Letter to the Editor

Professor John Mingers
Kent Business School, University of Kent
Canterbury, Kent CT2 7PE, UK
+44 1227824008
j.mingers@kent.ac.uk

Dear Sir

Problems with the SNIP Indicator[1]

Introduction

As is well known, citation practices differ across academic fields, especially between the science, social science and arts and humanities domains. Thus, when using citations as a measure of research impact, whether for journals, individuals or departments/institutions, it is necessary to normalise the raw data to the general citation potential (Garfield, 1972; Garfield, 1979) of the research area. Traditionally, this has been done by normalising to the number of *citations* generally in the field ("cited side" normalisation (Schubert & Braun, 1986). The most widely used method, developed by CWTS (van Raan, 2003; van Raan, 2005; van Raan et al., 2007; Moed, 2010a), known as the Leiden methodology, compared citation rates per paper to the mean of such rates across a defined research field. Typically, the list of journals in the field was provided by Web of Science (Mingers and Lipitakis, 2013).

More recently, an approach was developed (Moed 2010b; Zitt & Small, 2008) that normalised against the mean number of *references* per citing paper in the field ("citing side" normalisation). It was also innovatory in not using a pre-defined definition of the relevant field. Rather, the journal's [2] subject field "is defined as the collection of *papers* citing that journal" (Moed, 2010b, p 267). More specifically, the subject field is the set of papers that, in a particular year, cite at least one paper in the journal in the preceding ten years. (p. 275)[3]. This approach is called "source-normalised impact per paper (SNIP). This approach was developed in conjunction with Elsevier and was implemented in their Scopus database for all journals contained in it. It is essentially a normalised impact factor for a journal and is becoming widely used.

However, problems with this indicator were highlighted by Leydesdorff (2013; 2010; 2010) (discussed below) and CWTS itself recognised other difficulties. This led to a revised version of SNIP being presented in 2013 (Waltman et al., 2013). The purpose of this letter is to point out problems with SNIP both in its revised form and the older form.

[1] I would like to acknowledge helpful comments from Loet Leydesdorff
[2] We shall restrict ourselves to journal evaluation in this letter
[3] There is some confusion here as in a later paper revising the definition of SNIP (Waltman et al., 2013) it is stated that the period used was eight years – see note 8 p. 276.

The SNIP indicator

We wish to measure the impact of a particular journal in terms of its citations per paper. The original SNIP is based on a time frame that uses the citations made in one year (say 2010) to papers published in the journal in the preceding three years (2007-2009). The "raw impact per paper" (RIP) is the number of citations made in 2010 to the papers published in 2007-9, divided by the number of cited papers. This is essentially a three-year impact factor[4].

This is normalised with reference to the journals subject field, that is, all the papers in 2010 that cite papers published in the journal in the preceding ten year. For each of these papers, the number of references, *to publications in the relevant database (i.e.., Scopus)*, is counted and these figures are then averaged using the arithmetic mean. The result, the average number of active references in the subject field, is called the "database citation potential" (DCP) of the journal's subject field.

The next step is to normalise the DCP to become the "Relative DCP" (RDCP). The DCP can be calculated for every single journal within the database and there will be quite a large range of values. The journal with the *median* DCP is such that 50% of journals will have a larger DCP, and 50% a lower one. To relativise the target DCP it is divided by the median DCP. Thus

$$RDCP_j = \frac{DCP_j}{DCP_{med}} \quad \text{for journal j}$$

So, for journals in fields with high reference rates RDCP will be above 1; in fields with low citation density RDCP will be below 1.

Finally, SNIP is defined as the RIP divided by RDCP. In high density fields, where RDCP is above 1, the RIP will be reduced. In low density fields it will be increased. The RIP of the median journal will remain the same.

Problems with SNIP and the Revised Version

Leydesdorff and Opthof (2010) pointed out two problems with SNIP. As at metric it is complex – the numerator is a mean of numbers of citations and the denominator is a ratio of the mean of a number of references divided by the median of that mean.

- The first problem, which has been discussed in connection with cited side normalisation (Leydesdorff, 2008; Opthof and Leydesdorff, 2010), is that the order of calculations is wrong. In SNIP, numbers of citations/references are first aggregated and then normalised by division whereas mathematically division should be carried out before addition;
- This has the result that it is not possible to carry out statistical tests for the significance of differences in results.

Waltman et al (2013) show two further problems:

- That in some circumstances, if a journal receives an extra citation this may actually have a negative effect on its SNIP value;

[4] SNIP generally only uses paper types of article, conference paper or review.

- That SNIP is not consistent in that if two journals, with different SNIP values, were combined together into one, we would expect the combined SNIP to be somewhere between the two individual ones, but this is not always the case.

These problems led to the specification of a revised version of SNIP. The main differences are:

1. The DCP values (but not the RIP values) are calculated using the harmonic mean rather than the arithmetic mean of the number of references in citing papers;
2. DCP now takes into account the proportion of citing publications that have at least one active reference;[5]
3. The use of the median DCP, and thus RCDP, is abandoned;
4. The time window for defining the journals subject field is brought down to three years, the same as the citation window.

These changes, especially 1 and 3, are not minor but major and change the whole basis of the normalisation. It might therefore have been better to change the name of the indicator.

An Illustration of the New SNIP

In order to understand the effects of these changes, we will construct a simple example.

Cited papers in journal j over years t, t+1, t+2	**No. of citations received from year t+3 cites**
ed1	1
ed2	2
ed3	3
ed4	4
ed5	5
ed6	6
n (no of cites)	21
m (no. of papers)	6
RIP	3.5

Table 1 Cited Papers in Journal j

Table 1 shows a set of papers from journal j over three years with the citations they received. Six papers, 21 cites, giving an RIP of 3.5.

Citing papers in year t+3	**Paper cited**	**No. of references (r)**	**p**	**1/pr**
ing1	ed1	4	1	0.25
ing2	ed2	4	1	0.25
ing3	ed2	4	1	0.25

[5] This will not be discussed in this paper

ing4	ed3	4	1	0.25
ing5	ed3	4	1	0.25
ing6	ed3	4	1	0.25
ing7	ed4	4	1	0.25
ing8	ed4	4	1	0.25
ing9	ed4	4	1	0.25
ing10	ed4	4	1	0.25
ing11	ed5	4	1	0.25
ing12	ed5	4	1	0.25
ing13	ed5	4	1	0.25
ing14	ed5	4	1	0.25
ing15	ed5	4	1	0.25
ing16	ed6	4	1	0.25
ing17	ed6	4	1	0.25
ing18	ed6	4	1	0.25
ing19	ed6	4	1	0.25
ing20	ed6	4	1	0.25
ing21	ed6	4	1	0.25
Total		**84**		**5.25**

Table 2. Citing Papers constituting the Journal's Subject Field

Table 2 shows the 21 citing papers which thus constitute the journal's subject field. Initially, in this example, all the papers have the same number of active references – 4.

The harmonic mean is defined as the reciprocal of the average of the reciprocal of the numbers

$$Mean\ (harmonic) = \frac{n}{\frac{1}{p_1 r_1} + \cdots + \frac{1}{p_n r_n}}$$

where n = number of publication in the subject field, r_i = the number of references in the ith publication and p_i = proportion of publications with an active reference

For this data the harmonic mean is the same as the arithmetic mean, i.e., 4.

Calculating the old value of SNIP, we have:

RIP = 3.5 and DCP = 4 (ie the arithmetic mean of the references)

And if we assume for simplicity that this journal is in fact the median journal, then

DCP_{med} = 4 and RDCP = 1

Thus $SNIP_{Old}$ = RIP/RDCP = 3.5/1 = 3.5

The new version of SNIP is defined as RIP/DCP without the normalising effect of the median DCP. But the DCP is not just the harmonic mean, it is in fact:

DCP = 1/3 x harmonic mean = (1/3) x4 = 1.333

The reason for the 1/3 is not explained very clearly in Waltman et al (2013) but it is said to be so that the average SNIP value for all the journals in the database should be approximately 1 (p. 276). Later (p. 283), this correction then appears to be needed because there are three cited years and only one citing year.

Thus $SNIP_{New}$ =RIP/DCP = 3.5/1.333 = 2.63

Problem 1 – The basis of the normalisation

What is interesting to note from this example is that, with the old SNIP, the RCDP is 1 and so SNIP is the same as the RIP (3.5). However the new value (2.63) is not the same and the RIP is reduced. This is recognised in the paper by Waltman et al as they state that in their empirical tests the old value is 26% higher than the new value. This would seem to make the resulting figure somewhat arbitrary. The two mechanisms are essentially the same – the ratio of RIP to DCP is multiplied by a constant that is the same for all journals – the median DCP in the old measure and "3" in the new measure, so this is only a scaling factor but at least with the old version it is anchored to a particular value – the median DCP – and so the RIP for the median journal remains the same.

Problem 2 – The use of the harmonic mean

A more important problem comes with the use of the harmonic mean. The arithmetic mean depends only on the sum of the values and how many there are. However, the harmonic mean also depends on the spread or dispersion of the values.

We can illustrate this with Table 3. Here, the total number of references remains the same (84) but we have made them very unevenly spread. The harmonic mean here is 1.05 (compared with 4 before) and the value of the new SNIP is 10.01 (compared with 2.63), a huge difference.

Citing papers in year t+3	**Paper cited**	**No. of references (r)**	**p**	**1/pr**
ing1	ed1	1	1	0.25
ing2	ed2	1	1	0.25
ing3	ed2	1	1	0.25
ing4	ed3	1	1	0.25
ing5	ed3	1	1	0.25
ing6	ed3	1	1	0.25
ing7	ed4	1	1	0.25
ing8	ed4	1	1	0.25
ing9	ed4	1	1	0.25
ing10	ed4	1	1	0.25

ing11	ed5	1	1	0.25
ing12	ed5	1	1	0.25
ing13	ed5	1	1	0.25
ing14	ed5	1	1	0.25
ing15	ed5	1	1	0.25
ing16	ed6	1	1	0.25
ing17	ed6	1	1	0.25
ing18	ed6	1	1	0.25
ing19	ed6	1	1	0.25
ing20	ed6	1	1	0.25
ing21	ed6	64	1	0.015625
Total		**84**		**20.01563**

Table 3. Citing Papers constituting the Journal's Subject Field with High Dispersion

Thus we find that with the new SNIP the result can be significantly dependent on the dispersion of the number of references as well as their volume. We cannot see any justification for that – the point of normalisation is to make allowance for the volume of activity not its degree of variability.

Problem 3 – The overall effectiveness of the normalisation

As stated above, the point of normalisation is to make allowance for variations in citation potential, and generally this is particularly true in comparisons of science with social science and arts and humanities. If a normalisation method is successful, we would therefore expect that journals from these three domains would be put on an equal footing. However, data provided in Waltman et al's paper (2013) demonstrates that this is not the case. Tables 3 and 4 of that paper show the 30 journals with the highest values of the old and new SNIP. What is striking is that almost without exception the journals are from science. One management journal (Academy of Management Review) appears in both, and one economics journal (Quarterly Journal of Economics) appears in the new SNIP table. Apart from those every journal is either science or technology.

Also, as the authors point out, a major difference between the two tables is the significant number of computing and engineering journals that drop out with the new SNIP although they do not offer an explanation. One possibility would be the effects of the harmonic mean as discussed above. If computing and engineering tended to have references that were less dispersed then the values of the new SNIP would fall.

Overall, we feel that there needs to be considerable empirical work undertaken looking into these effects before metrics like this can be used with confidence. One obvious starting point is to look at the underlying factors that cause the citation differentials. Some of these factors will appertain to the actual domain behaviour, but some is a measurement effect of the citation databases being used. In terms of behaviour, there may well be longer reference lists, more papers produced, and a shorter obsolescence period. For example,

while citation rates for a paper in the science may peak after two or three years, in the social sciences the peak may not occur for six or seven years, and continues for many more (Baumgartner and Leydesdorff, 2014; Mingers, 2008). In terms of databases, many studies have shown that while they have a high coverage of science journals (often around 90%), the coverage for social science may only be around 50%, and for humanities even lower (Mingers and Lipitakis, 2010; Mingers and Lipitakis, 2013). Also, social science and humanities use research monographs much more and these are not included in WoS or Scopus (Larivière et al., 2006).

Conclusions

There were problems with the old version of SNIP, and we have shown that there are also problems with the new version, particularly the effects of using the harmonic mean. The problem is that these metrics, particularly when they are implemented in a major database such as Scopus, can quickly come to prominence without a proper understanding of their limitations. They are also relatively intransparent as they require significant access to the databases in order to test them, and yet this is not generally available.

References

Baumgartner, S. E., & Leydesdorff, L. (2014). Group-based trajectory modeling (GBTM) of citations in scholarly literature: Dynamic qualities of "transient" and "sticky knowledge claims". *Journal of the Association for Information Science and Technology, 65*(4), 797-811.

Garfield, E. (1972). Citation analysis as a tool in journal evaluation. *Science* **178**: 471-479.

Garfield, E. (1979). Is citation analysis a legitimate evaluation tool? *Scientometrics, 1*(4), 359-375.

Larivière, V., Archambault, É., Gingras, Y., & Vignola-Gagné, É. (2006). The place of serials in referencing practices: Comparing natural sciences and engineering with social sciences and humanities. *Journal of the American Society for Information Science and Technology, 57*(8), 997-1004.

Leydesdorff, L. (2008). Caveats for the use of citation indicators in research and journal evaluation. *Journal of the American Society for Information Science and Technology* **59**: 278-287.

Leydesdorff, L. (2013). The revised SNIP indicator of Elsevier's Scopus. *Journal of Informetrics* **7**: 859-860.

Leydesdorff, L. and Opthof, T. (2010). Scopus's source normalized impact per paper (SNIP) versus a journal impact factor based on fractional counting of citations. *Journal of the American Society for Information Science and Technology* **61**: 2365-2369.

Leydesdorff, L. and Opthof, T. (2010). Scopus SNIP indicator: Reply to Moed. *Journal of the American Society for Information Science and Technology* **62**: 214-215.

Leydesdorff, L. and Opthof, T. (2011). Remaining problems with the "New Crown Indicator" (MNCS) of the CWTS. *Journal of Informetrics* **5**: 224-225.

Mingers, J. (2008). Exploring the dynamics of journal citations: modelling with S-curves. *Journal Operational Research Society* **59**: 1013-1025.

Mingers, J. and Lipitakis, E. (2010). Counting the citations: A comparison of Web of Science and Google Scholar in the field of management. *Scientometrics* **85**: 613-625.

Mingers, J. and Lipitakis, E. (2013). .Evaluating a Department's Research: Testing the Leiden Methodology in Business and Management. *Information Processing & Management* **49**: 587-595.

Moed, H. (2010a). CWTS crown indicator measures citation impact of a research group's publication oeuvre. *Journal of Informetrics* **4**: 436-438.

Moed, H. (2010b). Measuring contextual citation impact of scientific journals. *Journal of Informetrics* **4**: 265-277.

Opthof, T., & Leydesdorff, L. (2010). *Caveats* for the journal and field normalizations in the CWTS ("Leiden") evaluations of research performance. *Journal of Informetrics, 4*(3), 423-430.

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

van Raan, A. (2003). The use of bibliometric analysis in research performance assessment and monitoring of interdisciplinary scientific developments. *Technology Assessment - Theory and Practice* **1**: 20-29.

van Raan, A. (2005). Fatal attraction: Conceptual and methodological problems in the ranking of universities by bibliometric methods. *Scientometrics* **62**: 133-143.

van Raan, F., Moed, H. and van Leeuwen, N. (2007). *Scoping study on the use of bibliometric analysis to measure the quality of research in UK higher education institutions.* Report, Centre for Science and Technology Studies, Leiden University, Leiden, http://www.hefce.ac.uk/pubs/rdreports/2007/rd18_07/.

Waltman, L., van Eck, N., van Leeuwen, T. and Visser, M. (2013). Some modifications to the SNIP journal impact indicator. *Journal of Informetrics* **7**: 272-285.

Zitt, M. and Small, H. (2008). Modifying the journal impact factor by fractional citation weighting: The audience factor. *Journal of the American Society for Information Science and Technology* **59**: 1856-1860.