CrossMark

Interpreting correlations between citation counts and other indicators

Mike Thelwall[1]

Abstract Altmetrics or other indicators for the impact of academic outputs are often correlated with citation counts in order to help assess their value. Nevertheless, there are no guidelines about how to assess the strengths of the correlations found. This is a problem because the correlation strength affects the conclusions that should be drawn. In response, this article uses experimental simulations to assess the correlation strengths to be expected under various different conditions. The results show that the correlation strength reflects not only the underlying degree of association but also the average magnitude of the numbers involved. Overall, the results suggest that due to the number of assumptions that must be made, in practice it will rarely be possible to make a realistic interpretation of the strength of a correlation coefficient.

Introduction

New indicators are often developed within scientometrics from new data sources, such as altmetrics (Kousha and Thelwall 2015; Thelwall and Kousha 2015a, b), or with a new method to process existing data, such as the h-index (Hirsch 2005). A standard technique to assess the value of any new quantitative indicator is to measure the extent to which it correlates with human judgements or an existing indicator of better known value. This has been proposed as a useful general approach for patent citations (Oppenheim 2000), hyperlink counts (Thelwall 2006) and altmetrics (Sud and Thelwall 2014). A positive

✉ Mike Thelwall
   m.thelwall@wlv.ac.uk

1 Statistical Cybermetrics Research Group, University of Wolverhampton, Wulfruna Street, Wolverhampton WV1 1ST, UK

correlation suggests that the new indicator at least partially reflects the quality that the better known indicator signifies. For example, a statistically significant positive correlation between a new indicator and citation counts suggests that the new indicator relates to scholarly impact or quality, at least if the correlation is for articles from a field for which citation counts tend to reflect scholarly quality. Nevertheless, statistical significance does not give evidence about the *extent* to which the new indicator signals scholarly quality. For this, the magnitude of the correlation coefficient is important. A very high correlation between two indicators suggests that they are essentially equivalent whereas a low correlation suggests that the new indicator predominantly reflects something other than scholarly quality but there are no guidelines about how to interpret specific values.

In contrast to the lack of empirically-grounded guidelines for interpreting correlation coefficients in informetrics, specific values have been suggested in the behavioural sciences. The most widely used guidelines are probably the minimum values of 0.1 for "small", 0.3 for "medium" and 0.5 for "large" (Cohen 1988, 1992). These terms are recommended only when more rigorous alternatives are not available (Cohen 1988). A preferable approach in the behavioural sciences is to compare any new correlation coefficient with other correlation coefficients obtained in similar contexts to see how large it is relative to them (Lipsey et al. 2012). Assuming that these correlation coefficients would all be affected to a similar extent by spurious factors, such as random noise and measurement error, this comparison would hint at the likely underlying strength of association. In psychology, this recommendation is sometimes put into practice in systematic review articles in the form of a table of the range of correlation coefficient values reported in the bottom, middle and upper thirds of (published and unpublished) studies about an issue (Hemphill 2003). From this table, new investigations can explicitly position their correlation coefficients within the range of those found in previous studies. Alternatively, the practical significance of correlation coefficients can sometimes be interpreted in real world contexts, such as to predict the number of lives that would be improved by a medical treatment (Ellis 2010).

At the most basic level, the magnitude of a correlation coefficient should be interpreted relative to the maximum possible value that it would be reasonable to expect from a perfect underlying relationship between the variables being correlated. This intuition underlies the widely used Cronbach's alpha coefficient of reliability (Cronbach 1951). The maximum theoretical size depends upon the natural variability within the data gathered and the presence of unavoidable spurious factors. If the measurements used are highly precise and the natural variation in the phenomenon being measured is small compared to the range of values of the data then a correlation close to 1 would be theoretically possible (e.g., 0.99 in physics: Ettori 2015; Liu et al. 2012). If the measurements are not precise or the natural variation in the phenomenon being measured (i.e., spurious factors that are impossible to control for) is large compared to the range of values of the data then small correlations are the biggest that could be hoped for. This probably applies to all experiments involving subjective assessments by human subjects, for example, because it would be impossible to control for the influence of the different life experiences of the individuals on their judgments.

Correlations between citation counts and alternative indicators are complicated by both being *indirect* indicators of the phenomenon of interest, which is research quality or impact. Because of this, a perfect correlation is not theoretically possible, unless both have the same biases. Mixing disciplines within a data set can also artificially reduce correlation strengths (Thelwall and Fairclough 2015). In addition, document properties that affect citation counts but not necessarily the quality of research can also weaken the relationship

between citation counts and research quality. These may include author nationality, collaboration type, number of co-authors, paper length, number of references, the pure/applied nature of the research, and abstract readability (Didegah and Thelwall 2013; Hartley and Sydes 1997; Kostoff 2007; Larivière and Gingras 2010; Onodera and Yoshikane 2015; Persson et al. 2004). It is not clear whether some of these factors, such as collaboration, tend to produce better research or whether they tend to produce research that is more highly cited for other reasons. A case in point is that multiple authors may generate additional publicity or self-citations for their articles (van Raan 1998). It therefore seems impossible to be sure about all of the factors that can weaken the relationship between research quality and citation counts. This problem is exacerbated by substantial disciplinary differences in citation practices (Hyland 1999). It is also exacerbated by the scarcity of evidence about the underlying quality or impact of academic articles. Although large scale peer review evaluations of the quality of academic articles is collected by some national research exercises, such as that of the UK and Italy, and these have been used for statistical analyses (e.g., HEFCE 2015; Franceschet and Costantini 2011), the data sets are not freely available and have not been used to conclusively identify the factors that influence the relationship between quality or impact and citation counts.

In the absence of comprehensive knowledge about the number and strength of factors that influence the relationship between citation counts and research quality it is impossible to give a convincing interpretation of the strength of a correlation between citation counts and any new indicator. Although it is logical to follow advice from behavioural sciences and compare such correlations with each other to determine whether a coefficient is large relative to other coefficients (Hemphill 2003) there may be too few scientometric studies with correlations and too few genuinely new indicators to make this approach robust for three reasons. First, there are substantial disciplinary differences in citation cultures, the variability of citation counts and the extent to which citation counts reflect scholarly quality or are influenced by spurious factors, as discussed above. Hence, similar correlations for different fields may have substantially different meanings. Second, properties of the data set examined may affect the magnitude of the correlation coefficient. It would be reasonable to expect lower correlation coefficients for uniform data sets (e.g., highly cited articles or articles from a single journal) than from non-uniform data sets (e.g., sets containing interdisciplinary research or all articles from a field of study). Third, the average number of citations may affect the size of a correlation coefficient because citation data is discrete and therefore unable to reveal small differences in impact at the individual article level. Because of these factors, a "one size fits all" approach to interpreting scientometric correlation coefficients is inappropriate and a more fine-grained strategy is needed that is sensitive to both the field of study and the properties of the data set analysed.

To reduce the level of uncertainty when interpreting correlation coefficients in scientometric studies, the current article assesses the influence of three factors on the strength of correlation between citation counts and alternative research quality indicators: The average number of citations per paper for the data set investigated; the variability in the distribution of citation counts in the data set investigated; and the strength of the relationship between research quality and indicator values. Although this is not an exhaustive list of relevant factors, it includes the main factors that can be experimentally controlled.

- RQ1: Does the average number of citations per paper in a data set affect the likely strength of a Spearman correlation with an alternative indicator?
- RQ2: Does the variability of the number of citations per paper in a data set affect the likely strength of a Spearman correlation with an alternative indicator?

- RQ3: Does the magnitude of the connection between research quality and expected citation counts in a data set affect the likely strength of a Spearman correlation with an alternative indicator?

Methods

This article uses an experimental simulation modelling approach to assess the influence of average citation counts, variability and research quality relationship strength on correlations between citation counts and alternative indicators. These three factors are investigated by generating simulated citation count and alternative indicator data sets and then calculating the correlation between them for different parameter values.

In order to simulate a set of citation counts, their statistical distribution needs to be known. Early research suggested that citation counts follow a power law (Clauset et al. 2009; Garanina and Romanovsky 2015; Redner 1998) but, with the possible exception of physics, this distribution only fits reasonably well if low-cited articles are removed. This is also true for a discrete version of the power law, the Yule-Simon distribution (Brzezinski 2015). Two better fitting distributions are the shifted or hooked power law (Thelwall and Wilson 2014; Thelwall 2016; see also: Pennock et al. 2002) and the discretised lognormal (Eom and Fortunato 2011; Radicchi et al. 2008; Thelwall and Wilson 2014; Thelwall 2016). Stopped sum distributions may fit citation data even better for some subjects, but have substantial parameter estimation problems (Low et al. 2015). Negative binomial distributions probably do not fit as well overall (Ajiferuke and Famoye 2015; Low et al. 2015) due to problems with predicting very high values.

Although there are few studies of the distributions of alternative indicators, one has shown that Mendeley readership counts for medical fields follow both the hooked/shifted power law and the discretised lognormal reasonably well (Thelwall and Wilson in press).

The discretised lognormal distribution was chosen here for the simulations because its parameters can be manipulated to set the mean and variance relatively independently of each other and this is necessary to address two of the issues investigated. The probability density function of the (continuous) lognormal distribution is $f(x) = \frac{1}{x\sigma\sqrt{2\pi}} e^{-\frac{(\ln(x)-\mu)^2}{2\sigma^2}}$. Its scale parameter σ and location parameter μ are the mean and standard deviations of the natural log of the data (Limpert et al. 2001). The mean of the untransformed distribution is also related the standard deviation, with formula $e^{\mu+\sigma^2/2}$. The variance of the distribution is $\left(e^{\sigma^2} - 1\right)e^{2\mu+\sigma^2}$. The (continuous) lognormal distribution be discretised to generate the discretised lognormal distribution $\overset{...}{\ln}\mathcal{N}(\mu, \sigma^2)$ with probability mass function $f(n) = \frac{1}{\int_{0.5}^{\infty} f(x)dx} \int_{n-0.5}^{n+0.5} f(x)dx$, for $n = 1, 2, \ldots$. Although this excludes zeros, it is standard to add 1 to citation counts before modelling, so that uncited articles are not excluded. The formulae for the mean and standard deviation for the continuous lognormal distribution are presumably reasonable approximations, but not exact, for the discretised version. The discretised lognormal distribution in the powerRlaw R package (Gillespie 2015) was used here.

The simulation modelling controlled three parameters: the location, scale and connection with research quality. Location parameters were varied between 0.1 and 4 in steps of 0.1 to ensure that the distribution means included the full range of average citation counts

normally found in scientometric studies (up to $e^{\mu+\sigma^2/2} = 70.1$ for a location of 4 and scale parameter 0.5—see below). Scale parameters for discretised lognormal distributions fitted to data from individual fields and years have varied from 0.67 to 1.53 (Thelwall and Wilson 2014) and so three values were chosen to encapsulate this range: 0.5, 1 and 2.

The relationship between the underlying research quality of a publication and its citation count or alternative indicator value was modelled by allowing the location parameter to vary with the quality value. Each data set was split into articles at four quality levels, with the location parameter of each determined by its quality level using a quality multiplier q. Thus if the base location parameter was μ then the location parameters of the four sets would be μ, μq, μq^2, μq^3 so that each level had its location parameter increased proportionately to the one before. The quality multiplier was allowed to vary between 1 (no effect) and 2 in steps of 0.1. This choice is relatively arbitrary since quality is not a numerical concept and therefore has no natural scale. It is broadly based upon the UK context, which suggests an exponential relationship between ratings and research quality. This is evident from the funding formula used in which a rating of 4 out of 4 is worth four times as much funding as a rating of 3 out of 4 (Else 2015; see also: Wilsdon et al. 2015).

Since $q \in \{1, 1.1, \ldots 2\}$, $\mu \in \{0.1, \ldots 4\}$, and $\sigma \in \{0.5, 1, 2\}$ were varied independently, there were $11 \times 40 \times 3 = 1320$ different parameter sets altogether. Each of the 1320 parameter sets was used to generate 1000 simulations of two datasets of size 400 each, as described above. Spearman correlations were then calculated for the datasets and the mean correlation was recorded out of the 1000 simulations as well as 95 % confidence intervals from the data (i.e., the 50th largest and 50th smallest correlations out of the 1000 calculated for each parameter set). For simplicity of reporting, results are presented only for cases where the two simulated distributions have identical parameters but results are available online (the file location is in the conclusions) for which they have different parameters.
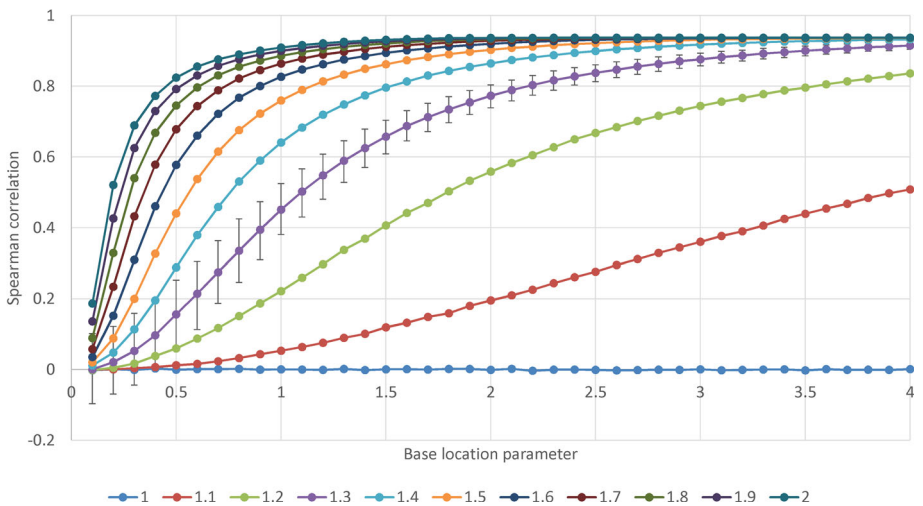
Fig. 1 Spearman correlations between two simulated discretised lognormal distributions with 100 data points at each of four quality levels, with a separate line for each quality differential q between these levels. All distributions have a scale parameter of 0.5. *Error bars* on one of the lines show 95 % confidence intervals

Results

(RQ1) If the base location parameter is small (i.e., the average number of citations per paper for the data set is low) then, irrespective of the other parameters, the Spearman correlation between the two data sets has a low maximum (points near the left axis in Figs. 1, 2, 3). For example, for parameter set $q = 2$, $\mu = 0.1$, and $\sigma = 0.5$, the expected mean is approximately $e^{\mu+\sigma^2/2} = 1.3$ and the average correlation is 0.2 (Fig. 1), despite the strong underlying relationship between quality and indicator values. Conversely, if the base location parameter is large so that the average number of citations per paper for the data set is high, then the average correlation is 0.9 or higher (points near the right of Figs. 1, 2 and 3), unless the quality relationship is quite small. Thus the average citation count has a substantial effect on the size of a correlation that could be expected.

(RQ2) If the scale parameter is small (Fig. 1), then the Spearman correlation between the two data sets tends to be higher than if the scale parameter is large (Fig. 3), although the difference is most evident with small or moderate location parameters.

(RQ3) If the relationship between quality and citation counts is small (the 1.1 line in each of Figs. 1, 2 and 3) then, irrespective of the other parameters, the Spearman correlation between the two data sets has a low maximum. The maximum can be increased by decreasing the scale parameter or increasing the location parameter.

Discussion and limitations

The main limitation in this study is that the connection between research quality and citation impact is unknown and has been modelled in a simple way with a single parameter q. There is no evidence about how research quality scores for sets of research articles are typically distributed, with the exception of self-selected article sets for the UK and Italy (HEFCE 2015; Franceschet and Costantini 2011). Without this information, the magnitude
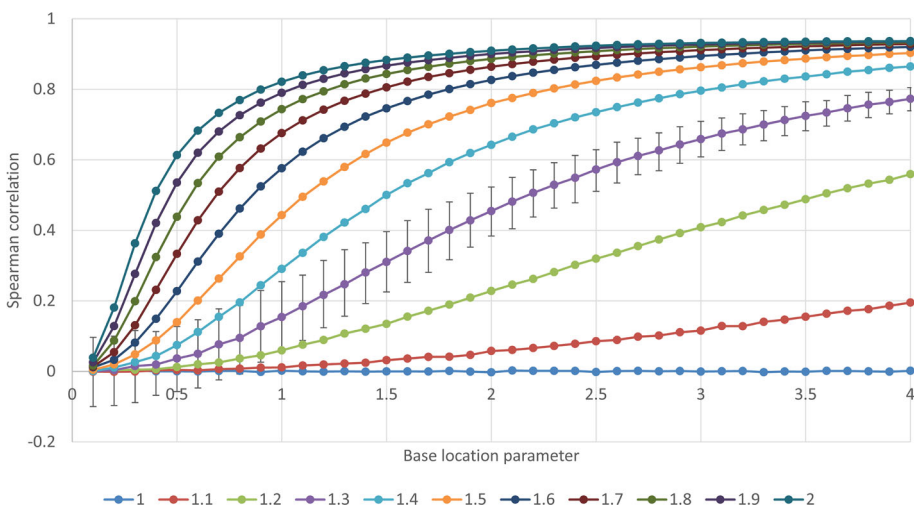
Fig. 2 This is the same as Fig. 1 except that all distributions have a scale parameter of 1
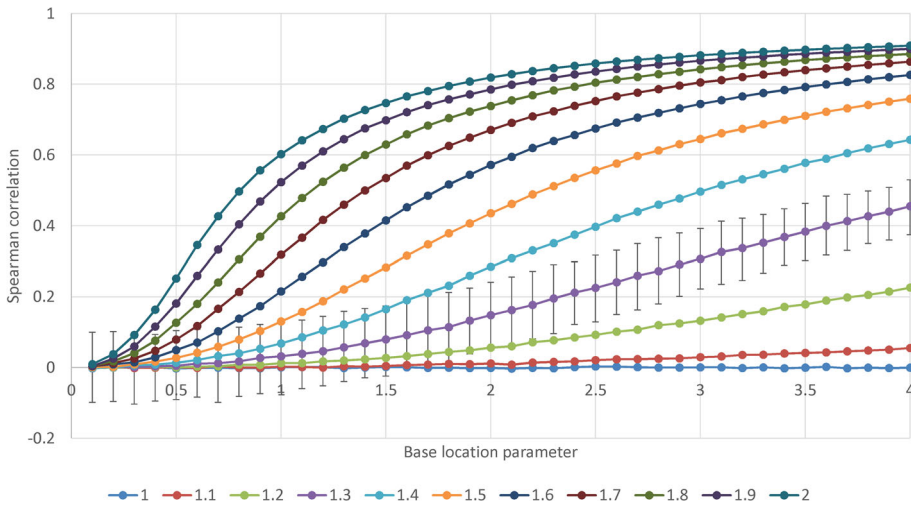
Fig. 3 This is the same as Fig. 1 except that all distributions have a scale parameter of 2

of the correlations can only be approximations, although the overall graph shapes are still useful to point to underlying trends.

A second major problem is that the method assumes that the two indicators correlated are independent of each other and this is rarely likely to be true. In an extreme case, an unknown fraction of Mendeley readerships for articles are recorded by people that use Mendeley as a device to manage references for their future journal articles and this is a source of partial dependence between citation counts and Mendeley readership counts. More generally, most indicators reflect some type of citation, in the most abstract sense, and this introduces a degree of dependence since the citation counts and indicator value will both be affected by citation-specific influences that are independent of quality, such as article type and subfield membership. This issue could be circumvented by replacing "quality" in the above discussion and methods by a term such as "citability". This would give more credible results at the expense of moving further away from the information that evaluators would like to know. This approach could also be used in more theoretical studies of observable properties of academic publications (e.g., Bosquet and Combes 2013).

Another limitation is that only the discretised lognormal distribution has been used for the simulation modelling, although the shifted power law fits some sets of citation counts better. Since the two distributions have reasonably similar overall shapes and the Spearman correlation is not a parametric test, this seems unlikely to affect the conclusions.

Although the answers to the second and third research questions are expected, the cause of the solution to the first research question is less transparent. For the continuous lognormal distributions, the location parameter would presumably effect the correlations exclusively by varying the standard deviation of the simulated data, since it would not affect the mean. With the use of discretisation, reducing the location parameter tends to increase the number of tied ranks in the data set because the ties tend to occur for lower citation counts. In particular, reducing the mean parameter increases the number of zeros. The ties in the data thus tend to reduce the Spearman correlation by making the data sets less different.

The results also support the argument made in the introduction that it is not reasonable to specify specific correlation coefficient ranges as being universally small, medium or large. Thus there can be no scientometric equivalent of the Cohen (1988) table of recommendations for interpreting behavioural sciences correlation coefficients. Moreover, extreme caution must be used when comparing correlations between different studies in the literature. Even if the studies cover publications from the same field, the above results show that the citation window can affect the correlation coefficients.

In theory, the simulation results could be used to generate a corrected correlation coefficient by dividing the correlation for the real data by the simulated correlation coefficient. This is a standard technique in psychology to deal with measurement error in instruments using Cronbach's alpha (Cronbach 1951; Ellis 2010). The corrected correlation coefficient would then report the correlation as a proportion of the maximum possible value. This corrected figure would be credible if there was good evidence about the key parameters for the simulation, such as the quality distribution and the connection between quality and indicator values, but this seems unlikely to happen often in practice. Without this credibility, a set of corrected values should be calculated based upon a range of reasonable assumptions and then this range of corrected values should be reported. See the conclusions for the location of software to run simulations and an extended table of results that can be used for standard benchmark comparisons.

Conclusions

The simulation modelling results show that the location and scale parameters influence the strength of correlation to be expected between citation count data and alternative indicators, assuming that both are connected to underlying research quality. The results also confirm that stronger connections to research lead to higher correlations, other factors being equal. The main implication for interpreting correlation coefficients is that their magnitude is not a simple reflection of the underlying relationship with research quality but is also related to the average citation counts and the variability of the data. For example, if the same correlation test is carried out separately for each of a number of disciplines (e.g., as in Mohammadi and Thelwall 2014) then higher correlations should be expected for disciplines that attract more citations, irrespective of the underlying connection between quality and research. Thus, in this context, a higher correlation does not necessarily imply a stronger connection between the indicators and research quality, unless the variability and means of the data sets are similar. Thus, in future when reporting correlation coefficients, average numbers of citations and variance should also be reported and disciplines should only be compared if they have similar values on these two parameters.

In theory, it would be possible to assess the strength of a correlation by comparing it to the correlation expected if a perfect relationship was present between quality and citation counts, and assuming that citation counts and the other indicator were independent of each other. This could be achieved by fitting a discretised lognormal distribution to the citation count data and the alternative indicator and then using the two fitted location and scale parameters to generate two simulated data sets, from which a Spearman correlation could be calculated. This would be possible if the exact relationship between research quality and citation counts was to be known (e.g., the value of q in the simulations above). Since the relationship between research quality and citation counts is unknown then an alternative strategy is to conduct multiple simulations with apparently reasonable parameters and then

interpret the real value in the context of this range of values. If the two data sets analysed have similar properties then the theoretical maximum correlation may be read from Figs. 1, 2 and 3. If they have the same scale parameter but a different location parameter then the maximum expected value may be read from the online data associated with the current paper (doi:10.6084/m9.figshare.3184687.v1). If the scale parameters are also different, or if more precise values are needed then new simulation models can be run using R code placed online to help future studies (doi:10.6084/m9.figshare.3184687.v1). This information can also be used to calculate a range of corrected values, as described at the end of the Discussion section, if coefficients need to be compared between different data sets.

These conclusions apply only to the case of correlating raw citation counts and raw scores from another integer metric. The method above would need to be substantially adapted for correlations involving indicators derived from such data (e.g., Chakraborty et al. 2015; Finardi 2013) as well as for direct correlations between indicators or citation data and human quality ratings (Ahlgren and Waltman 2014; Wainer and Vieira 2013). In practice, however, the exact distribution of quality scores is rarely likely to be known, even if the concept of quality is operationalised in a useful way, such as through a numerical score on a scale rating given by expert judges. Moreover, two indicators are rarely likely to be fully independent and in many cases may be highly dependent. Hence, it seems impractical to expect to be able to gain a realistic impression of the underlying strength of a correlation (but see the discussion above for a "citability" alternative). Nevertheless, it may help to use the methods here to get very approximate guidelines as long as they are reported alongside a statement of limitations.

References

Ahlgren, P., & Waltman, L. (2014). The correlation between citation-based and expert-based assessments of publication channels: SNIP and SJR vs. Norwegian quality assessments. *Journal of Informetrics, 8*(4), 985–996.

Ajiferuke, I., & Famoye, F. (2015). Modelling count response variables in informetric studies: Comparison among count, linear, and lognormal regression models. *Journal of Informetrics, 9*(3), 499–513.

Bosquet, C., & Combes, P. P. (2013). Are academics who publish more also more cited? Individual determinants of publication and citation records. *Scientometrics, 97*(3), 831–857.

Brzezinski, M. (2015). Power laws in citation distributions: Evidence from Scopus. *Scientometrics, 103*(1), 213–228.

Chakraborty, T., Tammana, V., Ganguly, N., & Mukherjee, A. (2015). Understanding and modeling diverse scientific careers of researchers. *Journal of Informetrics, 9*(1), 69–78.

Clauset, A., Shalizi, C. R., & Newman, M. E. (2009). Power-law distributions in empirical data. *SIAM Review, 51*(4), 661–703.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Abingdon: Lawrence Erlbaum Associates.

Cohen, J. (1992). A power primer. *Psychological Bulletin, 112*(1), 155–159. doi:10.1037/0033-2909.112.1.155.

Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika, 16*(3), 297–334.

Didegah, F., & Thelwall, M. (2013). Which factors help authors produce the highest impact research? Collaboration, journal and document properties. *Journal of Informetrics, 7*(4), 861–873.

Ellis, P. D. (2010). *The essential guide to effect sizes: Statistical power, meta-analysis, and the interpretation of research results*. Cambridge: Cambridge University Press.

Else, H. (2015). Research funding formula tweaked after REF 2014 results. https://www.timeshighereducation.com/news/research-funding-formula-tweaked-after-ref-2014-results/2018685.article.

Eom, Y. H., & Fortunato, S. (2011). Characterizing and modeling citation dynamics. *PLoS ONE, 6*(9), e24926.

Ettori, S. (2015). The physics inside the scaling relations for X-ray galaxy clusters: Gas clumpiness, gas mass fraction and slope of the pressure profile. *Monthly Notices of the Royal Astronomical Society, 446*(3), 2629–2639.

Finardi, U. (2013). Correlation between journal impact factor and citation performance: An experimental study. *Journal of Informetrics, 7*(2), 357–370.

Franceschet, M., & Costantini, A. (2011). The first Italian research assessment exercise: A bibliometric perspective. *Journal of Informetrics, 5*(2), 275–291.

Garanina, O. S., & Romanovsky, M. Y. (2015). Citation distribution of individual scientist: Approximations of stretch exponential distribution with power law tails. In A. A. Salah, Y. Tonta, A. A. Akdag Salah, C. Sugimoto, & U. Al (Eds.), *Proceedings of ISSI 2015* (pp. 272–277). Turkey: Bogaziçi University Printhouse.

Gillespie, C.S. (2015). Fitting heavy tailed distributions: the poweRlaw package. *Journal of Statistical Software*, 64(2), 1–16. http://www.jstatsoft.org/v64/i02/paper.

Hartley, J., & Sydes, M. (1997). Are structured abstracts easier to read than traditional ones? *Journal of Research in Reading, 20*(2), 122–136.

HEFCE. (2015). The metric tide: Correlation analysis of REF2014 scores and metrics. Supplementary Report II to the Independent review of the role of metrics in research assessment and management. Bristol: Hefce. http://www.hefce.ac.uk/pubs/rereports/Year/2015/metrictide/Title,104463,en.html.

Hemphill, J. F. (2003). Interpreting the magnitudes of correlation coefficients. *American Psychologist, 58*(1), 78–79.

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.

Hyland, K. (1999). Academic attribution: Citation and the construction of disciplinary knowledge. *Applied Linguistics, 20*(3), 341–367.

Kostoff, R. (2007). The difference between highly and poorly cited medical articles in the journal Lancet. *Scientometrics, 72*, 513–520.

Kousha, K., & Thelwall, M. (2015). Web indicators for research evaluation, part 3: Books and non-standard outputs. *El Profesional de la Información, 24*(6), 724–736. doi:10.3145/epi.2015.nov.04.

Larivière, V., & Gingras, Y. (2010). On the relationship between interdisciplinarity and scientific impact. *Journal of the American Society for Information Science and Technology, 61*, 126–131.

Limpert, E., Stahel, W. A., & Abbt, M. (2001). Lognormal distribution across sciences: Key and clues. *BioScience, 51*(5), 341–351.

Lipsey, M.W., Puzio, K., Yun, C., Hebert, M.A., Steinka-Fry, K., Cole, M.W., et al. (2012). Translating the statistical representation of the effects of education interventions into more readily interpretable forms. Washington, DC: US Dept of Education, National Center for Special Education Research, Institute of Education Sciences, NCSER 2013-3000.

Liu, G., Qi, X. L., Robert, N., Dick, A. J., & Wright, G. A. (2012). Ultrasound-guided identification of cardiac imaging windows. *Medical Physics, 39*(6), 3009–3018.

Low, W. J., Thelwall, M., & Wilson, P. (2015). Stopped sum models for citation data. In A. A. Salah, Y. Tonta, A. A. AkdagSalah, C. Sugimoto, & U. Al (Eds.), *Proceedings of ISSI 2015 Istanbul: 15th international society of scientometrics and informetrics conference* (pp. 184–194). Istanbul: Bogaziçi University Printhouse.

Mohammadi, E., & Thelwall, M. (2014). Mendeley readership altmetrics for the social sciences and humanities: Research evaluation and knowledge flows. *Journal of the American Society for Information Science and Technology, 65*(8), 1627–1638.

Onodera, N., & Yoshikane, F. (2015). Factors affecting citation rates of research articles. *Journal of the Association for Information Science and Technology, 66*(4), 739–764.

Oppenheim, C. (2000). Do patent citations count? In B. Cronin & H. B. Atkins (Eds.), *The web of knowledge: A festschrift in honor of Eugene Garfield* (pp. 405–432). Metford: Information Today Inc. ASIS Monograph Series.

Pennock, D. M., Flake, G. W., Lawrence, S., Glover, E. J., & Giles, C. L. (2002). Winners don't take all: Characterizing the competition for links on the web. *Proceedings of the National Academy of Sciences, 99*(8), 5207–5211.

Persson, O., Glänzel, W., & Danell, R. (2004). Inflationary bibliometric values: The role of scientific collaboration and the need for relative indicators in evaluative studies. *Scientometrics, 60*(3), 421–432.

Radicchi, F., Fortunato, S., & Castellano, C. (2008). Universality of citation distributions: Toward an objective measure of scientific impact. *Proceedings of the National Academy of Sciences, 105*(45), 17268–17272.

Redner, S. (1998). How popular is your paper? An empirical study of the citation distribution. *The European Physical Journal B-Condensed Matter and Complex Systems, 4*(2), 131–134.

Sud, P., & Thelwall, M. (2014). Evaluating altmetrics. *Scientometrics, 98*(2), 1131–1143. doi:10.1007/s11192-013-1117-2.

Thelwall, M. (2006). Interpreting social science link analysis research: A theoretical framework. *Journal of the American Society for Information Science and Technology, 57*(1), 60–68.

Thelwall, M. (2016). The discretised lognormal and hooked power law distributions for complete citation data: Best options for modelling and regression. *Journal of Informetrics, 10*(2), 336–346. doi:10.1016/j.joi.2015.12.007.

Thelwall, M., & Fairclough, R. (2015). The influence of time and discipline on the magnitude of correlations between citation counts and quality scores. *Journal of Informetrics, 9*(3), 529–541. doi:10.1016/j.joi.2015.05.006.

Thelwall, M., & Kousha, K. (2015a). Web indicators for research evaluation, Part 1: Citations and links to academic articles from the web. *El Profesional de la Información, 24*(5), 587–606. doi:10.3145/epi.2015.sep.08.

Thelwall, M., & Kousha, K. (2015b). Web indicators for research evaluation, Part 2: Social media metrics. *El Profesional de la Información, 24*(5), 607–620. doi:10.3145/epi.2015.sep.09.

Thelwall, M., & Wilson, P. (2014). Distributions for cited articles from individual subjects and years. *Journal of Informetrics, 8*(4), 824–839.

Thelwall, M., & Wilson, P. (in press). Mendeley readership altmetrics for medical articles: An analysis of 45 fields. Journal of the Association for Information Science and Technology. doi:10.1002/asi.23501.

van Raan, A. (1998). The influence of international collaboration on the impact of research results: Some simple mathematical considerations concerning the role of self-citations. *Scientometrics, 42*(3), 423–428.

Wainer, J., & Vieira, P. (2013). Correlations between bibliometrics and peer evaluation for all disciplines: the evaluation of Brazilian scientists. *Scientometrics, 96*(2), 395–410.

Wilsdon, J., Allen, L., Belfiore, E., Campbell, P., Curry, S., Hill, S., et al. (2015). The metric tide: Report of the independent review of the role of metrics in research assessment and management. http://www.hefce.ac.uk/pubs/rereports/Year/2015/metrictide/Title,104463,en.html.