*, *metric*, measur*, criteri*, citation*, unconscious bias, rank**) with '*' used as an unlimited truncation to capture variation in search terms, as seen in online supplementary appendix 1. These two searches were combined (using 'and'), and results were downloaded into EndNote,²² the reference management software.

Study selection

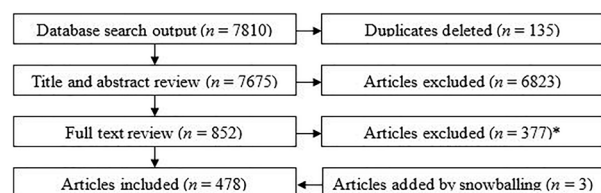
After removing duplicate references in EndNote, articles were allocated among pairs of reviewers (MB-JL, CP-CB, KL-JH and KC-LAE) for screening against inclusion criteria. Following established procedures,^{23 24} each pair was randomly assigned 5% of their allocation to review concurrently against inclusion criteria, with inter-rater reliability assessed using Cohen's kappa (κ). The κ statistic was calculated for pairs of researchers, with agreement ranging from moderate to almost perfect (0.4848–0.9039).²⁵ Following the abstract and title screen, selected articles underwent full text review. Reasons for exclusion were recorded.

Inclusion criteria

The following inclusion criteria were operationalised: (1) English language, (2) published in the last 10 years (2007–2017), (3) full text for the article was available, and (4) the article discussed an approach to the assessment of an individual researcher's achievements (at the researcher or singular output-level). The research followed the Preferred Reporting Items for Systematic Review and Meta-Analysis Protocols framework.²⁶ Empirical and non-empirical articles were included because many articles proposing new approaches to assessment, or discussing the limitations of existing ones, are not level one evidence or research based. Both quantitative and qualitative studies were included.

Data extraction

Data from the included articles were extracted, including: the country of article origin, the characteristics of the



*Reasons for exclusion are noted below

Reason for exclusion at the full text level	Number of articles excluded
Not in English language	47
Full text not available	62
Does not discuss assessment of an individual researcher	268
Total	377

Figure 1 Data screening and extraction process for academic articles.

models or metrics discussed, the perspective the article presented on the metric or model (positive, negative and indeterminable) including any potential benefits or limitations of the assessment model (and if these were perceived or based on some form of evidence). A customised data extraction sheet was developed in Microsoft Excel, trialled among members of the research team and subsequently refined. This information was synthesised for each model and metric identified through narrative techniques. The publication details and classification of each paper are contained in online supplementary appendix 2.

Appraisal of the literature

Due to the prevalence of non-empirical articles in this field (eg, editorial contributions and commentaries), it was determined that a risk of bias tool such as the Quality Assessment Tool could not be applied.²⁷ Rather, assessors were trained in multiple meetings (24 October, 30 October and 13 November 2017) to critically assess the quality of articles. Given the topic of the review (focusing on the publication process), the type of models and metrics identified (ie, more metrics that use publication data metrics) may influence the cumulative evidence and subsequently create a risk of bias. In addition, three researchers (JH, EM and CB) reviewed every included article to extract documented conflicts of interests of authors.

Patient and public involvement

Patients and the public were not involved in this systematic review.

RESULTS

The final dataset consisted of 478 academic articles. The data screening process is presented in figure 1.

Of the 478 included papers (see online supplementary appendix 2 for a summary), 295 (61.7%) had an empirical component, which ranged from interventional studies that assessed researcher achievement as an

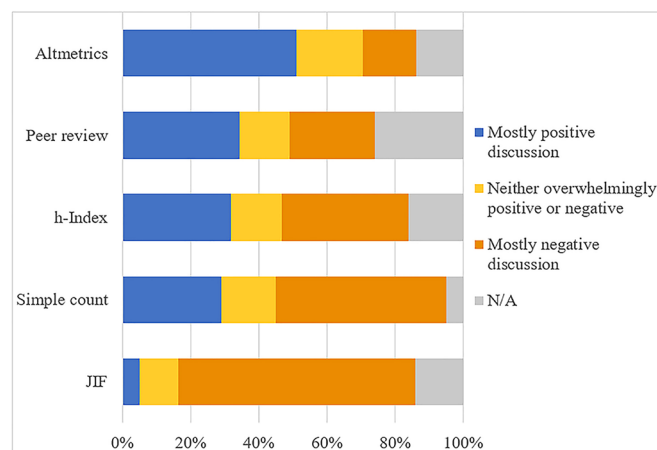


Figure 2 Percentages of positive and negative discussion regarding selected commonly used metrics for assessing individual researchers (n=478 articles). Positive discussion refers to articles that discuss the metric in favourable light or focus on the strengths of the metric; negative discussion refers to articles that focus on the limitations or shortcomings of the metric. JIF, journal impact factor.

outcome measure (eg, a study measuring the outcomes of a training programme),²⁸ as a predictor^{29–31} (eg, a study that demonstrated the association between number of citations early in one's career and later career productivity) or reported a descriptive analysis of a new metric.^{32 33} One hundred and sixty-six (34.7%) papers were not empirical, including editorial or opinion contributions that discussed the assessment of research achievement, or proposed models for assessing researcher achievement. Seventeen papers (3.6%) were reviews that considered one or more elements of assessing researcher achievements. The quality of these contributions ranged in terms of the risk of bias in the viewpoint expressed. Only for 19 papers (4.0%) did the authors declare a potential conflict of interest.

Across the study period, 78 articles (16.3%) involved authors purporting to propose new models or metrics. Most articles described or cited pre-existing metrics and largely discussed their perceived strengths and limitations. Figure 2 shows the proportion of positive or negative discussions of five of the most common approaches to assessing an individual's research achievement (altmetrics, peer-review, h-index, simple counts and JIF). The approach with most support was altmetrics (51.0% of articles mentioning altmetrics). The JIF was discussed with mostly negative sentiments in relevant articles (69.4%).

Citation-based metrics

Publication and citation counts

One hundred and fifty-three papers (32.0%) discussed the use of publication and citation counts for purposes of assessing researcher achievement, with papers describing them as a simple 'traditional but somewhat crude measure',³⁴ as well as the building blocks for other metrics.³⁵ A researcher's number of publications, commonly termed an n-index,³⁶ was suggested by some

to indicate researcher productivity,¹⁴ rather than quality, impact or influence of these papers.³⁷ However, the literature suggested that numbers of citations indicated the academic impact of an individual publication or researcher's body of work, calculated as an author's cumulative or mean citations per article.³⁸ Some studies found support for the validity of citation counts and publications in that they were correlated with other indications of a researcher's achievement, such as awards and grant funding,^{39 40} and were predictive of long-term success in a field.⁴¹ For example, one paper argued that having larger numbers of publications and being highly cited early in one's career predicted later high-quality research.⁴²

A number of limitations of using citation or publication counts was observed. For example, Minasny *et al*⁴³ highlighted discrepancies between publications and citations counts in different databases because of their differential structures and inputs.⁴³ Other authors^{38 44 45} noted that citation patterns vary by discipline, which they suggested can make them inappropriate for comparing researchers from different fields. Average citations per publication were reported as highly sensitive to change or could be skewed if, for example, a researcher has one heavily cited article.^{46 47} A further disadvantage is the lag-effect of citations^{48 49} and that, in most models, citations and publications count equally for all coauthors despite potentially differential contributions.⁵⁰ Some also questioned the extent to which citations actually indicated quality or impact, noting that a paper may influence clinical practice more than academic thinking.⁵¹ Indeed, a paper may be highly cited because it is useful (eg, a review), controversial or even by chance, making citations a limited indication of quality or impact.^{40 50 52} In addition to limitations, numerous authors made the point that focusing on citation and publication counts can have unintended, negative consequences for the assessment of researcher achievement, potentially leading to gaming and manipulation, including self-citations and gratuitous authorship.^{53 54}

Singular output-level approaches

Forty-one papers (8.6%) discussed models and metrics at the singular output or article-level that could be used to infer researcher achievement. The components of achievement they reported assessing were typically quality or impact.^{55 56} For example, some papers reported attempts to examine the quality of a single article by assessing its content.^{57 58} Among the metrics identified in the literature, the immediacy index focused on impact by measuring the average number of cites an article received in the year it was published.⁵⁹ Similarly, Finch²¹ suggested adapting the Source Normalized Impact per Publication (a metric used for journal-level calculations across different fields of research) to the article-level.

Many of the article-level metrics identified could also be upscaled to produce researcher-level indications of academic impact. For example, the sScientific currENcy Tokens (CENTs), proposed by Szymanski *et al*⁶⁰ involved

giving a 'cent' for each new non-self-citation a publication received; CENTs are then used as the basis for the researcher-level i-index, which follows a similar approach as the h-index but removes self-citations.⁶⁰ The temporally averaged paper-specific impact factor calculates an article's average number of citations per year combined with bonus cites for the publishing journal's prestige and can be aggregated to measure the overall relevance of a researcher (temporally averaged author-specific impact factor).⁶¹

Journal impact factor

The JIF, commonly recognised as a journal-level measure of quality,^{59 62} was discussed in 211 (44.1%) of the papers reviewed in relation to assessing singular outputs or individual researchers. A number of papers described the JIF being used informally to assess an individual's research achievement at the singular output-level and formally in countries such as France and China.⁶³ It implies article quality because it is typically a more competitive process to publish in journals with high impact factors.⁶⁴ Indeed, the JIF was found to be the best predictor of a paper's propensity to receive citations.⁶⁵

The JIF has a range of limitations when used to indicate journal quality,⁶⁶ including that it is disproportionately affected by highly cited, outlier articles^{41 67} and is susceptible to 'gaming' by editors.^{17 68} Other criticisms focused on using the JIF to assess individual articles or the researchers who author them.⁶⁹ Some critics claimed that using the JIF to measure an individual's achievement encourages researchers to publish in higher impact but less appropriate journals for their field, which ultimately means their article may not be read by relevant researchers.^{70 71} Furthermore, the popularity of a journal was argued to be a poor indication of the quality of any one article, with the citation distributions for calculating JIF found to be heavily skewed (ie, a small subset of papers receive the bulk of the citations, while some may receive none).¹⁸ Ultimately, many commentators argued that the JIF is an inappropriate metric to assess individual researchers because it is an aggregate metric of a journal's publication and expresses nothing about any individual paper.^{21 49 50 72} However, Bornmann and Pudovkin⁷³ suggested one case in which it would be appropriate to use JIF for assessing individual researchers: in relation to their recently published papers that had not had the opportunity to accumulate citations.⁷³

Researcher-level approaches

h-index

The h-index was among the most commonly discussed metrics in the literature (254 [53.1%] of the papers reviewed); in many of these papers, it was described by authors as more sophisticated than citation and publication counts but still straightforward, logical and intuitive.^{74–76} Authors noted its combination of productivity (h publications) and impact indicators (h citations) as being more reliable^{77 78} and stable than average citations per

publications,⁴¹ because it is not skewed by the influence of one popular article.⁷⁹ One study found that the h-index correlated with other metrics more difficult to obtain.⁷⁶ It also showed convergent validity with peer-reviewed assessments⁸⁰ and was found to be a good predictor of future achievement.⁴¹

However, because of the lag-effect with citations and publications, the h-index increases with a researcher's years of activity in the field, and cannot decrease, even if productivity later declines.⁸¹ Hence, numerous authors suggested it was inappropriate for comparing researchers at different career stages,⁸² or those early in their career.⁶⁸ The h-index was also noted as being susceptible to many of the critiques levelled against citation counts, including potential for gaming and inability to reflect differential contributions by coauthors.⁸³ Because disciplines differ in citation patterns,⁸⁴ some studies noted variations in author h-indices between different methodologies⁸⁵ and within medical subspecialties.⁸⁶ Some therefore argued that the h-index should not be used as the sole measure of a researcher's achievement.⁸⁶

h-Index variants

A number of modified versions of the h-index were identified; these purported to draw on its basic strengths of balancing productivity with impact while redressing perceived limitations. For example, the g-index measures global citation performance⁸⁷ and was defined similarly to the h-index but with more weight given to highly cited articles by assuming the top g articles have received at least g² citations.⁸⁸ Azer and Azer⁸⁹ argued it was a more useful measure of researcher productivity.⁸⁹ Another variant of the h-index identified, the m-quotient, was suggested to minimise the potential to favour senior academics by accounting for the time passed since a researcher has begun publishing papers.^{90 91} Other h-index variations reported in the articles reviewed attempted to account for author contributions, such as the h-maj index, which includes only articles in which the researcher played a core role (based on author order), and the weighted h-index, which assigns credit points according to author order.^{87 92}

Recurring issues with citation-based metrics

The literature review results suggested that no one citation-based metric was ideal for all purposes. All of the common metrics examined focused on one aspect of an individual's achievement and thus failed to account for other aspects of achievement. The limitations with some of the frequently used citation-based metrics are listed in box 1.

Non-citation-based approaches

Altmetrics

In contradistinction with the metrics discussed above, 54 papers (11.3%) discussed altmetrics (or 'alternative metrics'), which included a wide range of techniques to measure non-traditional, non-citation based usage of

Box 1 Common limitations in the use of citation-based metrics

1. Challenges with reconciling differences in citation patterns across varying fields of study.
2. Time-dependency issues stemming from differences in career length of researchers.
3. Prioritising impact over merit, or quality over quantity, or vice versa.
4. The lag-effect of citations.
5. Gaming and the ability of self-citation to distort metrics.
6. Failure to account for author order.
7. Contributions from authors to a publication are viewed as equal when they may not be.
8. Perpetuate 'publish or perish' culture.
9. Potential to stifle innovation in favour of what is popular.

articles, that is, influence.¹⁷ Altmetric measures included the number of online article views,⁹³ bookmarks,⁹⁴ downloads,⁴¹ PageRank algorithms⁹⁵ and attention by mainstream news,⁶³ in books⁹⁶ and social media, for example, in blogs, commentaries, online topic reviews or Tweets.^{97 98} These metrics typically measure the 'web visibility' of an output.⁹⁹ A notable example is the social networking site for researchers and scientists, ResearchGate, which uses an algorithm to score researchers based on the use of their outputs, including citations, reads and recommendations.¹⁰⁰

A strength of altmetrics lies in providing a measure of influence promptly after publication.^{68 101 102} Moreover, altmetrics allows tracking of the downloads of multiple sources (eg, students, the general public, clinicians, as well as academics) and multiple types of format (eg, reports and policy documents),¹⁰³ which are useful in gauging a broader indication of impact or influence, compared with more traditional metrics that solely or largely measure acknowledgement by experts in the field through citations.¹⁷

Disadvantages noted in the articles reviewed included that altmetrics calculations have been established by commercial enterprises such as *Altmetrics LLC* (London, UK) and other competitors,¹⁰⁴ and there may be fees levied for their use. The application of these metrics has also not been standardised.⁹⁶ Furthermore, it has been argued that, because altmetrics are cumulative and typically at the article-level, they provide more an indication of influence or even popularity,¹⁰⁵ instead of quality or productivity.¹⁰⁶ Hence, one study suggested no correlation between attention on Twitter and expert analysis of an article's originality, significance or rigour.¹⁰⁷ Another showed that Tweets predict citations.¹⁰⁸ Overall, further work needs to assess the value of altmetric scores in terms of their association with other traditional indicators of achievement.¹⁰⁹ Notwithstanding this, there were increasing calls to consider altmetrics alongside more conventional metrics in assessing researchers and their work.¹¹⁰

Past funding

A past record of being funded by national agencies was identified as a common measurement of individual academic achievement (particularly productivity, quality and impact) in a number of papers and has been argued to be a reliable method that is consistent across medical research.^{111–113} For example, the National Institute of Health's (NIH) Research Portfolio Online Reporting Tools system encourages public accountability for funding by providing online access to reports, data and NIH-funded research projects.^{111 114}

New metrics and models identified

The review also identified and assessed new metrics and models that were proposed during the review period, many of which had not gained widespread acceptance or use. While there was considerable heterogeneity and varying degrees of complexity among the 78 new approaches identified, there were also many areas of overlap in their methods and purposes. For example, some papers reported on metrics that used a PageRank algorithm,^{115 116} a form of network analysis based on structural characteristics of publications (eg, coauthorship or citation patterns).¹⁴ Metrics based on PageRank purported to measure both the direct and indirect impacts of a publication or researcher. Other approaches considered the relative contributions of authors to a paper in calculating productivity.¹¹⁷ Numerous metrics and models that built on existing approaches were also reported.¹¹⁸ For example, some developed composite metrics that included a publication's JIF alongside an author contribution measure¹¹⁹ or other existing metrics.¹²⁰ However, each of these approaches reported limitations, in addition to their strengths or improvements on other methods. For example, in focusing on productivity, a metric necessarily often neglected impact.¹²¹ Online supplementary appendix 3 provides a summary of these new or refashioned metrics and models, with details of their basis and purpose.

DISCUSSION

This systematic review identified a large number of diverse metrics and models for assessing an individual's research achievement that have been developed in the last 10 years (2007–2017), as evidenced in online supplementary appendix 3. At the same time, other approaches that pre-dated our study time period of 2007–2017 were also discussed frequently in the literature reviewed, including the h-index and JIF. All metrics and models proposed had their relative strengths, based on the components of achievement they focused on, and their sophistication or transparency.

The review also identified and assessed new metrics emerging over the past few decades. Peer-review has been increasingly criticised for reliance on subjectivity and propensity for bias,⁷ and there have been arguments that the use of specific metrics may be a more objective and

fair approach for assessing individual research achievement. However, this review has highlighted that even seemingly objective measures have a range of shortcomings. For example, there are inadequacies in comparing researchers at different career stages and across disciplines with different citation patterns.⁸⁴ Furthermore, the use of citation-based metrics can lead to gaming and potential ethical misconduct by contributing to a 'publish or perish' culture in which researchers are under pressure to maintain or improve their publication records.^{122 123} New methods and adjustments to existing metrics have been proposed to explicitly address some of these limitations; for example, normalising metrics with 'exchange rates' to remove discipline-specific variation in citation patterns, thereby making metric scores more comparable for researchers working in disparate fields.^{124 125} Normalisation techniques have also been used to assess researchers' metrics with greater recognition of their relative opportunity and career longevity.¹²⁶

Other criticisms of traditional approaches centre less on how they calculated achievement and more on what they understood or assumed about its constituent elements. In this review, the measurement of impact or knowledge gain was often exclusively tied to citations.¹²⁷ Some articles proposed novel approaches to using citations as a measure of impact, such as giving greater weight to citations from papers that were themselves highly cited¹²⁸ or that come from outside the field in which the paper was published.¹²⁹ However, even other potential means of considering scientific contributions and achievement, such as mentoring, were still ultimately tied to citations because mentoring was measured by the publication output of mentees.¹³⁰

A focus only on citations was widely thought to disadvantage certain types of researchers. For example, researchers who aim to publish with a focus on influencing practice may target more specialised or regional journals that do not have high JIFs, where their papers will be read by the appropriate audience and findings implemented, but they may not be well cited.⁵¹ In this regard, categorising the type of journal in which an article has been published in terms of its focus (eg, industry, clinical and regional/national) may go some way towards recognising those publications that have a clear knowledge translation intention and therefore prioritise real-world impact over academic impact.¹²² There were only a few other approaches identified that captured broader conceptualisations of knowledge gain, such as practical impact or wealth generation for the economy, and these too were often simplistic, such as including patents and their citations¹³¹ or altmetric data.⁹⁶ While altmetrics hold potential in this regard, their use has not been standardised,⁹⁶ and they come with their own limitations, with suggestions that they reflect popularity more so than real-world impact.¹⁰⁵ Other methodologies have been proposed for assessing knowledge translation and real-world impact, but these can often be labour intensive.¹³² For example, Sutherland *et al*¹³³ suggested that assessing individual

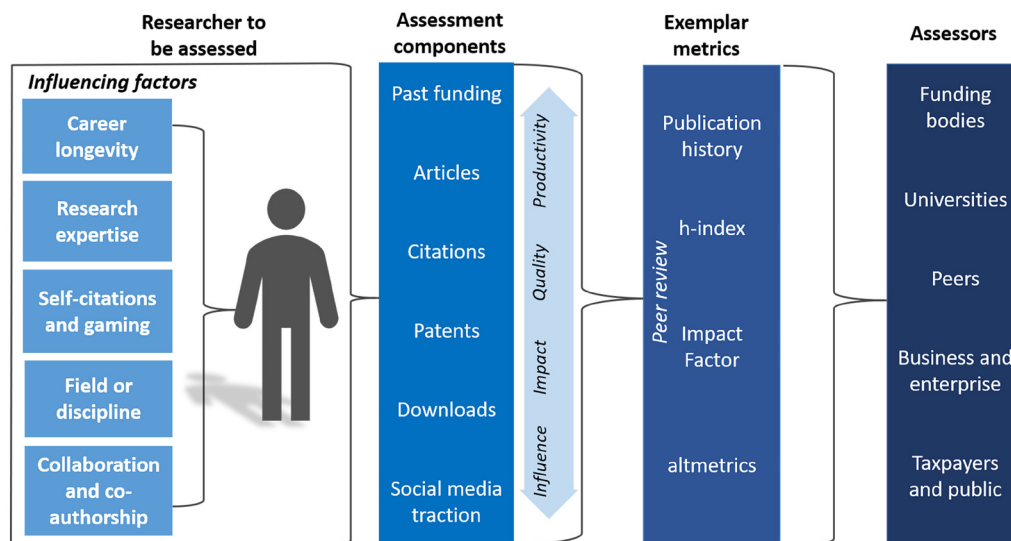


Figure 3 The Comprehensive Researcher Achievement Model.

research outputs in light of specific policy objectives through peer-review based scoring, may be a strategy, but this is typically not feasible in situations such as grant funding allocation, where there are time constraints and large applicant pools to assess.

In terms of how one can make sense of the validity of many of these emerging approaches for assessing an individual's research achievements, metrics should demonstrate their legitimacy empirically, as well as having a theoretical basis for their use and clearly differentiating what aspects of quality, achievement or impact they purport to examine.^{55 65} If the recent, well-publicised^{134–136} San Francisco Declaration on Research Assessment¹³⁷ is anything to go by, internationally, there is a move away from the assessment of individual researchers using the JIF and the journal in which the research has been published.

There is momentum, instead, for assessment of researcher achievements on the basis of a wider mix of measures, hence our proposed Comprehensive Researcher Achievement Model (CRAM) (figure 3). On the left-hand side of this model is the researcher to be assessed and key characteristics that influence the assessment. Among these factors, some (ie, field or discipline, coauthorship and career longevity) can be controlled for depending on the metric, while other components, such as gaming or the research topic (ie, whether it is 'trendy' or innovative), are less amenable to control or even prediction. Online databases, which track citations and downloads and measure other forms of impact, hold much potential and will likely be increasingly used in the future to assess both individual researchers and their outputs. Hence, assessment components (past funding, articles, citations, patents, downloads and social media traction) included in our model are those primarily accessible online.

Strengths and limitations

The findings of this review suggest assessment components should be used with care, with recognition of how they can be influenced by other factors, and what aspects of achievement they reflect (ie, productivity, quality, impact and influence). No metric or model singularly captures all aspects of achievement, and hence use of a range, such as the examples in our model, is advisable. CRAM recognises that the configuration and weighting of assessment methods will depend on the assessors and their purpose, the resources available for the assessment process and access to assessment components. Our results must be interpreted in light of our focus on academic literature. The limits of our focus on peer-reviewed literature were evident in the fact that some new metrics were not mentioned in articles and therefore not captured in our results. While we defined impact broadly at the outset, overwhelmingly, the literature we reviewed focused on academic, citation-based impact. Furthermore, although we assessed bias in the ways documented, the study design limited our ability to apply a standardised quality assessment tool. A strength of our focus was that we set no inclusion criteria with regard to scientific discipline, because novel and useful approaches to assessing research achievement can come from diverse fields. Many of the articles we reviewed were broadly in the area of health and medical research, and our discussion is concerned with the implications for health and medical research, as this is where our interests lie.

CONCLUSION

There is no ideal model or metric by which to assess individual researcher achievement. We have proposed a generic model, designed to minimise risk of the use of any one or a smaller number of metrics, but it is not proposed as an ultimate solution. The mix of assessment

components and metrics will depend on the purpose. Greater transparency in approaches used to assess achievement including their evidence base is required.³⁷ Any model used to assess achievement for purposes such as promotion or funding allocation should include some quantitative components, based on robust data, and be able to be rapidly updated, presented with confidence intervals and normalised.³⁷ The assessment process should be difficult to manipulate and explicit about the components of achievement being measured. As such, no current metric suitably fulfils all these criteria. The best strategy to assess an individual's research achievement is likely to involve the use of multiple approaches¹³⁸ in order to dilute the influence and potential disadvantages of any one metric while providing more rounded picture of a researcher's achievement^{83 139}; this is what the CRAM aims to contribute.

All in all, achievement in terms of impact and knowledge gain is broader than the number of articles published or their citation rates and yet most metrics have no means of factoring in these broader issues. Altmetrics hold promise in complementing citation-based metrics and assessing more diverse notions of impact, but usage of this type of tool requires further standardisation.⁹⁶ Finally, despite the limitations of peer-review, the role of expert judgement should not be discounted.⁴¹ Metrics are perhaps best applied as a complement or check on the peer-review process, rather than the sole means of assessment of an individual's research achievements.¹⁴⁰

Author affiliations

¹Australian Institute of Health Innovation, Macquarie University, North Ryde, New South Wales, Australia

²Division of Health Sciences, University of South Australia, Adelaide, South Australia, Australia

Contributors JB conceptualised and drafted the manuscript, revised it critically for important intellectual content and led the study. JH, KC and JCL made substantial contributions to the design, analysis and revision of the work and critically reviewed the manuscript for important intellectual content. CP, CB, MB, RC-W, FR, PS, AH, LAE, KL, EA, RS and EM carried out the initial investigation, sourced and analysed the data and revised the manuscript for important intellectual content. PDH and JW critically commented on the manuscript, contributed to the revision and editing of the final manuscript and reviewed the work for important intellectual content. All authors approved the final manuscript as submitted and agree to be accountable for all aspects of the work.

Funding The work on which this paper is based was funded by the Australian National Health and Medical Research Council (NHMRC) for work related to an assessment of its peer-review processes being conducted by the Council. Staff of the Australian Institute of Health Innovation undertook this systematic review for Council as part of that assessment.

Disclaimer Other than specifying what they would like to see from a literature review, NHMRC had no role in the conduct of the systematic review or the decision to publish.

Competing interests None declared.

Patient consent for publication Not required.

Provenance and peer review Not commissioned; externally peer reviewed.

Data sharing statement All data have been made available as appendices.

Open access This is an open access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially,

and license their derivative works on different terms, provided the original work is properly cited, appropriate credit is given, any changes made indicated, and the use is non-commercial. See: <http://creativecommons.org/licenses/by-nc/4.0/>.

REFERENCES

1. Ibrahim N, Habacha Chaibi A, Ben Ahmed M. New scientometric indicator for the qualitative evaluation of scientific production. *New Libr World* 2015;116:661–76.
2. Franco Aixelá J, Rovira-Esteva S. Publishing and impact criteria, and their bearing on translation studies: In search of comparability. *Perspectives* 2015;23:265–83.
3. Belter CW. Bibliometric indicators: opportunities and limits. *J Med Libr Assoc* 2015;103:219–21.
4. Frixione E, Ruiz-Zamarrípa L, Hernández G. Assessing individual intellectual output in scientific research: Mexico's national system for evaluating scholars performance in the humanities and the behavioral sciences. *PLoS One* 2016;11:e0155732.
5. Marzolla M. Assessing evaluation procedures for individual researchers: The case of the Italian National Scientific Qualification. *J Informetr* 2016;10:408–38.
6. Marsh HW, Jayasinghe UW, Bond NW. Improving the peer-review process for grant applications: reliability, validity, bias, and generalizability. *Am Psychol* 2008;63:160–8.
7. Kaatz A, Magua W, Zimmerman DR, et al. A quantitative linguistic analysis of National Institutes of Health R01 application critiques from investigators at one institution. *Acad Med* 2015;90:69–75.
8. Aoun SG, Bendok BR, Rahme RJ, et al. Standardizing the evaluation of scientific and academic performance in neurosurgery—critical review of the "h" index and its variants. *World Neurosurg* 2013;80:e85–90.
9. Hicks D, Wouters P, Waltman L, et al. Bibliometrics: the Leiden Manifesto for research metrics. *Nature* 2015;520:429–31.
10. King J. A review of bibliometric and other science indicators and their role in research evaluation. *J Inf Sci* 1987;13:261–76.
11. Abramo G, Cicero T, D'Angelo CA. A sensitivity analysis of researchers' productivity rankings to the time of citation observation. *J Informetr* 2012;6:192–201.
12. Arimoto A. Declining symptom of academic productivity in the Japanese research university sector. *High Educ* 2015;70:155–72.
13. Carey RM. Quantifying scientific merit is time to transform the impact factor? *Circ Res* 2016;119:1273–5.
14. Durieux V, Gevenois PA. Bibliometric indicators: quality measurements of scientific publication. *Radiology* 2010;255:342–51.
15. Selvarajoo K. Measuring merit: take the risk. *Science* 2015;347:139–40.
16. Wildgaard L, Schneider JW, Larsen B. A review of the characteristics of 108 author-level bibliometric indicators. *Scientometrics* 2014;101:125–58.
17. Maximin S, Green D. Practice corner: the science and art of measuring the impact of an article. *Radiographics* 2014;34:116–8.
18. Callaway E. Beat it, impact factor! Publishing elite turns against controversial metric. *Nature* 2016;535:210–1.
19. Hirsch JE. An index to quantify an individual's scientific research output. *Proc Natl Acad Sci USA* 2005;102:16569–72.
20. Bollen J, Crandall D, Junk D, et al. An efficient system to fund science: from proposal review to peer-to-peer distributions. *Scientometrics* 2017;110:521–8.
21. Finch A. Can we do better than existing author citation metrics? *Bioessays* 2010;32:744–7.
22. EndNote. *Clarivate analytics*, 2017.
23. Schlosser RW. Appraising the quality of systematic reviews. *Focus: Technical Briefs* 2007;17:1–8.
24. Braithwaite J, Herkes J, Ludlow K, et al. Association between organisational and workplace cultures, and patient outcomes: systematic review. *BMJ Open* 2017;7:e017708.
25. Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics* 1977;33:159–74.
26. Shamseer L, Moher D, Clarke M, et al. Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) 2015: elaboration and explanation. *BMJ* 2015;349:g7647.
27. Hawker S, Payne S, Kerr C, et al. Appraising the evidence: reviewing disparate data systematically. *Qual Health Res* 2002;12:1284–99.
28. Thorngate W, Chowdhury W. By the numbers: track record, flawed reviews, journal space, and the fate of talented authors. In: Kaminski B, Koloch G, eds. *Advances in Social Simulation: Proceedings of the 9th Conference of the European Social Simulation Association. Advances in Intelligent Systems*

- and Computing. 229. Heidelberg, Germany: Springer Berlin, 2014:177–88.
29. Sood A, Therattil PJ, Chung S, *et al*. Impact of subspecialty fellowship training on research productivity among academic plastic surgery faculty in the United States. *Eplasty* 2015;15:e50.
 30. Mutz R, Bornmann L, Daniel H-D. Testing for the fairness and predictive validity of research funding decisions: a multilevel multiple imputation for missing data approach using ex-ante and ex-post peer evaluation data from the Austrian science fund. *J Assoc Inf Sci Technol* 2015;66:2321–39.
 31. Rezek I, McDonald RJ, Kallmes DF. Pre-residency publication rate strongly predicts future academic radiology potential. *Acad Radiol* 2012;19:632–4.
 32. Knudson D. Kinesiology faculty citations across academic rank. *Quest* 2015;67:346–51.
 33. Wang D, Song C, Barabási AL. Quantifying long-term scientific impact. *Science* 2013;342:127–32.
 34. Efron N, Brennan NA. Citation analysis of Australia-trained optometrists. *Clin Exp Optom* 2011;94:600–5.
 35. Perlin MS, Santos AAP, Imasato T, *et al*. The Brazilian scientific output published in journals: a study based on a large CV database. *J Informetr* 2017;11:18–31.
 36. Stallings J, Vance E, Yang J, *et al*. Determining scientific impact using a collaboration index. *Proc Natl Acad Sci USA* 2013;110:9680–5.
 37. Kreiman G, Maunsell JH. Nine criteria for a measure of scientific output. *Front Comput Neurosci* 2011;5.
 38. Mingers J. Measuring the research contribution of management academics using the Hirsch-index. *J Oper Res Soc* 2009;60:1143–53.
 39. Halvorson MA, Finlay AK, Cronkite RC, *et al*. Ten-year publication trajectories of health services research career development award recipients: collaboration, awardee characteristics, and productivity correlates. *Eval Health Prof* 2016;39:49–64.
 40. Stroebe W. The graying of academia: will it reduce scientific productivity? *Am Psychol* 2010;65:660–73.
 41. Agarwal A, Durairajanayagam D, Tatagari S, *et al*. Bibliometrics: tracking research impact by selecting the appropriate metrics. *Asian J Androl* 2016;18:296–309.
 42. Jacob JH, Lehl S, Henkel AW. Early recognition of high quality researchers of the German psychiatry by worldwide accessible bibliometric indicators. *Scientometrics* 2007;73:117–30.
 43. Minasny B, Hartemink AE, McBratney A, *et al*. Citations and the h index of soil researchers and journals in the Web of Science, Scopus, and Google Scholar. *Peer J* 2013;1:1.
 44. Gorraiz J, Gumpenberger C. Going beyond citations: SERUM — a new tool provided by a network of libraries. *LIBER Quarterly* 2010;20:80–93.
 45. van Eck NJ, Waltman L, van Raan AF, *et al*. Citation analysis may severely underestimate the impact of clinical research as compared to basic research. *PLoS One* 2013;8:e62395.
 46. Meho LI, Rogers Y. Citation counting, citation ranking, and h-index of human-computer interaction researchers: a comparison of Scopus and Web of Science. *J Assoc Inf Sci Technol* 2008;59:1711–26.
 47. Selek S, Saleh A. Use of h index and g index for American academic psychiatry. *Scientometrics* 2014;99:541–8.
 48. Kali A. Scientific impact and altmetrics. *Indian J Pharmacol* 2015;47:570–1.
 49. Neylon C, Wu S. Article-level metrics and the evolution of scientific impact. *PLoS Biol* 2009;7:e1000242.
 50. Sahel JA. Quality versus quantity: assessing individual research performance. *Sci Transl Med* 2011;3:84cm13.
 51. Pinnock D, Whittingham K, Hodgson LJ. Reflecting on sharing scholarship, considering clinical impact and impact factor. *Nurse Educ Today* 2012;32:744–6.
 52. Eyre-Walker A, Stoletzki N. The assessment of science: the relative merits of post-publication review, the impact factor, and the number of citations. *PLoS Biol* 2013;11:e1001675.
 53. Ferrer-Sapena A, Sánchez-Pérez EA, Peset F, *et al*. The Impact Factor as a measuring tool of the prestige of the journals in research assessment in mathematics. *Res Eval* 2016;25:306–14.
 54. Moustafa K. Aberration of the citation. *Account Res* 2016;23:230–44.
 55. Abramo G, D'Angelo CA. Refrain from adopting the combination of citation and journal metrics to grade publications, as used in the Italian national research assessment exercise (VQR 2011–2014). *Scientometrics* 2016;109:2053–65.
 56. Pál-Gergely B. On the confusion of quality with impact: a note on Pyke's M-Index. *BioScience* 2015;65:117.
 57. Niederkrötenhaler T, Dorner TE, Maier M. Development of a practical tool to measure the impact of publications on the society based on focus group discussions with scientists. *BMC Public Health* 2011;11:588.
 58. Kreines EM, Kreines MG. Control model for the alignment of the quality assessment of scientific documents based on the analysis of content-related context. *J Comput Syst Sci* 2016;55:938–47.
 59. DiBartola SP, Hinchcliff KW. Metrics and the scientific literature: deciding what to read. *J Vet Intern Med* 2017;31:629–32.
 60. Szymanski BK, de la Rosa JL, Krishnamoorthy M. An internet measure of the value of citations. *Inf Sci* 2012;185:18–31.
 61. Bloching PA, Heinzl H. Assessing the scientific relevance of a single publication over time. *S Afr J Sci* 2013;109:1–2.
 62. Benchimol-Barbosa PR, Ribeiro RL, Barbosa EC. Additional comments on the paper by Thomas et al: how to evaluate "quality of publication". *Arq Bras Cardiol* 2011;97:88–9.
 63. Slim K, Dupré A, Le Roy B. Impact factor: an assessment tool for journals or for scientists? *Anaesth Crit Care Pain Med* 2017;36:347–8.
 64. Diem A, Wolter SC. The use of bliometrics to measure research performance in education sciences. *Res High Educ* 2013;54:86–114.
 65. Bornmann L, Leydesdorff L. Does quality and content matter for citedness? A comparison with para-textual factors and over time. *J Informetr* 2015;9:419–29.
 66. Santangelo GM. Article-level assessment of influence and translation in biomedical research. *Mol Biol Cell* 2017;28:1401–8.
 67. Ravenscroft J, Liakata M, Clare A, *et al*. Measuring scientific impact beyond academia: an assessment of existing impact metrics and proposed improvements. *PLoS One* 2017;12:e0173152.
 68. Trueger NS, Thoma B, Hsu CH, *et al*. The altmetric score: a new measure for article-level dissemination and impact. *Ann Emerg Med* 2015;66:549–53.
 69. Welk G, Fischman MG, Greenleaf C, *et al*. Editorial board position statement regarding the Declaration on research assessment (DORA) recommendations with respect to journal impact factors. *Res Q Exerc Sport* 2014;85:429–30.
 70. Taylor DR, Michael LM, Klimo P. Not everything that matters can be measured and not everything that can be measured matters response. *J Neurosurg* 2015;123:544–5.
 71. Christopher MM. Weighing the impact (factor) of publishing in veterinary journals. *J Vet Cardiol* 2015;17:77–82.
 72. Jokic M. H-index as a new scientometric indicator. *Biochem Med* 2009;19:5–9.
 73. Bornmann L, Pudovkin AI. The journal impact factor should not be discarded. *J Korean Med Sci* 2017;32:180–2.
 74. Franceschini F, Galetto M, Maisano D, *et al*. The success-index: an alternative approach to the h-index for evaluating an individual's research output. *Scientometrics* 2012;92:621–41.
 75. Prathap G. Citation indices and dimensional homogeneity. *Curr Sci* 2017;113:853–5.
 76. Saad G. Applying the h-index in exploring bibliometric properties of elite marketing scholars. *Scientometrics* 2010;83:423–33.
 77. Duffy RD, Jadidian A, Webster GD, *et al*. The research productivity of academic psychologists: assessment, trends, and best practice recommendations. *Scientometrics* 2011;89:207–27.
 78. Prathap G. Evaluating journal performance metrics. *Scientometrics* 2012;92:403–8.
 79. Lando T, Bertoli-Barsotti L. A new bibliometric index based on the shape of the citation distribution. *PLoS One* 2014;9:e115962.
 80. Bornmann L, Wallon G, Ledín A. Is the h index related to (standard) bibliometric measures and to the assessments by peers? An investigation of the h index by using molecular life sciences data. *Res Eval* 2008;17:149–56.
 81. Pepe A, Kurtz MJ. A measure of total research impact independent of time and discipline. *PLoS One* 2012;7:e46428.
 82. Haslam N, Laham S. Early-career scientific achievement and patterns of authorship: the mixed blessings of publication leadership and collaboration. *Res Eval* 2009;18:405–10.
 83. Ioannidis JP, Klavans R, Boyack KW. Multiple citation indicators and their composite across scientific disciplines. *PLoS Biol* 2016;14:e1002501.
 84. van Leeuwen T. Testing the validity of the Hirsch-index for research assessment purposes. *Res Eval* 2008;17:157–60.
 85. Ouimet M, Bédard PO, Gélinau F. Are the h-index and some of its alternatives discriminatory of epistemological beliefs and methodological preferences of faculty members? The case of social scientists in Quebec. *Scientometrics* 2011;88:91–106.
 86. Kshetry VR, Benzel EC. Research productivity and fellowship training in neurosurgery. *World Neurosurg* 2013;80:787–8.

87. Biswal AK. An absolute index (Ab-index) to measure a researcher's useful contributions and productivity. *PLoS One* 2013;8:e84334.
88. Tschudy MM, Rowe TL, Dover GJ, et al. Pediatric academic productivity: Pediatric benchmarks for the h- and g-indices. *J Pediatr* 2016;169:272–6.
89. Azer SA, Azer S. Bibliometric analysis of the top-cited gastroenterology and hepatology articles. *BMJ Open* 2016;6:e009889.
90. Joshi MA. Bibliometric indicators for evaluating the quality of scientific publications. *J Contemp Dent Pract* 2014;15:258–62.
91. Danielson J, McElroy S. Quantifying published scholarly works of experiential education directors. *Am J Pharm Educ* 2013;77:167.
92. Ion D, Andronic O, Bolocan A, et al. Tendencies on traditional metrics. *Chirurgia* 2017;112:117–23.
93. Suiter AM, Moulaison HL. Supporting scholars: An analysis of academic library websites' documentation on metrics and impact. *J Acad Librariansh* 2015;41:814–20.
94. Butler JS, Kaye ID, Sebastian AS, et al. The evolution of current research impact metrics: From bibliometrics to altmetrics? *Clin Spine Surg* 2017;30:226–8.
95. Krapivin M, Marchese M, Casati F. Exploring and understanding scientific metrics in citation networks. Zhou J, ed. *Complex Sciences, Pt 2. Lecture Notes of the Institute for Computer Sciences Social Informatics and Telecommunications Engineering*. 2009 5:1550–63.
96. Carpenter TA. Comparing digital apples to digital apples: background on NISO's effort to build an infrastructure for new forms of scholarly assessment. *Inf Serv Use* 2014;34:103–6.
97. Gasparyan AY, Nurmashev B, Yessirkepov M, et al. The journal impact factor: moving toward an alternative and combined scientometric approach. *J Korean Med Sci* 2017;32:173–9.
98. Moed HF, Halevi G. Multidimensional assessment of scholarly research impact. *J Assoc Inf Sci Technol* 2015;66:1988–2002.
99. Chuang K-Y, Olaiya MT. Bibliometric analysis of the Polish Journal of Environmental Studies (2000–11). *Pol J Environ Stud* 2012;21:1175–83.
100. Vinyard M. Altmetrics: an overhyped fad or an important tool for evaluating scholarly output? *Computers in Libraries* 2016;36:26–9.
101. van Noorden R. A profusion of measures. *Nature* 2010;465:864–6.
102. Van Noorden R. Love thy lab neighbour. *Nature* 2010;468:1011.
103. Dinsmore A, Allen L, Dolby K. Alternative perspectives on impact: the potential of ALMs and altmetrics to inform funders about research impact. *PLoS Biol* 2014;12:e1002003.
104. Cress PE. Using altmetrics and social media to supplement impact factor: maximizing your article's academic and societal impact. *Aesthet Surg J* 2014;34:1123–6.
105. Moreira JA, Zeng XH, Amaral LA. The distribution of the asymptotic number of citations to sets of publications by a researcher or from an academic department are consistent with a discrete lognormal model. *PLoS One* 2015;10:e0143108.
106. Waljee JF. Discussion: are quantitative measures of academic productivity correlated with academic rank in plastic surgery? A National Study. *Plast Reconstr Surg* 2015;136:622–3.
107. Fazel S, Wolf A. What is the impact of a research publication? *Evid Based Ment Health* 2017;20:33–4.
108. Eysenbach G. Can tweets predict citations? Metrics of social impact based on Twitter and correlation with traditional metrics of scientific impact. *J Med Internet Res* 2011;13:e123.
109. Hoffmann CP, Lutz C, Meckel M. Impact factor 2.0: applying social network analysis to scientific impact assessment. 2014 47th Hawaii International Conference on System Sciences. *Proceedings of the Annual Hawaii International Conference on System Sciences*, 2014:1576–85.
110. Maggio LA, Meyer HS, Artino AR. Beyond citation rates: a real-time impact analysis of health professions education research using altmetrics. *Acad Med* 2017;92:1449–55.
111. Raj A, Carr PL, Kaplan SE, et al. Longitudinal analysis of gender differences in academic productivity among medical faculty across 24 medical schools in the United States. *Acad Med* 2016;91:1074–9.
112. Markel TA, Valsangkar NP, Bell TM, et al. Endangered academia: preserving the pediatric surgeon scientist. *J Pediatr Surg* 2017;52:1079–83.
113. Mirnezami SR, Beaudry C, Larivière V. What determines researchers' scientific impact? A case study of Quebec researchers. *Sci Public Policy* 2016;43:262–74.
114. Napolitano LM. Scholarly activity requirements for critical care fellowship program directors: what should it be? How should we measure it? *Crit Care Med* 2016;44:2293–6.
115. Bai X, Xia F, Lee I, et al. Identifying anomalous citations for objective evaluation of scholarly article impact. *PLoS One* 2016;11:e0162364.
116. Gao C, Wang Z, Li X, et al. PR-Index: Using the h-index and PageRank for determining true impact. *PLoS One* 2016;11:e0161755.
117. Assimakis N, Adam M. A new author's productivity index: p-index. *Scientometrics* 2010;85:415–27.
118. Petersen AM, Succi S. The Z-index: A geometric representation of productivity and impact which accounts for information in the entire rank-citation profile. *J Informetr* 2013;7:823–32.
119. Claro J, Costa CAV. A made-to-measure indicator for cross-disciplinary bibliometric ranking of researchers performance. *Scientometrics* 2011;86:113–23.
120. Sahoo BK, Singh R, Mishra B, et al. Research productivity in management schools of India during 1968–2015: A directional benefit-of-doubt model analysis. *Omega* 2017;66:118–39.
121. Aragón AM. A measure for the impact of research. *Sci Rep* 2013;3:1649.
122. Shibayama S, Baba Y. Impact-oriented science policies and scientific publication practices: the case of life sciences in Japan. *Res Policy* 2015;44:936–50.
123. Tjldink JK, Schipper K, Bouter LM, et al. How do scientists perceive the current publication culture? A qualitative focus group interview study among Dutch biomedical researchers. *BMJ Open* 2016;6:e008681.
124. Crespo JA, Li Y, Li Y, et al. The measurement of the effect on citation inequality of differences in citation practices across scientific fields. *PLoS One* 2013;8.
125. Teixeira da Silva JA, da Silva JAT. Does China need to rethink its metrics- and citation-based research rewards policies? *Scientometrics* 2017;112:1853–7.
126. Devos P. Research and bibliometrics: a long history.... *Clin Res Hepatol Gastroenterol* 2011;35:336–7.
127. Slyder JB, Stein BR, Sams BS, et al. Citation pattern and lifespan: a comparison of discipline, institution, and individual. *Scientometrics* 2011;89:955–66.
128. Zhou Y-B, Lu L, Li M. Quantifying the influence of scientists and their publications: distinguishing between prestige and popularity. *New J Phys* 2012;14:033033.
129. Sorensen AA, Weedon D. Productivity and impact of the top 100 cited Parkinson's disease investigators since 1985. *J Parkinsons Dis* 2011;1:3–13.
130. Jeang KT. H-index, mentoring-index, highly-cited and highly-accessed: how to evaluate scientists? *Retrovirology* 2008;5:106.
131. Franceschini F, Maisano D. Publication and patent analysis of European researchers in the field of production technology and manufacturing systems. *Scientometrics* 2012;93:89–100.
132. Sibbald SL, MacGregor JC, Surmacz M, et al. Into the gray: a modified approach to citation analysis to better understand research impact. *J Med Libr Assoc* 2015;103:49–54.
133. Sutherland WJ, Goulson D, Potts SG, et al. Quantifying the impact and relevance of scientific research. *PLoS One* 2011;6:e27537.
134. Team NE. Announcement: Nature journals support the San Francisco Declaration on Research Assessment. *Nature* 2017;544:394.
135. Pugh EN, Gordon SE. Embracing the principles of the San Francisco Declaration of Research Assessment: Robert Balaban's editorial. *J Gen Physiol* 2013;142:175.
136. Zhang L, Rousseau R, Sivertsen G. Science deserves to be judged by its contents, not by its wrapping: Revisiting Seglen's work on journal impact and research evaluation. *PLoS One* 2017;12:e0174205.
137. DORA—ASCB San Francisco, US. San Francisco Declaration on Research Assessment (DORA). 2016 <http://www.ascb.org/dora/>.
138. Cabezas-Clavijo A, Delgado-López-Cózar E. [Google Scholar and the h-index in biomedicine: the popularization of bibliometric assessment]. *Med Intensiva* 2013;37:343–54.
139. Iyengar R, Wang Y, Chow J, et al. An integrated approach to evaluate faculty members' research performance. *Acad Med* 2009;84:1610–6.
140. Jacsó P. Eigenfactor and article influence scores in the *Journal Citation Reports*. *Online Information Review* 2010;34:339–48.