Scientific Research in the Indian Subcontinent: Comparing Bangladesh, Pakistan and Sri Lanka with India

Dilruba Mahbuba[1,2] and Ronald Rousseau[3,4,5]

[1] *ruba@icddrb.org*
Executive Director's Division, Library and Information Services Unit (LISU)
ICDDR,B; 68 Shahid Tajuddin Sarani, Mohakhali, Dhaka 1212 (Bangladesh)
[2] Antwerp University (UA), IBW, B-2000 Antwerp (Belgium)

[3] *ronald.rousseau@khbo.be*
KHBO (Association K.U.Leuven), Industrial Sciences and Technology, B-8400, Oostende (Belgium)
[4] Hasselt University, Universitaire Campus, B-3590 Diepenbeek (Belgium)
[5] K.U.Leuven, Department of Mathematics, Celestijnenlaan 200B, 3001 Leuven (Heverlee), (Belgium)

Abstract

As part of a research program to analyse research in Bangladesh we provide a comparison between research indicators related to India, Bangladesh, Pakistan and Sri Lanka. In this investigation we make use of Web of Science (WoS) data as well as Scopus data (using the SCImago website). It is shown that the number of publications of these countries is sometimes best described by an exponential curve, sometimes a power law and sometimes a linear relation. Special attention is given to the evolution of country h-indices. It is shown that in relative terms Sri Lanka is the strongest country of the four.

Introduction

We review developments in scientific research between 1973 and 2007 in the four South Asian countries Bangladesh, India, Pakistan and Sri Lanka. How do Pakistan, Bangladesh and Sri Lanka compare with each other and with the local giant, India? This article continues investigations published earlier (Mahbuba & Rousseau, 2008).

We intend to clarify some of these questions using basic scientometric tools, measuring the evolution of scientific research in these four countries during the period 1973-2007 (recall that Bangladesh became independent in 1971). The Web of Science and, to some extent, the SCImago Journal & Country Rank database, based on SCOPUS data (http://www.scimagojr.com/) are our main sources. However, we never tried to pool data from the two databases (WoS and SCOPUS) as this would have to be done on a paper by paper basis.

Several recent articles compare India with China, see e.g. (Arunachalam, 2008; Arunachalam & Viswanathan, 2008; Balaram, 2004; Guan & Ma, 2004). Also studies related to the target countries have been published in the past. Sadana et al. (2004) study health research in the region, including not only Bangladesh, India, Pakistan and Sri Lanka, but also Afghanistan, Nepal, Bhutan and the Maldives. They note that collaboration on health research across the region must be strengthened, e.g. by developing networks of researchers, policymakers and institutions. Because of the large number of people involved (India on its own is the second most populous country in the world) and the high disease burden on the population, health related research is very important for these countries. B.M. Gupta et al. (2002) studied collaborations of India with Bangladesh, Pakistan, Sri Lanka and Nepal in the period 1992-1999, based on WoS data. The strongest collaborative linkages are with Bangladesh. India-Bangladesh collaborative articles have also the highest impact. These authors also found that geographical proximity played an important role. Of the 79 institutional collaborative linkages between India and Bangladesh 32 were between neighbouring West Bengal and Bangladesh.

Proceedings of ISSI 2009, pages 702–713.
Edited by B. Larsen and J. Leta.

The strongest collaboration ties, however, were between ICDDR, B (Dhaka) and NICED (Kolkata). Recently, Glänzel and Gupta (2008) published a study focussing on India's research. However, as it goes with (local) giants, no attention is paid to India's scientific relation to its smaller neighbours Pakistan, Bangladesh and Sri Lanka.

WOS publication data: each country separately

Bangladesh

The yearly number of publications (all types but no conference proceedings) with a Bangladesh address increased during the period 1973-2007 from about 50 (a year) to almost 800, leading to a total of almost 10,400. Another 950 articles in conference proceedings (since 1990) can be added to this total. The yearly number of publications can be described by an exponential function: $y = a.e^{bt}$, where y denotes the number of publications, a = 77.78, b = 0.066, and t(ime) = 0 in 1973 (R^2 for a nonlinear least squares regression is equal to 0.98). This corresponds to a doubling time of about 10.5 years. Figure 1 shows the yearly number of Bangladeshi publications and the best fitting exponential curve. Note that, as the fitted curve begins higher than the observed data and ends lower, the observed doubling time is lower. We collected data on the percentage of 'normal" articles among the published Bangladeshi documents in the WoS. Although this percentage fluctuates between 70 and 90 % there is a clear increasing trend (see Fig. 2).

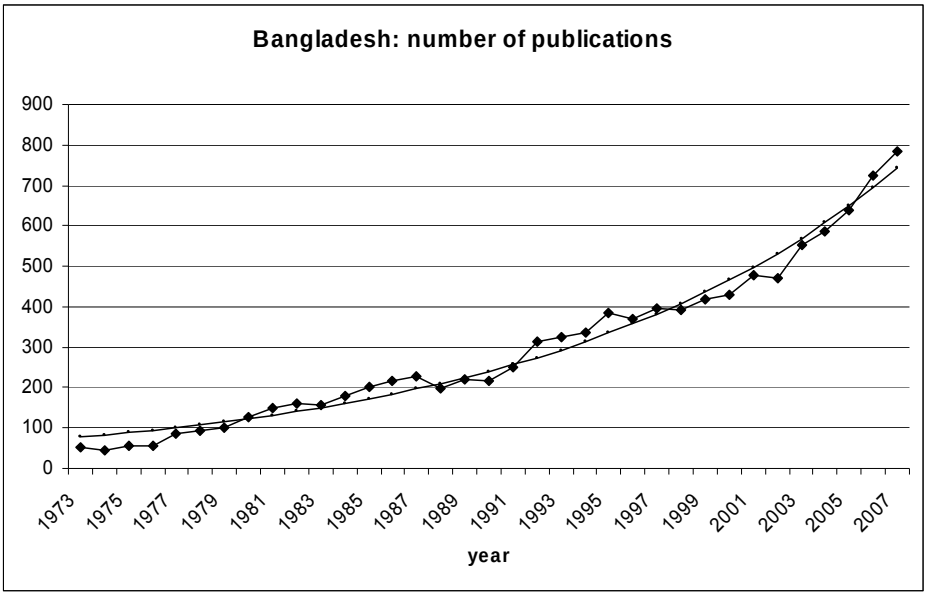

Figure 1. Best fitting exponential curve for Bangladesh (WoS data)

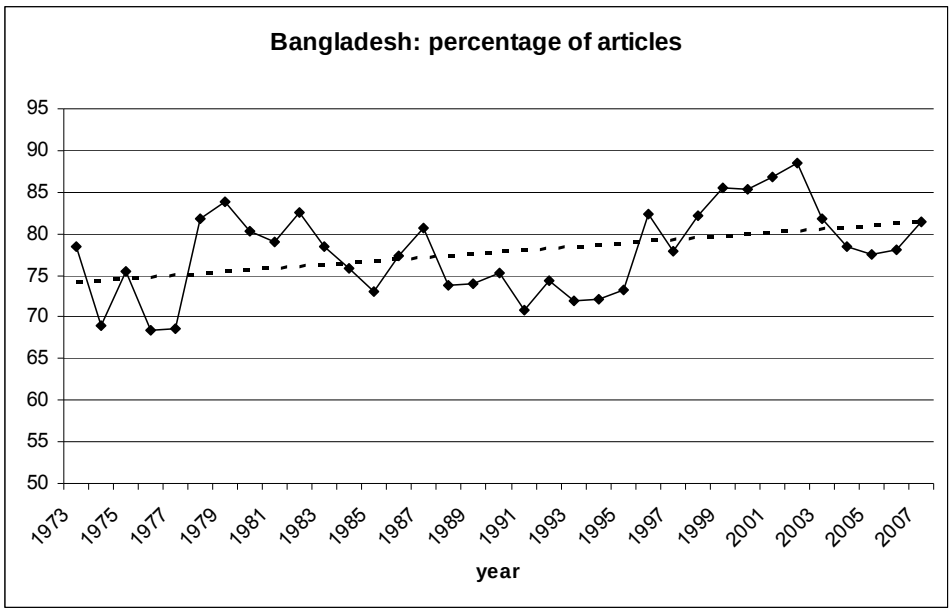

Figure 2. Bangladesh: percentage of 'normal' articles among all publications (WoS data)

We also investigated the percentage of uncited publications (among all documents) and the percentage of uncited 'normal' articles. As expected, the data related to the latter case lies below (less uncited articles) the former, and uncitedness values increase at the end of the period (as publications have less chance to be cited). When we restrict data to the year 2002, we see a clear decreasing trend (see Fig.3).

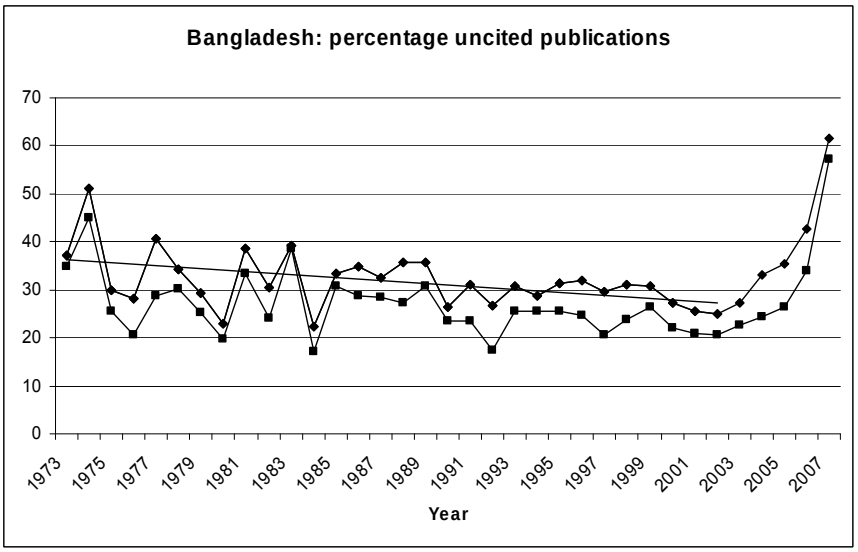

Figure 3. Uncited publications (November 2008): all documents (upper curve) and normal articles (lower one) (WoS data)

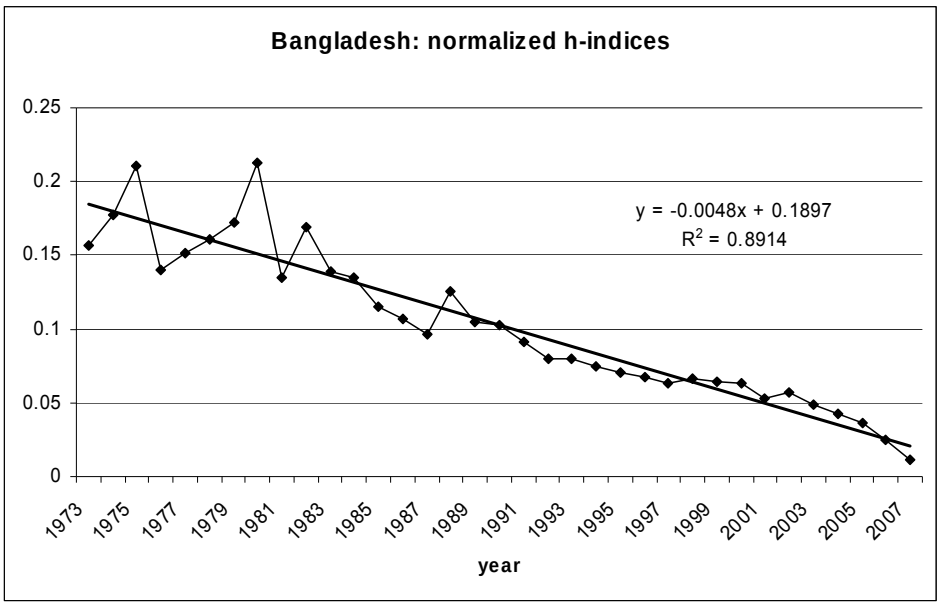

Figure 4. Normalized h-index series based on all publications (WoS data)

The h-index for all Bangladeshi publications over the period 1973-2007 (data collected in November 2008) is 87. We calculated the yearly h-index for all publications (see Fig.4) and for normal articles alone (not shown). When normalizing with respect to the number of publications, i.e. we calculate the h-index corresponding to a particular year and then divide by the number of publications in that year, we see a clear decreasing trend (Fig.4). Such as decreasing trend is expected as the h-index depends on the number of publications and the number of citations. When normalized for the number of publications, only citations remain. As the oldest articles have a longer period than the younger over which to accrue citations a decreasing trend is indeed expected.

Pakistan

The number of publications (all types but not including conference proceedings) with a Pakistan address increased during the period 1973-2007 from about 140 to more than 2,400 a year, leading to a total of more than 19,000. Another 1,772 articles in conferences can be added to this total. Over the period 1973-2000 the yearly number of publications can be described by a linear function (see Fig.5); from then on an exponential function fits best: $y = a.e^{bt}$, where y denotes the number of publications, a = 553.6, b = 0.237, and t= 0 in 2001 ($R^2 = 0.96$ for a nonlinear least squares regression). This corresponds to a doubling time of about 2.9 years. Figure 6 shows the yearly number of Pakistani documents (period 2001-2007) and the best fitting exponential curve.

We also collected data on the percentage of 'normal" articles among the published Pakistani documents in the WoS. Although this percentage fluctuates heavily between 65 and 85 % there is a clear decreasing trend (not shown).

We further investigated the percentage of uncited documents (all) and the percentage of uncited 'normal' articles. Also here data related to the latter case lies below (less uncited articles) the former, and data increase at the end of the period. For normal articles there is a plateau over the period 1993-2002 of about 24% uncited articles, for all publications there is a similar plateau of about 35% (see Fig.7).

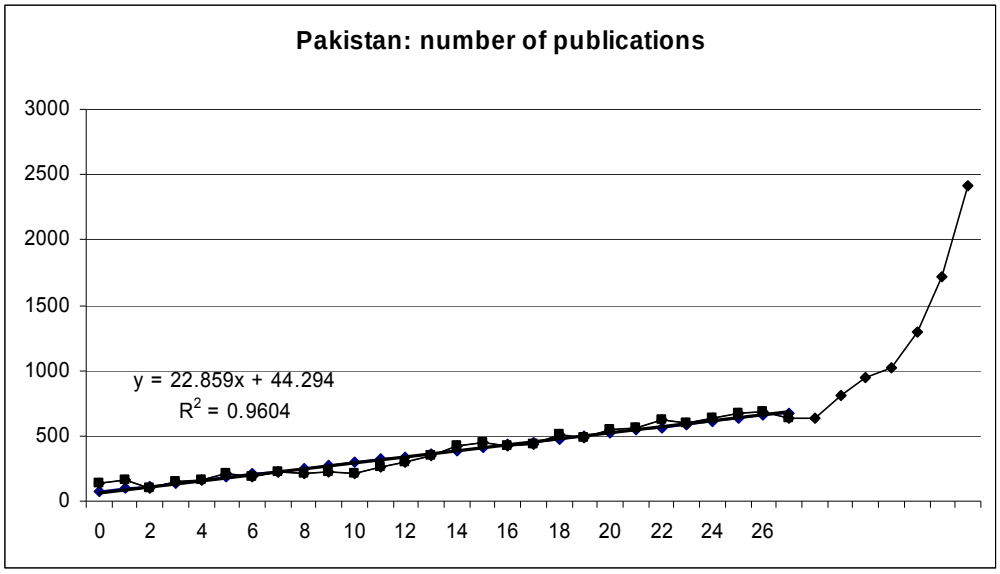

Figure 5. Number of Pakistani publications and best fitting linear line between 1973 and 2000 (the first 28 years), based on WoS data

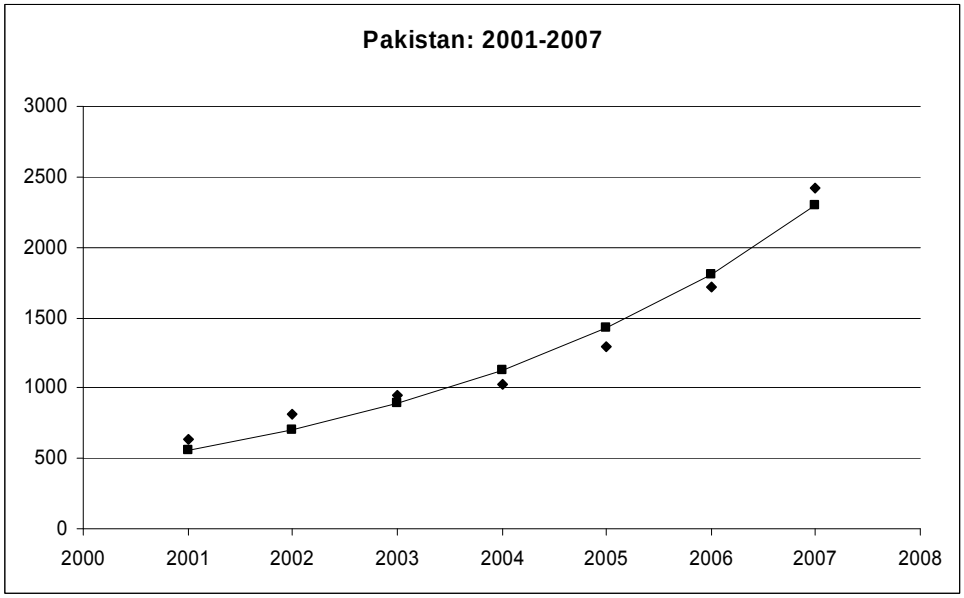

Figure 6. Best fitting exponential curve for Pakistan (period 2001-2007)

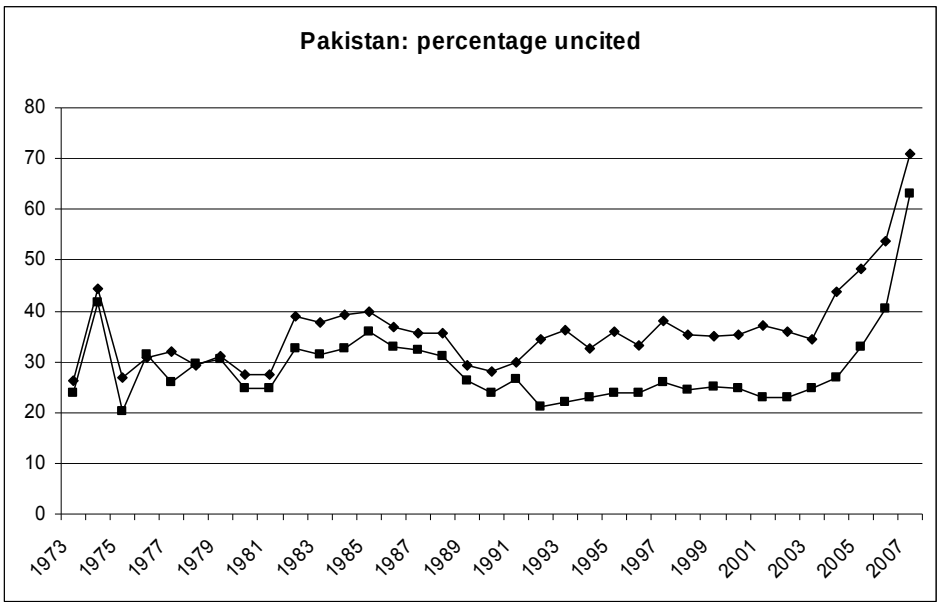

Figure 7. Pakistan's percentage of uncited documents (November 2008): all documents (upper curve) and normal articles (lower one) (WoS data)

The h-index for all Pakistani publications over the period 1973-2007 (data collected in November 2008) is 77. We obtained the yearly h-index for all publications and for normal articles alone (not shown). When normalizing with respect to the number of publications we see a clear decreasing trend.

India

Now we come to the local giant. The yearly number of publications (all types but not including conference proceedings) with an Indian address increased during the period 1973-2007 from about 8500 to more than 36,000, leading to a total of almost 600,000. Another 26,300 articles in conference proceedings can be added to this. Over the period 1973-2000 the yearly number of publications increases first following a convex curve, which then becomes concave and increases slowly; from about 2002 the curve increases much more vigorously. We show the yearly number of publications since 1993 and a best fitting power function: $y = 16{,}796 + 3.06 \cdot t\ 3.31$, with $t = 0$ in 1993 ($R^2 = 0.996$ for a nonlinear least squares regression). Figure 8 illustrates our findings. The fact that during the period 1993-1998 India's yearly number of publications increased only slowly (in absolute terms) resulted in a decrease in terms of global share (Glänzel and Gupta, 2008).

As we did for the other countries we collected data on the percentage of 'normal" articles among the published Indian documents in the WoS. Although this percentage fluctuates heavily between 65 and 90 % there is an increasing trend (not shown).

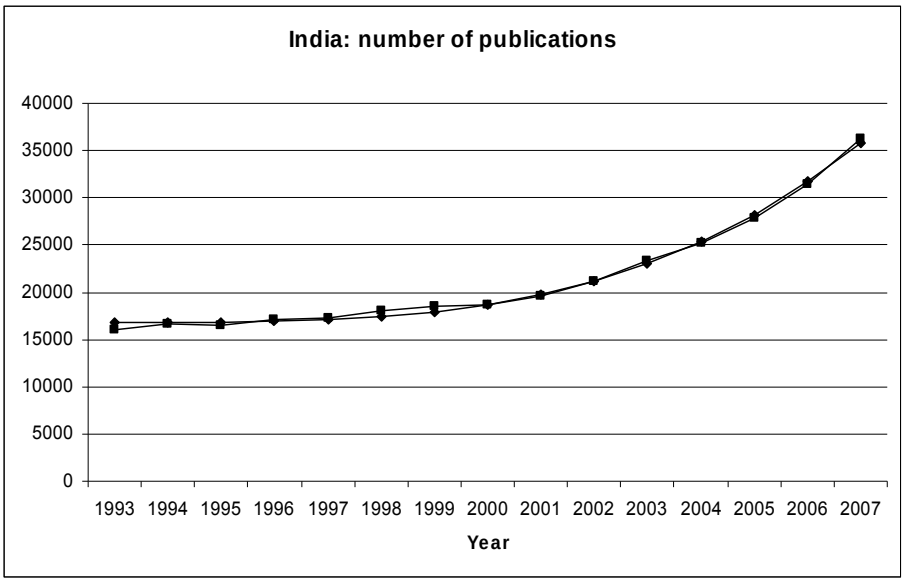

Figure 8. Yearly number of Indian publications and best fitting power law, based on WoS data

We also investigated the percentage of uncited documents (all) and the percentage of uncited 'normal' articles. Also here data related to the latter case lies below (less uncited articles) the former, and data increase at the end of the period. Between 1979 and 2003 there is a decrease in the number of uncited publications.

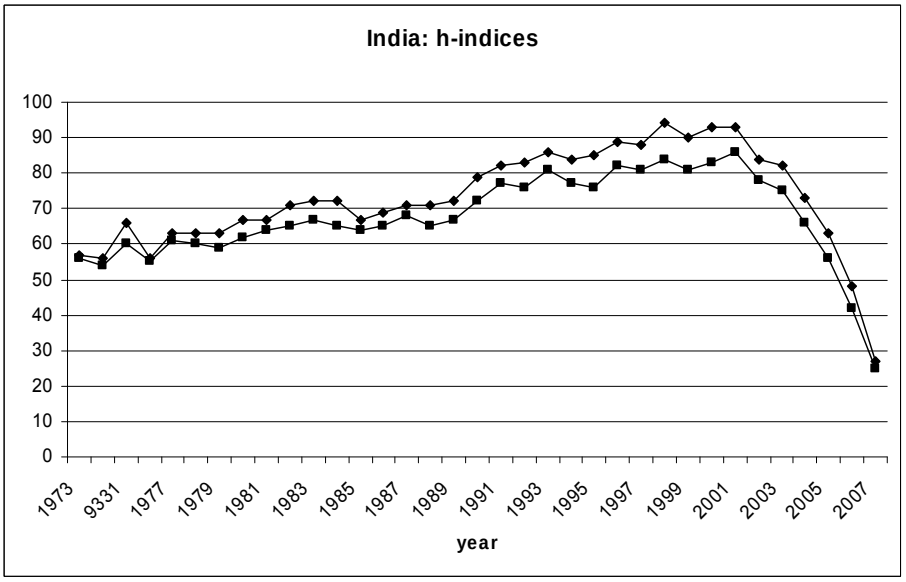

Figure 9. h-Index series based on all publications (upper curve) and on normal articles (lower curve), WoS data

The h-index for all Indian publications over the period 1973-2007 (data collected in November 2008) is 235. We obtained the yearly h-index for all publications and for normal articles alone (see Fig.9). When normalizing with respect to the number of publications, this trend curve becomes essentially flat between 1983 and 2000 (not shown).

Sri Lanka

Of the four countries we study here Sri Lanka is by far the country with the smallest population. Yet, we will show that relatively speaking it is scientifically the strongest of the

four. Until 1972 the country was officially known as Ceylon. For that reason we searched for "Sri Lanka" AND Ceylon, as most of the articles published in 1973 and 1974 still carried the name Ceylon in the address. The number of publications (all types but not including conference proceedings) with a Sri Lankan address increased during the period 1973-2007 from about 75 to more than 400 a year, leading to a total of almost 6,000. Another 460 articles in conference proceedings can be added to this total. Over the period 1973-2000 the yearly number of publications can best be described by a power law: y = a + b* t c, where y denotes the number of publications, a = 102.45, b= 0.0074 and c = 2.97 where and t = 0 in 1973 (R^2 = 0.85 for a nonlinear least squares regression). Figure 10 shows the yearly number of Sri Lankan documents and the best fitting power law.

We collected data on the percentage of 'normal" articles among the published Sri Lankan documents in the WoS. Although this percentage fluctuates heavily between 70 and 87 % there is a clear decreasing trend. We also investigated the percentage of uncited documents (all) and the percentage of uncited 'normal' articles. Also here data related to the latter case lies below (less uncited articles) the former, and data increase at the end of the period. There seems to be a slight decreasing trend between 1987 and 2001. The number of uncited articles is generally lower than for the other three countries.

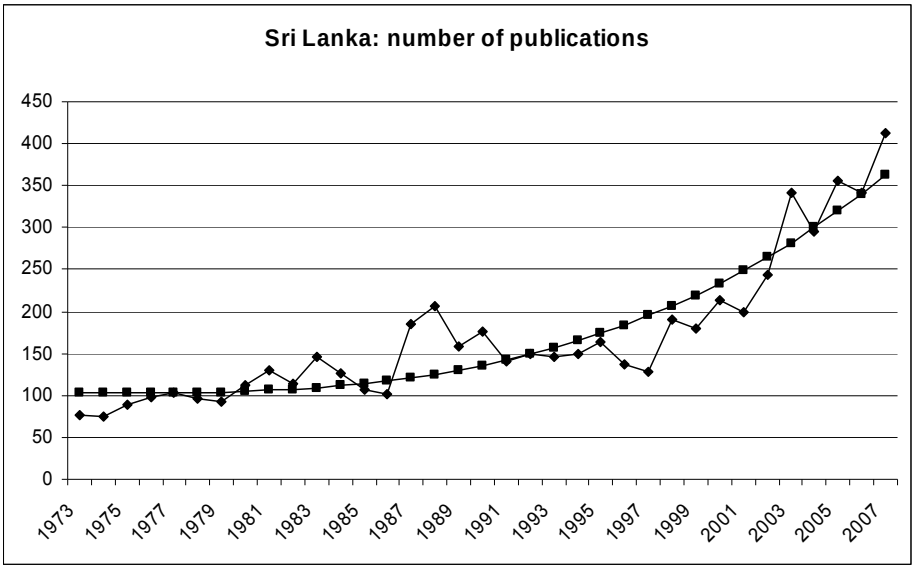

Figure 10. Best fitting power law for Sri Lankan publications (WoS data)

The h-index for all Sri Lankan publications over the period 1973-2007 (data collected in December 2008) is 62. We obtained the yearly h-index for all publications and for normal articles alone (not shown). When normalizing with respect to the number of publications we see two decreasing trends, one, slowly decreasing over the period 1973-1997, and one faster decreasing over the period 1998-2007 (see Fig.11).

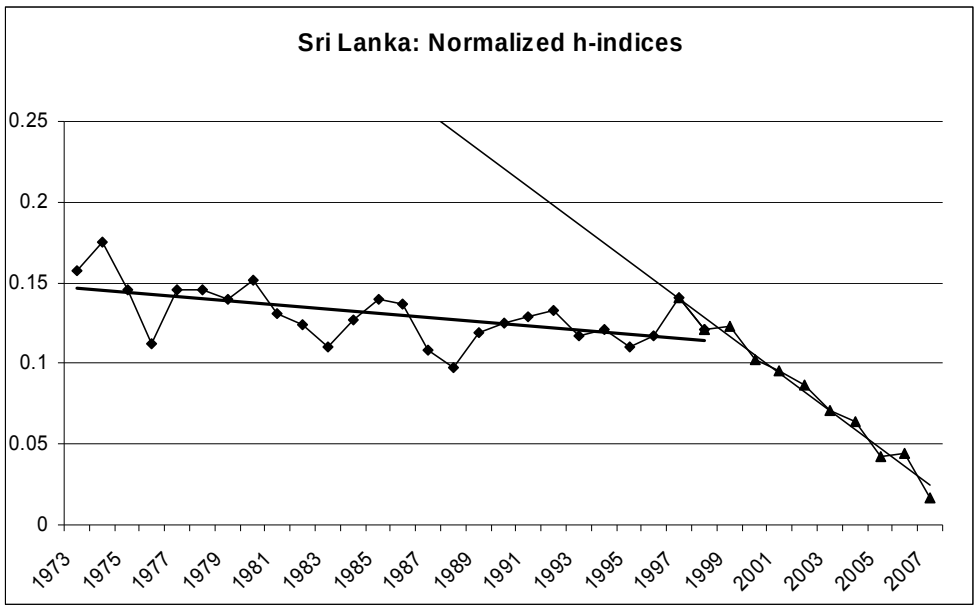

Figure 11. Normalized h-index series based on all publications (WoS data)

Comparisons

Table 1 recapitulates some indicator values (based on WoS). In absolute terms India's number of WoS publications is about 30 times larger than Pakistan's, which in turn is almost double that of Bangladesh. Finally Sri Lanka's number of publications is about 60% of Bangladesh'. In SCOPUS (see Table 2) India's number of publications is about 18 times larger than Pakistan's, which has more than twice the number of articles as Bangladesh. Sri Lanka has less than half the number of publications of Bangladesh.

The h-index depends on the absolute number of publications, and hence favours large countries. This is clear from Tables 1 and 2. Note that Bangladesh and Pakistan change ranks according to the database used. For this reason we next consider three indicators that are not or less dependent on the absolute number of publications. The first is the percentage of uncited articles, (see Fig.12).

Table 1. WoS scientometric indicators

Country	Documents 1973-2007	Documents + articles in conference proceedings	Trend in percentage of 'normal' articles	h index 1973-2008
Bangladesh	10,383	11,331	Increase	87
India	599,076	625,403	Increase	235
Pakistan	19,375	21,147	Decrease	77
Sri Lanka	5,975	6,437	Decrease	62

Table 2. SCOPUS scientometric indicators

Country	Documents	Citable Documents	Cites	Self-Cites	Cites per Doc.	h index 1996-2007
Bangladesh	7,835	7,638	31,821	6,424	4.64	52
India	336,429	322,168	1,347,950	485,084	4.59	170
Pakistan	19,025	18,112	51,229	14,781	3.38	54
Sri Lanka	3,608	3,393	18,046	2,251	6.17	45

Source URL: http://www.scimagojr.com. SCImago Research Group.
Copyright 2007.Data Source: Scopus®

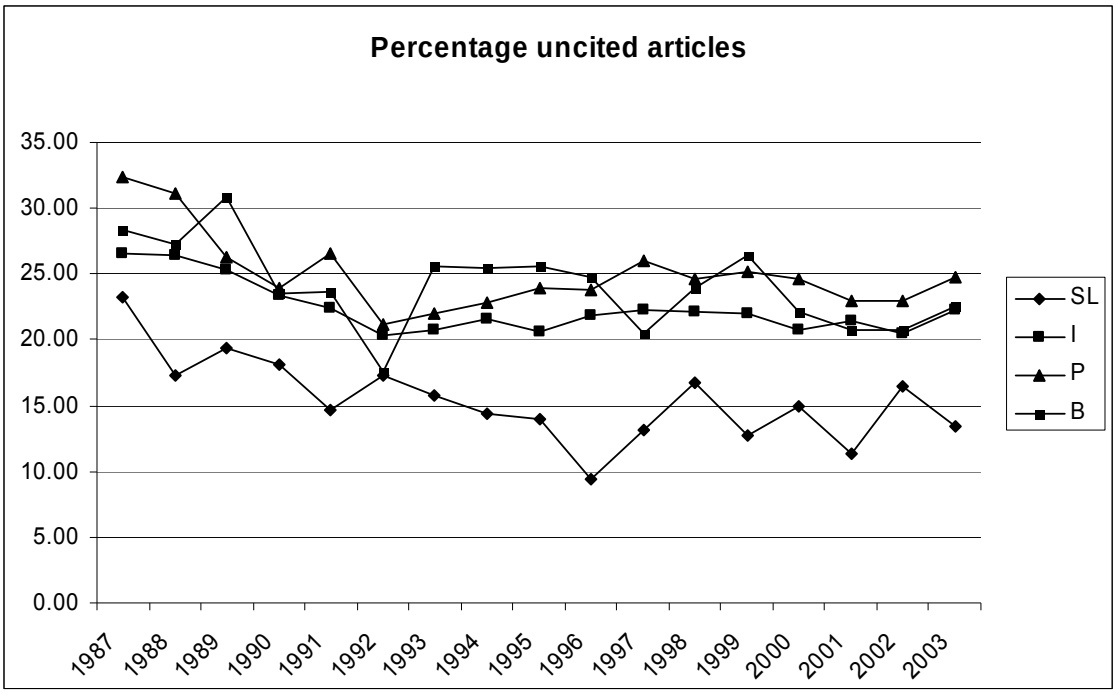

Figure 12. Percentage of uncited articles (WoS data)

Sri Lanka clearly has the lowest rate of uncited articles. For most of the period Pakistan has the largest (or second largest) rate. There are six years in which Bangladesh has the highest rate of uncited articles, while India's rate falls in between. So globally we have the following ranking: Sri Lanka < India < Bangladesh < Pakistan, where Sri Lanka occupies the position corresponding with the highest quality (lowest number of uncited documents), and Pakistan the one corresponding to the poorest quality (according to this indicator).

The next indicator comes from SCImago (based on SCOPUS), and is the number of citations per document in the SCOPUS database (see Table 2). According to this indicator Sri Lanka is again the best country, followed by Bangladesh, India and Pakistan. Finally, we consider an attempt to produce a size-independent h-index. Molinari and Molinari (2008) proposed a size-independent h-index, denote as hm. The size-independent h-index, hm, is defined as h / T 0.4, where T denotes the number of publications. Kinney (2007), making use of a preprint version of the Molinari article, confirmed the usefulness of the Molinari approach in an article studying national scientific facilities in the USA and published in the Proceedings of the National Academy of Sciences of the USA.

Following a suggestion by these authors we restrict our data to the period [1973-1998] so that articles are at least ten years old and have had time to reach their full citation potential. This leads to Table 3.

Table 3. WoS scientometric indicators; publication period [1973-1998]; h-index determined on March 28, 2009

Country	All publications	h index	h_m index
Bangladesh	5,528	82	2.61
India	385,433	155	0.90
Pakistan	9,812	69	1.75
Sri Lanka	3,577	55	2.08

When considering only normal articles these h_m-values become slightly higher. Preliminary investigations seem to indicate that Bangladesh is strongest in biomedical sciences. Hence, when restricting data to the sciences and engineering fields, the h_m-value for Sri Lanka seems to become the largest. This leads to a ranking:

$$\text{Sri Lanka} \approx \text{Bangladesh} > \text{Pakistan} > \text{India}.$$

Bringing all size-independent measures together suggests the following 'quality' ranking (starting with the 'best'):

$$\text{Sri Lanka} > \text{Bangladesh} > \text{India} > \text{Pakistan}$$

Based on all relative indicators used by us Sri Lanka's lead is undisputable.

Conclusion

A (partial) comparison has been provided between research indicators related to India, Bangladesh, Pakistan and Sri Lanka. Data originate from Thomson Scientific's Web of Science as well as from Scopus (using the SCImago website). It seems that in relative terms Sri Lanka is the strongest scientific country of the four. This observation would confirm the fact that in general terms Sri Lanka is the most developed country among these four South Asian countries: it has the highest life expectancy, a considerably higher GDP per capita than the other three countries and a higher health expenditure per capita (Sadana et al., 2004).

From a scientometric perspective we note the interesting results obtained by applying the Molinari approach.

Acknowledgement

We thank two anonymous reviewers for some pertinent and useful remarks.

References

Arunachalam, S. (2008). The science race continues in Asia. *Current Science*, 94, 848-849.

Arunachalam, S. & Viswanathan, B. (2008). South-South collaboration: the case of Indo-Chinese collaboration in scientific research. *Current Science*, 95, 311-313.

Balaram, P. (2004). Science and technology proficiency: China and India. *Current Science*, 86, 755-756.

Glänzel, W. & Gupta, B.M. (2008). Science in India. A bibliometric study of national research performance in 1991-2006. *ISSI Newsletter*, 4(3), 42-48.

Guan, JC. & Ma, N. (2004). A comparative study of research performance in computer science. *Scientometrics*, 61, 339-359.

Gupta B.M., Munshi, U.M. & Mishra, P.K. (2002). S&T collaboration of India with other South Asian countries. *Current Science*, 83, 1201-1209.

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

Kinney, A.L. (2007).National scientific facilities and their science impact on nonbiomedical research. *Proceedings of the National Academy of Sciences,* 104, 17943-17947.

Mahbuba, D. & Rousseau, R. (2008). Scientific research in Bangladesh and a comparison with India and Pakistan. Proceedings of WIS 2008, Berlin. Fourth International Conference on Webometrics, Informetrics and Scientometrics & Ninth COLLNET Meeting (H. Kretschmer and F. Havemann, eds.). http://www.collnet.de/Berlin-2008/MahbubaWIS2008srb.pdf

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

Sadana, R., D'Souza, C., Hyder, A.A. and Chowdhury, A.M.R. (2004). Importance of health research in South Asia. *BMJ,* 328, 826-830.

SCImago. (2007). SJR — SCImago Journal & Country Rank. Retrieved May 27, 2008, from http://www.scimagojr.com