How to measure academic impact

B. Morton[1], H. Ranson[2] and K. El-Boghdadly[3,4]

1 Department of Clinical Sciences, Liverpool School of Tropical Medicine, Liverpool, UK

2 Professor, Department of Vector Biology, Liverpool School of Tropical Medicine, Liverpool, UK

3 Consultant, Department of Anaesthesia, Guy's and St Thomas' NHS Foundation Trust, London, UK

4 Honorary Senior Lecturer, Centre for Human and Applied Physiological Sciences, King´s College London, London, UK

Correspondence: B. Morton

E-mail: Ben.Morton@lstmed.ac.uk

Running title: How to measure academic impact

Twitter: @BenJamesMorton; @hilary_ranson; @elboghdadly

ORCID: BM 0000-0002-6164-2854

HR 0000-0003-2332-8247

KE 0000-0002-9912-717X

Good research informs action, gathers evidence for theories and/or contributes to developing knowledge in a field of study. Research impact is the extent to which research has contributed both to academia and beyond to influence broader society, culture, our environment and the economy. Widespread research dissemination is a distinct but vital measure to generate impact, described in a separate article in this series [1]. Measurement of impact is vital across the research ecosystem including for funders, institutions and for individual researchers. Impact metrics are used in multiple ways: 1) benchmark researchers and institutions; 2) demonstrate productivity; 3) guide promotion; 4) quantify return on investment for funders; and 5) leverage additional funding for researchers. No single measure, however, exists to accurately represent the impact of a researcher or an individual article [2]. This issue has recently been highlighted by the removal of the ResearchGate (ResearchGate GmbH, Berlin, Germany) 'RG score' for individual researchers from their platform [3]; stating that this score does not meet all key criteria for an impact metric of being intuitive, transparent, robust and relevant. There remains uncertainty regarding how researchers can assess the impact of their work. The aim of this article is to outline the strengths and weaknesses of existing metrics at the individual, article, journal and institutional level to quantify or qualify academic impact. We do not suggest a single 'unifying' metric, but instead, that a holistic approach should be taken, drawing together multiple parameters to measure academic impact.

Individual impact

Multiple metrics have been proposed to quantify individual researcher impact based on article citations. Traditional measures include 'h-index'; 'i10-index'; and 'g-index' (see Table 1 for detailed description). However, these values are limited and do not account for less easily measured indicators of impact, such as leadership and vision; teaching skills; research quality; teamwork; and collaborations [4]. Recently, more integrated approaches have been suggested, most notably by the San Francisco Declaration on Research Assessment (DORA) [5]; an international initiative to improve the ways researchers and research outputs are evaluated. DORA recommendations include using specific statements on research accomplishments and plans, for example within the curriculum vitae; qualitative appraisal/assessment by peers and colleagues; and transparent and flexible indicators based on open data.

Individual and institutions are increasingly recognising and subscribing to recommendations made by DORA, including Wiley, the publisher of *Anaesthesia* [6]. Specific DORA recommendations for individual researchers to consider: use scientific content rather than publication metrics when submitting publications as evidence for funding and promotion decisions; cite primary literature rather than reviews to give scientific credit where credit is due; use a range of metrics, indicators and supporting statements as evidence of impact; and to challenge assessment practices that rely inappropriately on journal impact factor, instead focusing on the value and influence of research outputs.

Article impact

The traditional metric of article impact is the citation count. This value may vary based on the database used to access citations. For example, Journal Citation Report (JCR; Clarivate, Philadelphia, PA, USA) provides the most conservative values for these indices; Scopus a more balanced result as it includes a wider range of citing sources; and Google Scholar (Alphabet, Inc. Mountain View, CA, USA) the most liberal number of citations as it accounts for non-peer-reviewed sources such as websites.

There is no accepted threshold for what a 'well-cited' article is, though in general, zero citations suggest limited academic impact. As the citation count is known to differ between different fields of research [7], the field-weighted citation impact (FWCI) was designed. This is the ratio of citations for an article to the average number of citations within each field over a three-year window. An FWCI of 1 suggests that manuscript was as cited as expected, whereas > 1 indicates that the manuscript is more cited than expected. For example, the most highly-cited article in *Anaesthesia* in recent years, airway guidelines published in 2020 with ~500 citations (Scopus database) at the time of writing, has a FWCI of 102 [8]. This contrasts with, an original general medical article published in *New England Journal of Medicine* in 2019 with 423 citations on the Scopus with has a FWCI of 66 [9]. The FWCI metric may be a more useful measure than citations alone, but would require increased use by major databases and improved access for researchers with transparent interpretation guides to gain traction as a metric of choice.

Citations demonstrate that other researchers are interested in publication but do not account for other aspects of academic impact. For clinical research in particular, true impact

is measured by change in clinical practice. More widely, policy, behavioural or practice changes are the largest indicators for impact. However, these are all notoriously challenging to measure. Altmetrics (Altmetric, London, UK) and PlumX metrics (Plum Analytics, Philadelphia, PA, USA) are being used as alternative metrics of attention in both public and professional venues. Altmetrics are described in detail by Charlesworth and Selak [1]. The PlumX Metric uses similar methodology to Altmetrics but presents different weighted scores for usage (e.g. clicks, downloads, views); captures (e.g. readers, bookmarks); mentions (e.g. blog posts, news mentions, Wikipedia); social media (e.g. likes, shares, Tweets); and citations (e.g. citation indices, patents and policies). Whilst Altmetrics are more widely used and easier to interpret, PlumX are more granular, and directly includes manuscript citation count [10]. However, neither Altmetric nor PlumX Metrics assess scientific quality, such that low-quality articles have the potential to drive media and social media engagement [11]. Interestingly, there have been recent attempts to incorporate these article-level metrics into quantifying journal-level impact metrics [12], though it remains to be seen whether this will be widely adopted.

Journal impact

There are several metrics that can be used to determine the impact of a journal (Fig. 1), but no single metric is perfect and a combination of metrics is now recommended. Before DORA, the most applied and recognised metric was the Journal Impact Factor (JIF). This metric evolved from originally being a tool to inform library purchase of journal subscriptions to being a measure of 'importance' or 'quality' [13]. The JIF is published annually by Clarivate (Clarivate, Philadelphia, USA). Citable items are defined as original articles or reviews; thus, editorials and correspondence do not affect the denominator. An example of this calculation for *Anaesthesia*'s 2021 JIF is shown in Figure 2. Notably, as the 2021 JIF analyses all citations throughout the 2021 calendar year, it is not published until mid-2022. Despite being a cornerstone of academia for years, this metric has recently come under increasing criticism given its weaknesses (Table 1) [14]; and should not be used as an isolated metric. The research community is increasingly distancing from JIFs, and it is likely this metric will recede in importance in coming years.

Clarivate also publishes the "5-year Impact Factor" and the "immediacy index". The 5-year index is designed to measure consistency and sustained changes over time whilst the immediacy index is designed to measure more rapid citations, indicative that the journal publishes topical articles, disseminated widely at the point of publication. CiteSocre is an alternative journal metric, published by Elsevier (Amsterdam, Netherlands). This metric includes citations over a 4-year window and from a wider range of sources including articles, reviews, conference articles, book chapters and databases. An example of the latest 2021 CiteScore calculation for *Anaesthesia* is shown in Figure 2. The Eigenfactor score is another alternative metric, designed to account for the 'importance' of citations, weighted according to the prominence of the source journal. [15]. Thus, a citation from an article published in a journal that has many citations is 'worth' more than one from a journal that has very few citations. This metric has yet to become an established measure. The 'Article Influence Score' uses similar methodology to Eigenfactor but is normalised to account for the total number of articles in the cited journal (online Supporting Information Appendix S1).

Reliance on citations within each of these metrics is a major limitation as these do not include other important components of journal quality or impact. For example, author experience, such as time taken for peer-review, ease of manuscript submission and journal dissemination activities, are not quantified. Further, publishing articles that lead to changes in clinical practice or understanding is difficult to quantify.

Institutional impact

The Research Excellence Framework (REF) is used in the UK to assess the quality of research in UK higher education institutions. This information is vitally important in higher education institutes as it informs the allocation of ~£2 billion in public research funding investment annually. The REF defines research impact as "*as an effect on, change or benefit to the economy, society, culture, public policy or services, health, the environment or quality of life, beyond academia* [16]". The REF2021 exercise has recently been completed with an overall research quality profile produced for each submitting institution. This aggregated assessment is based on the research outputs (60%); impact (25%); and research environment (10%) for individual institutions. Individual researchers are 'returned' within the REF submission if they meet criteria as set out by their institution. Meeting criteria to be 'REF returned' is an important benchmark for UK-based researchers with academic contracts but may be less

relevant for clinical academics employed outside of higher education institutions and the UK. Criteria usually require that individuals hold an academic employment contract with research responsibilities and work as an 'independent' researcher. Individual scores for publications are determined by expert peer review and aggregated within the overall institutional REF submission. Scores are not made available to individual researchers.

The Knowledge Exchange Framework (KEF) is a newer measure for UK higher education institutions [17]. The aim of this metric is to assess how knowledge generated by higher education institutions is shared and used for the benefit of the economy and wider society. Since February 2022, a dashboard has been produced for individual institutions [17], ranking the strength of knowledge exchange activities across seven domains by decile (e.g. an institution may be in the top 10% or bottom 50% of all institutions for different domains). The assessment domains are research partnerships; working with business; working with the public and third sector; skills, enterprise and entrepreneurship; local growth and regeneration; intellectual property and commercialisation; and public and community engagement. Both KEF and REF have recently been refined to align with DORA principles. These include recommendations that institutions should use explicit and transparent criteria to reach hiring, tenure and promotion decisions; and to consider the value of all research outputs (e.g. data sets and software) in addition to publications when considering research outputs [5].

Conclusion

Measuring the impact of research is complex and multifactorial. A systematic approach using individual-, manuscript-, journal- and institutional-level metrics is likely to be the optimal approach to measuring impact of research outputs. For clinical research, there remains a dearth of metrics that demonstrate direct changes to patient care, and the true clinical impact is challenging to measure. When evaluating the academic impact of an individual, it is critical that holistic approaches are applied to recognise their work and that this should be considered within the wider research ecosystem is employed. No single metric should be used in isolation, and fundamentally, the principle of Goodhart's law should be employed: "*When a measure becomes a target, it ceases to be a good measure*."

Acknowledgements

BM and KE are editors of *Anaesthesia*. No external funding or competing interests declared.

References

1. Charlesworth M, Selak T. How to disseminate clinical research. *Anaesthesia* 2022; **77:** 1051-4.
2. Smith R. Measuring research impact—all the rage but hard to get right, 2018. https://blogs.bmj.com/bmj/2018/07/30/richard-smith-measuring-research-impact-rage-hard-get-right/ (accessed 12/08/2022).
3. ResearchGate. Why we're removing the RG Score (and what's next), 2022. https://www.researchgate.net/researchgate-updates/removing-the-rg-score (accessed 12/08/2022).
4. Centre for Science. and Technology Studies. Halt the H-index, 2021. https://doi.org/10.5281/zenodo.4635649 (accessed 12/08/2022).
5. Declaration on Research Assessement (DORA). About DORA, 2012. https://sfdora.org/about-dora/ (accessed 12/08/2022).
6. Wiley. Wiley Signs Declaration on Research Assessment, Deepens Commitment to Responsible Research Assessment 2022. https://newsroom.wiley.com/press-releases/press-release-details/2022/Wiley-Signs-Declaration-on-Research-Assessment-Deepens-Commitment-to-Responsible-Research-Assessment-/default.aspx (accessed 12/08/2022).
7. Aksnes DW, Langfeldt L, Wouters P. Citations, Citation Indicators, and Research Quality: An Overview of Basic Concepts and Theories. *SAGE Open* 2019; **9:** 2158244019829575.
8. Cook TM, El-Boghdadly K, McGuire B, McNarry AF, Patel A, Higgs A. Consensus guidelines for managing the airway in patients with COVID-19: Guidelines from the Difficult Airway Society, the Association of Anaesthetists the Intensive Care Society, the Faculty of Intensive Care Medicine and the Royal College of Anaesthetists. *Anaesthesia* 2020; **75:** 785-99.
9. DeFilipp Z, Bloom PP, Torres Soto M, et al. Drug-Resistant E. coli Bacteremia Transmitted by Fecal Microbiota Transplant. *N Engl J Med* 2019; **381:** 2043-50.
10. Tornberg HN, Moezinia C, Wei C, et al. Assessing the Dissemination of COVID-19 Articles Across Social Media With Altmetric and PlumX Metrics: Correlational Study. *J Med Internet Res* 2021; **23:** e21408.
11. Qiu C, Lin JC, Shi JM, et al. Association Between Epidural Analgesia During Labor and Risk of Autism Spectrum Disorders in Offspring. *JAMA Pediatr* 2020; **174:** 1168-75.
12. Fassoulaki A, Tsaroucha A, Micha G, Soulioti E. Bibliometric analysis of alternative performance metrics for peri-operative, critical care and pain medicine journals. *Anaesthesia* 2021; **76:** 856-7.
13. Garfield E. The history and meaning of the journal impact factor. *Jama* 2006; **295:** 90-3.
14. Paulus FM, Cruz N, Krach S. The Impact Factor Fallacy. *Front Psychol* 2018; **9:** 1487.
15. Bergstrom CT, West JD, Wiseman MA. The Eigenfactor metrics. *J Neurosci* 2008; **28:** 11433-4.
16. Research Excellence Framework 2021 (REF2021). Guidance on submissions (2019/01), 2019. https://www.ref.ac.uk/publications-and-reports/guidance-on-submissions-201901/ (accessed 12/08/2022).
17. Research England. Knowledge Exchange Framework, 2021. https://kef.ac.uk/dashboard (accessed 12/08/2022).

Tables

Table 1.

Metric	Calculation	Strengths	Weaknesses
Individual-level			
H-index	The number of articles (h) that have received (h) or more citations. An author with an h-index of 11 has 11 articles cited at least 11 times	1. Objective measure quantity and impact in a single measure. 2. Easy to calculate, understand and obtain	1. Career duration dependent, leading to early career researcher prejudice 2. Disciplinary variation in typical H values 3. Poorly weighted for highly cited articles 4. May encourage excessive self-citation
i10-index	The number of publications with at least 10 citations	1. Easy to calculate and understand 2. Simple to use and obtain	1. Google Scholar specific index, not used widely in other fora
g-index	With articles ranked in decreasing order of citation count, the *g*-index is the unique largest number where the top *g* articles have at least g^2 citations. Hence, a *g*-index of 10 indicates that the top 10 publications have been cited at least 100 times (10^2)	1. Greater weight to highly cited articles compared with H index 2. Allows higher cited articles to bolster lower cited articles (g index usually higher than h index for individual researchers)	1. Less known/accepted compared with H index 2. May be less intuitive metric compared with other indices
Journal-level			
Journal Impact Factor	Number of citations in a year attributed to articles published in that journal during the two preceding years divided by the total number of citable items published in that journal	1. Most established and commonly used 2. Easy to understand and obtain 3. Influential in the scientific community	1. Uncertain marker of scientific quality, rigour and reliability 2. Inaccurate estimate of citations to individual articles within a journal 3. Easily manipulated

	during those years. Published by Clarivate using the JCR database.		4. Can be skewed by a few highly cited articles 5. Miss the peak citation potential of articles at 3-4 years. 6. Not reproducible as dataset unavailable
CiteScore	Number of citations over four years divided by the total number of citable items published in that journal during those years. Published by Elsevier using the Scopus database.	1. Easy to understand and obtain 2. Captures peak citation potential of articles	1. Novel score, not yet widely used 2. Uncertain marker of scientific quality, rigour and reliability 3. Inaccurate estimate of citations to individual articles within a journal 4. Easily manipulated 5. Can be skewed by a few highly cited articles
Eigenfactor score	Total number of citations over previous five JCR years, weighted according to how highly-cited journals from which citations came from are. Thus, highly cited journals will influence the network more than lesser cited journals.	1. Not influenced by self-citation 2. Accounts for more than just number citations 3. Captures peak citation potential of articles	1. Not yet widely used 2. Uncertain marker of scientific quality, rigour and reliability 3. Inaccurate estimate of citations to individual articles within a journal 4. Easily manipulated 5. Can be skewed by a few highly cited articles 6. Complex calculation
Article-level			

Citations	Number of published articles in indexed peer-review journals referencing an article	1. Widely used and easily understood 2. Robust measure	1. Does not measure scientific quality 2. Differences in citations depending on database searched 3. Does not include citing source 4. May be manipulated
Altmetric	Composite score including mentions in public policy documents; mainstream media; online reference managers; post-publication peer-review platforms; Wikipedia; Open Syllabus Project; patents; blogs; Research highlights such as F1000; and social media and other online platforms.	3. Real-time reporting 4. Including a wide range of sources 5. Beyond citations only 6. Widely used and easily understood	5. Poor marker of scientific quality 6. Algorithms not open access 7. Not comparable across disciplines 8. May be manipulated
PlumX Metric	Weighted scores for usage; captures; mentions; social media; and citations	1. Real-time reporting 2. Including a wide range of sources 3. Beyond citations only	1. Does not measure scientific quality 2. Algorithms not open access 3. Not comparable across disciplines 4. May be manipulated 5. Less used and more complex than Altmetric

Figure Legends

Figure 1. Various impact metrics in anaesthesia journals over five years. a) Journal Impact Factor (yellow), CiteScore (red) and normalised Eigenfactor scores over 5 years for *Anaesthesia*. b) Journal Impact Factor for journals in the field of anaesthesia in 2017 (yellow), 2018 (red), 2019 (green), 2020 (blue) and 2021 (purple). It can be seen that metrics are dynamic and adjust over time.

Figure 2. Calculation of *Anaesthesia's* 2021 a) Journal Impact Factor (JIF); and b) CiteScore (CS).

Figure 3. Potential metrics for measuring research impact. This only provides a general framework for considering how to measure academic impact. Each metric is flawed, none should be used in isolation, and in within clinical practice, none reliably assess clinical impact.

Online Supporting Information
Appendix S1. Calculation of the Article Influence Score.