Clarivate
Web of Science™

Interpreting the citation performance of individual researchers with beamplots

By Martin Szomszor and David A. Pendlebury

Executive summary

This report presents a new visualization tool that showcases the range of a researcher's publication and citation impact in a single data exhibit – the Web of Science™ Author Impact Beamplots. Rather than drawing attention to a single point metric, such as the h-index, beamplots draw the user into the data and surface details that are usually obscured by summary indicators.

This presentation significantly enhances the researcher evaluation toolkit by providing deeper insights into an individual's publication portfolio and is well aligned with community efforts to reform research assessment and encourage the responsible use of metrics. Beamplots make use of field-normalized citation metrics and do not unduly penalize those with gaps in the publication record or those who work in fields with distinctly different publication activity.

We provide guidance on how to interpret beamplots using a hypothetical hiring scenario and real (anonymized) data. The guidance highlights that although publication and citation data are useful indicators of research activity and impact, they must be considered alongside a variety of contributions made by academics and suitably contextualized for the individual's particular history.

Beamplots make use of field-normalized citation metrics and do not unduly penalize those with gaps in the publication record or those who work in fields with distinctly different publication activity.

Context

Many researchers are rightly concerned about management approaches that reduce their work to performance scores (Hammarfelt & Rushforth, 2017). Publication and citation metrics have become more common and more prominent in academic appointments, promotions and funding despite the historic preeminence of peer review. When individual evaluation involves such metrics, the task is both delicate and dangerous (Edwards & Roy, 2017).

The h-index (Hirsch, 2005) is one of the most frequently used metrics of achievement because it appears to summarize both research output and influence in a single number. For example, a researcher with an h-index of 17 has 17 publications each cited at least 17 times. But it is important to ask, "What's behind the score?" The h-index of two individuals may be the same despite very different contributions: one may have some moderately cited and many uncited papers while another has fewer but much more highly cited papers.

The h-index has other defects (Waltman & van Eck, 2012). It is biased in favor of senior researchers with many publications and to the detriment of early career researchers with briefer bibliographies. It cannot compare disciplines that differ in their average citation rates.

Nor does it account for the time papers take to accrue citations. It remains steady and does not fall even if a researcher produces nothing new. Dependence on such simple and inadequate metrics is a sign of indicator impoverishment and lack of awareness of best practices.

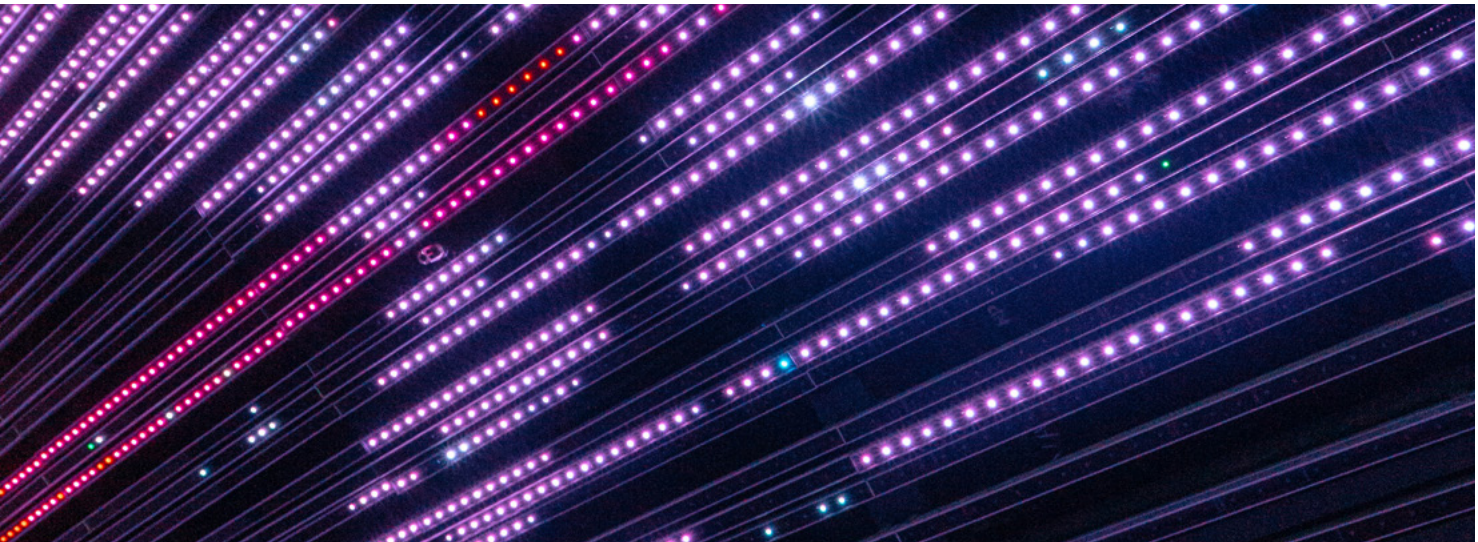
Researchers have proposed reforms and practices to address misguided and improper uses of metrics. The San Francisco Declaration on Research Assessment (DORA, 2012) condemns the use of the Journal Impact Factor™ to judge individual papers or people. Scientometricians themselves state explicitly: "Don't reduce individual performance to a single number" (Wouters et al, 2013). The Leiden Manifesto (Hicks et al, 2015) offers 10 principles for responsible research evaluation and Clarivate also concurs that "Indicators must not substitute for informed judgement."

Dubious evaluation practices also undermine research integrity (Szomszor & Quaderi, 2020) because they incentivize practices that can be gamed through manipulation and misrepresentation of the publication and citation record. Consequently, Responsible Research Assessment (RRA) – "…an umbrella term for approaches to assessment which incentivize, reflect and reward the plural characteristics of high-quality research, in support of diverse and inclusive research cultures" (Curry et al, 2020) – is a growing topic of interest and participation from institutions and funders alike.

Qualitative considerations, which require peer expertise, are essential to informed judgment, and deployment of more informed metrics that address the multidimensional nature of research activity is needed (Moed & Halevi, 2015). For example, percentile scores should be preferred in dealing with skewed citation distributions, and these should be normalized for field and year of each paper (Bornmann & Marx, 2014a). It should also be mandatory to drill down into a collection of papers to understand what went into a summary statistic.

Clarivate recommends that individual researcher evaluation should always include a variety of appropriate quantitative indicators and qualitative assessment, not least the careful reading of the researcher's best representative publications. For assessors, the goal should be to recognize and appreciate the gestalt of individuals. To this end, Clarivate emphasizes profiles over metrics (Adams et al, 2019).

A new researcher evaluation tool, the beamplot visualization, is an excellent alternative to a single point metric, such as the h-index, because it reveals the volume and citation impact of an individual's publication portfolio through time. Author Impact Beamplots are now available through the Web of Science™ Author Records. In this report, we show how to read a beamplot and provide guidelines on their interpretation and limitations.



Beamplots

It was Lutz Bornmann and Werner Marx of the Max Planck Society who first introduced and developed the use of beamplots for scientometric data in 2014 (Bornmann & Marx 2014a, 2014b; also, Bornmann & Haunschild, 2018, Haunschild et al, 2019).

A beamplot puts a researcher's publications into a context suitable for comparison and unpacks the citation performance of a publication portfolio. Each paper's citation count is normalized (for example, it is benchmarked against other similar publications) and measured as a percentile (see sidebar 'What are percentiles?').

An example beamplot is annotated in Figure 1. The x-axis plots citation performance as a percentile (0 on the left is lowest and means zero citations - 100 on the right is the maximum and corresponds to the highest citation count in the field), with a purple circle representing an individual paper that is placed on the beam according to its citation percentile. Multiple papers with the same percentile are shown with a larger marker. The y-axis shows publication years – with papers for each successive year plotted on a beam, most recent at the top and oldest at the bottom. In each year, the green circle is the median percentile value for publications in that year, and the overall average is shown using a vertical dotted line.

This plot was generated in 2021, so the current year and prior year of publications are not shown.

We suppress recent years from the plot for two reasons:

- it takes years (approximately three, longer for social sciences and humanities) for papers to accumulate meaningful citation impact (Wang,2013); and

- in recent years, where most publications have zero citations, there is a skew in percentiles towards 100. For example, the receipt of just one citation in the current year can elevate the percentile score to over 50 immediately. If the current and prior years were included in the beamplot, there would be a tendency to observe a skew towards 100 in the recent years that would unduly influence interpretation of the data.

These limitations mean that new researchers will not be able to see their publication record in a beamplot.

But as Bornmann and Marx note in the original publication, "…young scientists should not pay too much attention to bibliometric indicators and should just concentrate on their research. As a rule, good research results in attention and, later on, in a substantial citation impact" (Bornmann & Marx 2014b).

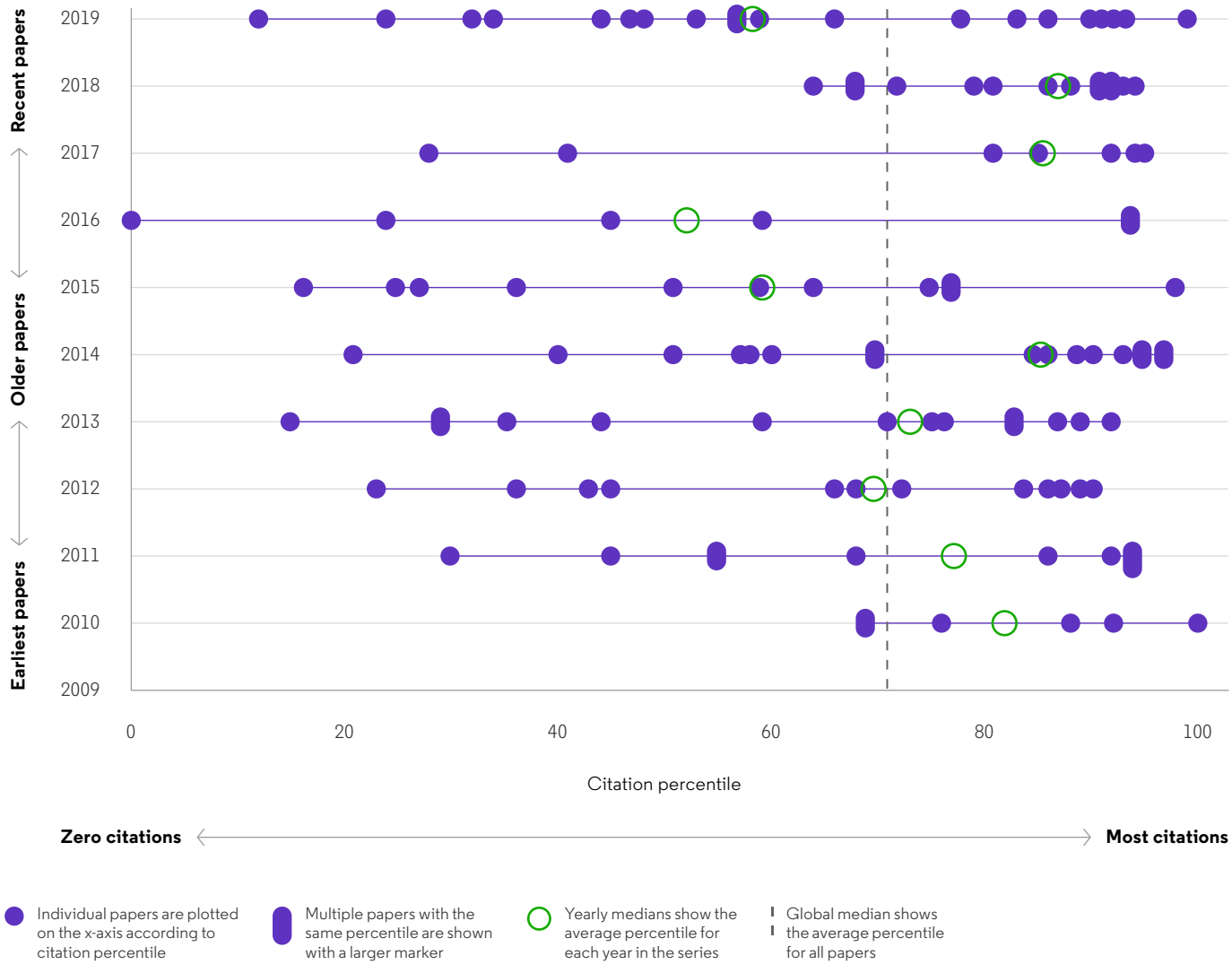
Beamplots provide an excellent way to unpack a single point metric. Because they are based on percentile measurements of citation impact, they are more amenable to comparison across fields than the h-index. The method of normalization used also caters for article type (i.e., an article or a review) so there is no implicit advantage associated with certain types of documents that on average are more cited, such as reviews.

Like the h-index, the use of beamplots is limited to individuals who have accumulated a significant publication portfolio over a number of years. Unlike the h-index, beamplots are not necessarily biased against individuals with career breaks or changes in volume of publication outputs – gaps in the record will be obvious in a beamplot, but the citation performance on either side will be accurately reflected.

The crucial feature of beamplots that positions them favorably in the context of responsible research assessment is that they steer us away from reduction to a single point metric and force us to consider why the citation performance is the way it is. As we outline further in the next section, the beamplot is a useful narrative tool that can refute or corroborate other evaluation criteria, and should be considered alongside contextual information about an individual, such as where they were working at the time, the nature of any collaborative projects and the type of research involved.

A beamplot puts a researcher's publications into a context suitable for comparison and unpacks the citation performance of a publication portfolio.

Figure 1: Example Beamplot



The crucial feature of beamplots that positions them favorably in the context of responsible research assessment is that they steer us away from reduction to a single point metric and force us to consider why the citation performance is the way it is.

Percentiles

Because citation distributions are strongly skewed, percentiles provide a good measure of central tendency.

90%+

score means that a paper is among the 10% most cited in the field

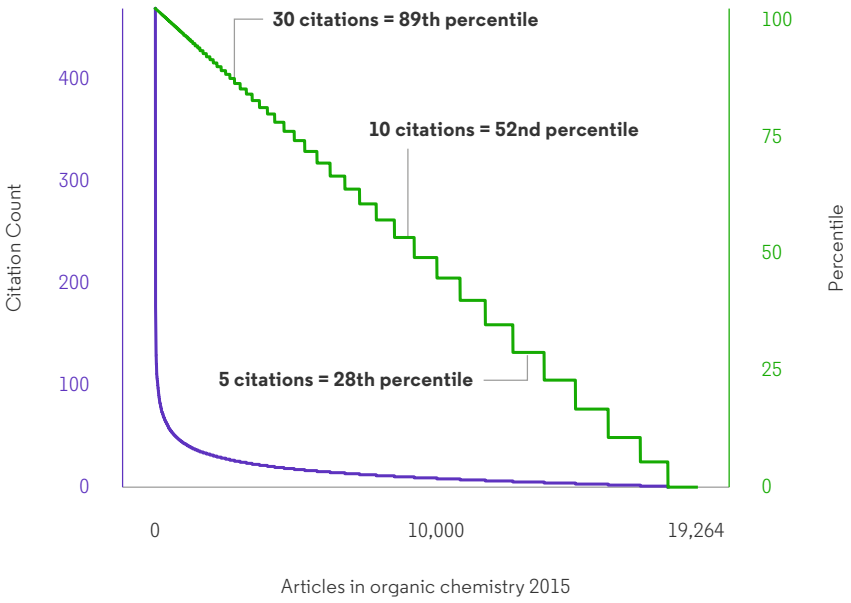
When comparing citation counts, it is important to account for variation in how citations accumulate across different fields. Differences in the volume of papers published, the average number of references made (citation density) and the tendency to cite older or more recent work all affect the rate at which citations accrue. Hence, it is common for bibliometric indicators to account for these differences using normalization – the comparison of citation counts to a reference set (or benchmark) that is based on:

- Papers published in the same year
- Papers of the same document type (to distinguish articles from reviews or proceedings, for example)
- Papers published in the same field or subject category

Using a reference set enables comparisons of citation counts across disciplines. However, the underlying skewness in the distribution of citations is also a problem. Many papers are never cited, most will only collect a few and a very small number will gather hundreds and perhaps thousands of them. This skew means traditional measurements, such as the mean or median, do not accurately convey the average (central tendency) because the data are not normally distributed.

A simple way to address this problem is to use percentiles instead of the raw citation counts. We illustrate this in Figure 2 where the distribution of citation counts is shown for organic chemistry articles published in the Web of Science in 2015. In purple (left axis), the number of citations received is plotted from 462 on the left (the highest citation count), all the way to 0 on the right and shows the characteristic long-tail distribution. In green (right axis), the corresponding percentile is shown, starting on the left at 100 (for example, 100% of other articles in the distribution have a lower citation count) and dropping regularly down to zero. The figure highlights three points (30, 10 and 5 citations) to show percentile values. The steps that become more noticeable toward the right convey the discrete drops in citation count, with the final step denoting the articles that received just one citation.

Figure 2: Citation counts (purple, left axis) for 19,264 articles published in organic chemistry in 2015 and their corresponding percentile values (crimson, right axis)



Interpretation

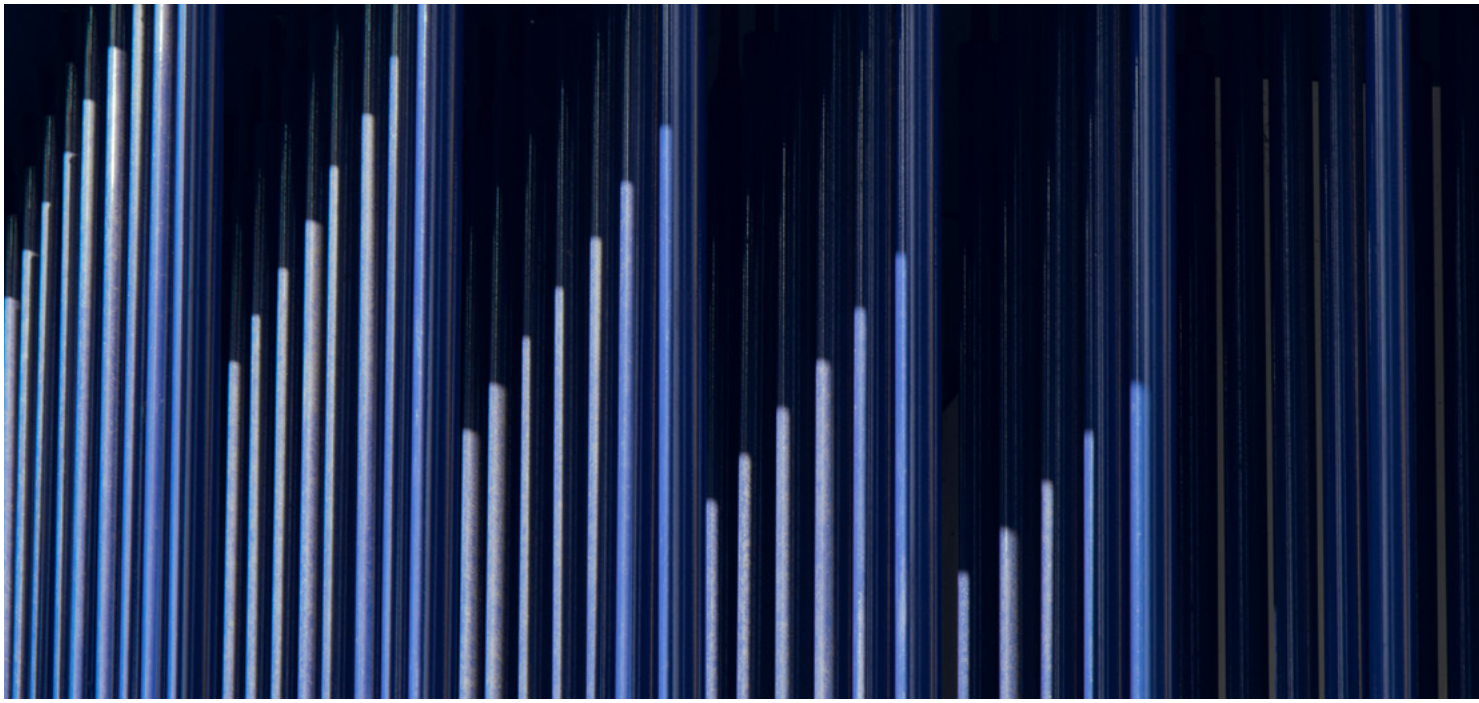
To describe the utility of the beamplot and to position its use within a wider evaluation context, we propose a hypothetical hiring scenario and review citation profiles for four potential candidates. We suppose the vacancy is for a junior teaching and research position (for example, junior lecturer or assistant professor) in the field of organic chemistry.

When hiring for such a role, many factors would be considered that might include aspects relating to professional skills, such as experience and success in applying for grants and knowledge of the grant award system; teaching experience and possible quality metrics that may have been collected, peer-review research contributions and editorial positions; engagement activities (such as speaking events, invited talks and panels, consultations, industry links); and, socio-economic impact of research (spin-out companies, patenting, policy influence), among others. Such factors are usually assessed in relation to the overall strategy of the organization and the potential for the candidate to fit well within the local research environment.

To find sample candidates, we used ResearcherID data to search for individuals whose discipline was 'organic chemistry', who began publishing in 2008 or later and who had at least 20 papers indexed in the Web of Science Core Collection™. This set of candidates was then refined until four representative profiles from the same geographic region were selected, referred to herein as researcher A, B, C and D. Basic summary metrics for each are listed in Table 1 – Bibliometric indicators for four sample researchers in organic chemistry. Each researcher has produced a similar number of papers, published in high quality journals (Q1 papers in Journal Citation Reports™), but there is variation in their citation metrics – Times Cited, h-index, Mean CNCI (the average Category Normalized Citation Impact) and Mean Percentile. Researcher A has the largest number of papers with international collaborators (50%). Based on this information alone, one might rank Researcher A highest.

Table 1: Bibliometric indicators for four sample researchers in organic chemistry

	A	B	C	D
Paper count	28	33	23	21
Times cited	698	354	384	345
H-index	15	13	12	11
Mean CNCI	1.17	0.52	0.86	1.09
Mean percentile	63	41	56	50
% inter. collab	50%	24%	22%	14%
Q1 JIF papers	13	13	13	11
% papers in top 10%	21%	3%	4%	19%



In Figure 3, we provide beamplots for the same set of researchers. With this view, new details on the trends in publication and citation performance for each are apparent. For example, Researcher A (who has the highest overall citation performance) has an excellent record between 2012 and 2015, but exhibits a marked drop-off in average percentile for the years 2016 to 2018. Researcher B also has a similar overall trend (for example, better citation performance in earlier publication years). Researcher D has a different profile, with a year-on-year increase in citation performance.

Of course, there are a variety of reasons for changes in citation performance, many of which are expected as careers progress. To use the beamplot in a responsible evaluation setting, it would be necessary to bear in mind the following:

Location – Where were they based? Locality will influence publication opportunity, both at a national level and at a specific institution. As researchers move around, there is likely to be an impact on their publication profile.

Role – What job were they doing? Different researchers are able to devote different amounts of time to research owing to other responsibilities (for example, teaching, commercialization or public engagement). Those who supervise Ph.D. students have additional opportunities to publish, although the quality of that research may vary. Doctoral and post doctoral students may have different prospects depending on where they study and who their supervisor is.

Collaboration – Who were they working with? Participation in a large international consortium provides additional routes to publication that may be hard to achieve otherwise. When internationally collaborative papers are published, they generally attract more citations than domestic research and earn higher percentile scores (Adams et al, 2019; Potter et al, 2020). It may be important to consider industry collaboration as that can be used as a proxy for engagement in some applied disciplines.

Nature of research – What kind of research was conducted? Whether basic or applied, the type of research may influence citation performance. Interdisciplinary research may lead to publications across a broad set of fields that could make interpretation of metrics more complicated. Many academics change disciplines over the course of their career and that could explain changes in performance.

In fact, the beamplot provides a basis to clarify these questions and an opportunity to investigate when changes may have had a positive or negative impact. It is possible to imagine a beamplot serving as the focus for discussion between evaluators and evaluatees, either in a formative or summative review.

It would be wrong to form an opinion based solely on the data presented in a beamplot, just as it is to depend on a single metric. As outlined above, context is essential to understanding trends in performance. In our hypothetical hiring scenario, finding out that Researcher D had in fact trained in a less well known and poorly networked institution (2010 to 2013) before moving on to a more prestigious appointment would likely affect the evaluation.

Therefore, it should be apparent that neither single point metrics nor a beamplot should be used as a crude evaluation instrument. However, beamplots do offer advantages because they surface variation in the data that should be investigated and compared against other quantitative and qualitative indicators, leading to a more responsible use of metrics. Importantly, they reveal the data behind composite scores such as the h-index, show the underlying data on a paper-by-paper basis and provide a picture of performance over time.

It is possible to imagine a beamplot serving as the focus for discussion between evaluators and evaluatees, either in a formative or summative review.

Figure 3: Beamplots for four sample researchers in organic chemistry

Figure 3a

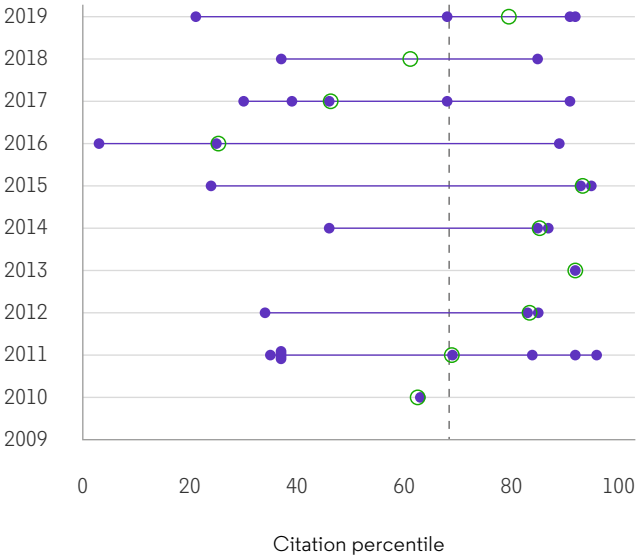


Figure 3b

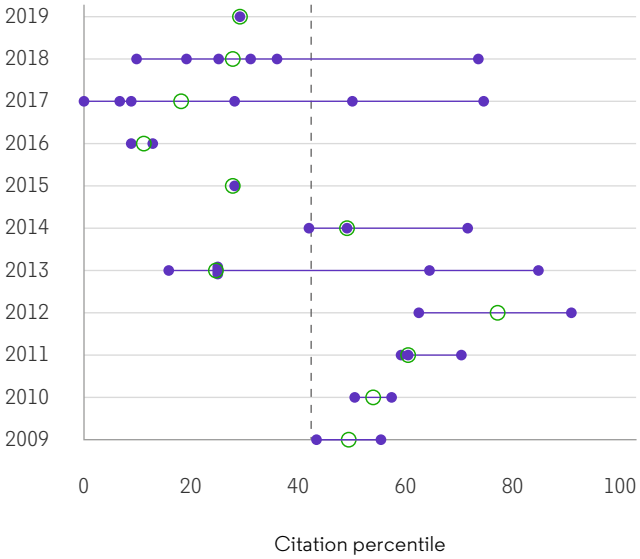


Figure 3c

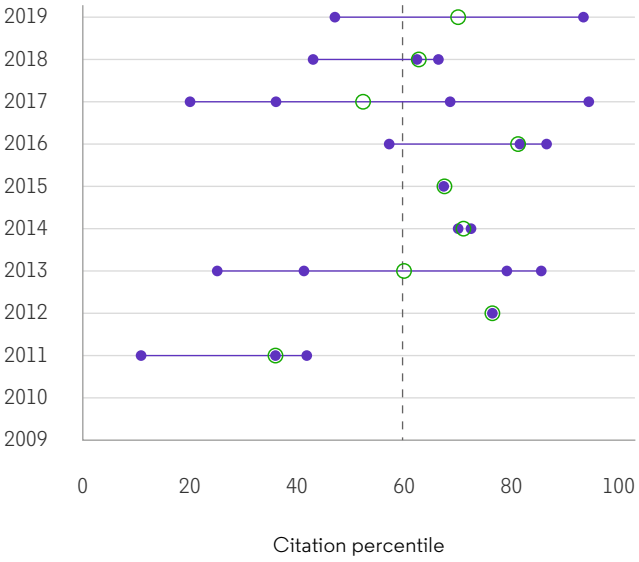
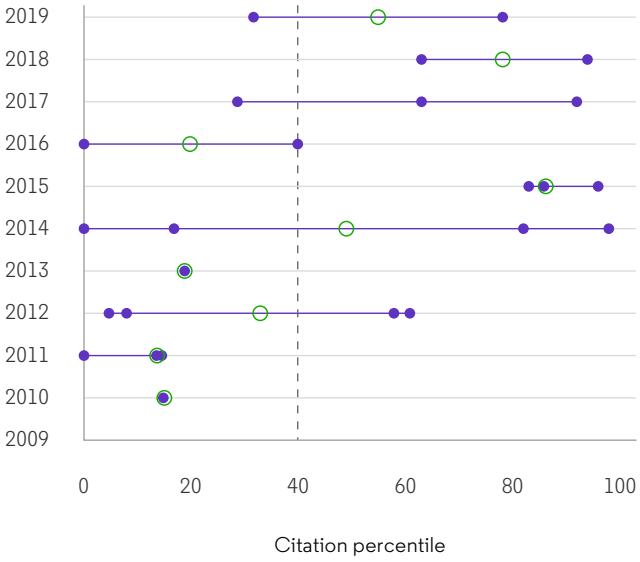


Figure 3d



Conclusions and future work

We believe that the addition of beamplots to the Web of Science Author Records will provide researchers and evaluators with valuable insights relating to the performance of a publication portfolio to help remove the current dependence on existing single-point metrics, eliminate indicator impoverishment and raise awareness of responsible research evaluation practices.

The introduction of field-normalized percentiles to measure citation impact is a crucial aspect of beamplots that addresses many criticisms of the h-index relating to comparability across disciplines and biases to portfolio size and continuity.

But perhaps more importantly, we believe that providing such a tool will encourage users of researcher metrics to consider what actually makes a metric and to engage more actively with the data. Assuming that users consider the interpretive guidance given in this report, beamplots will provide new opportunities to conduct research assessment in a responsible manner.

We foresee a number of future developments that would enhance the capability of beamplots. In this first iteration, only a single person's beamplot is displayed. Of course, it is possible to compare them side-by-side, but a feature that provides this functionality explicitly may be desirable. This would be a challenging design task due to the complex nature of the visualization.

Currently, all publications are displayed, but filters could be applied to segregate the portfolio according to certain facets such as funder acknowledgements, open access status, collaboration type (for example domestic, international or industrial), author position (first, corresponding, or last) or named co-authors and collaborating institutions.

This kind of improvement would allow researchers to learn more about the publication performance of certain projects and collaborations.

See for yourself – Go to the new Web of Science Author Search to look up a researcher and check out their author record to discover their beamplot.

The introduction of field-normalized percentiles to measure citation impact is a crucial aspect of beamplots that addresses many criticisms of the h-index relating to comparability across disciplines and biases to portfolio size and continuity.

References

Adams, J., Pendlebury, D., Potter, R., & Szomszor, M. (2019). Global research report: Multi-authorship and research analytics. Philadelphia and London: Institute for Scientific Information, Clarivate. ISBN 978-1-9160868-6-9

Adams, J., McVeigh, M., Pendlebury, D., & Szomszor, M. (2019). Profiles, not metrics. Philadelphia and London: Institute for Scientific Information, Clarivate Analytics, 6-7. ISBN 978-1-9160868-2-1

Bornmann, L., & Marx, W. (2014a). How to evaluate individual researchers working in the natural and life sciences meaningfully? A proposal of methods based on percentiles of citations. Scientometrics, 98(1), 487-509.

Bornmann, L., & Marx, W. (2014b). Distributions instead of single numbers: Percentiles and beam plots for the assessment of single researchers. Journal of the Association for Information Science and Technology, 65(1), 206-208.

Bornmann, L., & Haunschild, R. (2018). Plots for visualizing paper impact and journal impact of single researchers in a single graph. Scientometrics, 115(1), 385-394.

Curry, S., de Rijcke, S., Hatch, A., Pillay, D., van der Weijden, I., & Wilsdon, J. (2020). The changing role of funders in responsible research assessment: Progress, obstacles and the way ahead. Research on Research Institute. Report. https://doi.org/10.6084/m9.figshare.13227914.v1

Edwards, M. A., & Roy, S. (2017). Academic research in the 21st century: Maintaining scientific integrity in a climate of perverse incentives and hypercompetition. Environmental Engineering Science, 34(8), 51-61.

Hammarfelt, B., & Rushforth, A. D. (2017). Indicators as judgment devices: An empirical study of citizen bibliometrics in research evaluation. Research Evaluation, 26(3), 169-180.

Haunschild, R., Bornmann, L., & Adams, J. (2019). R package for producing beamplots as a preferred alternative to the h index when assessing single researchers (based on downloads from Web of Science). Scientometrics, 120(2), 925-927.

Hicks, D., Wouters, P., Waltman, L., de Rijcke, S., & Rafols, I. (2015). The Leiden Manifesto for research metrics. Nature, 520(7548), 429-431.

Hirsch, J. E. (2005). An index to quantify an individual's scientific research output. Proceedings of the National Academy of Sciences of the United States of America, 102(46), 16569-16572.

Moed, H. F., & Halevi, G. (2015). Multidimensional assessment of scholarly research impact. Journal of the Association for Information Science and Technology, 66(10), 1988-2002.

Potter, R. W. K., Szomszor, M., & Adams, J. (2020). Interpreting CNCIs on a country-scale: The effect of domestic and international collaboration type. Journal of Informetrics, 14(4), 101075.

San Francisco Declaration on Research Assessment (DORA). (http://am.ascb.org/dora/)

Szomszor, M., & Quaderi, N. (2020). Research Integrity: Understanding our shared responsibility for a sustainable scholarly ecosystem. London and Philadelphia: Institute for Scientific Information, Clarivate. ISBN 978-1-9160868-9-0

Wang, J. (2013). Citation time window choice for research impact evaluation. Scientometrics, 94(3), 851-872.

Waltman, L., & van Eck, N. J. (2012). The inconsistency of the h-index. Journal of the American Society for Information Science and Technology, 63(2), 406–415.

Wouters, P., Glänzel, W., Gläser, J., & Rafols, I. (2013). The dilemmas of performance indicators for individual researchers: An urgent debate in bibliometrics. ISSI Newsletter [no. 35], 9(3), 48-53.

WS622456320 / 04

About Clarivate

Clarivate™ is a global leader in providing solutions to accelerate the lifecycle of innovation. Our bold mission is to help customers solve some of the world's most complex problems by providing actionable information and insights that reduce the time from new ideas to life-changing inventions in the areas of science and intellectual property. We help customers discover, protect and commercialize their inventions using our trusted subscription and technology-based solutions coupled with deep domain expertise. For more information, please visit clarivate.com.

www.clarivate.com/webofsciencegroup