Towards a new crown indicator:
Some theoretical considerations

**Ludo Waltman, Nees Jan van Eck, Thed N. van Leeuwen,
Martijn S. Visser, and Anthony F.J. van Raan**

Universiteit Leiden

CWTS

Centrum voor Wetenschaps- en Technologie Studies

Towards a new crown indicator:
Some theoretical considerations

Ludo Waltman, Nees Jan van Eck, Thed N. van Leeuwen,
Martijn S. Visser, and Anthony F.J. van Raan

Centre for Science and Technology Studies, Leiden University, The Netherlands
{waltmanlr, ecknjpvan, leeuwen, visser, vanraan}@cwts.leidenuniv.nl

The crown indicator is a well-known bibliometric indicator of research performance developed by our institute. The indicator aims to normalize citation counts for differences among fields. We critically examine the theoretical basis of the normalization mechanism applied in the crown indicator. We also make a comparison with an alternative normalization mechanism. The alternative mechanism turns out to have more satisfactory properties than the mechanism applied in the crown indicator. In particular, the alternative mechanism has a so-called consistency property. The mechanism applied in the crown indicator lacks this important property. As a consequence of our findings, we are planning to move towards a new crown indicator, which relies on the alternative normalization mechanism.

1. Introduction

It is well known that in some scientific fields the average number of citations per publication (within a certain time period) is much higher than in other scientific fields. This is due to differences among fields in the average number of cited references per publication, the average age of cited references, and the degree to which references from other fields are cited. In addition, bibliographic databases such as Web of Science and Scopus cover some fields more extensively than others (e.g., Moed, 2005). Clearly, other things equal, one will find a higher average number of citations per publication in fields with a high database coverage than in fields with a low database coverage.

In citation-based research performance evaluations, it is crucial that one carefully controls for the above-mentioned differences among fields. This is especially the case for performance evaluations at higher levels of aggregation, such as at the level of countries, universities, or multi-disciplinary research groups. In performance evaluation studies, our institute, the Centre for Science and Technology Studies (CWTS) of Leiden University, uses a standard set of bibliometric indicators (Van Raan, 2005). Our best-known indicator, which we usually refer to as the crown indicator, relies on a normalization mechanism that aims to correct for the above-mentioned differences among fields. An indicator similar to the crown indicator is used by the Centre for R&D Monitoring (ECOOM) in Leuven, Belgium. ECOOM calls its indicator the normalized mean citation rate (e.g., Glänzel, Thijs, Schubert, & Debackere, 2009). Thomson Reuters uses an indicator to which it refers as the crown index or C-index (Thomson Reuters, 2008). This indicator relies on a similar normalization mechanism as our crown indicator, but it corrects for differences among journals rather than for differences among fields.

The normalization mechanism of the crown indicator basically works as follows. Given a set of publications, we count for each publication the number of citations it has received. We also determine for each publication its expected number of citations.

The expected number of citations of a publication equals the average number of citations of all publications of the same document type (i.e., article, letter, or review) published in the same field and in the same year. To obtain the crown indicator, we divide the sum of the actual number of citations of all publications by the sum of the expected number of citations of all publications.

The normalization mechanism of the crown indicator has been criticized by Lundberg (2007) and by Opthof and Leydesdorff (in press).[1] These authors have argued in favor of an alternative mechanism. According to the alternative mechanism, one first calculates for each publication the ratio of its actual number of citations and its expected number of citations and one then takes the average of the ratios that one has obtained. Lundberg refers to an indicator that uses this mechanism as the item-oriented field-normalized citation score average. This indicator is used by Karolinska Institute in Sweden (Rehn & Kronman, 2008). Similar indicators are used by Science-Metrix in the US and Canada (e.g., Campbell, Archambault, & Côté, 2008, p. 12) and by the SCImago research group in Spain (SCImago Research Group, 2009).

In this paper, we present a theoretical comparison between the normalization mechanism of the crown indicator and the alternative normalization mechanism discussed by Lundberg (2007) and others. We first consider two fictitious examples that provide some insight into the differences between the mechanisms. We then study the consistency (Waltman & Van Eck, 2009a, 2009b) of the mechanisms. We also pay some attention to the way in which overlapping fields should be handled. The main finding of the paper is that the alternative normalization mechanism has a more solid theoretical basis than the normalization mechanism currently applied in the crown indicator. As a consequence of this finding, CWTS is planning to move towards a new crown indicator, which relies on the alternative mechanism. An extensive empirical comparison between the two normalization mechanisms is still underway but will be reported soon.

2. Definitions of indicators

In this section, we provide formal mathematical definitions of the CPP/FCSm indicator and of the MNCS indicator. The CPP/FCSm indicator has been used as the crown indicator of CWTS for more than a decade. The MNCS indicator, where MNCS is an acronym for mean normalized citation score, is the new crown indicator that CWTS is planning to adopt. The two indicators differ from each other in the normalization mechanism they use. Throughout this paper, we focus on the issue of normalization for differences among fields. We do not consider the issue of normalization for differences among document types or for differences among publications of different ages. However, at the end of the paper, we will make some brief comments on the latter issue.

Consider a set of n publications, denoted by 1, …, n. Let c_i denote the number of citations of publication i, and let e_i denote the expected number of citations of publication i given the field in which publication i has been published. Hence, e_i equals the average number of citations of all publications published in the same field as publication i. The CPP/FCSm indicator, where CPP and FCSm are acronyms for, respectively, citations per publication and mean field citation score, is defined as

---

[1] See also our reply to Opthof and Leydesdorff (Van Raan, Van Leeuwen, Visser, Van Eck, & Waltman, 2010).

$$\text{CPP/FCSm} = \frac{\sum_{i=1}^{n} c_i \big/ n}{\sum_{i=1}^{n} e_i \big/ n} = \frac{\sum_{i=1}^{n} c_i}{\sum_{i=1}^{n} e_i}. \qquad (1)$$

The CPP/FCSm indicator was introduced by De Bruin, Kint, Luwel, and Moed (1993) and Moed, De Bruin, and Van Leeuwen (1995). A similar indicator, the normalized mean citation rate, was introduced somewhat earlier by Braun and Glänzel (1990). The normalization mechanism of the CPP/FCSm indicator goes back to Schubert and Braun (1986) and Vinkler (1986).

We now turn to the MNCS indicator. We define the MNCS indicator as

$$\text{MNCS} = \frac{1}{n} \sum_{i=1}^{n} \frac{c_i}{e_i}. \qquad (2)$$

The MNCS indicator uses the same normalization mechanism as the item-oriented field-normalized citation score average indicator introduced by Lundberg (2007). Comparing (1) and (2), it can be seen that the CPP/FCSm indicator normalizes by calculating a ratio of averages while the MNCS indicator normalizes by calculating an average of ratios.[2]

Interestingly, the CPP/FCSm indicator can be regarded as a kind of weighted version of the MNCS indicator. To see this, notice that (1) can be rewritten as

$$\text{CPP/FCSm} = \frac{1}{n} \sum_{i=1}^{n} w_i \frac{c_i}{e_i}, \qquad (3)$$

where w_i is given by

$$w_i = \frac{e_i}{\sum_{j=1}^{n} e_j \big/ n}. \qquad (4)$$

Hence, like the MNCS indicator, the CPP/FCSm indicator can be written as an average of ratios. However, unlike the MNCS indicator, the CPP/FCSm indicator does not weigh all ratios equally. Instead, it gives more weight to ratios corresponding with publications that have a higher expected number of citations. In other words, fields with a high average number of citations per publication have more weight in the calculation of the CPP/FCSm indicator than fields with a low average number of citations per publication. This has also been noted by Lundberg (2007). In the calculation of the MNCS indicator, all fields have the same weight, regardless of their average number of citations per publication. We will come back to this difference between the CPP/FCSm indicator and the MNCS indicator later on in this paper.

The CPP/FCSm indicator and the MNCS indicator are both size independent. These indicators are intended to measure the average performance of a set of

---

[2] In a somewhat different context, formulas similar to (1) and (2) were also studied by Egghe and Rousseau (1996). Egghe and Rousseau refer to (1) as a globalizing quotient and to (2) as an averaging quotient.

publications.[3] Although in performance evaluation studies one usually focuses on the average performance of a set of publications, the total performance of a set of publications can be of interest as well. A natural approach to measuring the total performance of a set of publications is to first measure the average performance of the set of publications and to then multiply the average performance by the total number of publications involved (Waltman & Van Eck, 2009a). When average performance is measured using the CPP/FCSm indicator, this approach yields

$$CPP/FCSm \times n = \frac{\sum_{i=1}^{n} c_i}{\sum_{i=1}^{n} e_i / n}. \tag{5}$$

At CWTS, we refer to the indicator in (5) as the brute force indicator. This indicator is for example used in our Leiden Ranking of universities (CWTS, n.d.). When instead of the CPP/FCSm indicator the MNCS indicator is used for measuring average performance, one obtains

$$TNCS = MNCS \times n = \sum_{i=1}^{n} \frac{c_i}{e_i}. \tag{6}$$

We refer to this indicator as the TNCS indicator, where TNCS is an acronym for total normalized citation score. The TNCS indicator is similar to what Lundberg (2007) refers to as the total field-normalized citation score indicator.

3. Example 1

The following fictitious example provides some insight into the differences between the CPP/FCSm indicator and the MNCS indicator. Suppose we want to compare the performance of two research groups, research group A and research group B. Both research groups are active in the same field. This field consists of two subfields, subfield X and subfield Y. Research groups A and B have the same number of publications, and they both have half of their publications in subfield X and half of their publications in subfield Y. The number of publications and citations of the two research groups in the two subfields is reported in Table 1. For each subfield, the expected number of citations of a publication is also reported in the table.

Table 1. Number of publications (P) and citations (C) of research groups A and B in subfields X and Y.

	Expected cit. per pub.	Research group A	Research group B
Subfield X	10	P = 100, C = 1000	P = 100, C = 2200
Subfield Y	20	P = 100, C = 4000	P = 100, C = 2400

As can be seen in Table 1, research group B outperforms research group A in subfield X while research group A outperforms research group B in subfield Y. The question that we want to answer is which research group has a higher overall

---

[3] Citation-based indicators in fact measure only one aspect of research performance, namely the aspect of citation impact. Throughout this paper, we use the term *performance* to refer specifically to the citation impact of publications rather than to research performance in general.

performance. The CPP/FCSm indicator and the MNCS indicator turn out to answer this question differently.

According to the CPP/FCSm indicator, the overall performance of research group A is higher than the overall performance of research group B. This is shown in Table 2. Values of the CPP/FCSm indicator for each subfield separately are also shown in the table. Notice that research group B's performance in subfield X is higher than research group A's performance in subfield Y and that research group B's performance in subfield Y is higher than research group A's performance in subfield X. Despite of this, the CPP/FCSm indicator states that research group B has a lower overall performance than research group A.

Table 2. Values of the CPP/FCSm indicator for research groups A and B.

	Research group A	Research group B
Subfield X	1.00	2.20
Subfield Y	2.00	1.20
Both subfields together	1.67	1.53

According to the MNCS indicator, the overall performance of research group B is higher than the overall performance of research group A. This is shown in Table 3. Notice that for each subfield separately the MNCS indicator yields exactly the same results as the CPP/FCSm indicator. For both subfields together, however, the indicators yield different results. In fact, they even yield opposite rankings of the two research groups.

Table 3. Values of the MNCS indicator for research groups A and B.

	Research group A	Research group B
Subfield X	1.00	2.20
Subfield Y	2.00	1.20
Both subfields together	1.50	1.70

Why does the CPP/FCSm indicator favor research group A over research group B? This is because the CPP/FCSm indicator gives more weight to subfield Y than to subfield X while the MNCS indicator weighs both subfields equally. This difference can be seen by comparing (2) with (3) and (4) (see Section 2). The CPP/FCSm indicator and the MNCS indicator agree with each other that an appropriate measure of the performance of a single publication is the ratio of the publication's actual number of citations and the publication's expected number of citations. As indicated by (2), the MNCS indicator calculates the performance of a set of publications as an unweighted average of the performance of the individual publications in the set. Since in the case of research groups A and B the number of publications in subfield X equals the number of publications in subfield Y, the MNCS indicator weighs both subfields equally. Unlike the MNCS indicator, the CPP/FCSm indicator calculates the performance of a set of publications as a weighted average of the performance of the individual publications in the set. As indicated by (3) and (4), publications with a higher expected number of citations have a higher weight. Since in subfield Y the expected number of citations of a publication is higher than in subfield X, the CPP/FCSm indicator gives more weight to subfield Y than to subfield X. In subfield Y, research group A outperforms research group B, and therefore the CPP/FCSm

indicator states that research group A has a higher overall performance than research group B.

Should publications be weighed differently depending on their field, like the CPP/FCSm indicator does? In general, we do not believe this to be desirable. Indicators such as the CPP/FCSm indicator and the MNCS indicator aim to correct for differences among fields. To achieve this aim, the number of citations of a publication should be normalized for differences among fields. However, after this normalization has been performed, there seems to be no reason to treat publications from different fields differently. Instead, after normalization, publications from different fields should be treated equally. This is exactly what the MNCS indicator does. By treating publications from different fields differently, the CPP/FCSm indicator introduces a bias towards fields with high a expected number of citations.

To further illustrate this point, suppose the number of publications and citations of research groups A and B in subfields X and Y is given by Table 4 rather than by Table 1. Notice that the only thing that has changed is that the actual and expected numbers of citations in subfield Y have been divided by four. Since both for research group A and for research group B the performance in each subfield separately has not changed, it seems natural to also expect no changes in the overall performance of the research groups. In the case of the MNCS indicator, there are indeed no changes. In the case of the CPP/FCSm indicator, however, research group A's value decreases from 1.67 to 1.33 while research group B's value increases from 1.53 to 1.87 (see Table 5). This seems a counterintuitive result. Although nothing substantive has changed, the ranking of the two research groups according to the CPP/FCSm indicator has reversed.

Table 4. Number of publications (P) and citations (C) of research groups A and B in subfields X and Y. (Modified version of Table 1.)

	Expected cit. per pub.	Research group A	Research group B
Subfield X	10	P = 100, C = 1000	P = 100, C = 2200
Subfield Y	5	P = 100, C = 1000	P = 100, C = 600

Table 5. Values of the CPP/FCSm indicator for research groups A and B. (Modified version of Table 2.)

	Research group A	Research group B
Subfield X	1.00	2.20
Subfield Y	2.00	1.20
Both subfields together	1.33	1.87

4. Example 2

We now turn to another fictitious example that demonstrates some of the differences between the CPP/FCSm indicator and the MNCS indicator. The example also illustrates the policy relevant consequences of the differences between the indicators. Suppose the faculty of natural sciences of some university finds itself in the following situation. The faculty is doing research in two broad fields, chemistry and physics. (For simplicity, we do not break down these fields into subfields.) When differences among fields are corrected for, the chemists and the physicists working at

the faculty turn out to perform equally well. This can be seen from the second and third column of Table 6.

Table 6. Number of publications (P) and citations (C) of the chemists and the physicists in the current situation and in two future scenarios.

	Expected cit. per pub.	Current situation	Scenario 1	Scenario 2
Chemistry	5	P = 100, C = 500	P = 100, C = 900	P = 100, C = 500
Physics	10	P = 100, C = 1000	P = 100, C = 1000	P = 100, C = 1600

To increase the performance of the faculty, a limited amount of money is available. The faculty wants to invest this money in new equipment for either the chemists or the physicists. The new equipment is expected to increase the average performance of the publications of the faculty. The expected effect is shown in the last two columns of Table 6. Scenario 1 shows what is expected to happen if the money is invested in new equipment for the chemists, and scenario 2 shows what is expected to happen if the money is invested in new equipment for the physicists.

Taking into account that in physics the expected number of citations of a publication is twice as high as in chemistry, it seems that an investment in new equipment for the chemists is preferable over an investment in new equipment for the physicists.[4] However, if an investment decision is made based on the expected effect on the overall CPP/FCSm indicator of the faculty, the available money will be invested in new equipment for the physicists. This can be seen in Table 7. It follows from this that the way in which the CPP/FCSm indicator reflects the effects of the two investment opportunities does not seem completely satisfactory.

Table 7. Values of the CPP/FCSm indicator in the current situation and in two future scenarios.

	Current situation	Scenario 1	Scenario 2
Chemistry	1.00	1.80	1.00
Physics	1.00	1.00	1.60
Both fields	1.00	1.27	1.40

Suppose now that an investment decision is made based on the expected effect on the overall MNCS indicator of the faculty. As can be seen in Table 8, the available money will then be invested in new equipment for the chemists. Given the information that is available, this indeed seems the best decision.

Table 8. Values of the MNCS indicator in the current situation and in two future scenarios.

	Current situation	Scenario 1	Scenario 2
Chemistry	1.00	1.80	1.00
Physics	1.00	1.00	1.60
Both fields	1.00	1.40	1.30

---

[4] An investment in new equipment for the physicists yields a larger increase in the absolute number of citations of the faculty than an investment in new equipment for the chemists (600 vs 400). However, when looking at the relative number of citations (i.e., the number of citations after correcting for field differences), an investment in new equipment for the chemists has a larger effect than an investment in new equipment for the physicists (400 / 5 = 80 vs 600 / 10 = 60).

Why does the CPP/FCSm indicator favor an investment in new equipment for the physicists over an investment in new equipment for the chemists? This is again due to the bias of the CPP/FCSm indicator towards fields with a high expected number of citations. In the above example, publications in physics have a higher expected number of citations than publications in chemistry. In the calculation of the CPP/FCSm indicator, publications in physics are therefore overweighted compared with publications in chemistry. As a consequence, a relatively small increase in the performance of publications in physics can lead to a relatively large increase of the CPP/FCSm indicator.

5. Consistency of indicators

In this section, we study the consistency of our indicators of interest. Consistency is a mathematical property that bibliometric indicators may or may not have. In earlier research (Waltman & Van Eck, 2009a, 2009b), it was pointed out that the well-known h-index (Hirsch, 2005) does not have the property of consistency.

We first introduce some mathematical notation. Let the multiset S be given by $S = \{(c_1, e_1), \ldots, (c_n, e_n)\}$, where n is a positive integer, $c_1, \ldots, c_n$ are non-negative integers, and $e_1, \ldots, e_n$ are positive rational numbers.[5] S denotes a set of n publications, and c_i and e_i denote, respectively, the actual and expected number of citations of publication i in this set. Let Σ be defined as the set of all multisets S. Hence, Σ denotes the set of all possible (non-empty) sets of publications. In this paper, we define a bibliometric indicator as a function from Σ to the set of non-negative rational numbers.

We make a distinction between on the one hand consistency of indicators of the average performance of a set of publications and on the other hand consistency of indicators of the total performance of a set of publications. We first consider the latter type of consistency. We define this type of consistency as follows.

Definition 1. Let f denote an indicator of the total performance of a set of publications. f is said to be *consistent* if and only if

$$f(S_1) \geq f(S_2) \Leftrightarrow f(S_1 \cup \{(c,e)\}) \geq f(S_2 \cup \{(c,e)\}) \qquad (7)$$

for all $S_1, S_2 \in \Sigma$, all non-negative integers c, and all positive rational numbers e.

Informally, the definition states that an indicator of total performance is consistent if adding the same publication to two different sets of publications never changes the way in which the indicator ranks the sets of publications relative to each other. This idea of consistency was also discussed by Waltman and Van Eck (2009a, 2009b). A similar idea was discussed by Marchant (2009a, 2009b), who referred to it as independence rather than consistency.

It seems very natural to expect that an indicator of total performance is consistent. It can be readily seen that the TNCS indicator defined in (6) is indeed consistent. However, the brute force indicator defined in (5) is not consistent. To see this, consider the following example. Let $S_1 = \{(3, 1)\}$ and $S_2 = \{(12, 6)\}$, and suppose a publication with $(c, e) = (0, 2)$ is added to both S_1 and S_2. Before adding the publication, the brute force indicator has a value of 3 for S_1 and 2 for S_2. After adding

---

[5] Since S is a multiset rather than an ordinary set, the elements of S need not be unique. Hence, it is possible that $(c_i, e_i) = (c_j, e_j)$ for $i \neq j$.

the publication, the brute force indicator has a value of 2 for S_1 and 3 for S_2. Hence, adding the same publication to both S_1 and S_2 causes a reversal of the way in which the brute force indicator ranks the two sets of publications. This shows the inconsistency of the brute force indicator.[6]

We now turn to consistency of indicators of the average performance of a set of publications. For indicators of average performance, we use a slightly different definition of consistency than for indicators of total performance.

Definition 2. Let f denote an indicator of the average performance of a set of publications. f is said to be *consistent* if and only if

$$f(S_1) \geq f(S_2) \Leftrightarrow f(S_1 \cup \{(c,e)\}) \geq f(S_2 \cup \{(c,e)\}) \qquad (8)$$

for all S_1, $S_2 \in \Sigma$ such that $|S_1| = |S_2|$ and for all non-negative integers c and all positive rational numbers e.

According to this definition, an indicator of average performance is consistent if adding the same publication to two different but equally large sets of publications never changes the way in which the indicator ranks the sets of publications relative to each other. A similar idea, referred to as independence rather than consistency, was discussed by Bouyssou and Marchant (2010).

Like for indicators of total performance, consistency also seems an appealing property for indicators of average performance. It is not difficult to see that the MNCS indicator indeed has the property of consistency. The CPP/FCSm indicator, however, does not have this property. This can be seen using the same example as given above for the brute force indicator. In this example, adding the same publication to both S_1 and S_2 causes the value of the CPP/FCSm indicator to decrease from 3 to 1 for S_1 and from 2 to 3/2 for S_2. Hence, adding the publication leads to a reversal of the ranking of S_1 and S_2 relative to each other. This violates the property of consistency.

Are there, apart from the MNCS indicator, any other indicators of average performance that normalize for differences among fields and that are also consistent? The following theorem provides a negative answer to this question.

Theorem 1. Let f denote an indicator of the average performance of a set of publications. For all $S = \{(c_1, e_1), \ldots, (c_n, e_n)\} \in \Sigma$ such that $e_1 = \ldots = e_n = e$, let $f(S)$ be equal to

$$f(S) = \frac{\sum_{i=1}^{n} c_i / n}{e}. \qquad (9)$$

f is then consistent if and only if f is the MNCS indicator defined in (2).

A proof of the theorem is provided in the appendix. The theorem can be interpreted as follows. For any indicator of average performance that normalizes for differences among fields, it is reasonable to require that, when the indicator is calculated for a set

---

[6] Notice also that adding the publication leads to a decrease of the value of the brute force indicator for S_1. It seems natural to expect that an indicator of total performance never decreases when a publication is added (Waltman & Van Eck, 2009a). However, as the example shows, the brute force indicator does not have this property.

of publications that all belong to the same field, the indicator equals the average number of citations per publication divided by the field's expected number of citations per publication. Given this requirement, there turns out to be only one indicator of average performance that normalizes for differences among fields and that is also consistent. This indicator is the MNCS indicator.

6. How to handle overlapping fields?

In the previous sections, we have shown that the MNCS indicator has attractive theoretical properties. In this section, we therefore focus exclusively on the MNCS indicator. We study how the indicator should be calculated in the case of overlapping fields.

A nice property that we would like the MNCS indicator to have is that the indicator has a value of one when calculated for the set of all publications published in all fields. If there are no publications that belong to more than one field, it is easy to see that the MNCS indicator indeed has this property. However, at CWTS we normally define fields based on subject categories in the Web of Science database and these subject categories are overlapping. Many publications therefore belong to more than one field. Special care then needs to be taken to ensure that the MNCS indicator has the above-mentioned property.

Consider the following example. Suppose the scientific universe consists of just three fields, field X, field Y, and field Z, and suppose just five publications have been published in these fields during a certain time period. For each publication, the field in which it has been published as well as the number of citations it has received is listed in Table 9. Notice that publication 5 belongs both to field X and to field Y. Hence, fields X and Y are overlapping.

Table 9. Overview for each publication of the field in which it has been published and the number of citations it has received.

	Field	Citations
Publication 1	X	2
Publication 2	X	3
Publication 3	Y	8
Publication 4	Z	6
Publication 5	X and Y	5

Because publications 1, 2, 3, and 4 each belong to only one field, it is straightforward to calculate their expected number of citations. Publications 1 and 2 belong to field X, and their expected number of citations therefore equals the average number of citations of all publications published in field X. This yields

$$e_1 = e_2 = \frac{2+3+5/2}{1+1+1/2} = 3. \tag{10}$$

As can be seen, publication 5 has a weight of 1/2 in this calculation. This is because publication 5 belongs half to field X and half to field Y. The expected number of citations of publication 3 is given by

$$e_3 = \frac{8+5/2}{1+1/2} = 7, \tag{11}$$

where publication 5 again has a weight of 1/2. Obviously, for publication 4 we obtain $e_4 = 6$.

How should the expected number of citations of publication 5 be calculated? One approach is to take the arithmetic average of (10) and (11). This results in $e_5 = 5$. Calculating the value of the MNCS indicator for the set of all publications published in all fields, we then obtain

$$\text{MNCS} = \frac{1}{5}\left(\frac{2}{3} + \frac{3}{3} + \frac{8}{7} + \frac{6}{6} + \frac{5}{5}\right) = \frac{101}{105}. \qquad (12)$$

Notice that the MNCS indicator does not have a value of one. This means that the property formulated at the beginning of this section is violated. Because of this, calculating the expected number of citations of publication 5 by taking the arithmetic average of (10) and (11) does not seem a completely satisfactory approach.

We now discuss an alternative approach that does yield satisfactory results. The calculations in (10) and (11) are based on the idea that publication 5 belongs half to field X and half to field Y. The same idea can also be applied in the calculation of the MNCS indicator. This results in

$$\text{MNCS} = \frac{1}{5}\left(\frac{2}{3} + \frac{3}{3} + \frac{8}{7} + \frac{6}{6} + \frac{1}{2}\frac{5}{3} + \frac{1}{2}\frac{5}{7}\right) = 1. \qquad (13)$$

In this case, the MNCS indicator does have the desired value of one. An equivalent way to obtain this result is to calculate the expected number of citations of publication 5 as the harmonic (rather than the arithmetic) average of (10) and (11). We then have

$$e_5 = \frac{2}{1/3 + 1/7} = \frac{21}{5}, \qquad (14)$$

which gives

$$\text{MNCS} = \frac{1}{5}\left(\frac{2}{3} + \frac{3}{3} + \frac{8}{7} + \frac{6}{6} + \frac{5}{21/5}\right) = 1. \qquad (15)$$

The use of harmonic averages ensures that the MNCS indicator always has a value of one when calculated for the set of all publications published in all fields. This therefore seems the most appropriate approach to deal with overlapping fields. The approach leads to a convenient interpretation of the MNCS indicator. When the indicator has a value above one, one's publications on average perform above world average. When the indicator has a value below one, one's publications on average perform below world average. As shown above, this interpretation is not valid when arithmetic rather than harmonic averages are used in the calculation of the MNCS indicator.

7. Conclusions

We have presented a theoretical comparison between two normalization mechanisms for bibliometric indicators of research performance. One normalization

mechanism is implemented in the CPP/FCSm indicator, also referred to at CWTS as the crown indicator. The other normalization mechanism is implemented in what we call the MNCS indicator. The examples that we have given show that the CPP/FCSm indicator sometimes yields counterintuitive results, which is not the case for the MNCS indicator. The counterintuitive results of the CPP/FCSm indicator are due to the unequal weighing of publications from different fields. Unlike the MNCS indicator, the CPP/FCSm indicator gives more weight to publications from fields with a high expected number of citations. We have also studied the consistency of both the CPP/FCSm indicator and the MNCS indicator. Consistency is a mathematical property based on the idea that a ranking should not change when everyone makes the same improvement. As we have pointed out, the MNCS indicator is consistent whereas the CPP/FCSm indicator is not. This is another reason why we consider the MNCS indicator preferable over the CPP/FCSm indicator. Finally, we have discussed how overlapping fields should be dealt with in the case of the MNCS indicator. Contrary to what one might expect, harmonic rather than arithmetic averages should be used to calculate the expected number of citations of a publication that belongs to multiple fields.

Based on the findings reported in this paper, CWTS is currently planning to adopt the MNCS indicator as its new crown indicator. However, there are some issues that still need to be addressed. In particular, the question needs to be answered whether the normalization mechanism of the MNCS indicator should be used not only for normalizing for differences among fields but also for normalizing for differences among document types and for differences among publications of different ages. In the latter case, an important issue is the way in which very recent publications should be handled. A very recent publication (e.g., less than one year old) usually has a rather low expected number of citations (quite close to zero in many cases). Hence, even if such a publication has been cited only once, the ratio of its actual number of citations and its expected number of citations may already be quite high. Because of this, very recent publications may cause the MNCS indicator to become unstable. This may be regarded as undesirable. There is a similar issue in the case of publications of document type letter. These publications typically also have a low expected number of citations and may therefore also cause the MNCS indicator to become unstable. We are currently working on an empirical paper in which issues such as these will be investigated in detail. In this paper, we will also present an extensive empirical comparison between the CPP/FCSm indicator and the MNCS indicator.

References

Bouyssou, D., & Marchant, T. (2010). *Bibliometric rankings of journals based on impact factors: An axiomatic approach*. Retrieved March 10, 2010, from http://users.ugent.be/~tmarchan/IFvNM.pdf.

Braun, T., & Glänzel, W. (1990). United Germany: The new scientific superpower? *Scientometrics*, *19*(5–6), 513–521.

Campbell, D., Archambault, E., & Côté, G. (2008). *Benchmarking of Canadian Genomics – 1996–2007*. Retrieved March 10, 2010, from http://www.science-metrix.com/pdf/SM_Benchmarking_Genomics_Canada.pdf.

CWTS (n.d.). *The Leiden Ranking 2008*. Retrieved March 10, 2010, from http://www.cwts.nl/ranking/.

De Bruin, R.E., Kint, A., Luwel, M., & Moed, H.F. (1993). A study of research evaluation and planning: The University of Ghent. *Research Evaluation*, *3*(1), 25–41.

Egghe, L., & Rousseau, R. (1996). Averaging and globalising quotients of informetric and scientometric data. *Journal of Information Science*, *22*(3), 165–170.

Glänzel, W., Thijs, B., Schubert, A., & Debackere, K. (2009). Subfield-specific normalized relative indicators and a new generation of relational charts: Methodological foundations illustrated on the assessment of institutional research performance. *Scientometrics*, *78*(1), 165–188.

Hirsch, J.E. (2005). An index to quantify an individual's scientific research output. *Proceedings of the National Academy of Sciences*, *102*(46), 16569–16572.

Lundberg, J. (2007). Lifting the crown—citation *z*-score. *Journal of Informetrics*, *1*(2), 145–154.

Marchant, T. (2009a). An axiomatic characterization of the ranking based on the *h*-index and some other bibliometric rankings of authors. *Scientometrics*, *80*(2), 327–344.

Marchant, T. (2009b). Score-based bibliometric rankings of authors. *Journal of the American Society for Information Science and Technology*, *60*(6), 1132–1137.

Moed, H.F. (2005). *Citation analysis in research evaluation*. Springer.

Moed, H.F., De Bruin, R.E., & Van Leeuwen, T.N. (1995). New bibliometric tools for the assessment of national research performance: Database description, overview of indicators and first applications. *Scientometrics*, *33*(3), 381–422.

Opthof, T., & Leydesdorff, L. (in press). Caveats for the journal and field normalizations in the CWTS ("Leiden") evaluations of research performance. *Journal of Informetrics*.

Rehn, C., & Kronman, U. (2008). *Bibliometric handbook for Karolinska Institutet*. Retrieved March 10, 2010, from http://ki.se/content/1/c6/01/79/31/bibliometric_handbook_karolinska_institutet_v_1.05.pdf.

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.

SCImago Research Group (2009). *SCImago Institutions Rankings (SIR): 2009 world report*. Retrieved March 10, 2010, from http://www.scimagoir.com/pdf/sir_2009_world_report.pdf.

Thomson Reuters (2008). *Using bibliometrics: A guide to evaluating research performance with citation data*. Retrieved March 10, 2010, from http://science.thomsonreuters.com/m/pdfs/325133_thomson.pdf.

Van Raan, A.F.J. (2005). Measuring science: Capita selecta of current main issues. In H.F. Moed, W. Glänzel, & U. Schmoch (Eds.), *Handbook of quantitative science and technology research* (pp. 19–50). Springer.

Van Raan, A.F.J., Van Leeuwen, T.N., Visser, M.S., Van Eck, N.J., & Waltman, L. (2010). *Rivals for the crown: Reply to Opthof and Leydesdorff*. Manuscript submitted for publication.

Vinkler, P. (1986). Evaluation of some methods for the relative assessment of scientific publications. *Scientometrics*, *10*(3–4), 157–177.

Waltman, L., & Van Eck, N.J. (2009a). A taxonomy of bibliometric performance indicators based on the property of consistency. In B. Larsen, & J. Leta (Eds.), *Proceedings of the 12th International Conference on Scientometrics and Informetrics* (pp. 1002–1003).

Waltman, L., & Van Eck, N.J. (2009b). A simple alternative to the *h*-index. *ISSI Newsletter*, *5*(3), 46–48.

Appendix

In this appendix, we provide a proof of Theorem 1. Proving sufficiency is trivial. We therefore focus on proving necessity.

Let f denote an indicator of average performance. Let f be field normalized. Hence, f equals (9) for all $S = \{(c_1, e_1), \ldots, (c_n, e_n)\} \in \Sigma$ such that $e_1 = \ldots = e_n = e$. Let f also be consistent.

We first prove that

$$f(S \cup \{(c,e)\}) = \frac{|S| \, f(S) + c/e}{|S| + 1} \tag{16}$$

for all $S \in \Sigma$, all non-negative integers c, and all positive rational numbers e. Let α and β be non-negative integers such that $f(S)e = \alpha / \beta$. Since $f(S)$ and e are non-negative rational numbers, α and β are guaranteed to exist. Let $S' \in \Sigma$ denote a multiset of $|S|$ identical elements $(\alpha, \beta e)$. f is field normalized, and it therefore follows from (9) that $f(S') = f(S)$. Consistency of f then implies that $f(S' \cup \{(c, e)\}) = f(S \cup \{(c, e)\})$. Due to field normalization, $f(\{(c, e)\}) = f(\{(\beta c, \beta e)\})$, which means that, as a consequence of consistency, $f(S' \cup \{(c, e)\}) = f(S' \cup \{(\beta c, \beta e)\})$. Again due to field normalization, $f(S' \cup \{(\beta c, \beta e)\})$ equals (16). Hence, (16) has been proven.

It is now straightforward to prove that f is the MNCS indicator defined in (2). For $|S| = 1$, field normalization implies that $f(S)$ equals (2). For $|S| > 1$, mathematical induction using (16) implies that $f(S)$ equals (2). This completes the proof of the theorem.