

# Comparing People with Bibliometrics

Michael J. Kurtz<sup>1,\*</sup>

<sup>1</sup>Harvard-Smithsonian Center for Astrophysics

**Abstract.** Bibliometric indicators, citation counts and/or download counts are increasingly being used to inform personnel decisions such as hiring or promotions. These statistics are very often misused. Here we provide a guide to the factors which should be considered when using these so-called quantitative measures to evaluate people. Rules of thumb are given for when begin to use bibliometric measures when comparing otherwise similar candidates.

## 1 Some things you should know when using bibliometrics to compare people

This is a serious business. Careers can be greatly affected by the decisions which flow from bibliometric evaluations. [1, 2]

It is important to know how accurate these measures are. Measures such as citation or download counts are not important in themselves, they are proxies for evaluating the (research) capabilities of an individual. [1, 3, 4]

While the relative completeness and accuracy of some sources of bibliometric information are frequently discussed, this is not particularly important to comparisons of people. The statistical scatter, even with “perfect” input data in these measures is quite large, and puts limits on any attempt to use the data to evaluate people. [1]

Before using citations or downloads to compare individuals these measures must be adjusted to account for the known systematic errors due to age, discipline, and co-authorship. [5]

The total number of citations a person’s work receives increases approximately quadratically with the “age” of that individual, where “age” means time as an active researcher. [5, 6] The correction is that the square root of the number of citations, divided by “age” is an approximate constant for persons with uniform productivity. For younger researchers determining the effective “age” is problematic, and makes this correction uncertain. [1]

Different disciplines, and subdisciplines, have different citation cultures, citation counts cannot be directly compared without correcting for this. [5, 7]

\*e-mail: [kurtz@cfa.harvard.edu](mailto:kurtz@cfa.harvard.edu) ORCID: 0000-0002-6949-0090

Multi-author papers are now the norm. Bibliometric measures must take the degree of co-authorship into account. [5, 8, 9]

For comparisons with individuals measured at different times one must account for the steady increase in the number of papers, which double about every 15 years [10, 11], and in the number of references per paper, which is experiencing a similar increase [12, 13].

The TORI statistic [5], which normalizes citations by both the number of authors on the cited document and the number of references in the citing document, is intended to remove the effects of co-authorship and citation culture. The RIQ statistic [5] is intended to remove the effects of age. Both are available on the ADS metrics pages.

Downloads, or reads, are also a useful bibliometric indicator [4, 9, 14]. Downloads by researchers have very similar properties to, and can predict citations [11, 14, 15]. Great care must be taken when using download information, however, as downloads by persons who are not researchers have very different properties, and can dominate the statistic [4, 7, 13].

The ADS metrics pages show the Read10 statistic [1, 9], which is the yearly sum of co-author number normalized reads by scientists via ADS of papers written in the preceding ten years. Because the use of internet based services has grown rapidly in the recent past care must be taken when comparing download counts from different time frames.

Using samples and measures designed to minimize the systematic errors in the measurements we can determine the intrinsic scatter in the use of bibliometric techniques to evaluate people. We do this by comparing different measures of exactly the same sets of papers by exactly the same people. Because the actual impact of these sets of papers is identical (because they are exactly the same) the differences in impact as measured by different techniques must be due to the intrinsic measurement error in the techniques. Complete details are in [1].

Because citations and reads of people form a log-normal distribution [1, 17] the error is in the logarithm of the measurements. For the comparisons using active mid-career astronomers the one standard deviation error is 0.17 dex, a multiplicative factor of 1.48. [1]

Citations (and downloads, by extension) are also used to predict performance; essentially this is why they are used in hiring decisions. For individuals five years past the PhD, about the time when assistant professor positions are granted, even using an a posteriori selected sample, substantially more prescient than any contemporaneously selected sample could be, the one standard deviation error is 0.395 dex, a multiplicative factor of 2.5. [1]

## 2 Conclusions

Bibliometric measures, citation and download counts, have been shown to be useful and accurate in a number of areas. When aggregated over large entities, such as universities [13, 18], countries [13, 19], or journals [20] bibliometric measures can provide accurate measures of the quantity and quality of research. Additionally, persons with very high citation counts are, almost always, very well regarded researchers. This means that citations can be accurately used to recognize these individuals,

and predict high honors [3].

These successes have led some to use bibliometric measures to evaluate people. In most cases this constitutes a misuse of the statistics. Decision aids such as citation or download statistics may well be able to identify most of the top people in a field, but they have almost no power in making the more common 41st chair [21] type of personnel decision: deciding between a set of similarly qualified candidates.

Because in any real instance the various corrections for age, co-authorship, subfield, etc will not be as clear as in the near ideal case analyzed in [1] it is reasonable to require that, when comparing two individuals, the difference in their citation or download measures be greater than two standard deviations (95% confidence) before ascribing any difference to them on the basis of those counts. This would be about a factor of two when comparing the current status of mid-career scientists, and about a factor of five when predicting the future productivity of possible assistant professor hires. Outside of these bounds one would want to understand why the counts were so different.

## References

- [1] M.J. Kurtz, E.A. Henneken, *Journal of the Association for Information Science and Technology* **68**, 695–708 (2017)
- [2] L. Bornmann, W. Marx, *Zeitschrift fuer Evaluation* **12**, 103-127 (2013)
- [3] E. Garfield, M.V. Malin, *Can Nobel Prize Winners be Predicted?* Paper presented at the 135th Annual Meeting of the American Association for the Advancement of Science, (1968)
- [4] M.J. Kurtz, J. Bollen, *Annual Reviews of Information Science and Technology* **44(1)**, 1-64 (2010)
- [5] A. Pepe, M.J. Kurtz, *A Measure of Research Impact Independent of Time and Discipline*, PLoSONE e46428, (2012)
- [6] J.E. Hirsch, *An index to quantify an individual's scientific research output* in *Proceedings of the National Academy of Science* (2005), Vol. 102(46), p. 16569–16572
- [7] D. Nicholas, P. Huntington, T. Dobrowolski, I. Rowlands, M. Jamali, P. Polydoratou, P., *Information Processing & Management* **41**, 1441–1461 (2005)
- [8] D.de S. Price, *Journal of the American Society for Information Science* **27 (5-6)**, 292-306 (1976)
- [9] M.J. Kurtz, G. Eichhorn, A. Accomazzi, C.S. Grant, M. Demleitner, S.S. Murray, N. Martimbeau, B. Elwell, *Journal of the American Society for Information Science and Technology* **56(2)**, 111-128. (2005)
- [10] D.de S. Price, *Little Science, Big Science*, New York: Colombia University Press, (1963), ISBN-10: 0231049560
- [11] M.J. Kurtz, G. Eichhorn, A. Accomazzi, C.S. Grant, S.S. Murray, J.M. Watson, *Astronomy and Astrophysics Supplement Series* **143**, 41-59 (2000)
- [12] E.A. Henneken, M.J. Kurtz, A. Accomazzi, C.S. Grant, D. Thompson, E. Bohlen, S.S. Murray, *Journal of Informetrics* **1**, 1-8 (2009)
- [13] E.A. Henneken, M.J. Kurtz, *Usage Bibliometrics as a Tool to Measure Research Activity* to appear in: *Handbook of Quantitative Science and Technology Research*, Springer, ArXiv e-prints arXiv:1706.02153. (2017)

- [14] S. Haustein, *Readership Metrics in Beyond Bibliometrics*, edited by B. Cronin and Cassidy R. DSugimoto, Cambridge:MIT Press (2014), 327-344
- [15] H.F. Moed, G. Halevi, *Journal of the Association for Information Science and Technology* **67**, 412–431. (2016)
- [16] T. Stankus, B.A. Rice, *Collection Management* **4(1/2)**, 95–110 (1982)
- [17] W. Shockley, *Proceedings of the Institute of Radio Engineers* tectbf45(3), 79–290 (1957)
- [18] Times Higher Education Ranking <https://www.timeshighereducation.com> retrieved (2017)
- [19] National Science Board, “Science & Engineering Indicators” <https://www.nsf.gov/statistics/2016/nsb20161/#/> retrieved (2017)
- [20] E. Garfield, *Science* **178(4060)**, 471–479 (1972)
- [21] R.K. Merton, *Science* **159**, 56 (1968)