Journal of Informetrics 6 (2012) 712–720

Contents lists available at SciVerse ScienceDirect

Journal of Informetrics

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

Comparison of the citation distribution and *h*-index between groups of different sizes

Maja Pohar Perme[a,*], Janez Stare[a], Rok Žaucer[b], Matjaž Žaucer[c]

[a] Institute for Biostatistics and Medical Informatics, Faculty of Medicine, University of Ljubljana, Slovenia
[b] University of Nova Gorica, Slovenia
[c] Central Technological Library, University of Ljubljana, Slovenia

ARTICLE INFO

Article history:
Received 23 March 2012
Received in revised form 18 July 2012
Accepted 19 July 2012

Keywords:
Research performance
Citation distribution
Reduced Hirsch index
Size-independent comparison
Stochastic method
Probability formulae

ABSTRACT

Evaluating the performance of institutions with different resources is not easy, any citation distribution comparisons are strongly affected by the differences in the number of articles published. The paper introduces a method for comparing citation distributions of research groups that differ in size. The citation distribution of a larger group is reduced by a certain factor and compared with the original distribution of a smaller group. Expected values and tolerance intervals of the reduced set of citations are calculated. A comparison of both distributions can be conveniently viewed in a graph. The size-independent reduced Hirsch index – a function of reducing factor that allows the comparison of groups within a scientific field – is calculated in the same way. The method can be used for comparing groups or units differing in full-time equivalent, funding or the number of researchers, for comparing countries by population, gross domestic product, etc. It is shown that for the calculation of the reduced Hirsch index, the upper part of the original citation distribution is sufficient. The method is illustrated through several case comparisons.

1. Introduction

The evaluation of individuals, teams, institutions and even countries is an important part of scientometric studies. Recognizing and comparing science performance is particularly important for governments, funding agencies and managers of research institutions. Peer reviews, the number of published documents, the quality of journals in which the documents are published and the number of citations received by these documents are standard indicators in the evaluation of the productivity and visibility of research work. Although many studies found only a weak positive correlation between peer reviews and bibliometric indicators (Aksnes & Taxt, 2004; Južnič et al., 2010), and although the use of citations as performance measure is controversial and seriously debated (Coryn, 2005), citation analysis is typically still the starting point of research evaluations.

In search for financial support for their work, researchers adapt their publication practice to funding agencies' policies, and so citation distributions may reflect the funding policies: if productivity is stimulated more than quality, authors tend to spread their research results over more documents, which are prone to lower citation rates (Butler, 2003).[1] This presents an important motivation for the comparison of citation distributions. However, due to the skewness of citation distributions, it is difficult to compare individual, team, institution or country performance if their research was backed by different resources.

* Corresponding author at: Vrazov trg 2, 1000 Ljubljana, Slovenia. Tel.: +386 1 543 77 85; fax: +386 1 543 77 71.
E-mail address: maja.pohar@mf.uni-lj.si (M.P. Perme).

[1] Compare also the effect of the A1 score in Slovenian Research Agency's "Rules on the Procedures of the (co)financing and Monitoring of Research Activities Implementation", http://www.arrs.gov.si/en/akti/prav-sof-ocen-sprem-razisk-dej-260111.asp.

1751-1577/$ – see front matter
http://dx.doi.org/10.1016/j.joi.2012.07.007

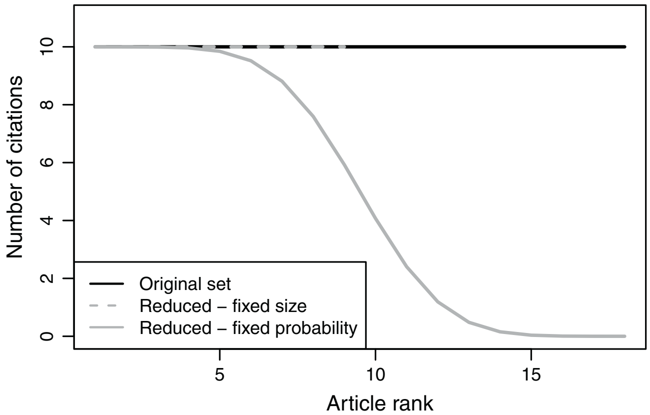

Fig. 1. A motivating example: the citation distribution of the original set versus the expected distributions on a reduced set following two different assumptions.

The distribution strongly depends on the number of articles, therefore any comparison is masked by the difference in the total number of articles.

There are methods for evaluating citation distributions by dividing them into subgroups of articles with different numbers of citations, such as uncited papers, poorly cited papers, fairly cited papers, remarkably cited papers and outstandingly cited papers (Schubert & Braun, 1986). Another method measures low and high impact in citation distribution (Albarrán, Ortuño, & Ruiz-Castillo, 2011). For comparisons of sets of different sizes with one another, the percentile-ranks approach has been proposed (Leydesdorff, Bornmann, Mutz, & Opthof, 2011). Recently, a new method for comparisons between research units of different sizes and fields was proposed (Crespo, Ortuño, & Ruiz-Castillo, 2011). It assesses the merit of any set of scientific papers in a given field with the probability that a randomly drawn sample of articles from a reference set would have a lower citation index.

Because of the highly skewed citation distributions (Tijssen, Visser, & Van Leeuwen, 2002), the simplest measures – average number of citations per document and total number of citations – have poor statistical properties. Consequently, a large number of other citation-based bibliometric indicators have been developed.

A widely accepted indicator of scientific performance is the h-index (Hirsch, 2005), which offers a simple measure of quantity and visibility and is easy to calculate, although it also has many disadvantages. For instance, large research institutions producing a large number of documents tend to have a larger h-index than smaller institutions, since the h-index also depends on the number of documents considered. Many properties of the h-index have been studied (Burrell, 2007; Egghe, 2008a, 2008b; Egghe & Rousseau, 2006; Glänzel, 2006). To overcome the drawbacks of h-index many of its variants have been proposed (Alonso, Cabrerizo, Herrera-Viedma, & Herrera, 2009; Egghe, 2010). Among them g-index is most popular, as it gives more weight to highly cited papers in contrast to the h-index and so it has greater discriminatory power (Egghe, 2006a, 2006b, 2006c). Egghe and Rousseau (2006) studied different h-indices for groups of authors. Bornmann, Mutz, Hug, and Daniel (2011) conducted a multilevel meta-analysis of studies reporting correlations between the h-index and 37 different variants of it. Despite this, such h-index variants are rarely used. A single measure cannot capture the complete information on the citation distribution over documents (Bornmann, Mutz, & Daniel, 2010).

For comparing the productivity and visibility of institutions of disparate size, a size-corrected index has been proposed based on the decomposition of the h-index into the product of an impact index and a factor that depends on the population size (Molinari & Molinari, 2008). Performance evaluations would greatly benefit from a general method of citation distribution reduction without empirical parameters and not limited to a specific citation distribution shape.

The goal of our research was to develop a method that will make it possible to compare citation distributions of disparately sized groups of authors. We calculated the expected values of a reduced set of citations and compared them with the original citation distribution of a smaller group. Using similar methodology, we obtained the expected value of a reduced (diminished) h_r-index for a certain reducing factor and compared it with the original h value of a smaller group.

2. Theory

2.1. A motivating example

Suppose a group has published n articles. Let c_i denote the number of citations of the ith article, $i = 1, \ldots, n$, and let the articles be ordered with respect to the number of citations, so that $c_1 \geq c_2 \geq \cdots \geq c_n$. Fig. 1 presents the citation curve joining the points (i, c_i) for a group A with $n = 18$ articles, all cited 10 times ($c_i = 10$ for each i): for each article rank i, the number of citations c_i is plotted and the values are joined in a curve. The h-index of this group is 10. Say we want to compare them to a research group B that is only half the size. Any comparison of the number of citations is strongly affected by the number of articles produced in the smaller group, for example, if the smaller group produced 9 articles that were each cited 20 times, their h-index could not exceed 9.

In order to make a comparison independent of sample size, we calculated the expected citation curve of group A if they were only half their size. To this end, we need to make an assumption. One option is to assume that if the group was half the size, they would write half of the papers, this would result in the dashed gray line as in Fig. 1. If the articles of the original set would differ in the number of citations, there of course exist many subgroups of 9, so the calculation of the expected citation curve would be a bit more complex. Another option is to assume that each of the articles would have the probability $p = 1/2$ of being written, this would result in the solid gray curve as in Fig. 1. This second assumption is less restrictive, since it does not fix the number of articles in the reduced set. The expected number of articles in the halved group is still 9, but there is a 0.41 probability of less than 9 articles being produced, hence the expected number of citations for their 9th best-cited article is only 5.9. It is also possible that the halved group would publish more than 9 articles, therefore, the expected curve does not end at 9, there is even a tiny probability that the diminished group would still publish all the 18 articles. Formulae for calculating the distributions are presented in the following sections.

Other, much more restrictive, options are to assume a certain fixed effect of reducing the size of a group. For example, to assume that a group half the size would only write every second article of the original sorted set. We take the view that each such assumption is just one special case out of many, and that probabilistic view is necessary to describe a possible range of outcomes. The important advantage of a probabilistic approach is thus an additional insight into the uncertainty of our knowledge of what would happen on a reduced set. We shall therefore not only report the expected reduced citation curve but also its uncertainty.

2.2. Fixed probability of choice

We assume that reducing the resources of a group for a certain factor would result in each of the articles having a fixed, lower probability of being written, we denote this probability by $p = (\text{factor of reduction})^{-1}$.

We order the articles in the subsample, so that the first article has the highest number of citations and let C_k denote the number of citations of the kth best-cited article in the sample, so that the reduced citation curves joins the points (k, C_k). The distribution of the C_k value for a given k is given by

$$P(C_k = c_j) = \binom{j-1}{k-1} p^k q^{j-k}, \quad j \geq k, \tag{1}$$

where $\binom{0}{0} = 1$ and $q = 1 - p$.

The idea behind the formula is as follows. The value of C_k is equal to c_k only if each of the first k articles is in the sample. The probability of this event is $P(C_k = c_k) = p^k$. The value of C_k equals c_{k+1} if our sample contains the $(k+1)$th article and all but one of the first k articles. The probability of this event is thus $P(C_k = c_{k+1}) = \binom{k}{k-1} p^{k-1} q \cdot p$. The value of C_k equals c_j if our sample contains $k-1$ articles from the first $j-1$ articles and the jth article of the original population, which gives the above general formula. Note that the sum of probabilities $\sum_{j \geq k}^{n} P(C_k = c_j)$ does not equal to 1, there always exists a positive probability $P(C_k = 0)$, which we do not need to calculate since it does not affect the expected value.

The expected value of C_k is thus

$$E(C_k) = \sum_{j=k}^{n} \binom{j-1}{k-1} p^k q^{j-k} \cdot c_j \tag{2}$$

and the variance can be calculated as $E(C_k^2) - E(C_k)^2$, where

$$E(C_k^2) = \sum_{j=k}^{n} \binom{j-1}{k-1} p^k q^{j-k} \cdot c_j^2$$

Since the whole distribution of C_k is known, tolerance intervals can be calculated in an exact way, using the percentiles of the distribution. For example, for calculation of the 95% tolerance intervals, we find indices l and u, for which

$$\sum_{j=k}^{l} P(C_k = c_j) \leq 0.025 \quad \text{and} \quad \sum_{j=k}^{u} P(C_k = c_j) \leq 0.975.$$

Note that due to the discrete nature of distribution, the size of the tolerance intervals may not be exactly as declared. Nevertheless the exact calculation is a better approach than a normal approximation, since the distributions are usually asymmetric.

For $p = 1/2$ we get the following expected values for the first two articles in the sample:

$$E(C_1) = \sum_{j=1}^{n} \binom{j-1}{0} \frac{1}{2}^{1} \frac{1}{2}^{j-1} \cdot c_j = \sum_{j=1}^{n} \frac{1}{2^j} \cdot c_j$$

$$E(C_2) = \sum_{j=2}^{n} \binom{j-1}{1} \frac{1}{2}^{2} \frac{1}{2}^{j-2} \cdot c_j = \sum_{j=2}^{n} (j-1) \frac{1}{2^j} \cdot c_j$$

2.3. *Fixed number of sampled articles*

An alternative for the sampling procedure of the previous section is to fix the number of articles chosen in the sample, i.e. to assume that diminishing the resources by a certain factor would diminish the number of articles by exactly the same factor.

Say we want to consider a sample of size d from our population. The value of C_k is then equal to c_k only if all of the first k individuals are in the sample. Since the sample size is fixed, exactly $d - k$ individuals must be chosen from the rest of the population ($n - k$ articles). Similarly, C_k is equal to c_j ($j \geq k$, $k \leq d$), if the following holds: $k - 1$ articles were chosen from the first $j - 1$ articles of the population, the jth article was chosen and $d - k$ articles were chosen from the last $n - j$ articles. The probability of this event therefore equals:

$$P(C_k = c_j) = \frac{\binom{j-1}{k-1} \cdot 1 \cdot \binom{n-j}{d-k}}{\binom{n}{d}}, \quad j \geq k \tag{3}$$

The expected values, the variance of C_k and the tolerance intervals can be calculated as described in the previous section.

2.4. *Comparison of the two sampling methods*

As we have seen in Fig. 1, the two assumptions produce quite different results. Here, we compare the results in general.

Say p and d are chosen so that $n = pd$. Formulae (1) and (3) have one common factor, we rewrite the rest of (3) as

$$\frac{\binom{n-j}{d-k}}{\binom{n}{d}} = \frac{d(d-1)\cdots(d-k+1)\cdot(n-d)(n-d-1)\cdots(n-d-(j-k)+1)}{n(n-1)\cdots(n-j+1)}$$

$$= \frac{d(d-1)\cdots(d-k+1)}{n(n-1)\cdots(n-k+1)} \cdot \frac{(n-d)(n-d-1)\cdots(n-d-(j-k)+1)}{(n-k)\cdots(n-j+1)} \doteq \left(\frac{d}{n}\right)^k \cdot \left(\frac{n-d}{n}\right)^{j-k} = p^k q^{j-k}$$

The similarity in the above equation of course depends on values of n, d and k. If d forms a rather large or rather small proportion of n, C_k values calculated with the two methods may differ considerably, but if d/n is around 0.5, practically all C_k values are almost equal.

In practice the two approaches give rather similar results, but since the assumption of fixed probability is more general, and also, as we will show, less sensitive to incomplete data, we choose the first method as the tool for comparisons in the rest of the paper.

2.5. *h-Index*

The calculated distributions of C_k can be used to estimate the expected value of the h-index. Since the h-index of any curve can only take integer values and the expected citation curves are right-continuous step functions, such an estimate is always rounded down. Furthermore, the confidence interval calculation for h-index estimated from the reduced citation curve is not straightforward. Therefore, we add an exact and direct calculation of the reduced h-index value for the fixed probability case. Let h denote a value and let $na(h)$ denote the number of articles with more than h citations in the original set, whereas $nb(h)$ is used to denote the number of articles with exactly h citations:

$$na(h) = \sum_{j=1}^{n} I[c_j > h]$$

$$nb(h) = \sum_{j=1}^{n} I[c_j = h]$$

Let the random variable $N_a(h)$ denote the number of selected articles with more than h citations and let $N_b(h)$ be the number of sampled articles with exactly h citations. Since all articles have the same probability p of being chosen and the sampling process treats them as independent units, both of these random variables are binomially distributed, $N_a(h) \sim Bin(na(h), p)$, $N_b(h) \sim Bin(nb(h), p)$. The value of h-index on the sample equals h if the sample contains at least h from the $na(h)+nb(h)$ articles with at least h citations but not more than h articles from the $na(h)$ articles that have at least $h+1$ citations. The probability of such an event thus equals

$$P(H = h) = \sum_{k=0}^{h} P(N_a = k)P(N_b \geq h - k).$$

With the given distribution of h-index, the expected value of the h_r-index can be calculated for a given factor p as

$$h_r = \sum_{h=1}^{n} hP(H = h). \tag{4}$$

Similarly, the given distribution can be used to calculate the variance and its percentiles can be used to give the tolerance intervals.

2.6. Incomplete data

In practice, it may be common to have data only for the more cited articles. For example, information on all articles with zero or perhaps less than five citations may be missing. In this section, we study how the measures may be affected by this missingness.

It is obvious that the fixed probability method does not need articles with zero citations for estimation of either the expected value or the standard error of the C_k distribution. If articles with less than 5 citations are missing, the sum in (2) does not contain all the summands and is thus biased downwards, i.e. the expected values of C_k are too small. The size of the bias depends on the probability that $C_k = c_j$, where c_j is one of the missing articles. The higher the difference between k and j, the less bias we can expect (less bias at the beginning of the curve). To get an idea about the bias, we can compare the worst scenario (no articles with less than 5 citations) with the best one (many articles with 4 citations), the true curve shall be in between.

The fixed number method is more susceptible to missing articles, since the probabilities in (3) depend on total number of articles n and thus the non-cited articles also contribute to the results. If articles with few citations are omitted, the subgroup is sampled from articles with higher citations and thus the bias is in the other direction.

In either example, the h-index is not affected much by the missing articles, since its standard error is typically rather small, see Section 3.

2.7. Properties of h-index and our methods

The most natural choice of the scaling factor when comparing two units is to diminish the larger group to the size of the smaller (with regard to population size, resources, etc.). When comparing more than two units, one must keep in mind that comparisons are only reasonable when performed on the same scale, so all units should be diminished either to the size of the smallest or to any other size that is lower than the smallest of the units. When choosing the factor of reduction, the order of h_r-index ranking may differ if a different factor is chosen. This is not a property of our method, but rather of the h-index itself. The citation curves of two units may cross – one unit may have a few highly cited articles, while the other may have many articles but lower citation rates. Since h-index compares the two distributions only at one point, the ranking of the two units according to h-index depends on where that point lies.

As an illustration, consider an artificial example of two units (of comparable size) given in Fig. 2. Unit A is better in the left tail and has some heavily cited articles, but unit B is better in the right tail. Fig. 2a and b presents identical curve shapes, but the scale is different – the most cited article has 480 citations in Fig. 2a compared to 48 citations in Fig. 2b. The h-index in Fig. 2a is 60 and 69 for A and B, respectively, compared to 16 and 14 for A and B, respectively, in Fig. 2b.

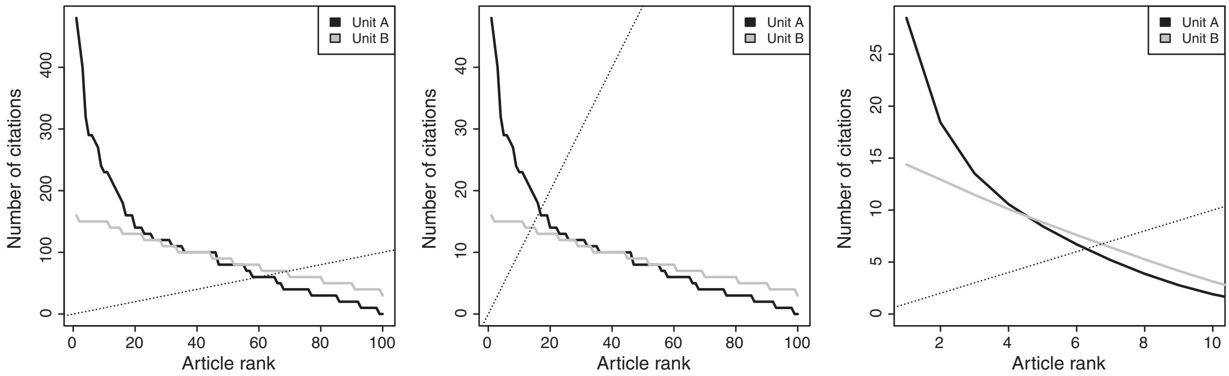

Fig. 2. Comparison of curves in an artificial example – the ranking with respect to h-index depends on where the curves cross.

If article citations are high with respect to the number of articles published, the h-index point moves to the right tail of the distribution and the same is true when a unit is diminished by a certain factor: the citation numbers get higher compared to the number of articles, hence the h_r-index depends more on the tail of the distribution. An example of this is given in Fig. 2c, where both units from graph in Fig. 2b are decreased with factor 0.1. The diminished unit B has higher h_r-index (6.6) than the diminished unit A (6).

The important thing to note is that the question: "which unit is better" is not well defined and has more than one answer. To enable comparison, the units should be of the same size. The approach taken in all our examples that follow is to diminish the larger unit to the size of the smaller and thus answer the question "would unit A be better than B if it was of the size of B?" But of course, both units could be diminished even further, say to the size of a third unit. If the citation distributions for two units of the same size do not cross, the ranking with respect to h-index is the same regardless of the factor of diminishment, if not, the factor used determines the ranking of the h-index.

Our proposed method enables reduction of the citation distribution and, consequently, of any of the measures based on it, one example being the h-index. However, as explained in this section, care must be taken to correctly interpret the reduced values of such measures, since they may be affected by different parts of the citation distribution depending on the reduction factor. As explained already in Marchant (2009), there is no objectively right ranking of the groups and various parts of the citation distribution may be of interest. There is a number of papers carefully studying the properties of various measures (see e.g. Marchant, 2009; Waltman & van Eck, 2011; Woeginger, 2008) and size-independence is one of the desired properties that many measures lack. With our approach we add this property to all the measures based on the citation curve.

3. Examples

Slovenian science evaluators often compare their country to neighboring Austria, whose population is approximately four times larger.

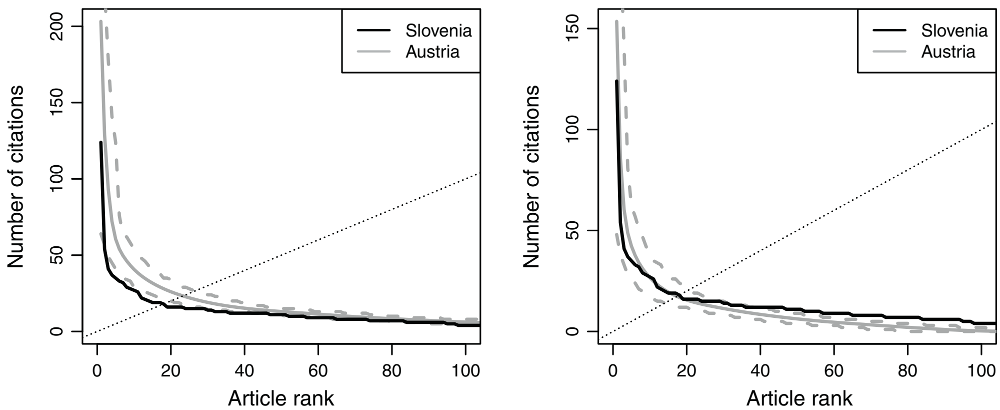

Fig. 3. A comparison of physics articles: Slovenia compared to an estimate of diminished Austria with 95% pointwise tolerance intervals (dashed curves). Left: factor 4; right: factor 8. Dotted symmetry line approximates the h-index.

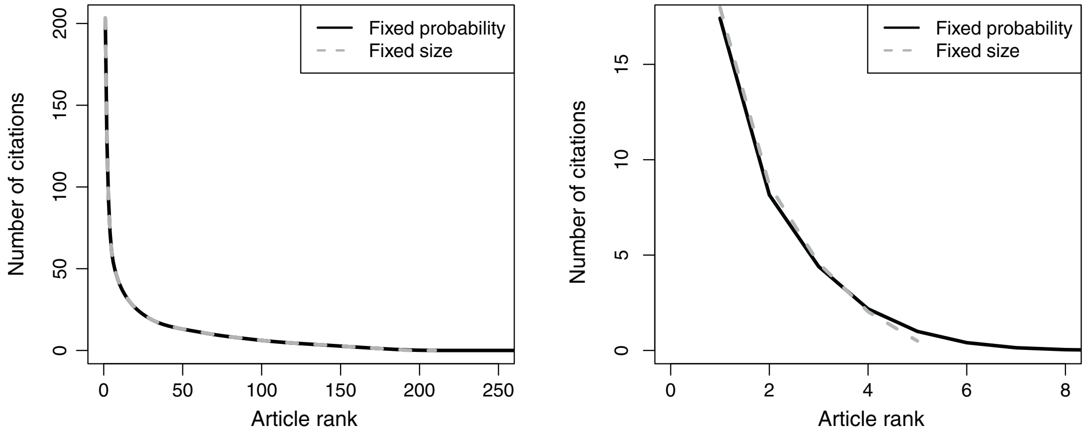

Fig. 4. Comparison of the sets diminished by factor 4 with the two sampling schemes. Left: a quarter of Austrian physics citation distribution, right: a quarter of an artificial example of 20 articles.

As an example, we compare Slovenia's physics articles that were published between 2002 and 2006 and cited up to mid-2008 to a quarter and an eighth of their Austrian counterparts in the same period – Fig. 3[2] presents the expected citation curve calculated with Eq. (1) for the diminished Austria. Austria published 838 articles in total, Slovenia 212, the h-indices are 39 and 18, respectively. The population of Slovenia is approximately a quarter of the Austrian and we thus choose the factor 4 to reduce the Austrian citation distribution. Austria is still better (Fig. 3, left graph), with Slovenia being approximately on the limits of the 95% tolerance interval.

The expected h_r-index using formula (4) for a quarter of Austria is 23. Around the value of its h-index, Slovenia's physics-articles citation rate was, even when accounted for its smaller population, still only at 2/3 of the Austrian. On the other hand, Slovenia is slightly better than one eighth of Austria (Fig. 3, right graph), which in this case has an expected h_r-index equal to 17.

To study how strongly the results are affected by our reducement assumption, we study diminishing by factor 4 using both methods proposed in this paper – we look at $p = 1/4$ or $d = 210$. The resulting expected citation curves given by the two schemes are compared in the left graph of Fig. 4. As a second example we look at a simulated group with only 20 articles in total and draw the reduced citation curve for $p = 1/4$ and $d = 5$. It is clear that departures are practically negligible in both cases.

As a second example, we compare the data on the USA's, Germany's and Canada's linguistics articles between 1996 and 2005 cited up to mid-2007. We compare the three data sets with respect to population sizes (in 2001) of the respective countries. The US data are thus reduced for a factor of 9.09, the German for a factor of 2.63. The three curves (with tolerance intervals) are compared in Fig. 5. The US curve is consistently higher than the German one, whereas Canada is lower (=less cited) at the very beginning, but higher (=more cited) later on. The tolerance intervals for the US curve are very large (due to the large factor), thus not much can be said about what would happen in the reduced set. When comparing the h-index only, Canada is best with 8, the h_r-index of the diminished US data equals 6.5 (95% CI [4.1, 8.9]) and the diminished German h_r-index equals 6.0 (95% CI [4.1, 7.9]).

Fig. 6 presents the differences between curves in data with missing information on the cited articles. The left graph presents a quarter of Austria's physics articles (solid black line) with the quarter we would get if we only had information on Austria's articles that were cited 10 or more times and assumed they published no other articles (dotted line). In addition, the dashed line shows the quarter of Austria that we would get assuming many articles cited exactly 9 times (in our example 1000). Note that the total number of Austria's physics articles in the sampled period is 838, of which 282 articles have 10 or more citations. A quarter of this number is 70.5. In practice, the solid black curve may not be known, so that all we know is that it is somewhere between the dashed and the dotted line. As we can see, the curves for the best and worst scenario overlap for most of the interval; on the interval $k \in (1, 40)$, where the values C_k are above 15, the bias is less than 0.0001. The gap then increases quickly, but is of course meaningless when the C_k values fall below 10, since this is the area for which we have no information. As the curves only start differentiating well after the h-index point, the difference in the h-index is negligible.

The right-hand graph in Fig. 6 compares the data of the USA's and Germany's physics articles. The comparison is made relative to the population size: Germany's population is 29% of that of the USA. The available data comprise only articles with 81 or more citations for Germany and 84 or more citations for the USA. The true curve for the reduced USA distribution now stays unknown, but we can see that the best and worst scenario (dashed and dotted curve) again differ only after the

[2] The data were exported from Thomson Reuter's SCI EXPANDED database, updated 2008-08-02, TS=physics, Document type "article", All languages, 2002–2006.

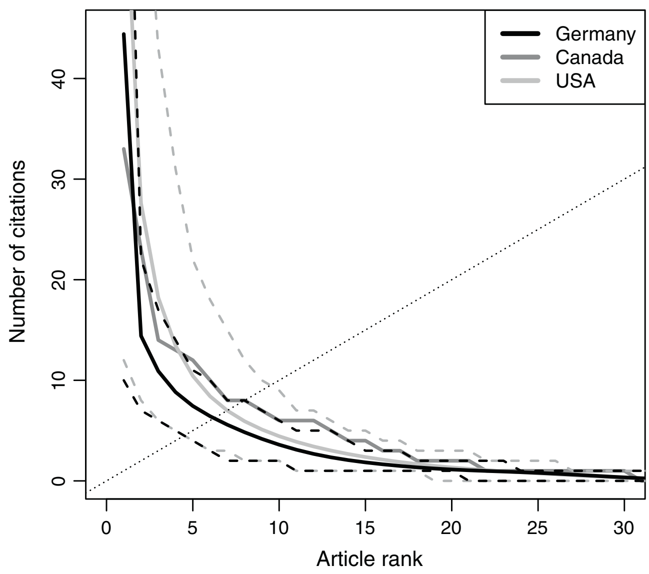

Fig. 5. A comparison of articles in linguistics: Canada, diminished US (factor 9.09) and diminished Germany (factor 2.63). Dashed curves denote the 95% pointwise tolerance intervals. Symmetry line approximates the h-index.

end of the German data (black curve) and thus do not contribute to any bias in comparisons. As we can see, the diminished US data are still significantly better than the German data, except perhaps for the first 15 most cited articles. The h_r-index for the diminished US data equals 94 (95% CI [88.7, 99.2]), whereas the h-index of Germany is 87.

All examples in this paper were calculated in R statistical software (Development Core Team, 2011), all the required functions are available at the website http://ibmi.mf.uni-lj.si/ibmi-english/biostat-center/programje/cdis.r.

4. Discussion

Calculation of a reduced (diminished) citation distribution for any diminishing factor has many advantages. Graphical presentation of two or more citation curves offers the most straightforward and easiest way for even people unfamiliar with statistical methods to be able to see the relationship between them. From the perspective of national research authorities, it makes it easy to compare citation distributions for different parameters, such as population, gross domestic product, funding, full-time equivalent of researchers, etc. It also makes it easy to see intervals of citation curves where differences are significant. From the point of view of its practical use, an important advantage of the method is that for the calculation of reduced citation distribution it does not require a complete original citation distribution, since non-cited articles do not contribute to diminished distribution and the contribution of low-cited articles is usually negligible. As the consequence, the export of original citation distributions from databases such as Thomson Reuter's Web of Knowledge or Elsevier Scopus

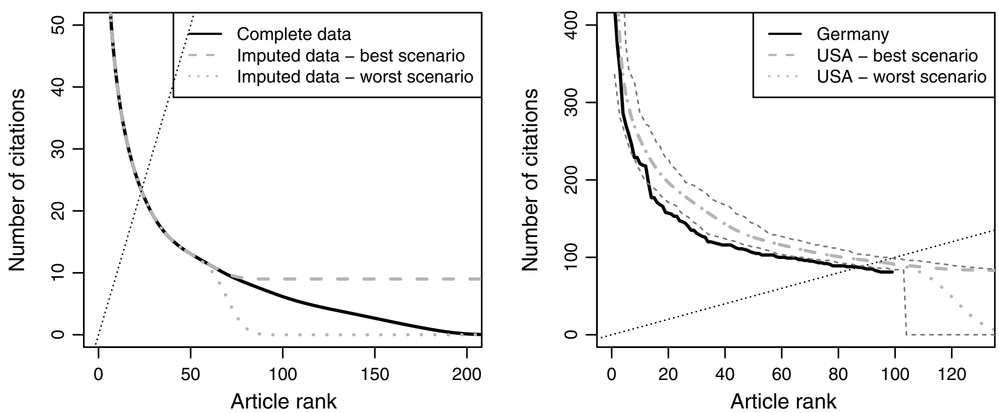

Fig. 6. Comparison of curves with missing data, left graph: a quarter of Austrian physics articles compared to the quarter we would get with missing data; right graph: comparison of German and reduced US data (factor 3.47, with 95% CI), only data on articles with 81 or more citations are available (84 or more for USA). The dotted symmetry line approximates the h-index.

is easier. This represents a significant advantage compared to quantile analysis. Especially for the calculation of the reduced h_r-index, we do not need more data than approximately down to the original h-index of the smaller group.

The reduced h_r-index is not another h-index variant but an extension of the h-index to comparison of units of different sizes. It is not an absolute figure but a function of the reducing factor.

An important advantage of the method is that the value of the reduced h_r-index is accompanied with tolerance intervals, which give an indication of how much uncertainty we have. The variance, which increases with the reducing factor, thus signalizes that the consequences of a huge reduction are hard to predict. The reduced h_r-index should always be reported with the tolerance intervals to prevent overinterpretation such as ranking the units in cases where differences are only within the tolerance intervals.

The variance of the reduced value was calculated pointwise for each ordered article separately, and it may turn out to be even smaller due to the smoothing effect of the ranking; we leave this issue for future research.

It is important to note that even if we reduce different units to the same size, this act alone does not yet mean they are directly comparable. In particular, such a comparison would make little sense if not corrected for the different citation habits, as described in the paper of Iglesias and Pecharromán (2007). If comparisons across fields are desired, any reduction from h-index to h_r-index can be further corrected using the method of Iglesias and Pecharromán (2007), thus making it possible to compare h-indices of research groups from different fields and of different sizes.

The reduced h_r-index complements the original size-dependent h-index and thus expands its use to size-independent comparisons.

5. Conclusions

Any performance comparison of different groups using the citation curve is obscured by its strong dependency on the size of the groups and the same is true for the measures based on this curve such as the popular h-index. By calculating the expected values of the reduced distribution and h_r-index with our approach, this disadvantage is eliminated. A graphical comparison of the citation distributions for groups of different sizes represents a very straightforward procedure and allows for wide usage.

References

Aksnes, D. W., & Taxt, R. E. (2004). Peer reviews and bibliometric indicators: A comparative study at a Norwegian university. *Research Evaluation*, *13*, 33–41.

Albarrán, P., Ortuño, I., & Ruiz-Castillo, J. (2011). The measurement of low- and high-impact in citation distributions: Technical results. *Journal of Informetrics*, *5*, 48–63.

Alonso, S., Cabrerizo, F. J., Herrera-Viedma, E., & Herrera, F. (2009). *h*-Index: A review focused in its variants, computation and standardization for different scientific fields. *Journal of Informetrics*, *3*, 273–289.

Bornmann, L., Mutz, R., & Daniel, H. D. (2010). The h index research output measurement: Two approaches to enhance its accuracy. *Journal of Informetrics*, *4*, 407–414.

Bornmann, L., Mutz, R., Hug, S. E., & Daniel, H. D. (2011). A multilevel meta-analysis of studies reporting correlation between the h index and 37 different h index variants. *Journal of Informetrics*, *5*, 346–359.

Burrell, Q. L. (2007). Hirsch's *h*-index: A stochastic model. *Journal of Informetrics*, *1*, 16–25.

Butler, L. (2003). Explaining Australia's increased share of ISI publications-the effects of a funding formula based on publication counts. *Research Policy*, *32*, 143–155.

Coryn, C. L. S. (2005). The use and abuse of citations as indicators of research quality. *Journal of Multidisciplinary Evaluation*, *4*, 115–121.

Crespo, J. A., Ortuño, I., & Ruiz-Castillo, J. (2011). The citation merit of scientific publications. *Working Paper, Economic Series*, 11–36.

Egghe, L. (2006). An improvement of the *h*-index: The *g*-index. *ISSI Newsletter*, *2*, 8–9.

Egghe, L. (2008). Examples of simple transformations of the *h*-index: Qualitative and quantitative conclusions and consequences for other indices. *Journal of Informetrics*, *2*, 136–148.

Egghe, L. (2006). How to improve the *h*-index. *The Scientist*, *20*, 14.

Egghe, L. (2008). Study of different h-indices for groups of authors. *Journal of the American Society for Information Science and Technology*, *59*, 1276–1281.

Egghe, L. (2010). The Hirsch index and related impact measures. *Annual Review of Information Science and Technology*, *44*, 65–114.

Egghe, L. (2006). Theory and practice of the *g*-index. *Scientometrics*, *69*, 131–152.

Egghe, L., & Rousseau, R. (2006). An informetric model for the Hirsch-index. *Scientometrics*, *69*, 121–129.

Glänzel, W. (2006). On the *h*-index – A mathematical approach to a new measure of publication activity and citation impact. *Scientometrics*, *67*, 315–321.

Hirsch, J. E. (2005). An index to quantify an individual's scientific research output. *Proceedings of the National Academy of Sciences of the United States of America*, *102*, 16569–16572.

Iglesias, J. E., & Pecharromán, C. (2007). Scaling the *h*-index for different scientific ISI fields. *Scientometrics*, *73*, 303–320.

Južnič, P., Pečlin, S., Žaucer, M., Mandelj, T., Pušnik, M., & Demšar, F. (2010). Scientometric indicators: Peer-review, bibliometric methods and conflict of interests. *Scientometrics*, *85*, 429–441.

Leydesdorff, L., Bornmann, L., Mutz, R., & Opthof, T. (2011). Turning the tables on citation analysis one more time: Principles for comparing sets of documents. *Journal of American Society for Information Science and Technology*, *62*, 1370–1381.

Marchant, T. (2009). An axiomatic characterization of the ranking based on the *h*-index and some other bibliometric rankings of authors. *Scientometrics*, *80*, 335–342.

Molinari, J. F., & Molinari, A. (2008). A new methodology for ranking scientific institutions. *Scientometrics*, *75*, 163–174.

Development Core Team, R. (2011). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. ISBN: 3-900051-07-0

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

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

Waltman, L., & van Eck, N. J. (2011). The inconsistency of the *h*-index. *Journal of the American Society for Information Science and Technology*, *63*, 406–415.

Woeginger, G. J. (2008). An axiomatic characterization of the Hirsch-index. *Mathematical Social Sciences*, *56*, 224–232.