CrossMark

# The influence of highly cited papers on field normalised indicators

Mike Thelwall[1]

## Abstract

Field normalised average citation indicators are widely used to compare countries, universities and research groups. The most common variant, the Mean Normalised Citation Score (MNCS), is known to be sensitive to individual highly cited articles but the extent to which this is true for a log-based alternative, the Mean Normalised Log Citation Score (MNLCS), is unknown. This article investigates country-level highly cited outliers for MNLCS and MNCS for all Scopus articles from 2013 and 2012. The results show that MNLCS is influenced by outliers, as measured by kurtosis, but at a much lower level than MNCS. The largest outliers were affected by the journal classifications, with the Science-Metrix scheme producing much weaker outliers than the internal Scopus scheme. The high Scopus outliers were mainly due to uncitable articles reducing the average in some humanities categories. Although outliers have a numerically small influence on the outcome for individual countries, changing indicator or classification scheme influences the results enough to affect policy conclusions drawn from them. Future field normalised calculations should therefore explicitly address the influence of outliers in their methods and reporting.

**Keywords** Highly cited papers · Citation outliers · Field normalised indicators · MNCS · MNLCS

## Introduction

Research evaluations that use citation count data are complicated by average citation counts differing between fields and years. This can be resolved by only comparing publications from the same field and year (and document type) or by using field-normalised indicators. The standard normalising approach is exemplified by the Mean Normalised Citation Score (MNCS), which divides the citation count for each article by the world average for the article's field and year (Waltman et al. 2011a, b). These normalised valued are then averaged for each group (e.g., country or research unit), with values above 1 indicating more citations than the world average. This calculation is, in theory, sensitive to outliers because citation data is highly skewed (Clauset et al. 2009; Price 1976; Thelwall 2016a).

✉ Mike Thelwall
 m.thelwall@wlv.ac.uk

1  University of Wolverhampton, Wulfruna Street, Wolverhampton WV1 1LY, UK

The Mean Normalised Log Citation Score (MNLCS) has been designed to deal with this issue with an extra step: applying a log transformation count $\ln(1 + x)$ to each citation count before any other calculations (Thelwall 2017a, b). For example, one study found that confidence intervals are narrower for MNLCS than for MNCS for funded medical research (Thelwall 2017b). Another found that the choice of MNCS or MNLCS affected the outcome of a test of which gender's research tended to be most highly cited (Thelwall 2018). Since the MNCS type of indicator is widely used (including Elsevier's Source Normalised Impact per Paper: SNIP; Clarivate Analytics' Category Normalized Citation Impact: CNCI), it is important to assess the influence of outliers on this calculation and on the MNLCS, which is designed to reduce outliers. A unit may have an above average MNCS value solely because it publishes a few highly cited articles, whereas the remainder of its research is average. Irrespective of whether this is regarded as a drawback or advantage of MNCS calculations, policy makers using the calculation would benefit from knowing if this is likely to occur in practice.

Highly cited articles are presumably much more influential than typical articles. They are more likely to be the result of international collaboration, especially for small countries (Aksnes 2003; Persson 2009), and so might represent an unusual type of research to some extent. It has been argued that highly cited papers should be analysed separately from a unit's typical output because the two can give different results (Tijssen et al. 2002). It is not clear whether the capacity to produce highly cited papers is separate from the capacity to produce good typical research or whether it could be conceived as an occasional almost accidental research by-product. The ability of a country to produce highly cited papers differs between nations and seems to be fairly stable over time (Bornmann et al. 2015), suggesting that it could be a side-effect of producing higher average quality or impact research.

Since citations follow an imitation pattern to some extent (Merton 1968), it seems likely that the citation counts for highly cited papers overestimate their value because a greater proportion of the citations will be imitative. A study of physics and physical chemistry suggested that 40% of citations to the 23 highly cited papers examined did not reflect active use (Oppenheim and Renn 1978). Conversely, some influential research may become standard enough to not need citations (McCain 2011), so extremely high citation counts may underestimate the influence of seminal works. One of the few studies to empirically assess the usefulness of highly cited studies found that 83% of 35 biomarker studies overestimated the effect that they claimed to measure (Ioannidis and Panagiotou 2011), and so highly cited studies are not necessarily excellent.

A previous analysis of the influence of highly cited articles on the citation impact of Norway in 24 fields 1981–1994 with a 5-year citation window found that the national citation average was predominantly due to a small number of highly cited articles in some fields. Substantial variations over time in individual field averages occurred because of these few highly cited articles and so field-specific average citation counts were not useful indicators of underlying research capacity or excellence (Aksnes and Sivertsen 2004). It is not clear whether these issues would be ameliorated for cross-disciplinary field normalised citation impact indicators, however. The great influence of individual highly cited publications on MNCS for individual universities has also been acknowledged (Waltman et al. 2012).

Highly cited articles tend to be indexed by scholarly databases (Martín-Martín et al. 2018) and so the main cause of any variations in outlier influence between databases is likely to be the classification scheme. Field normalised indicators are known to be affected by the field classification scheme used, with one study suggesting that thousands of categories may be needed for optimal results (Ruiz-Castillo and Waltman 2015). The standard

journal classifications from Scopus and the Web of Science contain errors (Wang and Waltman 2016), which may create outliers. It is therefore important to assess the influence of classification schemes on the influence of outliers in field normalised calculations.

Despite the findings discussed above, no previous study seems to have directly analysed the influence of outliers on field normalised calculations. The goal of this paper is to assess the influence of highly cited papers on field normalised calculations at the national level. As argued above, whilst it is theoretically possible that a small number of highly cited papers can exert a great influence on overall field normalised scores, it is not clear whether this is likely to be a problem in practice. The focus here is on the MNCS and MNLCS indicators because the former is standard in scientometrics and the latter is a MNCS variant that may reduce the influence of highly cited papers.

Motivated by the above discussion, the following research questions drive this study.

- *RQ1* How much influence do outliers exert on the field normalised citation indicators MNCS and MNLCS for individual countries?
- *RQ2* Does the choice of field classification scheme affect the influence of outliers on MNCS and MNLCS for individual countries?

## Methods

The research design was to obtain a large coherent collection of academic journal articles and their citation counts, assess them for the presence of outliers for both MNCS and MNLCS, and then check whether the country-level results differ between the two indicators. As an additional check, the country level results are compared between two time intervals to assess whether the indicator that is less influenced by outliers is more stable. For the second research question, the presence of outliers and indicator values are compared between two different subject classification schemes.

### Data

Scopus journal articles from the year 2013 were chosen for the main data set. The year 2013 ensures that each article has about 5 years in which to attract citations. Citation counts with a 5-year citation window have a correlation of 0.9 with long term (31 years) citation counts (Wang 2013), which is adequate for the current study. Scopus was chosen in preference to the Web of Science (WoS) for its more international coverage (Falagas et al. 2008). Since the current article focuses on international differences, greater international coverage is a desirable property. The additional journals in Scopus would presumably tend to publish articles that attract few citations, which would inflate the significance of outliers. Thus, a similar study for WoS may produce less substantial outliers. Reviews and documents that are not journal articles were excluded since these can have different average citation counts. Article records were downloaded from Scopus in October 2018 using queries like the following. Queries were submitted for each of the 304 Scopus narrow fields, excluding the overlapping general categories for each broad field. The example below is for *Classics*, with field code 1205, and the publication year was sent as a separate API parameter (date).

SUBJMAIN(1205) AND DOCTYPE(ar) AND SRCTYPE(j)

The Scopus API occasionally reported errors and so the searches were repeated for fields giving fewer results than reported as the maximum by the API. The final dataset should therefore be a complete list of all 3,387,576 documents of type journal article in Scopus with a publication year of 2013 (Table 1). Document types can sometimes be classified incorrectly in citation databases (Donner 2017) and so there may be some errors in the data.

Articles were allocated to countries based on the declared national affiliation of the first author, as recorded in Scopus. The first author contributes the largest average share of the work in all broad fields (Larivière et al. 2016). Alphabetical authoring in maths and economics undermines this to a limited extent (Levitt and Thelwall 2013). The last author is often senior in biomedical research but the first author still tends to do most of the work (Larivière et al. 2016).

The Scopus narrow field journal classification scheme of 304 fields (excluding general fields that are largely contained within subcategory fields) was used as the primary field classification (see the source title spreadsheet: www.elsevier.com/solutions/scopu s/how-scopus-works/content, the *ASJC classification codes* worksheet). This is a manual classification of journals into one or more subject categories, presumably with a primary information retrieval goal. Although document-level classifications are more desirable due to interdisciplinarity (Glänzel and Schubert 2003), journal-based classification schemes are currently more widely used. The Science-Metrix journal classification list of 176 narrow journal categories (www.science-metrix.com/en/classification, July 2018) was used as a second, independent scheme. This is partly based on classification schemes from WoS, Chi Research and the Australian Research Council Evaluation of Research Excellence (ERA) process (Archambault et al. 2011). The originally collected Scopus articles were fitted into this scheme, with articles in journals not in the Science-Metrix classification excluded from the Science-Metrix calculations (but retained for the Scopus classifications). The Science-Metrix classification is apparently designed for research impact calculations rather than information retrieval.

For comparison purposes, a second data set of articles published in 2012 was collected in October 2018 and processed with the same methods (Table 1). This has a different citation window (6 years rather than 5 years) but this should not affect the results much (Wang 2013). The purpose of this paper is not to focus on any time window but to examine long term citation behaviour for data that is recent enough to be relevant.

## Analyses

MNCS and MNLCS values were calculated for the Scopus classification (including duplicate articles that occur in multiple categories) and separately for the Science-Metrix classification (without duplicate articles). The individual normalised citation counts

**Table 1** Data sets and sample sizes used in this study

| Publication year | Scopus[a] | | | Science-Metrix[a] | | |
|---|---|---|---|---|---|---|
| | Articles | Unique articles | Fields | Matching articles | Unique articles | Fields |
| 2013 | 3,387,576 | 1,515,594 | 304 | 2,113,647 | 932,829 | 176 |
| 2012 | 3,131,494 | 1,403,495 | 304 | 1,999,417 | 882,047 | 176 |

[a]The Scopus dataset includes articles in multiple categories (Articles column) but the Science-Metrix dataset (Unique articles) does not

were retained for analysis. For article $i$ with citation count $c_i$ in field $f_i$, its normalised citation counts are as follows, where the divisor is the arithmetic mean of the values for all articles from the same field:

$$\text{MNCS}_i = c_i / \bar{c}_j|_{f_j = f_i}$$

$$\text{MNLCS}_i = \ln(1 + c_i) / \overline{\ln(1 + c_j)}|_{f_j = f_i}$$

MNCS and MNLCS values for countries were obtained by taking the arithmetic mean of the $\text{MNCS}_i$ and $\text{MNLCS}_i$ values for articles $i$ with a first author from the given country.

The extent to which outliers were present in each dataset was estimated using the kurtosis calculation. Although commonly thought of as a distribution shape (peakedness) measure, it is more accurately an assessment of the extent to which a distribution or sample has outliers (Westfall 2014). This is because the shape of the centre of the distribution has little influence on kurtosis compared to outliers, when present. Kurtosis values were calculated for each country separately (i.e., using the country average rather than the world average) so that the values would not be affected by whether the country tended to produce research that was differently cited than the world average. Kurtosis values are reported here rather than excess kurtosis (i.e., 3 was not subtracted). Although outliers can be either low or high cited articles in theory, in practice, even for the MNLCS calculation, the largest outliers are all above average and are therefore highly cited articles rather than uncited articles.

Correlations were used to assess whether the presence of outliers could systematically influence MNCS or MNLCS. A high positive correlation between kurtosis and MNCS or MNLCS would suggest that high country normalised indicators might be mainly due to the presence of outliers. Pearson correlations were not used because some of the samples failed a Kolmogorov–Smirnov univariate normality test. Spearman correlations were used instead, giving rank order comparisons. Lower correlations would suggest that outliers influence MNCS scores, since MNLCS is less prone to outliers, by design. All Spearman correlations reported are statistically significant from zero ($p = 0.000$, for a two-tailed hypothesis test that the underlying population, in the social sciences sense of repeated observations under similar conditions, correlation differs from 0), although this is not relevant to the current article.

The main analysis was conducted for the 20 countries with the most articles in Scopus in 2013 to assess the research questions for major research publishing nations. The analysis was repeated for the 50 and 75 countries with the most articles in Scopus in 2013 to encompass nations that publish moderate numbers of papers.

# Results

## Scopus classification scheme

The MNCS kurtosis values are extremely high for the top 20 countries and the Scopus classification scheme, as expected. They vary between 119 (Iran) and 518,540 (USA), with an average of 29,518. For reference, normally distributed data has a kurtosis of 3. Also as expected, the MNLCS kurtosis values are much lower, varying between 2.7 (India) and 8.4 (Sweden) with an average of 5.0. Some of the MNLCS country values

are therefore moderately outside the range expected for a sample taken from the normal distribution. Given that citation counts for a single field and year tend to follow the discretised lognormal distribution (Thelwall 2016a), values close to 3 might be expected for MNLCS because of its logarithmic formula.

Kurtosis values were plotted against MNLCS to check whether high MNLCS could be primarily due to the presence of outliers (Fig. 1). There is a tendency for higher MNLCS to be associated with higher MNLCS kurtosis (Spearman correlation: 0.817), although Poland is a prominent exception. Thus, highly cited articles may strongly influence the MNLCS, despite the outlier-reducing MNLCS formula.

Expanding the country set to the largest 75 countries in terms of Scopus journal articles in 2013, most of the extra countries have MNLCS kurtoses between 3 and 5 (Fig. 2). This is not true for all the extra countries, however, with Denmark, Philippines, Finland, Belgium and Austria all having higher MNLCS kurtoses than the top 20. The MNLCS against MNLCS kurtosis correlation is weaker for these 75 countries (Spearman correlation: 0.521).

Median MNLCS kurtosis values calculated separately for each country and field are much lower than cross-field MNLCS kurtosis values for each country, although the two correlate (Fig. 3; Spearman correlation: 0.765). The high overall MNLCS kurtosis values are therefore caused by merging fields with differing individual field MNLCS numbers.

The MNLCS individual article outliers for 2013 (i.e., the highest article $MNLCS_i$ values) were examined to determine their causes. For the Scopus classification scheme, 97% of the 100 largest MNLCS outliers were from Literature and Literary Theory or Visual Arts and Performing Arts. These two narrow fields were therefore the main cause of the high outlier values. For example, one US article in the first field had 89 citations but a log normalised score of 13.0. This extremely high value (the highest article $MNLCS_i$ in the dataset) is due to the low average for the category because it contains many prestigious but rarely cited large literary magazines that are not in English. The world geometric mean citation count for Literature and Literary Theory is only 0.41 as a result. The USA and
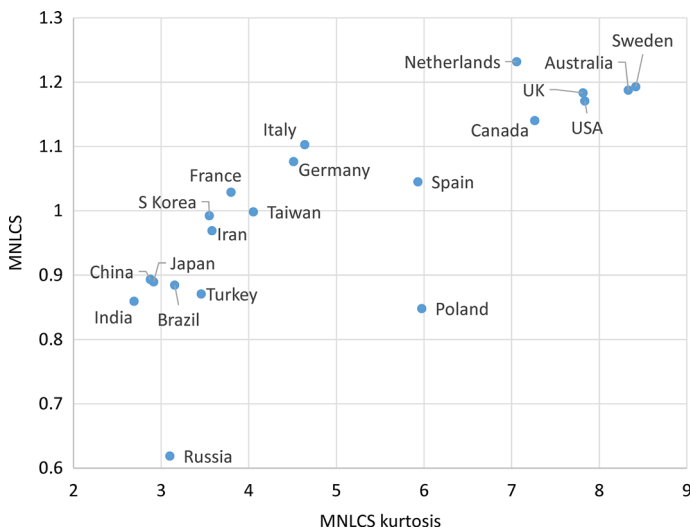


**Fig. 1** Kurtosis values calculated separately for the 20 countries with the most Scopus 2013 journal articles using the field normalised log citation counts and the Scopus classification scheme. MNLCS values are plotted on the y axis for comparison

**Fig. 2** Kurtosis values calculated separately for the 75 countries with the most Scopus 2013 journal articles using the field normalised log citation counts and the Scopus classification scheme. MNLCS values are plotted on the y axis for comparison. *Not all countries are labelled
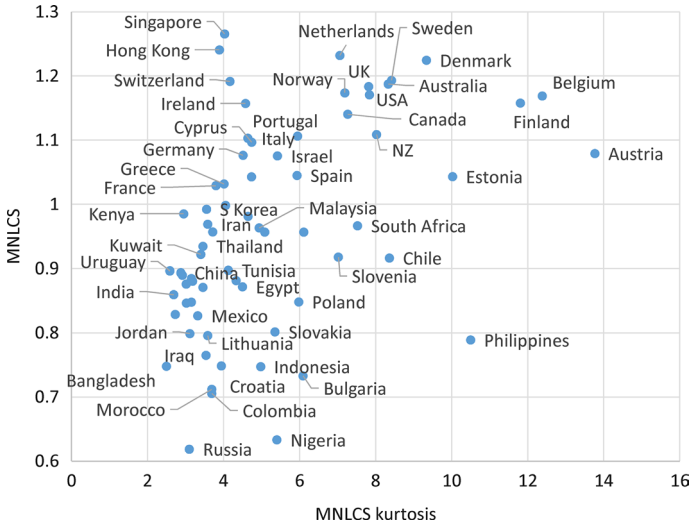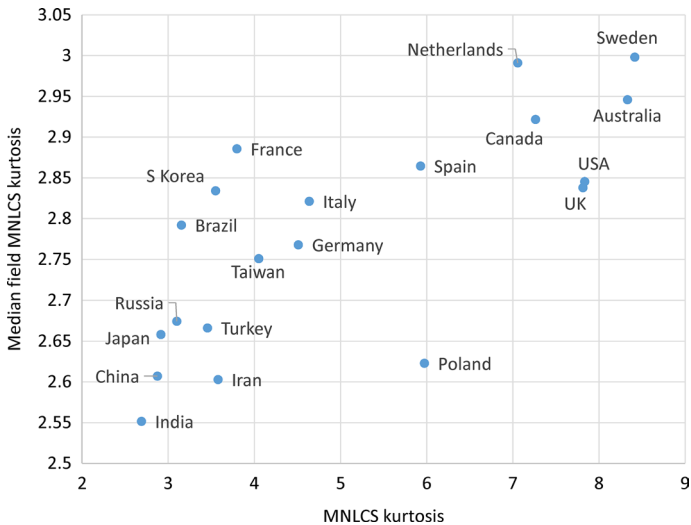


**Fig. 3** Kurtosis values calculated separately for the 20 countries with the most Scopus 2013 journal articles, using the field normalised log citation counts and the Scopus classification scheme. *The median of the individual field kurtosis values is plotted against the overall kurtosis value

UK scores have benefitted from avoiding non-English literary magazines but publishing in mainstream literary journals. The problem of essentially uncitable articles in Scopus has been previously shown to give some fields untypical (zero-inflated) citation distributions (Thelwall 2016b).

The eight countries in the top 20 with high MNLCS kurtosis values were investigated to find the cause. This entailed inspecting the 100 articles from each country with the

highest MNLCS$_i$ values because these will have contributed most to the MNLCS kurtosis score. In addition to Literature and Literary Theory and Visual Arts and Performing Arts, the Religious Studies category was prominent in these. There are outliers in Religious Studies due to the combination of magazines (e.g., *The Expository Times*, with all 53 articles uncited, including "All-Age Worship Resources for July"; *Parabola* with all 43 articles uncited, including "The night I died") and entirely uncited journals (e.g., *Svensk Teologisk Kvartalskrift*; *Bibel und Kirche*; *Bulletin de Litterature Ecclesiastique*) together with well-cited interdisciplinary journals with a main focus outside religion (e.g., *Psychology of Religion and Spirituality*: 97% cited; *Christian Bioethics*: 95% cited; *Journal of Religion and Health*: 85% cited). The following summarises the results for the largest 100 MNLCS$_i$ articles for each country. The list is ordered in decreasing order of country MNLCS kurtosis.

- Sweden: 24% in Literature and Literary Theory or Visual Arts and Performing Arts, including the largest 10 MNLCS values; 8% in Religious Studies; 57% in the Arts and Humanities broad category.
- Australia: 59% in Literature and Literary Theory or Visual Arts and Performing Arts, including the largest 24 MNLCS values; 6% in Religious Studies; 80% in the Arts and Humanities broad category.
- USA: 89% in Literature and Literary Theory or Visual Arts and Performing Arts, including the largest 28 MNLCS values; 7% in Religious Studies; 96% in the Arts and Humanities broad category.
- UK: 76% in Literature and Literary Theory or Visual Arts and Performing Arts, including the largest 21 MNLCS values; 4% in Religious Studies; 93% in the Arts and Humanities broad category.
- Canada: 46% in Literature and Literary Theory or Visual Arts and Performing Arts, including the largest 16 MNLCS values; 12% in Religious Studies; 78% in the Arts and Humanities broad category.
- Netherlands: 26% in Literature and Literary Theory or Visual Arts and Performing Arts, including the largest 5 MNLCS values; 15% in Religious Studies; 70% in the Arts and Humanities broad category.
- Poland: 13% in Literature and Literary Theory or Visual Arts and Performing Arts, including the largest 4 MNLCS values; 6% in Religious Studies; 40% in the Arts and Humanities broad category. A further 17% were from Computer Science (miscellaneous).
- Spain: 58% in Literature and Literary Theory or Visual Arts and Performing Arts, including the largest 9 MNLCS values; 3% in Religious Studies; 81% in the Arts and Humanities broad category.

In summary, relatively highly cited articles in the Literature and Literary Theory and/or Visual Arts and Performing Arts categories were the main causes of high MNLCS kurtosis for Australia, Canada, Spain, the UK and the USA. Considering the lower output for The Netherlands and Sweden, this, together with the wider Arts and Humanities, accounts for their high kurtosis values. For Poland, Computer Science (miscellaneous) articles were a significant contributory factor. The 17 articles in this category were all from *International Journal of Applied Mathematics and Computer Science*, from 13 different Polish research institutions. This journal seems to reflect a high citation specialism with the category. This is exacerbated by the inclusion of the huge rarely-cited *Information Technology Journal* from China that accounted for 44% (1439) of the articles in Computer Science (miscellaneous). Only 27% of its articles were cited in contrast with 71% of the remaining articles.

Changing between MNLCS and MNCS makes little difference to the relative sizes of country scores (Spearman correlation: 0.971) but the difference is substantial enough to change the policy conclusions that might be drawn from the results (see "Appendix" section, Table 7 for values). For example, the USA is ranked 2nd according to MNCS but 5th according to MNLCS (Fig. 4).

## Science-Metrix classification scheme

The results for the Science-Metrix classification scheme are like those for the Scopus classification scheme, but with approximately half the kurtosis. The MNCS kurtosis values vary between 59 (Poland) and 32,784 (Japan), with an average of 2660. The MNLCS kurtosis values range from 2.6 (China) to 3.9 (Netherlands) with an average of 3.3 and are therefore close to the normal distribution values (Figs. 5, 6).

The Science-Metrix $MNLCS_i$ article-level outliers are substantially more moderate than those for the Scopus classifications, with the highest being 6.8 for 2013 (Psychoanalysis, Canada, 60 citations) and 5.6 for 2012 (Literary Studies, USA, 90 citations). The lack of higher outliers is due to the Science-Metrix classification scheme not including many non-English literary magazines (in its Literary Studies category), increasing the average log citation score for the category.

Median MNLCS kurtosis values calculated separately for each country and field are slightly lower (average 2.9) than cross-field kurtosis values and the two correlate (Spearman correlation: 0.789), but the difference is much smaller (Fig. 7).

Changing between MNLCS and MNCS makes little difference to the relative sizes of country scores (Spearman correlation: 0.943) but, again, enough to change the conclusions that might be drawn from them. For example, China has a substantially higher MNCS than Brazil but a slightly lower MNLCS (Fig. 8).
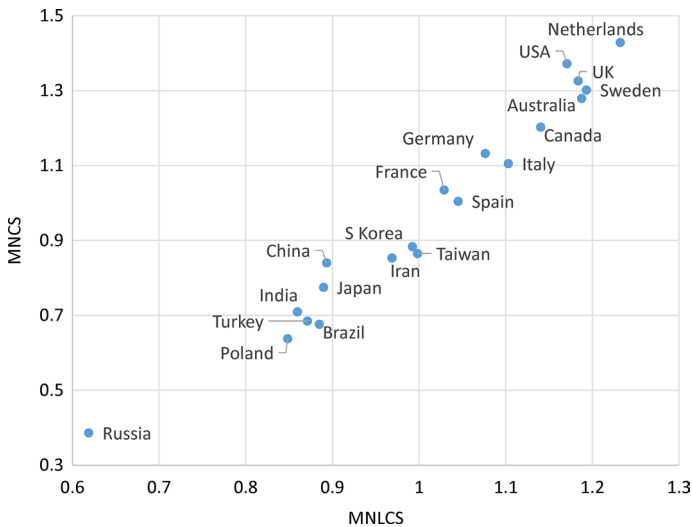


**Fig. 4** MNCS against MNLCS for the 20 countries with the most Scopus 2013 journal articles, using the Scopus classification scheme

**Fig. 5** Kurtosis values calculated separately for the 20 countries with the most Scopus 2013 journal articles using the field normalised log citation counts and the Science-Metrix classification scheme. MNLCS values are plotted on the y axis for comparison



**Fig. 6** Kurtosis values calculated separately for the 75 countries with the most Scopus 2013 journal articles using the field normalised log citation counts and the Science-Metrix classification scheme. MNLCS values are plotted on the y axis for comparison. *Not all countries are labelled

## Scopus versus Science-Metrix

The choice of classification scheme influences the MNCS and MNLCS results by a small but significant amount in all country sets (Table 2). Even when using the relatively outlier-resistant MNLCS, changing the classification scheme alters the country ranks. The change
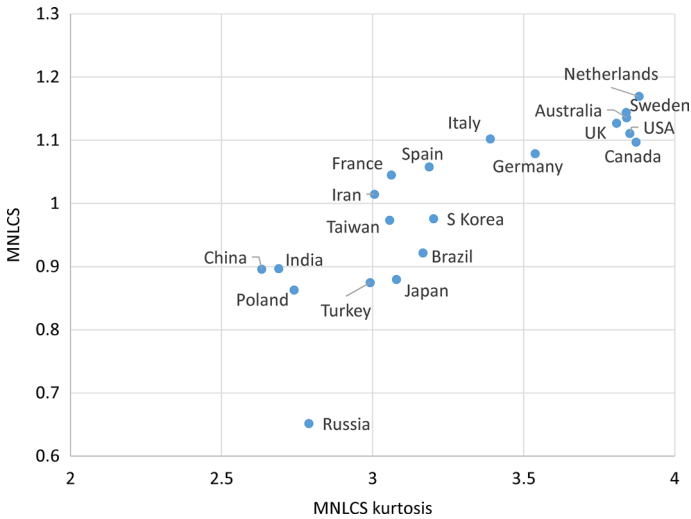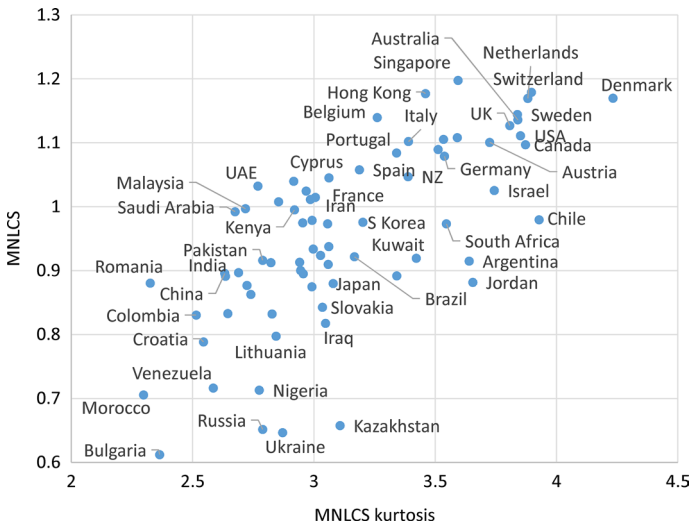
**Fig. 7** Kurtosis values calculated separately for the 20 countries with the most Scopus 2013 journal articles, using the field normalised log citation counts and the Science-Metrix classification scheme. The median of the individual field kurtosis values is plotted against the overall kurtosis value
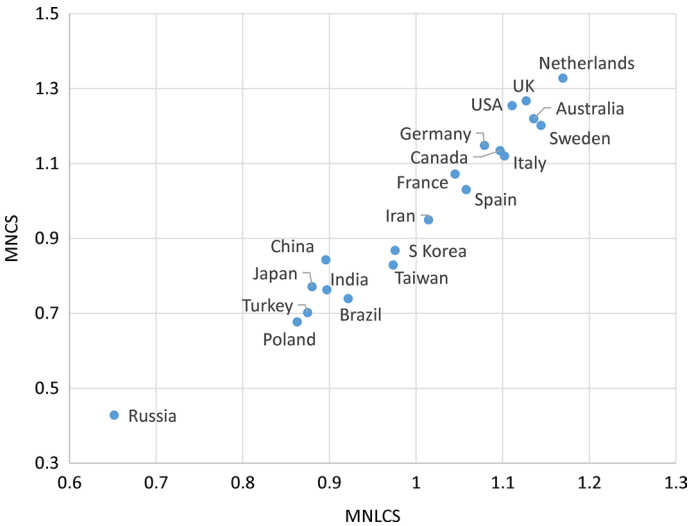


**Fig. 8** MNCS against MNLCS for the 20 countries with the most Scopus 2013 journal articles, using the Science-Metrix classification scheme

is larger for the top 50 and 75. MNCS is more resistant to classification scheme changes than is MNLCS.

The MNCS and MNLCS results for journal articles from 2013 were compared with those from 2012. Comparing two different source years is a better test of the influence of outliers than comparing within years because it uses a different set of data with none of the

**Table 2** Spearman correlations between MNCS and MNLCS based upon Scopus and Science-Metrix classifications for the 20, 50 and 75 countries with the most articles in Scopus in 2013

| Science-Metrix Scopus | Top 20 | | Top 50 | | Top 75 | |
|---|---|---|---|---|---|---|
| | MNLCS | MNCS | MNLCS | MNCS | MNLCS | MNCS |
| MNLCS (Spearman) | 0.976 | 0.965 | 0.976 | 0.968 | 0.972 | 0.948 |
| MNCS (Spearman) | 0.956 | 0.986 | 0.963 | 0.985 | 0.966 | 0.969 |

Comparison with 2012 data

same outliers. If there is unlikely to be a dramatic change in the research capability of a country between 2012 and 2013, then a robust field normalised average impact calculation should give results for 2012 that highly correlate with the results from 2013.

The Scopus MNLCS individual article outliers from 2012 were again mainly due to the two narrow fields Literature and Literary Theory and Visual Arts and Performing Arts, which together accounted for 94% of the 100 largest MNLCS$_i$ article values, with a highest individual score of 13.0 again.

For both the Scopus (Fig. 9) and Science-Metrix (Fig. 10) classification schemes, and for all three sets of countries investigated, the MNLCS and MNCS values from 2012 correlate very strongly with the corresponding values from 2013 (bold figures in Tables 3, 4, 5). Since the bold correlations are all very high and similar in magnitude within the same tables, they do not point to one indicator or classification scheme being more robust over time.

Kurtosis values were calculated for the 2012 data to check that the 2013 data was not unusual. The MNLCS but not MNCS averages were similar in all cases (Table 6). Sweden had the highest MNLCS kurtosis for MNLCS in 2012, echoing the situation for 2013. In contrast, whilst Canada had the highest MNLCS kurtosis for the Science-Metrix classifications in 2013, The Netherlands had the highest for Science-Metrix in 2012.
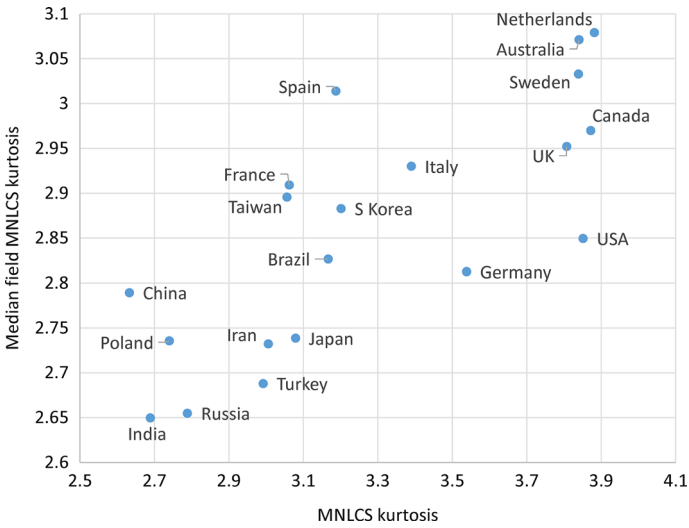


**Fig. 9** Kurtosis values from 2012 calculated separately for the 75 countries with the most Scopus 2013 journal articles using the field normalised log citation counts and the Scopus classification scheme. MNLCS values are plotted on the y axis for comparison. *Not all countries are labelled
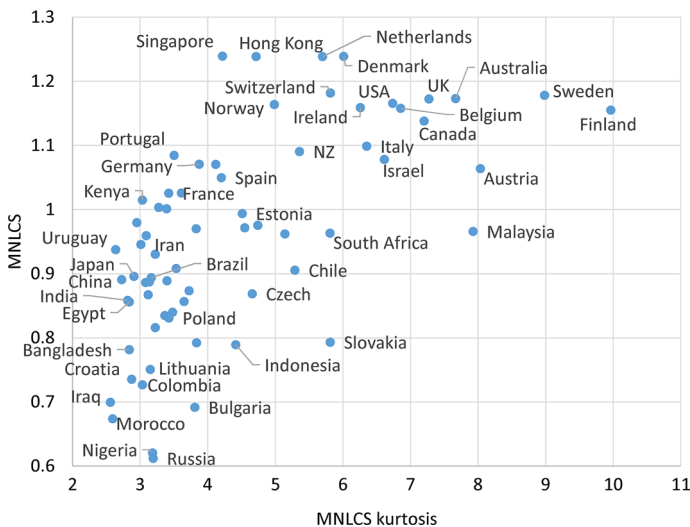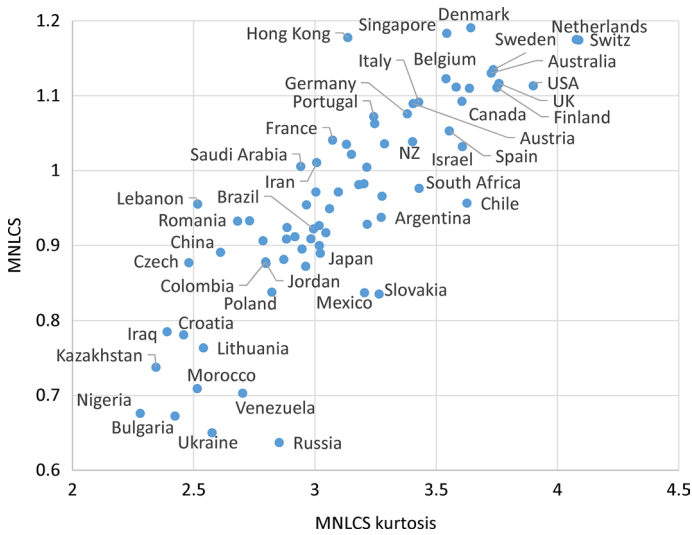
**Fig. 10** Kurtosis values from 2012 calculated separately for the 75 countries with the most Scopus 2013 journal articles using the field normalised log citation counts and the Science-Metrix classification scheme. MNLCS values are plotted on the y axis for comparison. *Not all countries are labelled

**Table 3** Spearman correlations between MNCS and MNLCS based upon Scopus and Science-Metrix classifications for the 20 countries with the most articles in Scopus in 2013

| Top 20 | | Scopus 2013 | | Science-Metrix 2013 | |
|---|---|---|---|---|---|
| | | MNLCS | MNLS | MNLCS | MNLS |
| Scopus 2012 | MNLCS | **0.990** | 0.983 | 0.968 | 0.949 |
| Scopus 2012 | MNCS | 0.981 | **0.988** | 0.961 | 0.958 |
| Science-Metrix 2012 | MNLCS | 0.939 | 0.938 | **0.976** | 0.962 |
| Science-Metrix 2012 | MNCS | 0.943 | 0.952 | 0.966 | **0.968** |

Correlations between years for the same indicator are bold

**Table 4** Spearman correlations between MNCS and MNLCS based upon Scopus and Science-Metrix classifications for the 50 countries with the most articles in Scopus in 2013

| Top 50 | | Scopus 2013 | | Science-Metrix 2013 | |
|---|---|---|---|---|---|
| | | MNLCS | MNCS | MNLCS | MNCS |
| Scopus 2012 | MNLCS | **0.989** | 0.977 | 0.971 | 0.960 |
| Scopus 2012 | MNCS | 0.977 | **0.991** | 0.959 | 0.978 |
| Science-Metrix 2012 | MNLCS | 0.962 | 0.955 | **0.987** | 0.968 |
| Science-Metrix 2012 | MNCS | 0.963 | 0.978 | 0.973 | **0.987** |

Correlations between years for the same indicator are bold

**Table 5** Spearman correlations between MNCS and MNLCS based upon Scopus and Science-Metrix classifications for the 75 countries with the most articles in Scopus in 2013

| Top 75 | | Scopus 2013 | | Science-Metrix 2013 | |
|---|---|---|---|---|---|
| | | MNLCS | MNCS | MNLCS | MNCS |
| Scopus 2012 | MNLCS | **0.994** | 0.970 | 0.979 | 0.959 |
| Scopus 2012 | MNCS | 0.973 | **0.997** | 0.947 | 0.985 |
| Science-Metrix 2012 | MNLCS | 0.970 | 0.952 | **0.994** | 0.950 |
| Science-Metrix 2012 | MNCS | 0.967 | 0.991 | 0.968 | **0.994** |

Correlations between years for the same indicator are bold

**Table 6** Average and extreme kurtosis values for the field normalised citation counts of the top 20 countries (i.e., with the most articles in Scopus in 2013) in 2012 and 2013

| Set | Kurtosis | 2012 kurtosis | | | | 2013 kurtosis | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | Scopus | | Science-Metrix | | Scopus | | Science-Metrix | |
| | | MNLCS | MNCS | MNLCS | MNCS | MNLCS | MNCS | MNLCS | MNCS |
| Top 20 | Min | 2.7 | 53 | 2.6 | 36 | 2.7 | 119 | 2.6 | 59 |
| | Max | 9.0 | 9007 | 4.1 | 4886 | 8.4 | 518,540 | 3.9 | 32,784 |
| | Average | 4.6 | 1996 | 3.3 | 1088 | 5.0 | 29,518 | 3.3 | 2660 |
| Top 50 | Min | 2.6 | 18 | 2.5 | 16 | 2.7 | 22 | 2.5 | 15 |
| | Max | 10.0 | 9007 | 4.1 | 4886 | 13.8 | 518,540 | 4.2 | 32,784 |
| | Average | 4.7 | 1050 | 3.2 | 622 | 5.4 | 11,967 | 3.2 | 1165 |
| Top 75 | Min | 2.6 | 11 | 2.3 | 9 | 2.5 | 11 | 2.3 | 8 |
| | Max | 10.0 | 9007 | 4.1 | 4886 | 13.8 | 518,540 | 4.2 | 32,784 |
| | Average | 4.3 | 721 | 3.1 | 424 | 5.1 | 7999 | 3.1 | 790 |

Correlations between years for the same indicator are bold

## Discussion

The results are limited by the classification schemes and years analysed. Different patterns may have been obtained from WoS (possibly with weaker outliers if WoS has more balanced journal coverage) or article-based classification schemes (Perianes-Rodriguez and Ruiz-Castillo 2017). Newer or much older data may also display different characteristics. The analysis is restricted to a technical discussion, without known correct values for the underlying research impacts of the countries. The influence of outliers may also be different for smaller types of unit, such as research groups, where they seem likely change the results more. Of course, citation counting is only one way of attempting to assess scholarly influence and it ignores may important ways in which scholarship can be useful to other researchers (MacRoberts and MacRoberts 2010) and society (Priem et al. 2010).

The high MNLCS kurtosis values for some countries with the Scopus classification scheme were mainly due to values being inflated for two fields with many essentially uncitable articles. This would be less of a problem for percentile indicators. These report the percentage of a unit's articles that are within the top X% (e.g., 1%, 5%, 10%, 50%) for

each field separately (Waltman and Schreiber 2013). The uncitable articles could inflate this percentile for units not publishing uncitable articles but individual outliers have little affect (Waltman, et al. 2012). Nevertheless, sets of outliers (as in the two fields above) may still influence the overall results. This may be the reason why one study found classification schemes to influence the top 1% results more than the top 10% results for universities (Perianes-Rodriguez and Ruiz-Castillo 2018). Two other citation indicators, the number of highly cited articles and the proportion of highly cited articles (Waltman 2016), have similar issues to percentile indicators.

## Conclusions

The country-level data for 2013 and 2012 shows, as expected, that MNCS has extreme outliers. It also shows that MNLCS can have moderately stronger outliers than the normal distribution, if the Scopus classification scheme is used and slightly stronger if the Science-Metrix scheme is used. The stronger Scopus outliers are due to the inclusion of essentially uncitable articles in a few Scopus categories, lowering the world average citation rate. The slightly stronger outliers for MNLCS with the Science-Metrix scheme than the normal distribution, despite citations approximately following the discretised lognormal distribution (Thelwall 2016a), could be due to countries having differing research impacts between fields, which would inflate kurtosis values calculated using the overall average rather than individual field averages. This conclusion is supported by the lower values found when taking the median of the within-field kurtoses rather than a single cross-field kurtosis.

Although the MNLCS was designed to reduce the impact of individual highly cited articles, the above findings show that it is important to check for uncitable articles in any category before producing MNLCS to inform policy decisions. This is even more important for MNCS calculations because of the much larger outliers. Surprisingly, however, the higher MNCS outliers do not make MNCS values more stable over time, nor does the use of a classification scheme (from Science-Metrix) that reduces the largest outliers. This is probably due to the long-term root cause of the highest outliers (at the country level) being classification scheme issues that influence countries consistently, despite differing between countries.

The differences between MNLCS and MNCS values and between classification schemes were numerically small in all cases, in the sense of producing very high correlations between them, especially for larger countries. Nevertheless, in a context where small differences can have implications for policy decisions, especially if they change the rank order of nations, these small variations are worrying. For example, a country believing that a recent policy decision had helped its researchers to overtake a near competitor in average research impact might not realise that this indicator-based conclusion would be reversed if a different indicator or classification scheme had been used. Thus, indicator producers should be careful to ensure that the categorisation schemes are appropriate and that the end users are aware of the potential influence of outliers and category schemes on the conclusions drawn.

## Appendix

See Table 7.

**Table 7** MNLCS and MNCS values for the 75 countries with the most articles in Scopus in 2013, calculated using data from 2012 or 2013 and either the Scopus or Science-Metrix classification scheme

| First author country | Scopus 2013 | | Science-Metrix 2013 | | Scopus 2012 | | Science-Metrix 2012 | |
|---|---|---|---|---|---|---|---|---|
| | MNLCS | MNCS | MNLCS | MNCS | MNLCS | MNCS | MNLCS | MNCS |
| USA | 1.171 | 1.373 | 1.111 | 1.255 | 1.166 | 1.365 | 1.113 | 1.271 |
| China | 0.894 | 0.841 | 0.896 | 0.843 | 0.891 | 0.823 | 0.891 | 0.826 |
| UK | 1.184 | 1.326 | 1.127 | 1.268 | 1.172 | 1.314 | 1.116 | 1.236 |
| Japan | 0.890 | 0.775 | 0.880 | 0.771 | 0.896 | 0.755 | 0.889 | 0.747 |
| Germany | 1.076 | 1.133 | 1.079 | 1.148 | 1.071 | 1.152 | 1.076 | 1.145 |
| India | 0.860 | 0.710 | 0.897 | 0.763 | 0.859 | 0.702 | 0.906 | 0.774 |
| France | 1.029 | 1.035 | 1.045 | 1.072 | 1.026 | 1.037 | 1.041 | 1.056 |
| Italy | 1.103 | 1.105 | 1.102 | 1.120 | 1.099 | 1.099 | 1.092 | 1.103 |
| S Korea | 0.992 | 0.884 | 0.976 | 0.868 | 1.004 | 0.907 | 0.983 | 0.881 |
| Canada | 1.140 | 1.203 | 1.097 | 1.135 | 1.138 | 1.200 | 1.093 | 1.147 |
| Spain | 1.045 | 1.005 | 1.058 | 1.031 | 1.050 | 1.016 | 1.053 | 1.042 |
| Australia | 1.188 | 1.279 | 1.136 | 1.220 | 1.173 | 1.277 | 1.130 | 1.221 |
| Brazil | 0.885 | 0.676 | 0.922 | 0.740 | 0.894 | 0.674 | 0.922 | 0.723 |
| Iran | 0.969 | 0.854 | 1.015 | 0.950 | 0.946 | 0.804 | 1.011 | 0.924 |
| Taiwan | 0.998 | 0.865 | 0.973 | 0.830 | 1.001 | 0.887 | 0.972 | 0.842 |
| Russia | 0.619 | 0.387 | 0.652 | 0.428 | 0.612 | 0.389 | 0.637 | 0.405 |
| Netherlands | 1.232 | 1.429 | 1.169 | 1.328 | 1.239 | 1.446 | 1.174 | 1.330 |
| Turkey | 0.871 | 0.685 | 0.875 | 0.702 | 0.887 | 0.694 | 0.900 | 0.720 |
| Poland | 0.848 | 0.638 | 0.863 | 0.678 | 0.831 | 0.603 | 0.838 | 0.630 |
| Sweden | 1.193 | 1.302 | 1.144 | 1.202 | 1.178 | 1.277 | 1.135 | 1.213 |
| Switzerland | 1.192 | 1.416 | 1.179 | 1.383 | 1.182 | 1.459 | 1.175 | 1.396 |
| Belgium | 1.169 | 1.287 | 1.139 | 1.257 | 1.158 | 1.262 | 1.123 | 1.252 |
| Malaysia | 0.963 | 0.832 | 0.997 | 0.924 | 0.966 | 0.823 | 1.035 | 0.992 |
| Portugal | 1.106 | 1.055 | 1.084 | 1.053 | 1.085 | 1.048 | 1.072 | 1.059 |
| Mexico | 0.827 | 0.603 | 0.832 | 0.622 | 0.840 | 0.632 | 0.837 | 0.650 |
| Israel | 1.075 | 1.020 | 1.025 | 0.948 | 1.078 | 1.004 | 1.032 | 0.961 |
| Denmark | 1.225 | 1.345 | 1.170 | 1.270 | 1.239 | 1.397 | 1.191 | 1.360 |
| Greece | 1.032 | 0.953 | 1.025 | 0.963 | 1.026 | 0.932 | 1.022 | 0.947 |
| Czech Republic | 0.881 | 0.716 | 0.877 | 0.746 | 0.869 | 0.699 | 0.877 | 0.727 |
| Austria | 1.079 | 1.087 | 1.101 | 1.131 | 1.064 | 1.076 | 1.090 | 1.106 |
| Finland | 1.158 | 1.144 | 1.108 | 1.085 | 1.155 | 1.142 | 1.111 | 1.108 |
| Norway | 1.174 | 1.160 | 1.105 | 1.098 | 1.164 | 1.155 | 1.110 | 1.112 |
| Singapore | 1.266 | 1.664 | 1.198 | 1.481 | 1.239 | 1.710 | 1.183 | 1.470 |
| South Africa | 0.967 | 0.829 | 0.973 | 0.876 | 0.963 | 0.803 | 0.976 | 0.858 |
| Argentina | 0.876 | 0.667 | 0.915 | 0.731 | 0.908 | 0.694 | 0.938 | 0.733 |
| Egypt | 0.872 | 0.677 | 0.913 | 0.760 | 0.856 | 0.639 | 0.927 | 0.751 |
| NZ | 1.109 | 1.122 | 1.047 | 1.061 | 1.091 | 1.086 | 1.039 | 1.068 |
| Thailand | 0.935 | 0.741 | 0.924 | 0.780 | 0.970 | 0.805 | 0.971 | 0.865 |
| Ireland | 1.157 | 1.210 | 1.089 | 1.127 | 1.159 | 1.244 | 1.112 | 1.167 |
| Saudi Arabia | 0.957 | 0.875 | 0.992 | 0.961 | 0.980 | 0.935 | 1.006 | 0.967 |
| Pakistan | 0.882 | 0.716 | 0.916 | 0.777 | 0.874 | 0.654 | 0.895 | 0.727 |
| Ukraine | 0.541 | 0.319 | 0.647 | 0.426 | 0.530 | 0.299 | 0.650 | 0.422 |

**Table 7** (continued)

| First author country | Scopus 2013 | | Science-Metrix 2013 | | Scopus 2012 | | Science-Metrix 2012 | |
|---|---|---|---|---|---|---|---|---|
| | MNLCS | MNCS | MNLCS | MNCS | MNLCS | MNCS | MNLCS | MNCS |
| Hungary | 0.848 | 0.659 | 0.910 | 0.729 | 0.857 | 0.683 | 0.924 | 0.747 |
| Chile | 0.916 | 0.793 | 0.980 | 0.880 | 0.906 | 0.729 | 0.957 | 0.821 |
| Serbia | 0.829 | 0.596 | 0.895 | 0.692 | 0.835 | 0.593 | 0.912 | 0.699 |
| Slovakia | 0.802 | 0.551 | 0.843 | 0.629 | 0.793 | 0.552 | 0.836 | 0.657 |
| Tunisia | 0.898 | 0.669 | 0.900 | 0.711 | 0.886 | 0.683 | 0.872 | 0.684 |
| Slovenia | 0.918 | 0.737 | 0.934 | 0.782 | 0.976 | 0.822 | 0.981 | 0.882 |
| Colombia | 0.706 | 0.492 | 0.831 | 0.635 | 0.727 | 0.518 | 0.878 | 0.694 |
| Croatia | 0.749 | 0.543 | 0.789 | 0.608 | 0.735 | 0.503 | 0.781 | 0.552 |
| Hong Kong | 1.241 | 1.430 | 1.177 | 1.353 | 1.239 | 1.530 | 1.178 | 1.361 |
| Nigeria | 0.633 | 0.372 | 0.713 | 0.462 | 0.621 | 0.359 | 0.676 | 0.413 |
| Algeria | 0.846 | 0.679 | 0.892 | 0.797 | 0.867 | 0.658 | 0.909 | 0.729 |
| Romania | 0.550 | 0.388 | 0.880 | 0.712 | 0.537 | 0.377 | 0.932 | 0.734 |
| Morocco | 0.712 | 0.519 | 0.706 | 0.528 | 0.674 | 0.445 | 0.709 | 0.553 |
| Bulgaria | 0.733 | 0.480 | 0.612 | 0.379 | 0.692 | 0.442 | 0.672 | 0.408 |
| Lithuania | 0.796 | 0.597 | 0.798 | 0.634 | 0.751 | 0.498 | 0.764 | 0.541 |
| Indonesia | 0.748 | 0.536 | 0.891 | 0.784 | 0.789 | 0.568 | 0.949 | 0.816 |
| Jordan | 0.799 | 0.596 | 0.882 | 0.723 | 0.792 | 0.586 | 0.876 | 0.696 |
| Bangladesh | 0.748 | 0.515 | 0.833 | 0.629 | 0.782 | 0.567 | 0.881 | 0.687 |
| Viet Nam | 0.981 | 0.878 | 0.979 | 0.930 | 0.960 | 0.785 | 0.917 | 0.780 |
| Estonia | 1.043 | 0.900 | 1.011 | 0.902 | 0.994 | 0.856 | 1.036 | 0.975 |
| Cuba | 0.491 | 0.286 | 0.478 | 0.284 | 0.544 | 0.331 | 0.521 | 0.308 |
| UAE | 1.043 | 0.948 | 1.032 | 0.938 | 0.972 | 0.804 | 0.966 | 0.795 |
| Venezuela | 0.581 | 0.378 | 0.716 | 0.489 | 0.570 | 0.338 | 0.703 | 0.455 |
| Cyprus | 1.097 | 1.040 | 1.040 | 0.955 | 1.070 | 1.042 | 1.063 | 1.129 |
| Lebanon | 0.957 | 0.777 | 0.938 | 0.756 | 0.962 | 0.815 | 0.955 | 0.805 |
| Belarus | 0.475 | 0.235 | 0.475 | 0.258 | 0.496 | 0.268 | 0.501 | 0.277 |
| Ethiopia | 0.957 | 0.815 | 1.008 | 0.882 | 0.930 | 0.772 | 0.933 | 0.826 |
| Kenya | 0.985 | 0.865 | 0.995 | 0.889 | 1.015 | 0.877 | 1.004 | 0.890 |
| Kazakhstan | 0.397 | 0.216 | 0.658 | 0.401 | 0.581 | 0.421 | 0.738 | 0.559 |
| Philippines | 0.789 | 0.630 | 0.913 | 0.816 | 0.816 | 0.648 | 0.928 | 0.814 |
| Iraq | 0.765 | 0.543 | 0.818 | 0.647 | 0.700 | 0.439 | 0.785 | 0.528 |
| Kuwait | 0.922 | 0.700 | 0.920 | 0.683 | 0.889 | 0.723 | 0.909 | 0.837 |
| Uruguay | 0.897 | 0.734 | 0.975 | 0.838 | 0.938 | 0.734 | 0.954 | 0.780 |

Countries are in decreasing order of number of journal articles in Scopus 2013

# References

Aksnes, D. W. (2003). Characteristics of highly cited papers. *Research Evaluation, 12*(3), 159–170.

Aksnes, D. W., & Sivertsen, G. (2004). The effect of highly cited papers on national citation indicators. *Scientometrics, 59*(2), 213–224.

Archambault, É., Beauchesne, O. H., & Caruso, J. (2011). Towards a multilingual, comprehensive and open scientific journal ontology. In *Proceedings of the 13th international conference of the international society for scientometrics and informetrics* (pp. 66–77). Durban.

Bornmann, L., Wagner, C., & Leydesdorff, L. (2015). BRICS countries and scientific excellence: A bibliometric analysis of most frequently cited papers. *Journal of the Association for Information Science and Technology, 66*(7), 1507–1513.

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

Donner, P. (2017). Document type assignment accuracy in the journal citation index data of Web of Science. *Scientometrics, 113*(1), 219–236.

Falagas, M. E., Pitsouni, E. I., Malietzis, G. A., & Pappas, G. (2008). Comparison of PubMed, Scopus, web of science, and Google scholar: Strengths and weaknesses. *The FASEB Journal, 22*(2), 338–342.

Glänzel, W., & Schubert, A. (2003). A new classification scheme of science fields and subfields designed for scientometric evaluation purposes. *Scientometrics, 56*(3), 357–367.

Ioannidis, J. P., & Panagiotou, O. A. (2011). Comparison of effect sizes associated with biomarkers reported in highly cited individual articles and in subsequent meta-analyses. *JAMA, 305*(21), 2200–2210.

Larivière, V., Desrochers, N., Macaluso, B., Mongeon, P., Paul-Hus, A., & Sugimoto, C. R. (2016). Contributorship and division of labor in knowledge production. *Social Studies of Science, 46*(3), 417–435.

Levitt, J., & Thelwall, M. (2013). Alphabetization and the skewing of first authorship towards last names early in the alphabet. *Journal of Informetrics, 7*(3), 575–582.

MacRoberts, M. H., & MacRoberts, B. R. (2010). Problems of citation analysis: A study of uncited and seldom-cited influences. *Journal of the American Society for Information Science and Technology, 61*(1), 1–12.

Martín-Martín, A., Orduna-Malea, E., & López-Cózar, E. D. (2018). Coverage of highly-cited documents in Google Scholar, Web of Science, and Scopus: A multidisciplinary comparison. *Scientometrics, 116*(3), 2175–2188.

McCain, K. W. (2011). Eponymy and obliteration by incorporation: The case of the "Nash Equilibrium". *Journal of the American Society for Information Science and Technology, 62*(7), 1412–1424.

Merton, R. K. (1968). The Matthew effect in science: The reward and communication systems of science are considered. *Science, 159*(3810), 56–63.

Oppenheim, C., & Renn, S. P. (1978). Highly cited old papers and the reasons why they continue to be cited. *Journal of the American Society for Information Science and Technology, 29*(5), 225–231.

Perianes-Rodriguez, A., & Ruiz-Castillo, J. (2017). A comparison of the Web of Science and publication-level classification systems of science. *Journal of Informetrics, 11*(1), 32–45.

Perianes-Rodriguez, A., & Ruiz-Castillo, J. (2018). The impact of classification systems in the evaluation of the research performance of the Leiden Ranking Universities. *Journal of the Association for Information Science and Technology, 69*(8), 1046–1053.

Persson, O. (2009). Are highly cited papers more international? *Scientometrics, 83*(2), 397–401.

Price, D. de Solla (1976). A general theory of bibliometric and other cumulative advantage processes. *Journal of the American Society for Information Science and Technology, 27*(5), 292–306.

Priem, J., Taraborelli, D., Groth, P., & Neylon, C. (2010). *Altmetrics: A manifesto.* http://altmetrics.org/manifesto/.

Ruiz-Castillo, J., & Waltman, L. (2015). Field-normalized citation impact indicators using algorithmically constructed classification systems of science. *Journal of Informetrics, 9*(1), 102–117.

Thelwall, M. (2016a). Are the discretised lognormal and hooked power law distributions plausible for citation data? *Journal of Informetrics, 10*(2), 454–470.

Thelwall, M. (2016b). Are there too many uncited articles? Zero inflated variants of the discretised lognormal and hooked power law distributions. *Journal of Informetrics, 10*(2), 622–633.

Thelwall, M. (2017a). *Web indicators for research evaluation: A practical guide.* San Rafael, CA: Morgan & Claypool.

Thelwall, M. (2017b). Three practical field normalised alternative indicator formulae for research evaluation. *Journal of Informetrics, 11*(1), 128–151.

Thelwall, M. (2018). Do females create higher impact research? Scopus citations and Mendeley readers for articles from five countries. *Journal of Informetrics, 12*(4), 1031–1041.

Tijssen, R. J., Visser, M. S., & Van Leeuwen, T. N. (2002). Benchmarking international scientific excellence: Are highly cited research papers an appropriate frame of reference? *Scientometrics, 54*(3), 381–397.

Waltman, L. (2016). A review of the literature on citation impact indicators. *Journal of Informetrics, 10*(2), 365–391.

Waltman, L., Calero-Medina, C., Kosten, J., Noyons, E. C., Tijssen, R. J., van Eck, N. J., et al. (2012). The Leiden Ranking 2011/2012: Data collection, indicators, and interpretation. *Journal of the American Society for Information Science and Technology, 63*(12), 2419–2432.

Waltman, L., & Schreiber, M. (2013). On the calculation of percentile-based bibliometric indicators. *Journal of the American Society for Information Science and Technology, 64*(2), 372–379.

Waltman, L., van Eck, N. J., van Leeuwen, T. N., Visser, M. S., & van Raan, A. F. (2011a). Towards a new crown indicator: Some theoretical considerations. *Journal of Informetrics, 5*(1), 37–47.

Waltman, L., van Eck, N. J., van Leeuwen, T. N., Visser, M. S., & van Raan, A. F. (2011b). Towards a new crown indicator: An empirical analysis. *Scientometrics, 87*(3), 467–481.

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

Wang, Q., & Waltman, L. (2016). Large-scale analysis of the accuracy of the journal classification systems of Web of Science and Scopus. *Journal of Informetrics, 10*(2), 347–364.

Westfall, P. H. (2014). Kurtosis as peakedness, 1905–2014. RIP. *The American Statistician, 68*(3), 191–195.