An Evaluation of Percentile Measures of Citation Impact, and a Proposal for Making Them Better

Lutz Bornmann*
Richard Williams**

* Division for Science and Innovation Studies
Administrative Headquarters of the Max Planck Society
Hofgartenstr. 8,
80539 Munich, Germany.
Email: bornmann@gv.mpg.de

** Department of Sociology
University of Notre Dame
4058 Jenkins Nanovic Hall
Notre Dame, IN 46556 USA
Email: rwilliam@nd.edu
Abstract
Percentiles are statistics pointing to the standing of a paper’s citation impact relative to other papers in a given citation distribution. Percentile Ranks (PRs) often play an important role in evaluating the impact of scholars, institutions, and lines of study. Because PRs are so important for the assessment of scholarly impact, and because citation practices differ greatly across time and fields, various percentile approaches have been proposed to time- and field-normalize citations. Unfortunately, current popular methods often face significant problems in time- and field-normalization, including when papers are assigned to multiple fields or have been published by more than one unit (e.g., researchers or countries). They also face problems for estimating citation counts for pre-defined PRs (e.g., the 90th PR). We offer a series of guidelines and procedures that, we argue, address these problems and others and provide a superior means to make the use of percentile methods more accurate and informative. In particular, we argue that two approaches, CP-IN and CP-EX, should be preferred in bibliometric studies because they consider the complete citation distribution. Both approaches are based on cumulative frequencies in percentages (CPs). The paper further shows how bar graphs and beamplots can present PRs in a more meaningful and accurate manner.

Key words
bibliometrics; citation analysis; percentile; percentile rank; percentile point; plotting position; field-normalization
1 Introduction

Citation analyses are frequently used for cross-field and cross-time comparisons. For example, universities are compared in university rankings (e.g., the Leiden Ranking, see https://www.leidenranking.com) which publish in different fields and publication years. This leads to challenges for citation analyses, because of two reasons: first, citation counts depend on the length of the citation window: the longer the time papers can be cited, the more citations can be expected. Second, publication and citation behaviors are different in the fields, why a single citation has a different meaning in each field. Since the 1980s, various methods have been introduced (Tahamtan & Bornmann, 2018) to time- and field-normalize citations. In one of the most frequently used methods – the relative citation rate – expected values are calculated for every combination of publication year and subject category in databases such as Web of Science (WoS, Clarivate Analytics) or Scopus (Elsevier). These expected values are mean numbers of citations. Then, the citation impact of every focal paper in the combination of publication year and subject category is normalized by dividing the citation counts of the focal paper by the corresponding expected value (Schubert & Braun, 1986).

This approach of generating normalized citation impact values which can be used for cross-time and cross-field comparisons has been frequently criticized, since it is based on arithmetic averages of citations and citations as a rule are skewed distributed. In case of skewed distributions, the arithmetic average should not be used as a measure for the central tendency of the distribution. As an alternative to this normalization approach, various percentile approaches (plotting positions and percentile ranks, PRs) have been proposed. Plotting positions are quantiles of an empirical (or theoretical) distribution whereby quantiles are defined as specific cut points partitioning distributions into subsets. For example, the median is the 4-quantile separating distributions in four parts. Quantiles can be plotted in so
called Q-Q (quantile-quantile) plots – a method for comparing two empirical (or theoretical) distributions. In other words, quantiles of two distributions are plotted against each other in Q-Q plots.

Since a percentile is defined as “a statistic that gives the relative standing of a numerical data point when compared to all other data points in a distribution” (Lavrakas, 2008), plotting positions can be interpreted as percentiles. PR x is defined as the citation count (at or) below which x% (e.g., 90%) of the papers in the combination of publication year and subject category falls. Two papers from different combinations of subject category and publication year with exactly the same PR may have different citation counts. The advantage of PRs (and plotting position) is that they are not affected by outliers (highly cited papers) and their interpretation is simple and clear: if a focal paper has a PR of 90, then 90% of the papers in the publication year and subject category have a citation impact which is (at or) below the impact of the focal paper. This interpretation of PRs makes their use in citation analysis attractive, since it clearly shows the position of the paper in the combination of subject category and publication year which can be compared with the position of other papers.

It is not only possible to calculate plotting positions and PRs for every single paper in a database such as WoS. It is popular in bibliometrics (Bornmann, 2014) to identify in every combination of publication year and subject category the papers which belong to the 10% most frequently cited papers (Bornmann, de Moya Anegón, & Leydesdorff, 2012; Narin, 1987; Waltman et al., 2012). These are the papers in the citation distribution at or above the 90th PR. The number of these papers can be counted for various units (e.g., for a journal, researcher, university or country). P(top 10%) is the number of top-10% papers and PP(top 10%) is the proportion of top-10% papers published by a unit. Both indicators are presented for every university in the Leiden Ranking. Waltman et al. (2012) regard the PP(top 10%) indicator “as the most important impact indicator in the Leiden Ranking” (p. 2425). The
Leiden Ranking also provides institutional results based on the 1%, 5%, and 50% most frequently cited papers.

In recent years, various percentile approaches have been introduced in bibliometrics. In this study, the different approaches are presented, and their advantages and disadvantages explained. Our discussion will show that widely used percentile measures are problematic in some fairly common situations, such as when a paper is classified in more than one field or has been published by more than one unit (e.g., researchers or countries). We then offer an optimized PR approach that maintains most of the advantages of PR measures while overcoming or minimizing their most serious weaknesses. We further show how graphical approaches such as bar graphs and beamplots can make the presentation of PR results more meaningful and accurate.

2 Methods

For discussing the different percentile approaches, two datasets are used: (1) a fictitious small dataset including 21 papers with various citation counts, and (2) all papers published between 2000 and 2005 with the document type “article” and their citation counts until the end of 2018. The publication and citation data are from the Max Planck Society’s in-house database which is based on the WoS. WoS subject categories, which are sets of similar journals, have been used to compute field-normalized citation impact values (plotting positions and PRs). The publication set consists of 6,973,937 articles. However, these are more papers than have been published between 2000 and 2005, since papers which have been assigned to more than one subject category have been considered multiple times (for calculating plotting positions and PRs in every combination of publication year and subject category). Without multiple mentions of papers, the publication set consists of 4,416,554 articles. The articles received between 0 and 67,582 citations until the end of 2018 (median = 14).
3 Counting highly cited papers

Let us start the discussion of the various percentile approaches with the family of P(top \(x\%\)) indicators: the number of papers belonging to the \(x\%\) most-frequently cited papers. The use of P(top \(x\%\)) in research evaluation might be interpreted as unsatisfying, since the diverse citation impact of a unit (e.g., a researcher) is transformed into a binary information. Imagine researcher A has published many papers which are all in the range of P(top 11\%) and P(top 31\%) and another researcher B who has published all papers in the range of P(top 71\%) and P(top 91\%). Based on these numbers, one can conclude that researcher A has a better performance than researcher B. However, if only the number of P(top 10\%) are counted, both researchers would receive the same assessment with P(top 10\%) = 0.

Furthermore, the calculation of P(top \(x\%\)) is affected by the problem of citation ties at the threshold for separating the \(x\%\) most-frequently cited papers from the rest: suppose five papers with 20 citations, 20 papers with 10 citations, and 75 papers with 1 citation. It is not clear with this citation distribution whether the 20 papers with 10 citations should be assigned to the P(top 10\%) or the bottom 90\%. Although Waltman and Schreiber (2013) found an elegant solution for that problem (leading to a fractional assignment of papers at the threshold to the group of highly cited papers), it leads to data which are no longer binary: the papers at the threshold are counted with a value less than 1. The consequence is, for instance, that the data can no longer be analyzed with logistic regression analyses, although the nature of the indicator (papers belonging to the top-\(x\%\) or not) would suggest that this is the appropriate method.

As the example above with researchers A and B reveal, it is desirable to have a percentile solution which is able to reflect the whole range of citation impact received by the papers in a certain publication set. With the integrated impact indicator (I3), Leydesdorff and Bornmann (2011, 2012) proposed a solution going beyond the binary classification of impact.
Here, the papers in a set (of a journal or a university) are assigned to more than two impact classes based on PRs [e.g., six classes; P(top 1, 5, 10, 25, 50, 75%]. The formula for calculating the indicator is

\[ I3 = \sum_{i=1}^{y} x_i \times PC_i \]

whereby \( y \) is the number of PR classes considered (e.g., six classes as mentioned above) and \( PC_i \) denotes PR class \( i \). \( x_i \) is the number of papers published by a unit in \( PC_i \). The PR class including the papers with the most citation impact should receive the maximum weight. The formula can be used very flexible. For example, papers in the highest impact class can be given little more (e.g., six in the case of six classes) or significantly more weight than lowly cited papers. For example, Leydesdorff, Bornmann, and Adams (2019) proposed to weight the number of papers in \( P(\text{top } 1\%) \) with 100 and the number of papers in \( P(\text{top } 10\%) \) with 10. For the purpose of notifying the number of PR classes and weights used for calculating \( I3 \) in a study, Leydesdorff et al. (2019) proposed to use the general notation

\[ I3(PC_1 - W_1, PC_2 - W_2 \ldots PC_n - W_n) \]

whereby \( PC \) is the lower threshold of the PR class, e.g. 99 in case of \( P(\text{top } 1\%) \), and \( W \) the corresponding weight (e.g., 100). \( n \) defines the number of classes and weights, respectively. The flexibility in the use of PR classes and weights might be an advantage of \( I3 \), since the user can adapt the indicator to certain evaluation tasks (see here Bornmann & Marewski, 2019). The disadvantage of this flexibility is, however, that there is no standardized use of \( I3 \) (and the results may not be comparable). Another problem is that the indicator is still based on classes – \( P(\text{top } x\%) \) actually is an \( I3 \) indicator which can be expressed with the \( I3 \) notation: \( I3(90 - 1) \) – and does not consider the complete information of citation impact distributions. It is a decisive disadvantage of \( I3 \) that it can be calculated only on the aggregated level, i.e. for groups of papers. That means, it is not possible to calculate \( I3 \)
for all papers included in a database such as the WoS and to use the preprocessed data for
citation analyses of various units later.

4 Plotting positions

Bornmann, Leydesdorff, and Mutz (2013) discussed several possibilities to calculate
plotting positions for receiving time- and field-normalized citation impact values on the single
paper level. They preferred the calculation of plotting position based on the rule proposed by
Hazen (1914) using the formula

\[
\frac{i - 0.5}{n}
\]

For calculating the plotting positions, the papers (published in one publication year
and subject category) are sorted in decreasing order of citation counts and ranking positions
are assigned, whereby \( i \) is the rank of the paper and \( n \) is the total number of papers in the set.
As the example in Table 1 shows, the formula returns values which are between 0 and 1; for
receiving percentages, the values can be multiplied by 100. Ties in citation data do not pose a
problem for the calculation, since the corresponding papers simply receive the same (mean)
rank and plotting position. Plotting positions can be calculated for the papers in all subject
categories and publication years in a database such as WoS, whereby one receives comparable
time- and field-normalized citation impact values. For example, the results by Bornmann and
Marx (2015) reveal favorable results of plotting positions based on the rule proposed by
Hazen (1914) compared to other time- and field-normalization methods (e.g., methods based
on mean citations, see above).

Table 1. Example set of papers for calculating plotting positions (21 papers)

| Paper | Citation count | Rank | Hazen | In percent |
|-------|----------------|------|-------|------------|
| A     | 20             | 20.5 | 0.95  | 95.24      |
| B     | 20             | 20.5 | 0.95  | 95.24      |
| C     | 13             | 18.5 | 0.86  | 85.71      |
Cox (2005) outlined that it is an important advantage of plotting positions based on the rule proposed by Hazen (1914) that the formula leads to a value of 0.5 (or 50 as percentage) for the single middle value in the citation distribution. However, this is not always the case as Table 1 demonstrates. Another problem of the plotting positions concerns their interpretation: a usual definition of a PRₙ is that it represents the citation count at or below which ₙ percent of the papers falls. Four papers in Table 1 have zero citations and a plotting position of 9.52. Thus, one could assume that around 10% of the papers in the table have zero citations; however, there are around 20% of the papers with zero citations. Plotting positions have been initially proposed and are calculated for the comparison of two empirical distributions (or an empirical distribution with a theoretical distribution), but not for using them for relative assessments. However, the problem with the interpretation of plotting positions especially concerns small publication sets with only a few papers (fewer than 100 papers). Suppose that there are 100 papers in a set with different citation counts each. Then, the plotting positions correspond approximately with the percentage of papers at or below the citation impact of the
focal paper. For example, the paper with the 10\textsuperscript{th} rank position will have the plotting position 0.095.

Figure 1 shows a Q–Q plot of quantiles which have been calculated based on the rule proposed by Hazen (1914). For all papers published between 2000 and 2005 and assigned to six subject categories in WoS ($n = 6,070$ papers), the quantiles are shown resulting from the first and sixth subject category. Since the general trend of the Q–Q plot is on the line $y = x$ which follows the 45° line, the quantiles in both subject categories 1 and 6 are similar. Thus, most of the papers seem to have the same citation impact relative to other papers in the corresponding subject categories.

Figure 1. Q–Q plot of quantiles based on the rule proposed by Hazen (1914). For all papers published between 2000 and 2005 and assigned to six subject categories, the quantiles are shown resulting from the first and sixth subject category.

Several other rules have been proposed in the past which can be used instead of Hazen (1914) which lead, however, to similar plotting positions (Cox, 2005, discusses some rules).
5 Approaches based on size-frequency distributions

In recent years, some other approaches have been proposed for citation analyses which might lead to time- and field-normalized citation impact values. Since these approaches are not used for calculating plotting positions, they are explained in this section based on the size-frequency distribution (Egghe, 2005). This distribution shows the frequencies of papers with certain citation counts in a set of papers (see the first two columns in Table 2).

The column “InCites” in Table 2 refers to the approach used in the InCites tool which is a citation-based evaluation tool to analyze institutional performance provided by Clarivate Analytics (see https://clarivate.com/webofsciencegroup/solutions/incites). InCites percentiles are cumulative percentages of the size-frequency distribution starting with the percentage of papers with the highest citations. For example, 19.05% of the papers in Table 2 have 13 or more citations; 9.52% of the papers have at least 20 citations. Clarivate Analytics defines InCites percentiles as the “percentage of papers at each level of citation, i.e., the percentage of papers cited more often than the paper of interest” (see https://clarivate.libguides.com/incites_ba/alpha-indicators). Other than the Hazen approach, the InCites approach provides normalized values which can be interpreted as an exact percentage of papers. The problem with the InCites approach is that one does not immediately know, how good the paper is compared to the other papers in the set. Only the subtraction from 100, i.e. \((100 – x)\), reveals the percentage of papers performing worse than the focal paper.

Table 2. Example set of papers for calculating various time- and field-normalized values based on size-frequency distributions (based on the same 21 papers as in Table 1)

| Citation count | Number of papers | Rank \(k\) | InCites | Rank \(i\) | P100 | Rank \(j\) | P100' |
|----------------|------------------|------------|--------|-----------|------|------------|-------|
| 0              | 4                | 21         | 100.00 | 0         | 0.00 | 0          | 0.00  |
| 1              | 3                | 17         | 80.95  | 1         | 11.11| 4          | 21.05 |
| 2              | 1                | 14         | 66.67  | 2         | 22.22| 7          | 36.84 |
Bornmann, Leydesdorff, and Wang (2013) introduced the P100 approach which does not use the size-frequency distribution as the InCites approach, but the distribution of unique citation values (see Table 2). Thus, the frequencies of papers with certain citation counts are not considered. The formula is

$$\frac{i}{i_{\text{max}}} \times 100$$

whereby $i$ is the rank of the citation in the distribution of unique citation values (see Table 2). According to Bornmann and Mutz (2014) P100, however, “has undesirable properties which should be avoided … [for example,] the scale value of a paper can increase as a result of the fact that another paper receives an additional citation” (p. 1940). Another problem with this approach is similar to that of the Hazen approach: a P100 value does not refer to the citation count at or below which $x$ percent of the papers in the combination of publication year and subject category falls. Thus, Bornmann and Mutz (2014) introduced P100’ as an alternative to P100 which is calculated using the formula

$$\frac{j}{j_{\text{max}}} \times 100$$

whereby $j$ are ranks based on the size-frequency distribution (see Table 2). It is a decisive advantage of P100’ that the indicator always has the maximum value 100 and the minimum value 0. However, it remains the problem (as with the P100 and plotting position indicators) that it cannot be interpreted properly: P100’ = 21.05 does not mean that 21.05% of the papers in the publication set are below 1 citation (or equal to that citation count).

|   |   |   |   |   |   |   |
|---|---|---|---|---|---|---|
| 3 | 1 | 13 | 61.90 | 3 | 33.33 | 8 | 42.11 |
| 7 | 4 | 12 | 57.14 | 4 | 44.44 | 9 | 47.37 |
| 8 | 2 | 8 | 38.10 | 5 | 55.56 | 13 | 68.42 |
| 9 | 1 | 6 | 28.57 | 6 | 66.67 | 15 | 78.95 |
| 10 | 1 | 5 | 23.81 | 7 | 77.78 | 16 | 84.21 |
| 13 | 2 | 4 | 19.05 | 8 | 88.89 | 17 | 89.47 |
| 20 | 2 | 2 | 9.52 | 9 | 100.00 | 19 | 100.00 |
6 Citation percentiles

The problem of proper interpretations of percentiles can be solved by calculating cumulative frequencies in percentages (CPs) as demonstrated in Table 3 (and by the InCites approach). The table includes two variants: CP-IN is the cumulative percentage of the size-frequency distribution of papers. For CP-EX, the first possible percentage is set at 0 rather than considering its actual cumulative percentage. Then, the calculation of the cumulative percentage starts with the percentage of the lowest citation count. In this way, CP-EX reveals exactly the percentage of papers with lower citation impact: for example, CP-EX = 90.48 means that 90.48% of the papers in the set received a citation impact which is below 20 citations; 19.05% of the papers received less than one citation. CP-IN has a slightly other interpretation: 90.48 means that 90.48% of the papers in the set received a citation impact which is at or below 13 citations; 19.05% of the papers received zero citations.

Table 3. Cumulative percentages including (CP-IN) or excluding (CP-EX) the number of papers in the row (based on the same 21 papers as in the previous tables)

| Citation count | Number of papers | Percent (including papers in row) | Cumulative percentage (CP-IN) | Percent (excluding papers in row) | Cumulative percentage (CP-EX) |
|----------------|------------------|-----------------------------------|-------------------------------|-----------------------------------|-------------------------------|
| 0              | 4                | 19.05                             | 19.05                         | 0.00                              | 0.00                          |
| 1              | 3                | 14.29                             | 33.33                         | 19.05                             | 19.05                         |
| 2              | 1                | 4.76                              | 38.10                         | 14.29                             | 33.33                         |
| 3              | 1                | 4.76                              | 42.86                         | 4.76                              | 38.10                         |
| 7              | 4                | 19.05                             | 61.90                         | 4.76                              | 42.86                         |
| 8              | 2                | 9.52                              | 71.43                         | 19.05                             | 61.90                         |
| 9              | 1                | 4.76                              | 76.19                         | 9.52                              | 71.43                         |
| 10             | 1                | 4.76                              | 80.95                         | 4.76                              | 76.19                         |
| 13             | 2                | 9.52                              | 90.48                         | 4.76                              | 80.95                         |
| 20             | 2                | 9.52                              | 100.00                        | 9.52                              | 90.48                         |
| Total          | 21               | 100.00                            |                               |                                   |                               |

CP-IN and CP-EX have been calculated for all articles in the dataset of this study published between 2000 and 2005. Figure 2 (upper left side) shows the distribution of citation...
counts for the six years using boxplots. It is clearly visible that the distributions are very skewed and characterized by outliers (by a few highly cited articles).

Figure 2. Boxplots showing the distributions of citation counts, CP-INs, and CP-EXs (left side) as well as histograms of CP-INs and CP-EXs (right side) for articles published between 2000 and 2005.

The boxplots of CP-INs and CP-EXs in Figure 2 (left side) show that the distributions changed compared to citation counts: they are not characterized by outliers and are in the range between about 0 and 100. The median CP-IN is between 51 (in 2005) and 52 (in 2000); the median CP-EX is approximately 46.6 (in all publication years). Figure 2 (right side) shows histograms of CP-INs and CP-EXs for articles published between 2000 and 2005. Both figures reveal that the PRs are uniformly distributed between values from around 10 to 100. PRs below 10 are significantly less frequent than the other PR values. The reason is that PRs with low values are underrepresented, because the subject categories usually include a lot of papers with zero or only a few citations. The exception, however, is the PR zero which is the
most frequent value for CP-EX. Since the lowest citation impact value in every subject category receives the PR zero, this result can be expected for CP-EX.

PRs such as CP-IN and CP-EX can be calculated for all papers in every combination of subject category and publication year in databases such as WoS. Then, these papers receive time- and field-normalized citation impact values which can be used for various cross-time and cross-field comparisons (e.g., for the comparison of universities or countries). For these and other practical uses of PRs (CP-IN and CP-EX), however, three problems have to be solved.

6.1 First problem

The first problem is that the PRs which have been calculated for different combinations of publication years and subject categories are frequently not comparable. As the results in Table 2 demonstrate, we know the citation counts for the PRs 90.48 and 42.86, but we do not know the citation counts for the PRs 90 and 50. In other tables, the citation counts for other PRs might be available (e.g., for 82.34 and 23.45). However, for comparing the impact differences between two subject categories, it is necessary to know these citation counts for predefined PRs (e.g., 90 and 50). Barrett (2003) describes an approach which can be used to solve this problem of comparability. This approach requires that one takes into account the lower and upper bounds for every citation count in a citation distribution assuming that “each exact integer score is actually the middle score of an interval extending 0.5 either side” (p. 6). Furthermore, the observable list of citation counts between the minimum and maximum value in a publication set is complemented by the missing citation counts (with zero numbers of papers).

Table 4 includes CP-IN and shows the same dataset as in the previous tables but considers the lower and upper bounds for each citation count. Furthermore, the missing citation counts (between two observable citation counts) with corresponding zero numbers of
papers are added. Thus, a complete list of citation counts beginning with the minimum and ending with the maximum citation counts from the initial publication set with the corresponding numbers of papers is available now. In Table 4, the observed citation counts from the previous tables are expanded to citation-intervals with equal sizes (i.e., 1 citation).

Table 4. Expanding the set of papers for calculating “theoretical” citation counts for certain PRs (CP-IN, based on the same 21 papers as in the previous tables)

| Citation count | Lower and upper bounds | Number of papers | Cumulative frequencies | Percent | CP-IN |
|----------------|-------------------------|------------------|-----------------------|---------|-------|
| 0              | -0.5 to 0.5             | 4                | 4                     | 19.05   | 19.05 |
| 1              | 0.5 to 1.5              | 3                | 7                     | 14.29   | 33.33 |
| 2              | 1.5 to 2.5              | 1                | 8                     | 4.76    | 38.10 |
| 3              | 2.5 to 3.5              | 1                | 9                     | 4.76    | 42.86 |
| 4              | 3.5 to 4.5              | 0                | 9                     | 0.00    | 42.86 |
| 5              | 4.5 to 5.5              | 0                | 9                     | 0.00    | 42.86 |
| 6              | 5.5 to 6.5              | 0                | 9                     | 0.00    | 42.86 |
| 7              | 6.5 to 7.5              | 4                | 13                    | 19.05   | 61.90 |
| 8              | 7.5 to 8.5              | 2                | 15                    | 9.52    | 71.43 |
| 9              | 8.5 to 9.5              | 1                | 16                    | 4.76    | 76.19 |
| 10             | 9.5 to 10.5             | 1                | 17                    | 4.76    | 80.95 |
| 11             | 10.5 to 11.5            | 0                | 17                    | 0.00    | 80.95 |
| 12             | 11.5 to 12.5            | 0                | 17                    | 0.00    | 80.95 |
| 13             | 12.5 to 13.5            | 2                | 19                    | 9.52    | 90.48 |
| 14             | 13.5 to 14.5            | 0                | 19                    | 0.00    | 90.48 |
| 15             | 14.5 to 15.5            | 0                | 19                    | 0.00    | 90.48 |
| 16             | 15.5 to 16.5            | 0                | 19                    | 0.00    | 90.48 |
| 17             | 16.5 to 17.5            | 0                | 19                    | 0.00    | 90.48 |
| 18             | 17.5 to 18.5            | 0                | 19                    | 0.00    | 90.48 |
| 19             | 18.5 to 19.5            | 0                | 19                    | 0.00    | 90.48 |
| 20             | 19.5 to 20.5            | 2                | 21                    | 9.52    | 100.00|

In the table, each exact citation count is the middle value of an interval extending 0.5 either side. Using the formula

\[ CC_i = lb + \left(\frac{np - cf}{f_i}\right) \times w \]

the citation count for a certain PR \( i \) can be calculated, where

\( CC_i \) = citation count for the \( i^{th} \) PR

\( lb \) = the exact lower bound of the interval containing the PR
\[ n = \text{the total number of papers in the publication set} \]
\[ p = \text{the proportion corresponding to the desired PR (between 0 and 1, instead of 0 and 100)} \]
\[ cf = \text{the cumulative frequency of papers in the interval containing the PR} \]
\[ f_i = \text{the frequency of papers in the interval containing the } i^{\text{th}} \text{ PR} \]
\[ w = \text{the width of the class interval} \]

The resulting citation counts for pre-specified PRs are “estimates of hypothetical real-valued continuous numbers” (Barrett, 2003, p. 9). These estimates can be calculated for PRs which are between the minimum and maximum CP-IN in Table 4 (i.e., 19.05 and 100). Thus, the citation count for the 4\(^{th}\) PR cannot be estimated.

Using the frequency distribution in Table 4, the citation counts for the 90\(^{th}\), 75\(^{th}\), and 50\(^{th}\) PRs are exemplarily calculated in the following. Let us start with the 90\(^{th}\) PR. Looking at CP-IN in the table, one can see that the 90\(^{th}\) PR is positioned between 12 citations (CP-IN = 80.95) and 13 citations (CP-IN = 90.48). We can expect that the 90\(^{th}\) PR is in the interval between 12.5 and 13.5 citations; 13 citations refer to the PR 90.48 which is close to 90. The lower bound of the interval is 12.5, i.e. \(lb = 12.5\). We are interested in the 90\(^{th}\) PR, thus \(p = 0.9\), and there are 21 papers in the publication set \((n = 21)\). The width of the class interval is 1 citation \((w = 1)\). The frequency and cumulative frequency of papers containing the PR is \(f_i = 2\) and \(cf = 17\). Filling these values in the formula lead to an estimate of 13.45 citations. Since 13 citations correspond to the PR 90.48 and the upper bound of the interval containing this PR is 13.5, 13.45 citations seems to be a realistic value.

The calculation of the estimated citation count for the 75\(^{th}\) PR is similar. This PR is between 8 and 9 citations in Table 4. Thus, \(lb = 8.5\). The interval, in which 8.5 is the lower bound, refers to the PR 76.19. The other values for the formula are \(n = 21, f_i = 1, cf = 15, w = 1, \text{ and } p = 0.75\). The estimated result for the 75\(^{th}\) PR is 9.25 which is close to the upper bound of the interval (9.5). This upper bound is from the PR 76.19 which is somewhat higher than
75. The last example is the 50th PR which is between 7 and 8 citations. Thus, the values for the formula are $lb = 6.5$, $n = 21$, $fi = 4$, $cf = 9$, $w = 1$, and $p = 0.5$. The estimated citation count is 6.875 – a realistic value with 7.5 as upper bound for the PRs 61.9 and 5.5 as lower bound for the PR 42.86.

Using the formula

$$PR_x = \left[ \frac{(cf + (\frac{x - lb}{w}) * fi)}{n} \right] * 100$$

and the values as specified above, the corresponding PRs ($PR_x$) can be calculated. The only new parameter is $x$. This is the citation count for which the PR is calculated. Filling in $ll = 6.5$, $n = 21$, $fi = 4$, $cf = 9$, $w = 1$, and $x = 6.875$, we receive the PR 50. The formula can be used to determine (estimate) the different PRs for the same citation count in different combinations of publication year and subject category. Thus, differences in citation counts between two combinations of publication years and subject categories can be made visible based on exact estimations: what is the PR for the same citation count in two different combinations of publication years and subject categories?

The same calculations as with CP-IN can be done with CP-EX but with slightly different formulas. Table 5 shows the same dataset as in Table 4 but it includes CP-EX instead of CP-IN.

| Citation count | Lower and upper bounds | Number of papers | Cumulative frequencies | Percent | CP-EX |
|----------------|------------------------|------------------|-----------------------|---------|-------|
| 0              | -0.5 to 0.5            | 4                | 0                     | 19.05   | 0.00  |
| 1              | 0.5 to 1.5             | 3                | 4                     | 14.29   | 19.05 |
| 2              | 1.5 to 2.5             | 1                | 7                     | 4.76    | 33.33 |
| 3              | 2.5 to 3.5             | 1                | 8                     | 4.76    | 38.10 |
The formula for calculating the citation count for PR \( i \) is

\[
CC_i = lb + 1 + \left( \frac{np - cf}{f_i} \right) * w
\]

where

- \( CC_i \) = citation count for the \( i \)th PR
- \( lb \) = the exact lower bound of the selected interval containing the score \( x \)
- \( n \) = the total number of papers in the publication set
- \( p \) = the proportion corresponding to the desired PR (between 0 and 1, instead of 0 and 100)
- \( cf \) = the cumulative frequency of papers in the selected interval
- \( f_i \) = the frequency of papers in the selected interval
- \( w \) = the width of the class interval

Based on Table 5, the citation counts for the 90th PR is exemplarily calculated. This PR is positioned between 13 citations (CP-EX = 80.95) and 14 citations (CP-EX = 90.48). We select the interval in the table with the exact upper bound value 13.5. The lower bound of the interval is 12.5, i.e. \( lb = 12.5 \). The other values are \( p = 0.9 \), \( n = 21 \), \( w = 1 \), \( f_i = 2 \), and \( cf = 17 \).
Filling these values in the formula lead to an estimate of 14.45 citations. Since 14 citations corresponds to the PR 90.48, which is very close to the 90th PR, and the upper bound of the interval containing this PR is 14.5, 14.45 citations seems to be reasonable.

Using the formula

$$PR_x = \left[ \frac{(cf + \frac{x - lb - 1}{w}) \times f_i}{n} \right] \times 100$$

and the values from above, the corresponding PR ($PR_x$) can be calculated. Filling in $ll = 6.5$, $n = 21$, $f_i = 4$, cf = 9, $w = 1$, and $x = 7.875$ in the formula, we receive the PR 50.

6.2 Second problem

Using two formulas proposed by Barrett (2003) as well as two modified formulas, the first problem – missing comparability of PRs resulting from calculations for different combinations of publication year and subject category – can be solved. Let us go on with the second problem. Many papers in the WoS and Scopus databases (and in other databases) are assigned to more than one subject category. Between 2000 and 2005, the 4,416,554 articles in the dataset of this study are assigned to up to six subject categories (which leads to 6,973,937 articles including multiple occurrences).

Table 6. Frequency and percentage of articles with different numbers of subject categories

| Number of subject categories | Frequency  | Percent | Cumulative frequencies |
|-----------------------------|------------|---------|------------------------|
| 1                           | 2,583,330  | 58.49   | 58.49                  |
| 2                           | 1,257,957  | 28.48   | 86.97                  |
| 3                           | 456,562    | 10.34   | 97.31                  |
| 4                           | 94,588     | 2.14    | 99.45                  |
| 5                           | 18,047     | 0.41    | 99.86                  |
| 6                           | 6,070      | 0.14    | 100.00                 |
| Total                       | 4,416,554  | 100.00  |                        |
Table 6 shows that 58.49% of the articles are assigned to one subject category. For around 40% of the articles, more than one PR is calculated, and it is not clear how these PRs can be aggregated into one value for a single article. For the InCites tool, the minimum value is used: “the category in which the percentile value is closest to zero is used, i.e. the best performing value” (see https://clarivate.libguides.com/incites_ba/understanding-indicators). This approach, however, leads to an overestimation of performance if papers are assigned to more than one subject category and the resulting PRs are (very) different. For example, 1,833,224 articles in the dataset of this study are assigned to at least two subject categories. The minimum of the differences between two PRs of an article is 0, the maximum is 82.9; the mean difference is 7.84 and the median is 5.77.

Another solution for the aggregation into one value could be the median of the PRs. A challenge to this calculation is that subject categories have different numbers of papers. In the dataset of this study, there are 1528 different combinations of subject category and publication year, with a minimum number of papers of 1 and a maximum number of 43,456 (mean = 4564.1). We can assume that the same PR has a higher value in a combination of publication year and subject category with many papers than in a combination with only a few papers. Suppose a paper is assigned to two subject categories with 10 and 3000 papers. To be at the 50th PR in these subject categories would have very different meanings: in one case, the paper would be better than five other papers and in the other case, it would be better than 1500 other papers. A good estimation of an aggregated PR might be to find the median of PRs where each subject category’s PR is reflected several times proportional to the number of papers in that subject category. Thus, the solution might be the median PR which is weighted by the number of papers in the subject categories. This weighted median can be calculated using the formula

$$\frac{(PR_{SC1} \times n_{SC1}) + (PR_{SC2} \times n_{SC2}) + \ldots + (PR_{SCx} \times n_{SCx})}{n_{SC1} + n_{SC2} + \ldots + n_{SCx}}$$
whereby $PR$ is the PR, the subscripts $SC1$ to $SCx$ represent the different subject categories to which a paper is assigned and $n$ is the total number of papers in these subject categories.

Table 7. Size-frequency distribution of papers in subject category 1 with two focal papers

| Paper  | Citation count | Number of papers | Rank $i$ | CP-EX |
|--------|----------------|------------------|----------|-------|
| Paper A| 100            | 1                | 67       | 98.53 |
|        | 98             | 3                | 64       | 94.12 |
|        | 90             | 1                | 63       | 92.65 |
|        | 88             | 1                | 62       | 91.18 |
|        | 40             | 1                | 61       | 89.71 |
| Paper B| 20             | 5                | 56       | 82.35 |
|        | 8              | 7                | 49       | 72.06 |
|        | 7              | 5                | 44       | 64.71 |
|        | 6              | 9                | 35       | 51.47 |
|        | 4              | 13               | 22       | 32.35 |
|        | 1              | 22               | 0        | 0.00  |
| Total  |                |                  | 68       |       |

Table 8. Size-frequency distribution of papers in subject category 2 with two focal papers

| Paper  | Citation count | Number of papers | Rank $i$ | CP-EX |
|--------|----------------|------------------|----------|-------|
| Paper A| 1100           | 1                | 410      | 99.76 |
|        | 980            | 1                | 409      | 99.51 |
|        | 465            | 3                | 406      | 98.78 |
|        | 200            | 5                | 401      | 97.57 |
|        | 145            | 7                | 394      | 95.86 |
|        | 120            | 19               | 375      | 91.24 |
| Paper A| 100            | 2                | 373      | 90.75 |
|        | 90             | 4                | 369      | 89.78 |
|        | 67             | 6                | 363      | 88.32 |
|        | 55             | 5                | 358      | 87.10 |
|        | 34             | 8                | 350      | 85.16 |
|        | 34             | 1                | 349      | 84.91 |
|        | 23             | 4                | 345      | 83.94 |
|        | 23             | 5                | 340      | 82.73 |
|        | 21             | 12               | 328      | 79.81 |
| Paper B| 20             | 13               | 315      | 76.64 |
|        | 11             | 15               | 300      | 72.99 |
Table 7 and Table 8 show the size-frequency distributions (CP-EX) of two different subject categories. Two papers A and B are assigned to both subject categories. For example, paper A has 100 citations which means that the paper has CP-EX = 98.53 in subject category 1 and CP-EX = 90.75 in subject category 2. The calculation of a median PR with 98.53 + 90.75 = 94.64 gives the PR from subject category 1 too much weight, since this subject category has significantly fewer papers than subject category 2. Thus, the weighted PR median is 91.86 \[\frac{(98.53 \times 68) + (90.75 \times 411)}{(68 + 411)}\] which is lower than the unweighted PR median. The weighted median of paper B is 77.45 \[\frac{(82.35 \times 68) + (76.64 \times 411)}{(68 + 411)}\] which is also lower than the unweighted PR with 79.5.

### 6.3 Third problem

The third problem which must be solved with the use of PRs is their aggregation if an ‘average’ PR is desired for a certain publication set (e.g., of a researcher, university or country). Suppose one is interested in an ‘average’ PR of paper A (CP-EX = 91.86) and paper B (CP-EX = 77.45) from Table 7 and Table 8, since these are two papers from a certain unit. In these cases, where the aggregation of PRs is desired for a unit, a similar formula can be used as for single papers with different subject categories, whereby single papers with several subject categories are considered multiple times. This weighted median for single units in science can be calculated using the formula
whereby $PR_1^{ws}$ to $PR_y^{ws}$ are the sums of the subject-specific PRs for paper 1 to paper $y$ published by the unit which are weighted each by the publication numbers in the corresponding subject category (denoted by the superscript $ws$). $SC_1$ to $SC_y$ is the number of subject categories to which a paper has been assigned, and $n_1^s$ to $n_y^s$ is the sum (superscript $s$) of the number of papers ($n$) in the subject categories, paper 1 to paper $y$ has been assigned to. Thus, for the two papers A and B, the average PR is 84.65 $\left[ \frac{\left( (98.53 \times 68) + (90.75 \times 411) \right)}{2} + \left( (82.35 \times 68) + (76.64 \times 411) \right)/2 \right] / \left( \frac{68 + 411}{2} + (68 + 411)/2 \right)]$.

The formula

$$\left( \frac{PR_1^{ws}}{SC_1} + \frac{PR_2^{ws}}{SC_2} + \cdots + \frac{PR_y^{ws}}{SC_y} \right) \frac{n_1^s}{SC_1} + \frac{n_2^s}{SC_2} + \cdots + \frac{n_y^s}{SC_y}$$

extends the calculation by another weight besides the number of subject categories: if papers have been published by more than one unit (e.g., researchers, institutions or countries), the papers should be fractionally assigned to these units. The fractional assignment is included by the notation $FR$ for paper 1 to paper $y$.

7 Presenting percentile ranks (PRs)

The publication set used in this study consists of 4,416,554 articles published between 2000 and 2005. In this section, based on this dataset, some graphs are presented how PRs can
be presented for enabling meaningful interpretations of empirical findings. Authors’ country information has been added to the articles in the dataset of this study to enable exemplary analyses and comparisons of units. Instead of countries, any other units could have been used (e.g., researchers or institutions).

Figure 3. Weighted PRs (CP-IN and CP-EX) for articles published between 2000 and 2005 by six countries: USA \((n = 1,395,809)\), Switzerland \((n = 73,344)\), UK \((n = 370,227)\), Germany \((n = 343,571)\), Japan \((n = 372,581)\), and China \((n = 255,379)\).

The first proposal for presenting PRs is to show weighted median PRs as bar graphs. Using the formula in section 6.3, the weighted median PR has been calculated for some countries (by considering three weights: number of papers in a subject category, number of subject categories, and number of countries). The results are shown in Figure 3: the USA is the best performing country in this group with a weighted median PR of 58.68 (CP-IN) and 56.65 (CP-EX). Thus, the articles from this country performed (equal to or) better than about 57% (59%) of the articles published in the same publication year and subject category. China
achieved a weighted median PR of 43.54 (CP-IN) and 40.2 (CP-EX) which is below the expected value for an ‘average’ citation level (50). Thus, China’s performance is better than around 40% of the articles published in the same publication year and subject category.

The second proposal for presenting PRs integrate the distribution of PRs besides mean PRs. Bornmann and Marx (2014a, 2014b) and more recently Bornmann and Haunschild (2018) proposed to visualize percentiles using beamplots (Doane & Tracy, 2000). Recently, the proposal has been taken up by Adams, McVeigh, Pendlebury, and Szomszor (2019) – members of the Institute for Scientific Information (ISI) which is part of Clarivate Analytics: the authors recommend to use beamplots instead of single numbers aggregating percentile distributions (such as in Figure 3). It is an advantage of beamplots that not only annual distributions of percentiles can be presented, but also summary statistics (annual and overall medians). However, it should be considered in the use of beamplots that they are especially suitable for small publication sets (i.e., publication sets of single researchers), since beamplots become unreadable for large sets with many publications.

Figure 4 shows the beamplot of CP-EXs for articles published between 2000 and 2005 by Libyan authors. Libyan has been selected for this analysis, since there are only 294 articles in the publication set of this study from this country. This number is in that range of paper numbers published by single authors. In the figure, the PR of every individual article is visualized using grey diamonds; the annual weighted median PRs (by considering three weights: number of papers in a subject category, number of subject categories, and number of countries; see section 6.3) are displayed with black triangles. The vertical black line in Figure 4 shows the weighted median PR (CP-EX = 27.94) across all articles published between 2000 and 2005 by Libyan authors (by considering the three weights mentioned above). The grey dashed line in the figure marks the value 50 – the expected value for an ‘average’ citation level.
As the weighted median PR demonstrates, Libyan authors achieved a citation impact which is significantly below 50: on average, only 27.94% of the articles published between 2000 and 2005 in the corresponding subject categories and publication years received a citation impact which is below the impact of the Libyan authors’ articles. The distributions of
the PRs in all publication years demonstrate that the PRs are especially concentrated in the low citation impact area (below the 20th PR). Highly cited papers – papers which belong to the 10% most frequently cited articles – exist, but they are rare.

As Figure 3 and Figure 4 demonstrate, percentiles can be visualized very differently. Further possibilities of presenting and statistically analyzing percentiles can be found in Bornmann (2013) and Williams and Bornmann (2014).

8 Discussion

In a recent study on landmark publications in scientometrics, Tahamtan and Bornmann (2018) worked out that the first method used for time- and field-normalizing citation data was based on percentiles. The introduction of the percentile method to bibliometrics is associated with the name Francis Narin (retired president of CHI Research Inc.). Already at the beginning of the 1980s, McAllister, Narin, and Corrigan (1983) explained citation percentiles as follows: “the pth percentile of a distribution is defined as the number of citations \(X_p\) such that the percent of papers receiving \(X_p\) or fewer citations is equal to \(p\). Since citation distributions are discrete, the pth percentile is defined only for certain \(p\) that occur in the particular distribution of interest” (p. 207). Evered, Hamett, and Narin (1989) used the percentage of papers belonging to the 10% most frequently cited papers (named as ‘top decile citation performance’) to evaluate the citation impact of various institutional units. About 30 years later, Hicks, Wouters, Waltman, de Rijcke, and Rafols (2015) published ten principles to guide research evaluation (using bibliometrics data) in *Nature*. According to these authors, “normalized indicators are required, and the most robust normalization method is based on percentiles: each paper is weighted on the basis of the percentile to which it belongs in the citation distribution of its field (the top 1%, 10% or 20%, for example)” (p. 430).

In this study, various approaches have been presented for using percentiles in research evaluation. A very popular approach today is to present the percentage of papers for a unit
(e.g., an institution) which belong to the 10% most frequently cited papers: PP(top 10%) (see Bornmann, 2014). This indicator comes under the family of I3 indicators which combine the number of papers in different PR classes with specific weights. P(top 10%) counts the number of papers belonging to the PR class of the 10% most frequently cited papers with a weight of 1. Other I3 indicators have used up to six PR classes to measure the citation performance of units with various weights. In section 3, it has been argued that I3 has the disadvantage that information of citation distributions is lost when the data are grouped into (PR) classes. Approaches which consider the complete distribution of data should be preferred. Bornmann, Leydesdorff, and Mutz (2013) proposed to use plotting positions which apply the rule by Hazen (1914) to consider the complete citation distributions. The problem with this approach is, however, that plotting positions cannot always be (exactly) interpreted as the percentage of papers (at or) below a certain citation count (especially when the plotting positions are calculated based on only a few papers).

In recent years, some other percentile approaches have been introduced based on size-frequency distributions with varying advantages and disadvantages – as outlined in section 5. In this study, two further approaches (CP-IN and CP-EX) are explained which are oriented towards the usual percentile rank definition: PR \( x \) is defined as the citation count (at or) below which \( x\% \) of the papers in the combination of publication year and subject category falls. Both approaches can be used very flexible by computing (1) PRs for observed citation counts in distributions and (2) estimated citation counts for pre-defined PRs (e.g., the 90\(^{th}\) PR). It is one problem for the use of PRs in citation analyses that papers in databases such as WoS are frequently assigned to more than one subject category. This problem has been solved by weighting median PRs with the number of papers in corresponding subject categories. Other problems with PRs concern their aggregation: how should PRs for papers of various units (e.g., institutions) be aggregated? In section 6, it has been proposed to use median PRs whereby weights are applied to consider that a paper has been assigned to more than one
subject category and/or has been published by more than one unit (e.g., more than one country).

Section 7 addresses the presentation of PRs: bar graphs and beamplots can be used to present the results for various units. These graph types are only two examples and other options for visualizing results based on PRs exist. Some other options can be found in Bornmann (2013) and Williams and Bornmann (2014). Since the percentile approach has significant advantages against other time- and field-normalizing approaches especially those which are based on mean citation rates, it would be desirable that the percentile approach is more frequently used in bibliometrics studies. Although I3 indicators such as the popular PP(top 10%) are robust indicators compared to other time- and field-normalized indicators, they have the important disadvantage that information is lost from the citation distribution. Since CP-IN and CP-EX consider the complete citation distribution, they should be preferred in bibliometric studies.
Acknowledgements

The bibliometric data used in this paper are from an in-house database developed and maintained by the Max Planck Digital Library (MPDL, Munich) and derived from the Science Citation Index Expanded (SCI-E), Social Sciences Citation Index (SSCI), Arts and Humanities Citation Index (AHCI) prepared by Clarivate Analytics (Philadelphia, Pennsylvania, USA). We would like to thank Paul Barrett for his support in writing this manuscript.
References

Adams, J., McVeigh, M., Pendlebury, D., & Szomszor, M. (2019). Profiles, not metrics. Philadelphia, PA, USA: Clarivate Analytics.
Barrett, P. (2003). Percentiles and textbook definitions – confused or what? Retrieved November 11, 2019, from https://www.pbarrett.net/techpapers/percentiles.pdf
Bornmann, L. (2013). How to analyze percentile citation impact data meaningfully in bibliometrics: The statistical analysis of distributions, percentile rank classes, and top-cited papers. Journal of the American Society for Information Science and Technology, 64(3), 587-595. doi: 10.1002/asi.22792.
Bornmann, L. (2014). How are excellent (highly cited) papers defined in bibliometrics? A quantitative analysis of the literature. Research Evaluation, 23(2), 166-173. doi: 10.1093/reseval/rvu002.
Bornmann, L., de Moya Anegón, F., & Leydesdorff, L. (2012). The new Excellence Indicator in the World Report of the SCImago Institutions Rankings 2011. Journal of Informetrics, 6(2), 333-335. doi: 10.1016/j.joi.2011.11.006.
Bornmann, L., & Haunschild, R. (2018). Plots for visualizing paper impact and journal impact of single researchers in a single graph. Scientometrics, 115(1), 385-394. doi: 10.1007/s11192-018-2658-1.
Bornmann, L., Leydesdorff, L., & Mutz, R. (2013). The use of percentiles and percentile rank classes in the analysis of bibliometric data: Opportunities and limits. Journal of Informetrics, 7(1), 158-165. doi: 10.1016/j.joi.2012.10.001.
Bornmann, L., Leydesdorff, L., & Wang, J. (2013). Which percentile-based approach should be preferred for calculating normalized citation impact values? An empirical comparison of five approaches including a newly developed citation-ranking approach (P100). Journal of Informetrics, 7(4), 933-944. doi: 10.1016/j.joi.2013.09.003.
Bornmann, L., & Marewski, J. N. (2019). Heuristics as conceptual lens for understanding and studying the usage of bibliometrics in research evaluation. Scientometrics, 120(2), 419–459.
Bornmann, L., & Marx, W. (2014a). Distributions instead of single numbers: Percentiles and beam plots for the assessment of single researchers. Journal of the American Society of Information Science and Technology, 65(1), 206–208.
Bornmann, L., & Marx, W. (2014b). How to evaluate individual researchers working in the natural and life sciences meaningfully? A proposal of methods based on percentiles of citations. Scientometrics, 98(1), 487-509. doi: 10.1007/s11192-013-1161-y.
Bornmann, L., & Marx, W. (2015). Methods for the generation of normalized citation impact scores in bibliometrics: Which method best reflects the judgements of experts? Journal of Informetrics, 9(2), 408-418.
Bornmann, L., & Mutz, R. (2014). From P100 to P100': A new citation-rank approach. Journal of the Association For Information Science and Technology, 65(9), 1939-1943. doi: 10.1002/asi.23152.
Cox, N. J. (2005). Calculating percentile ranks or plotting positions. Retrieved May 30, 2019, from http://www.stata.com/support/faqs/stat/pcrank.html
Doane, D. P., & Tracy, R. L. (2000). Using beam and fulcrum displays to explore data. American Statistician, 54(4), 289-290. doi: 10.2307/2685780.
Egghe, L. (2005). Power laws in the information production process: Lotkaian informetrics. Kidlington, UK: Elsevier Academic Press.
Evered, D., Hamett, S., & Narin, F. (1989). The impact of different modes of research funding. In D. Evered & S. Hamett (Eds.), The evaluation of scientific research (pp. 120-140). Chichester, UK: John Wiley and Sons.
Hazen, A. (1914). Storage to be provided in impounding reservoirs for municipal water supply. *Transactions of American Society of Civil Engineers, 77*, 1539-1640.

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

Lavrakas, P. J. (Ed.). (2008). *Encyclopedia of survey research methods*. Thousand Oaks, CA, USA: Sage.

Leydesdorff, L., & Bornmann, L. (2011). Integrated impact indicators (I3) compared with impact factors (IFs): An alternative research design with policy implications. *Journal of the American Society of Information Science and Technology, 62*(11), 2133-2146.

Leydesdorff, L., & Bornmann, L. (2012). Percentile ranks and the integrated impact indicator (I3). *Journal of the American Society for Information Science and Technology, 63*(9), 1901-1902. doi: 10.1002/asi.22641.

Leydesdorff, L., Bornmann, L., & Adams, J. (2019). The integrated impact indicator (I3) revisited: A non-parametric alternative to the journal impact factor. *Scientometrics, 119*(3), 1669-1694.

McAllister, P. R., Narin, F., & Corrigan, J. G. (1983). Programmatic evaluation and comparison based on standardized citation scores. *IEEE Transactions on Engineering Management, 30*(4), 205-211.

Narin, F. (1987). Bibliometric techniques in the evaluation of research programs. *Science and Public Policy, 14*(2), 99-106. doi: 10.1093/spp/14.2.99.

Schubert, A., & Braun, T. (1986). Relative indicators and relational charts for comparative assessment of publication output and citation impact. *Scientometrics, 9*(5-6), 281-291.

Tahamtan, I., & Bornmann, L. (2018). Creativity in science and the link to cited references: Is the creative potential of papers reflected in their cited references? *Journal of Informetrics, 12*(3), 906-930.

Waltman, L., Calero-Medina, C., Kosten, J., Noyons, E. C. M., Tijssen, R. J. W., van Eck, N. J., . . . Wouters, P. (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.

Williams, R., & Bornmann, L. (2014). The substantive and practical significance of citation impact differences between institutions: Guidelines for the analysis of percentiles using effect sizes and confidence intervals. In Y. Ding, R. Rousseau & D. Wolfram (Eds.), *Measuring scholarly impact: Methods and practice* (pp. 259-281). Heidelberg, Germany: Springer.