Normalized International Collaboration Score: A Novel Indicator for Measuring International Co-Authorship

Adam Finch[1], Kumara Henadeera[2] and Marcus Nicol[3]

[1] *adam.finch@csiro.au*
CSIRO, Waite Campus, Science Excellence Team, Waite Road, Urrbrae, SA 5064 (Australia)

[2] *kumara.henadeera@nhmrc.gov.au*
National Health & Medical Research Council, Strategic Policy Group, 16 Marcus Clarke Street, Canberra, ACT 2601 (Australia)

[3] *marcus.nicol@arc.gov.au*
Australian Research Council, Research Excellence Branch, 11 Lancaster Place, Canberra ACT 2609 (Australia)

Abstract
International collaboration on research publications is increasingly evaluated as part of a raft of performance measures. Levels of international co-authorship have increased substantially over the last few decades and vary substantially by research field and publication type; however, these variations are not typically accounted for by international collaboration indicators. In this research-in-progress paper, we introduce a novel metric, the Normalised International Collaboration Score, which adjusts the number of countries appearing on publication records using baselines relevant to the subject, age and type of the publication. A pilot analysis shows that these baselines vary substantially and that the application of this metric yields very different results to a more common measure of international collaboration. The limitations of the metric are discussed, along planned extensions for the full version of the study, as well as the relationship between normalised collaboration and citation.

Conference Topic
Indicators

Background and Purpose

Measuring international co-authorship

The availability of author address metadata on publication indices such as Web of Science and Scopus allows the analysis of patterns in co-authorship, including the collaboration by authors from different countries on research outputs. This approach has been used in many studies for decades (such as Glänzel & De Lange, 1997; Narin, Stevens, & Whitlow, 1991; Nederhof & Moed, 1993) and metrics describing international collaboration now appear regularly in bibliometric handbooks (Colledge, 2014; Rehn, Kronman, & Wadskog, 2007) and in reporting tools such as Thomson Reuters' InCites, Elsevier's SciVal and SCImago's Journal & Country Ranking. Such publications tend to receive higher levels of citation, an effect that is not due to the increased propensity for self citation arising from additional authors (Van Raan, 1998), but likely rather shared experience, knowledge and equipment.

Analysis of international co-authorship metadata has highlighted other important aspects of collaboration. Firstly, levels of international collaboration have increased substantially over the last quarter century (Leydesdorff & Wagner, 2008); and secondly, levels of international collaboration vary by field of research (Frame & Carpenter, 1979). A report on Thomson Reuters' InCites (retrieved 7 January 2015) indicates that 2013 articles, reviews and proceedings papers in Tropical Medicine involved international collaboration 46.7% of the time, while for History, this was only 4.3% of the time. Even within Medicine, Emergency Medicine saw only 9.9% foreign collaboration, far lower than Tropical Medicine. Variation is significant over time, with Astronomy & Astrophysics international collaboration rising from 19.4% in 1993 to 45.0% in 2013. To these two aspects, we must add publication type; 2013

Astronomy & Astrophysics articles saw 51.4% international collaboration, but its Proceedings Papers only 0.2%. Such variations exist across the full gamut of subject, years and publication types but most metrics used to evaluate collaboration do not take account of them.

Existing metrics

Frequently, analyses use either the number or proportion of collaborative publications (see for example Boekholt et al., 2009; Colledge, 2014; Luukkonen et al., 1993). Glänzel and De Lange (2002) use a Multilateral Collaboration Index to measure the number of collaborative links compared to the number of collaborative papers, establishing the intensity of collaboration.

Beaudet, Campbell, Côté, Haustein, Lefebvre and Roberge (2014) use a regression model based on power law relationships to establish the expected level of collaboration for a country and an Affinity Index to identify key partners. Degelsegger et al. (2013) propose thematic assessment, normalized either by relating it to the output of the country in the subject, or by comparing it to co-authored output in the same subject but with a different partner. Ding, Yang and Liu (2013) propose using network metrics to evaluate collaboration impact, which is a sound approach within a subject and time frame. Pohl, Warnan and Baas (2014) go the greatest distance to normalizing for the three aforementioned influences, by adjusting the proportion of publications with international collaboration by the number of collaborating countries in each subject. This study only considered a single year, however, did not adjust for publication type and was based on adjusting the share of research with a binary attribute (either internationally collaborative or non-internationally collaborative). The properties and results of this alternative will be compared to our metric in the full version of our study.

The Normalised International Collaboration Score (NICS)

The Normalised International Collaboration Score uses fundamentally the same calculation as the "new" Crown Indicator by which it was inspired (Waltmann, van Eck, van Leeuwen, Visser, & van Raan, 2011). For each publication, a global baseline is constructed, representing the average number of countries contributing to publications of the same type, from the same year and appearing in the same subject area(s). The number of countries contributing to the publication in question is then divided by the relevant baseline to yield a ratio. This ratio is then averaged for all publications in a set (for an institution, country, journal, etc). Our exploratory analysis uses both the mean (as in the Crown Indicator) and the statistically preferable median (Bornmann, Mutz, Neuhaus, & Daniel, 2008), for the purposes of comparison. While the present study only includes a selection of publication types, years and subjects, our full study will include all subjects and publication types back to 1996.

Methodology

The Advanced Search function on Web of Science was used to isolate publications of the Article, Review and Proceedings Paper types with issue cover dates in 1993, 2003 and 2013, and allocated to the subject categories Dance, Engineering (Manufacturing), Evolutionary Biology, Gastroenterology & Hepatology, Political Science, Psychology (Educational), Soil Science and Tropical Medicine. These publication types were selected as those most likely to contain address data; these years as spaced such to demonstrate evolution in collaboration trends and aspects of the data; and those subjects as representing a broad spectrum across science, social science, and the arts and humanities. The selection of a single discipline of period would not have illustrated any variation over time or theme. Record metadata were downloaded, tagged with the relevant subject name and recombined into a single dataset. Individual addresses were broken out, the non-country information in each field deleted and duplicate country entries deleted. A count of unique country contributions per publication was made. The baselines were constructed by averaging all unique country contribution counts for each combination of year, publication type and subject, using the arithmetic mean and the median. These represented the denominator of the metric's article-level ratio.

The institutional data came from a database of Australian publication records from 2001 to 2014. A query extracted the unique identifier, selected subject areas, year, type and Crown Indicator of each publication, along with the author addresses. The addresses were subjected to a unique contributing country count, yielding the numerator of the metric's article-level ratio. The subject, publication type and publication year data were used to look up the mean and median baseline data (our ratio's denominator). Dividing the latter by the former yielded the article-level NICS, which was then averaged for each Australian institution – using the arithmetic mean and then the median, as appropriate for the baseline.

This gives the following notation for the mean form of NICS:

$$\frac{1}{p}\sum_{i=1}^{p}\frac{n_i}{g_i}$$

And the following notation for the median form of NICS:

$$\widetilde{\left(\frac{n_i}{m_i}\right)}$$

Where p denotes the number of publication produced by a unit of analysis, n_i denotes the number of countries contributing to the unit's publication i, g_i denotes the global mean number of countries contributing to publications of the same type, year and subject(s) as publication i and m_i denotes the global median number of countries contributing to publications of the same type, year and subject(s) as publication i. A third, "hybrid", version of was calculated, finding the median of article level ratios based on a mean:

$$\widetilde{\left(\frac{n_i}{g_i}\right)}$$

Results & Discussion

Table 1 shows the mean baselines for each year in each subject, combining publication types into a single entry. Several points are clear. Some subjects see a substantial increase in average country contributions over time – such as the increase from 1.15 to 1.71 for Evolutionary Biology – indicating a need to normalise for this change if fair comparisons are to be made among publication sets from different year ranges. There are also significant disparities between subjects, with the Engineering subject baseline 1.09 in 2013, compared to 1.78 for Tropical Medicine. It is also notable that, unlike citation counts, there does not seem to be a pattern of lower country contributions for social sciences as opposed to sciences, at least in this very limited dataset; Political Science has one of the higher baseline sets and Engineering, Manufacturing one of the lower. Lastly, some subjects, most likely those in the Arts & Humanities, may be difficult to assess using this metric, due to a paucity of address and a low publication count; the baselines would be based on too low a sample size and very prone to skew from outliers. It is also worth noting that, while country contributions are strongly positively skewed, the variance of the natural log of country contribution counts is lower than that of citation counts for publications of the same year, type and subject, in a all of a selection of the below instances that were considered.

Table 2 shows the number of publications missing address data in each of the three years for each subject. Coverage is a problem in Dance for all years and is more of a problem in the social science subjects than the sciences, but is an issue for all subjects in 1993. In the full analysis, work will be conducted to establish the point at which coverage is sufficient for robust analysis, but the institutional analysis in this pilot study exclude the 1993 publications.

Table 1. Mean Subject Country Contribution Baselines by Year.

	1993		*2003*		*2013*	
Table	*# Pubs*	*# Countries*	*# Pubs*	*# Countries*	*# Pubs*	*# Countries*
Dance	2	1.50	25	1.00	46	1.13
Engineering, Manuf.	1242	1.03	7935	1.06	14513	1.09
Evolutionary Biology	987	1.15	3900	1.38	5543	1.71
Gastroent. & Hepat.	3567	1.09	8595	1.15	11300	1.29
Political Science	987	1.15	3172	1.26	5549	1.76
Psychology, Educ.	483	1.05	1167	1.10	2253	1.20
Soil Science	1724	1.09	3890	1.23	4721	1.36
Tropical Medicine	798	1.45	1381	1.68	3128	1.78

Table 2. Instances of Publication Entries Missing Address Data by Year.

	1993			*2003*			*2013*		
Table	*No Address*	*Total Pubs*	*%*	*No Address*	*Total Pubs*	*%*	*No Address*	*Total Pubs*	*%*
Dance	245	247	99.2%	386	411	93.9%	184	230	80.0%
Eng., Manufact.	936	2178	43.0%	587	8522	6.9%	219	14732	1.5%
Evolutionary Biology	698	1685	41.4%	15	3915	0.4%	10	5553	0.2%
Gastro. & Hepat.	2080	5647	36.8%	158	8753	1.8%	68	11368	0.6%
Political Science	3421	4408	77.6%	965	4137	23.3%	636	6185	10.3%
Psych., Education.	573	1056	54.3%	20	1187	1.7%	47	2300	2.0%
Soil Science	1702	3426	49.7%	132	4022	3.3%	16	4737	0.3%
Tropical Medicine	475	1273	37.3%	11	1392	0.8%	24	3152	0.8%

Table 3 shows the mean baselines for each publication type in each subject, combining years into a single entry. It is clear that publication type is also a major factor for the baselines, with the Proceedings Papers consistently seeing fewer country contributions than other types. However, there is further variation; Political Science, for example, sees higher country counts for Articles than Reviews, while the reverse is true for Soil Science.

Table 3. Mean Subject Country Contribution Baselines by Publication Type.

	Articles		*Proceedings*		*Reviews*	
Table	*# Pubs*	*# Countries*	*# Pubs*	*# Countries*	*# Pubs*	*# Countries*
Dance	73	1.10	-	-	-	-
Engineering, Manuf.	9192	1.20	14398	1.00	100	1.23
Evolutionary Biology	9665	1.54	101	1.00	664	1.59
Gastroent. & Hepat.	20482	1.21	811	1.00	2169	1.24
Political Science	8650	1.57	790	1.18	268	1.28
Psychology, Educ.	3542	1.16	263	1.00	98	1.09
Soil Science	8547	1.31	1641	1.01	147	1.69
Tropical Medicine	5113	1.71	23	1.00	171	1.73

Table 4 shows a comparison of institutional collaboration analysis using the proportion of publications with international collaboration and each of the three variants of the NICS metric. The Median calculation appears the least useful; every baseline in each year, subject and document was 1, so this version essentially reports the median country contribution per article and cannot strongly differentiate among institutions. The version using mean baselines are more useful for ranking but, like the Crown Indicator, remains sensitive to outliers (as in the example of Flinders University, where performance was inflated by a single article with 35 contributing countries). Even though the full study will involve far larger sample sizes, which should be less susceptible to such outliers, it appears that the "hybrid" (median of ratios based on mean baselines) is the strongest option. This would preclude statistical analysis based on parametric data, but it is impossible to tell from the pilot study whether the article level results of the mean calculation would be normally distributed on a global scale either.

Table 4. Selected Australian Institution Ranking.

		% Collaboration		*NICS Mean*		*NICS Median*		*NICS 'Hybrid'*	
Table	*Pubs*	*Value*	*Rank*	*Value*	*Rank*	*Score*	*Rank*	*Score*	*Rank*
Queensland Inst Med Res	55	70.9%	1	1.61	3	2	1	1.19	3
James Cook Univ	98	65.3%	2	1.72	1	2	1	1.44	2
Charles Darwin Univ	40	62.5%	3	1.60	4	2	1	1.12	5
Univ Western Sydney	52	57.7%	4	1.55	6	2	1	1.45	1
Univ Western Australia	219	50.7%	6	1.26	15	2	1	1.12	5
Univ Melbourne	233	50.2%	7	1.48	7	2	1	1.12	5
Univ Adelaide	174	49.4%	9	1.24	22	1	9	0.86	19
Univ Sydney	286	48.6%	10	1.38	10	1	9	0.99	13
CSIRO	210	48.6%	11	1.24	21	1	9	0.99	12
Univ Queensland	271	48.3%	12	1.25	18	1	9	0.91	15
Queensland Univ Technol	58	48.3%	13	1.25	16	1	9	1.00	9
Murdoch Univ	45	46.7%	15	1.23	23	1	9	0.79	22
Univ Newcastle	48	45.8%	16	1.29	14	1	9	1.00	10
Australian Natl Univ	203	44.8%	17	1.08	27	1	9	0.79	22
Univ New S Wales	195	44.1%	18	1.24	20	1	9	0.88	16
Monash Univ	155	43.2%	19	1.35	11	1	9	0.88	16
Curtin Univ Technol	44	43.2%	20	1.20	24	1	9	1.00	11
Howard Florey Inst	48	35.4%	26	1.56	5	1	9	0.78	25
Flinders Univ S Australia	70	30.0%	27	1.65	2	1	9	0.78	25

Discussion

While only a few institutions see a large difference in ranking when applying NICS rather than proportion of international publications, the difference in results and the variations in baselines on which they are based suggest the metric has informational content. It is also worth noting that, at an article level, the Crown Indicator correlates positively and fairly strongly with NICS (Spearman's Rank r=0.384) and that at an institutional level, the two versions of NICS derived from mean baselines correlate more closely with NCI performance (r=0.289 and 0.148) than does share of publications with international collaboration (r=0.09).

There are clearly limitations to this approach. It does not account for collaboration intensity; eight co-authoring institutions in a specific foreign country count the same as one. The NICS baselines could be rescaled to count not only contributing foreign countries but also the numbers of institutions in those countries, and even potentially types of institutions. As it would require a set of baselines for each country, this would be computationally intensive but will be explored in the full study. This approach would also normalise for the propensity of a country to collaborate, which many of the above-mentioned metrics are aimed at doing. Lower collaboration levels can arise from several causes, including a lower advantage yielded and having a large share of global output (therefore limiting the avenues available for external collaboration); normalising for national collaboration levels may obscure these differences and render accurate national comparisons challenging. In its present form, NICS serves best as a metric to compare the collaboration of countries and institutions, variations in which may then be considered in the context of national motivation and propensity to collaborate.

Other criticisms leveled at the Crown Indicator apply to NICS, most notably a limited representation of global output in some subjects and of some publication types, and the reliance on a subject taxonomy designed for information retrieval rather than bibliometric analysis. In the pilot study, moreover, many articles analysed here appeared in more than one subject area, and yet were normalised only with the baselines for one of those subject areas.

The full study will apply a wide range of statistical tests to the properties of the baselines, the country contribution counts and the resultant ratios; for now, however, and even with the aforementioned caveats, this metric shows potential for robust and meaningful analysis of institutional and national research collaboration abroad.

References

Beaudet, A., Campbell, D., Côté, G., Haustein, S., Lefebvre, C., & Roberge, G. (2014) *Bibliometric Study in Support of Norway's Strategy for International Research Collaboration.* Report to the Science Council of Norway, Science-Metrix. March 2014.

Boekholt, P., Edler, J., Cunningham, P., & Flanagan, K. (2009). *Drivers of international collaboration in research*. Report to EC, DG Research. Technopolis BV,Netherlands, April 2009.

Bornmann, L., Mutz, R., Neuhaus, C. & Daniel, H-D. (2008). Citation counts for research evaluation: standards of good practice for analyzing bibliometric data and presenting and interpreting results. *Ethics in Science and Environmental Politics, 8*, 93-102.

Colledge, L. (2014). Snowball Metrics Recipe Book, Edition 2. Retrieved December 19, 2014 from: http://www.snowballmetrics.com/wp-content/uploads/snowball-recipe-book_HR.pdf

Degelsegger, A., Lampert, D., Büsel, K., Simon, J., Tschant, J., & Wagner, I. (2013) *Assessing international cooperation in S&T through bibliometric methods. Proc. ISSI*, 175-184.

Ding, J., Yang, L. & Liu, Q. (2013). Measuring the academic impact of researchers by combined citation and collaboration impact. *Proc. ISSI*, 1177-1187.

Frame, J.D., & Carpenter, M.P. (1979). International research collaboration. *Social Studies of Science,* 9, 481-497.

Glänzel, W., & De Lange, C. (1997). Modelling and measuring multilateral co-authorship in international scientific collaboration. Part II. A Comparative study on the extent and change of international scientific collaboration links. *Scientometrics*, *40*, 605-626.

Glänzel, W., & De Lange, C. (2002). A distributional approach to multinationality measures of international scientific collaboration. *Scientometrics*, *54*, 75-89.

Leydesdorff, L. & Wagner, C.S. (2008). International collaboration in science and the formation of a core group. *Journal of Informetrics, 2*, 317-325.

Luukkonen, T., Tijssen, R.J.W., Persson, O., & Sivertson, G. (1993). The measurement of international scientific collaboration. *Scientometrics, 28*, 15-36.

Narin, F., Stevens, K., & Whitlow, E.S. (1991). Scientific cooperation in europe and the citation of multinationally authored papers. *Scientometrics, 21*, 313-323.

Nederhof, A.J., & Moed, H.F. (1993). Modeling multinational publication - development of an online fractionation approach to measure national scientific output. *Scientometrics*, *27*, 39-52.

Pohl, H., Warnan, G., & Baas, J. (2014). Level the playing field in scientific international collaboration with the use of a new indicator: Field-Weighted Internationalization Score. *Research Trends, 9*, 3-8.

Rehn, C., Kronman, U., & Wadskog, D. (2007). *Bibliometric indicators – definitions and usage at Karolinska Institutet*. Retrieved December 19, 2014 from: http://kib.ki.se/sites/kib.ki.se/files/Bibliometric_indicators_definitions_1.0.pdf

van Raan, A.F.J. (1998). The influence of international collaboration on the impact of research results. *Scientometrics*, *43*, 423-428.

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