Journal of Informetrics 9 (2015) 529–541

Contents lists available at ScienceDirect

Journal of Informetrics

journal homepage: www.elsevier.com/locate/joi

The influence of time and discipline on the magnitude of correlations between citation counts and quality scores

Mike Thelwall*, Ruth Fairclough[1]

Statistical Cybermetrics Research Group, School of Mathematics and Computer Science, University of Wolverhampton, Wulfruna Street, Wolverhampton WV1 1LY, UK

ARTICLE INFO

Article history:
Received 2 February 2015
Received in revised form 11 May 2015
Accepted 11 May 2015

Keywords:
Citation analysis
Correlation
Disciplinary differences
Citation windows

ABSTRACT

Although various citation-based indicators are commonly used to help research evaluations, there are ongoing controversies about their value. In response, they are often correlated with quality ratings or with other quantitative indicators in order to partly assess their validity. When correlations are calculated for sets of publications from multiple disciplines or years, however, the magnitude of the correlation coefficient may be reduced, masking the strength of the underlying correlation. This article uses simulations to systematically investigate the extent to which mixing years or disciplines reduces correlations. The results show that mixing two sets of articles with different correlation strengths can reduce the correlation for the combined set to substantially below the average of the two. Moreover, even mixing two sets of articles with the same correlation strength but different mean citation counts can substantially reduce the correlation for the combined set. The extent of the reduction in correlation also depends upon whether the articles assessed have been pre-selected for being high quality and whether the relationship between the quality ratings and citation counts is linear or exponential. The results underline the importance of using homogeneous data sets but also help to interpret correlation coefficients when this is impossible.

1. Introduction

The ongoing controversy within the wider scientific community about the use of citation-based indicators in research evaluations underlines the importance of continuing to assess their validity. The most straightforward way is to correlate citation-based indicators with peer judgements about the value of individual publications. Studies taking this approach have used data from research assessment exercises (Franceschet & Costantini, 2011), surveys of experts (Gottfredson, 1978), the prestige of the publishing journal (Singh, Haddad, & Chow, 2007; Starbuck, 2005) or public peer review systems like F1000 (Bornmann & Leydesdorff, 2013; Li & Thelwall, 2012; Mohammadi & Thelwall, 2013; Waltman & Costas, 2014; Wardle, 2010). A statistically significant positive correlation with an independent measure of quality would suggest that the indicators reflect an aspect of quality to some extent. Moreover, it seems intuitively plausible that stronger correlations are likely to reflect stronger relationships between citations and quality. This is not necessarily true, however, because the strength of a correlation always partly reflects the extent to which the two numbers correlated are derived from homogenous sources. In

* Corresponding author. Tel.: +44 1902 321470; fax: +44 1902 321478.
E-mail addresses: m.thelwall@wlv.ac.uk (M. Thelwall), r.fairclough@wlv.ac.uk (R. Fairclough).
[1] Tel.: +44 1902 321000; fax: +44 1902 321478.

http://dx.doi.org/10.1016/j.joi.2015.05.006
1751-1577/

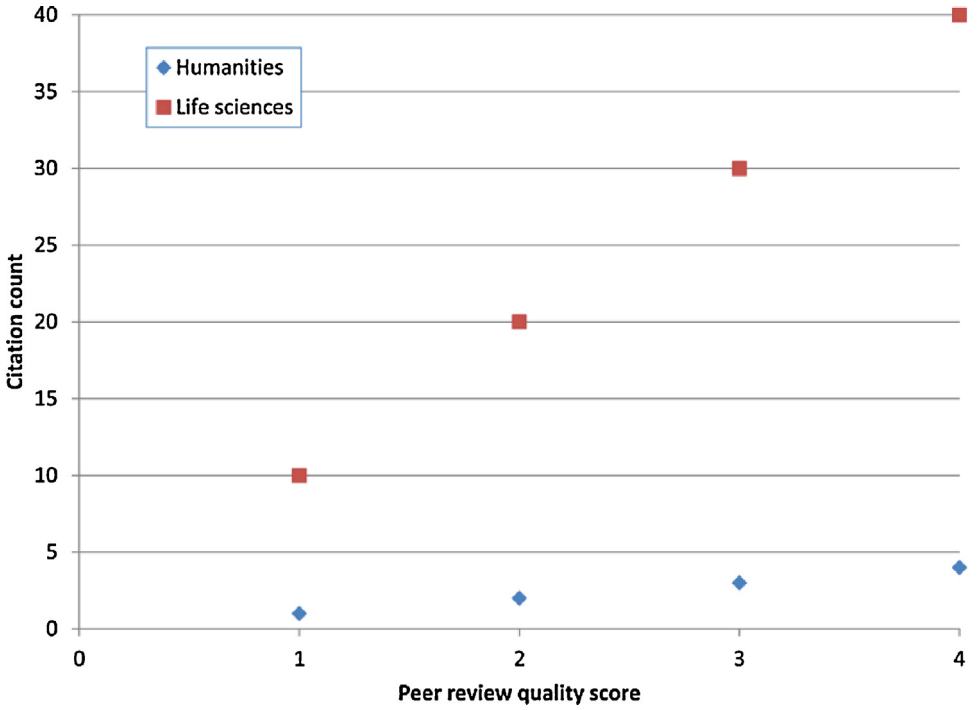

Fig. 1. Four humanities articles with a perfect linear relationship between quality and citations together with four life sciences articles with a perfect linear relationship between quality and citations. Whilst the mono-disciplinary Spearman correlations are both 1, the combined Spearman correlation is 0.488.

terms of scientometrics, this means that the homogeneity of the set of publications analysed, such as in terms of publication date and subject area, can affect the strength of any correlations derived from it. It is therefore important to understand the impact of a lack of homogeneity on the magnitude of correlation coefficients.

When conducting a scientometric study, the publications analysed can sometimes be classified only crudely by subject area. The consequent mixing may weaken the strength of any relationships between the citation counts and other variables. To illustrate this, suppose that four humanities publications are rated 1, 2, 3, and 4 for quality by a subject expert and receive 1, 2, 3, and 4 citations, respectively. Then there is a perfect relationship between quality and citation counts with a Spearman rank correlation coefficient of 1. Suppose that four life sciences publications are also rated 1, 2, 3, and 4 but, because life sciences articles tend to be more cited, they receive 10, 20, 30, and 40 citations, respectively. For the four life sciences articles there is also a perfect relationship and the correlation coefficient is 1. If the eight articles are collected together then the relationship is no longer perfect and the correlation coefficient is less than half: 0.488 (Fig. 1). The mono-disciplinary correlation of 1 for both disciplines is therefore not evident from the mixed set correlation. Thus, anyone calculating a correlation coefficient for a mixed data set could be misled about the underlying monodisciplinary rank correlation strengths. Nevertheless, correlations for combined sets of publications are not necessarily always low. For example, a study of articles rated by peers as part of an Italian research assessment exercise calculated correlations between these ratings and citations to papers published across three years and within quite broad disciplinary areas but still found correlations as strong as 0.8 (for both Physics and Earth Sciences) (Franceschet & Costantini, 2011).

If publications with different citation windows (periods of time for which citations are collected) are merged then this can also reduce the strength of correlation for the same reason as for mixing disciplines. This is because articles with a longer citation window will tend to have higher citation counts than articles with shorter citation windows. As the Fig. 1 example shows, combining sets with different average citation counts can substantially reduce the correlation for the mixed set. A previous altmetrics paper has even given an example of a negative correlation between variables (Tweets to papers) despite an underlying positive relationship. The cause of this was that one of the two variables decreased over time (due to a shrinking citation window) and the other increased over time (Thelwall, Haustein, Larivière, & Sugimoto, 2013). The problem of variable citation windows may be unavoidable in some circumstances, however. For example, a study of a narrow group of outputs may need to cover multiple years to get enough data for quantitative analyses. Moreover, combining data from different fields with the same citation window can also be problematic if one attracts citations more quickly than does the other. Although it is clear from a large number of previous scientometric studies that the rate at which articles attract citations varies substantially by field, the cause is not fully understood. The reason could be different citation practices to some extent, but partial coverage of disciplinary outputs by the Web of Science is probably more important (Marx & Bornmann, 2015).