

## **SUPPORTING INFORMATION**

# **On Bibliometric Analysis of Chinese Research on Cyclization, MALDI-TOF, and Antibiotics: Methodical Concerns**

*Petr Heneberg*

Half Century ago, Eugene Garfield introduced his concept of tracking the scientific knowledge using the database of citations of individual scientific papers. One of major outcomes of this concept is Web of Knowledge (WoK), recently encompassing several indexes with retrospective up to year 1899. Currently, it includes over 700 million of cited references, and covers over 23,000 journals in 256 scientific disciplines<sup>1</sup>. However, as any other commercial product, WoK has also its limits and caveats<sup>2-3</sup>. Thus the potential user must be cautious, whether the respective source of data has parameters he needs and expects.

## **SUPPLEMENTARY METHODS**

As a datasource I used the Thomson Reuters` WoK database, which included Web of Science (1899-present) with Conference Proceedings, Derwent Innovations Index (1983-present), BIOSIS Previews (1980-present), Medline (1950-present), and Journal Citation Reports (2000-2008).

Initial search for publications was done as described<sup>4</sup>. search for number of publications per year was performed for three topics (CYCLI?ATION\*, MALDI-TOF, or ANTIBIOTIC\*) with address field containing words CHINA, USA, GERMANY, or JAPAN. As an example, search for publications on cyclization in China was performed using the search statement:

Topic = (CYCLI?ATION\*) AND Address = (CHINA)

In this query, the question mark (?) represents any single character, and the asterisk (\*) represents any group of characters or no character. “Topic” refers to at least one of the fields Title, Abstract, Author Keywords, and Keywords Plus. Number of publications from each country in each year was evaluated. All the countries and topics chosen were identical with those reported by Li & Willett<sup>4</sup>.

The search confirmed unusual one-point increases in scientific productivity reported earlier (Supporting Info, Fig. S1-S2, Tab. S1), and shifted my focus on the experimental design of the search protocol used. I postulated and tested three hypotheses, which had a potential to decrease the risk of errors of omission and commission:

- 1) The one-point increase of publication output in 1991 may be caused by inclusion of conference proceedings in early 90<sup>th</sup> years<sup>3</sup>. Thus, I searched for change in relative frequency of document types “abstract” and “meeting” between years 1989 and 1991.
- 2) Part of the reported changes may be caused by nonrandom absence of abstracts and address fields in some references<sup>3</sup> References from 70<sup>th</sup> years of the 20<sup>th</sup> Century (and earlier) mostly do not contain abstract, which significantly reduces the probability that they would be included in search results when using field "Topic". Thus I newly applied the search phrases CYCLI?ATION\* and ANTIBIOTIC\* to the “title” field only, combined or not with the search for the address field. Phrase MALDI-TOF was not included in this search as its earliest hit was from 1992 at thus it was not affected by changes in abstracting policies of WoK<sup>3</sup>.
- 3) Supposed changes in productivity of German scientists may be caused by German reunification. Thus, search for country code GER was included in the protocol.

All the searches were carried out on February 2, 2010. Limitations of the searches performed include lack of country name or symbol in 14 % of recent WoK records<sup>5</sup>; this percentage was found to be even higher for the older records. Publications from Taiwan and Hong-Kong were not excluded.

## **SUPPLEMENTARY RESULTS**

### **“Topic” search**

The largest amount of publication on topic CYCLI?ATION was produced in the U.S.A., followed by Japan, Germany, and China (Tab. S1). The first papers on this topic appeared at the beginning of 20<sup>th</sup> Century, however the first Chinese paper on cyclization was published as late as in 1980. Topic search

suggests constant increase in number of Chinese papers, which resulted in overtaking of Germany and Japan in 2008 and 2009 (Fig. S1A).

Topic MALDI-TOF appeared only recently, in 1992. First publication of the U.S. authors appeared in 1992, followed two years later by Germans and Japanese, and by Chinese in 1997. The U.S. publication output on this topic peaked in 2005, publication output of Germany and Japan reached plateau in the same year. Chinese publication output on this topic was steadily increasing up to the end of the period tested. In 2009, China produced the largest amount of publications containing phrase MALDI-TOF in title, abstract or keywords (Fig. S1B).

Topic ANTIBIOTIC\* shown its first hit in year 1931. Among the four countries tested, the majority of publications were produced in the U.S.A., followed by Japan, Germany, and China (Tab. S1). Tracking of first publication date is not possible due to almost complete omission of address field in records beyond 1973. Based on the topic search, the German and Chinese production of papers on antibiotics is still rising, U.S. output reached plateau in 2001, and number of Japanese papers is decreasing since 1995 (Fig. S1C).

The results of “topic” search confirmed the previously published data<sup>4</sup>. Only the number of papers found was higher than reported previously, which results from undisclosed limited access to some indexes within the WoK database<sup>4</sup>. In 1991, the topic search revealed dramatic productivity increase of U.S., German and Japanese authors focusing on cyclization, and in the triple increase in productivity of U.S. authors focusing on antibiotics (Fig. S1A,C), similarly to the previously published results (Fig. S2).

### **Hypothesis 1: Inclusion of Conference Proceedings Citation Index (CPCI)**

First, I hypothesized that the one-point increase of publication output in 1991 may be caused by inclusion of conference proceedings in early 90<sup>th</sup> years. However, relative share of document type "meeting" and “abstract” on topic CYCLI?ATION\* decreased between years 1989 and 1991 from 9.7% to 5.9%. Within the identical time period, the total number of references on this topic increased by 93%. There was no major change in total number of references on the other two topics – CPCI covers the

whole timespan of MALDI-TOF papers, and there was only 2.5% increase in total number of references on ANTIBIOTIC\* between the above mentioned years.

## **Hypothesis 2: Nonrandom absence of abstracts and addresses**

Second, I hypothesized that part of the reported changes may be caused by nonrandom absence of abstracts and address fields in some references. Thus I newly applied the search phrases CYCLI?ATION\* and ANTIBIOTIC\* to the “title” field only, combined or not with the search for the address field. “Title” search search for phrases CYCLI?ATION\* and ANTIBIOTIC\* resulted in substantially lower growth of publication output when compared with the “topic” search (Fig. 1A,B). The difference between “topic” and “title” emerged in 1966, and completely eliminated the one-point increase in 1991 (Fig. 1A).

Combined search for “title” and “address” was found to have more limitations contrary to the “title” search only. There were published and indexed thousands of papers on cyclization and antibiotics until 1973 (Fig. 1A,B), which was not reflected in the data searches limited by the country criterion (Fig. 1C,D). The country information is almost completely absent in references indexed until 1973.

The new search design revealed continual decrease in number of Japanese publications on antibiotics since 1980, which is opposite to the results of the “topic” search. Virtual stagnation of all three developed countries tested highlighted even more the role of China, the publication number of which showed rapid increase both in “title” and in “topic” search. Combined “title” and “address” search completely eliminated one-point increase in production of U.S. and Japanese authors in 1991 (Fig. 1C). However the supposed change in number of German papers in years 1990/1991 remained unsolved (Fig. 1C,D).

## **Hypothesis 3: Observed numbers are affected by German reunification**

The supposed change in German scientific productivity coincides with German reunification in October 1990. When searching for East-German cities, such as Dresden, I found that the most common

identification of East Germany in WoK was GER DEM REP. Combined search for CYCLI?ATION\* or ANTIBIOTIC\* with address GERMANY OR GER proved that the supposed one-point increase (Fig. 1C,D) was caused by underrepresentation of studies under country code GER DEM REP or FED REP GER before year 1990 (Fig. 1E,F).

## SUPPLEMENTARY DISCUSSION

Scientometric research indicators are increasingly used by national and international agencies, including National Science Foundation, the World Bank, or the OECD<sup>5</sup>. Thus it is important to be aware of the critical elements of scientometric evaluation. On the example of recently published study<sup>4</sup>, I have shown how inappropriate search protocol may lead to confusing or false-positive results. Similarly to the previously published papers<sup>6-17</sup>, the respective study concluded that the publication output of Chinese research is improving in number, and to lesser extent also in the quality. Despite the recent multiplication of research output, the Chinese research is still lagging behind countries with long and uninterrupted tradition of support for science and innovation, which resembles situation in several other less developed or transitional economies such as South Korea, Turkey, or Cuba<sup>18-21</sup>. However some of their other findings were largely unexpected and are discussed below.

Using “topic” search, I confirmed that use of the published protocol<sup>4</sup> leads to supposed dramatic one-point increases in productivity in 1991 and 1973, which was originally explained by “large increase that year in the number of journals covered in WoK”. However, in 80<sup>th</sup> and 90<sup>th</sup> years, the WoK coverage was increasing gradually, with the exception of CPCI addition with coverage from 1990 onwards<sup>3</sup>.

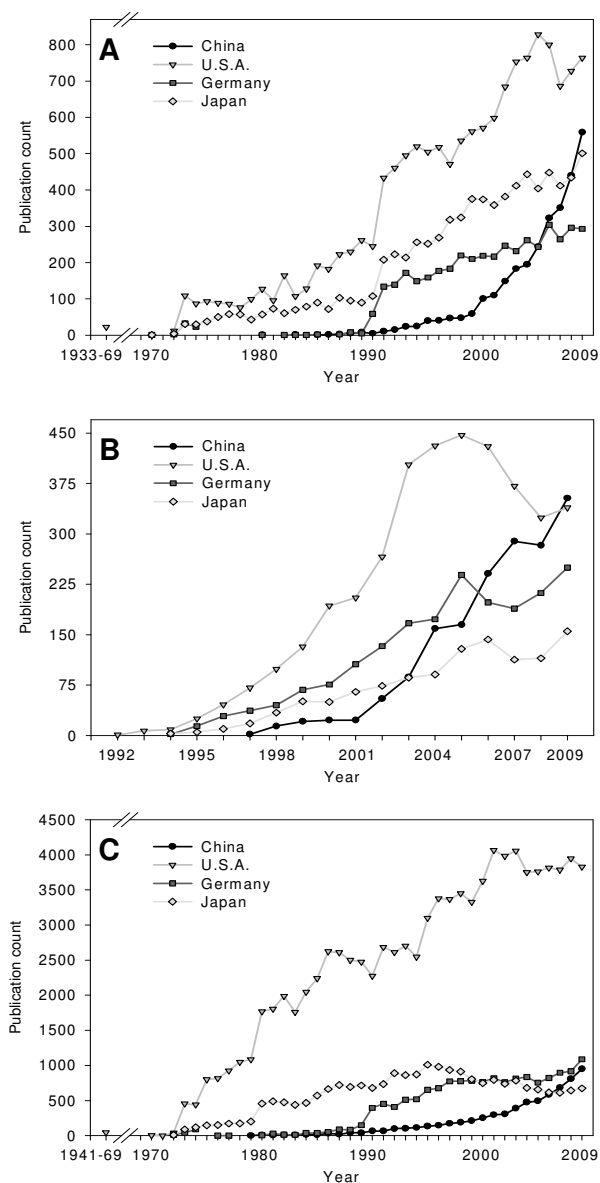
I verified, that the reported increases are not caused by changes in coverage (such as inclusion of CPCI), but rather by changes in abstracting policies. By replacing “topic” search with “title” search, I found that nonrandom absence of abstracts and keywords in some references is fully responsible for the one-point increase in productivity of all but German authors. Besides that, also the long-term gains in productivity appear to be much smaller than previously thought (Fig. 1A-D). In some areas of research, I detected long-term stagnation or even decrease of the publication output, which is in sharp contrast with

data obtained by “topic” search. Use of “title” search leads to lower number of hits than “topic” search, but the measure is more objective for comparisons of scientific productivity as it removes bias caused by nonrandom inclusion of abstracts and/or keywords since year 1966. Terms related to methods (e.g. cyclization) are expected to be influenced more than study subjects (e.g. antibiotics), as methods are usually not represented in the title of other than methodical papers.

Similarly, I tested, whether the results are affected by possibly nonrandom omission of the address field. Since late 70<sup>th</sup> years, the recent average country omission rate is 14%<sup>5</sup>. However here I found that the country information was almost completely missing in references indexed until 1973. This absence results in false-negative absence of records when using search for the “address” field, and in the false-positive increase of publication output in early 70<sup>th</sup> years (Fig. 1C,D), both of which are eliminated by any search excluding “address” field (Fig. 1A,B).

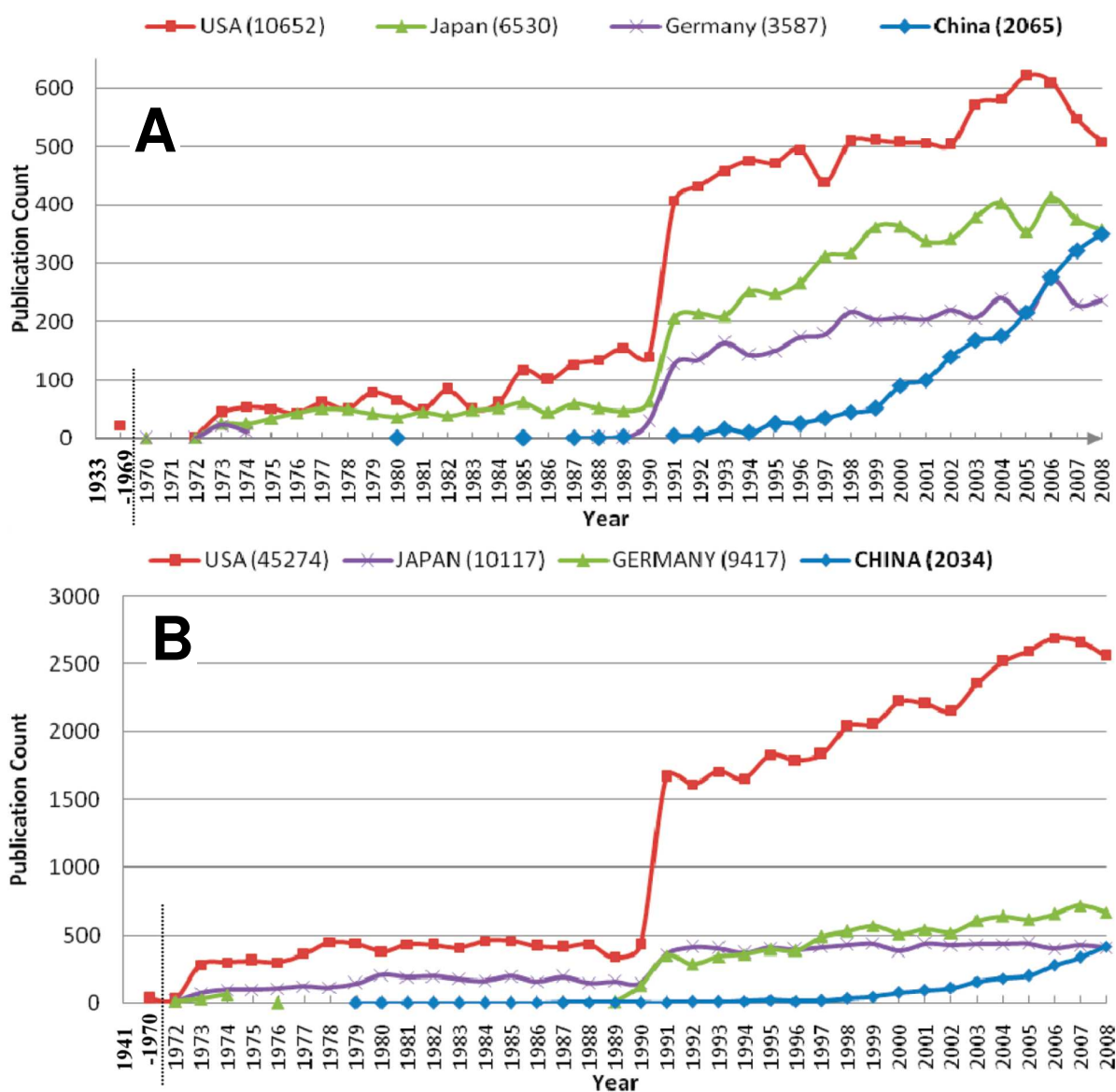
Third, the only remaining unexplained one-step increase was in the productivity of German scientists<sup>4</sup>. This error of omission was affected by German division and reunification, which were associated with change of country names and codes. Combined search for GERMANY OR GER led to its complete elimination (Fig. 1E,F).

## FIGURES AND FIGURE CAPTIONS



**Supplementary Figure S1.** WoK “topic” search for cyclization (A), MALDI-TOF (B), and antibiotics (C).





**Supplementary Figure S2.** Publication growth as shown previously<sup>4</sup> using the topic search. The individual curves represent number of publications in China, U.S.A., Germany, and Japan in cyclization (A) and antibiotics (B). Note the supposed dramatic increase between years 1990 and 1991, and the virtual absence of papers on antibiotics and cyclization before 1972. Reprinted from Li & Willett<sup>4</sup>.

**Supplementary Table S1.** Results of the “Topic” search.

Topic	CYCLIZATION*		MALDI-TOF		ANTIBIOTIC*	
Address	N	%	N	%	N	%
CHINA	3,017	4.6	1,743	12.3	6,794	1.8
USA	14,343	22.0	3,822	27.0	95,787	25.0
JAPAN	7,840	12.0	1,150	8.1	22,423	5.8
GERMANY	4,268	6.5	1,954	13.8	15,122	3.9
Σ all countries	65,309		14,134		383,726	

## SUPPLEMENTARY REFERENCES

- (1) <http://wokinfo.com/about/whatitis/> [cited as 2-Feb-2010]
- (2) Meho, L.I.; Sugimoto, C.R. Assessing the Scholarly Impact of Information Studies: A Tale of Two Citation Databases-Scopus and Web of Science. *J. Am. Soc. Inf. Sci. Technol.* **2009**, *60*, 2499-2508.
- (3) Bar-Ilan, J. A Closer Look at the Sources of Informetric Research. *Cybermetrics* **2009**, *13*, 4.
- (4) Li, J.; Willett, P. Bibliometric Analysis of Chinese Research on Cyclization, MALDI-TOF, and Antibiotics. *J. Chem. Inf. Model.* **2010**, *50*, 22-29.
- (5) Jacsó, P. Errors of omission and their implications for computing scientometric measures in evaluating the publishing productivity and impact of countries. *Online Inform. Rev.* **2009**, *33*, 376-385.
- (6) Makris, G.C.; Spanos, A.; Rafailidis, P.I.; Falagas, M.E. Increasing contribution of China in modern biomedical research. Statistical data from ISI Web of Knowledge. *Med. Sci. Monitor* **2009**, *15*, SR15-SR21.
- (7) Leydesdorff, L.; Zhou, P. Are the Contributions of China and Korea Upsetting the World System of Science. *Scientometrics* **2005**, *63*, 617-630.
- (8) Zhou, P.; Leydesdorff, L. The Emergence of China as a Leading Nation in Science. *Res. Policy* **2006**, *35*, 83-104.
- (9) Zhou, P.; Thijs, B.; Glanzel, W. Is China Also Becoming a Giant in Social Sciences. *Scientometrics* **2009**, *79*, 593-621.
- (10) Cyranoski, D. China Increases Share of Global Scientific Publications. *Nature* **2004**, *431*, 116.
- (11) Kumari, G.L. Synthetic Organic Chemistry Research: Analysis by Scientometric Indicators. *Scientometrics* **2009**, *80*, 559-570.

- (12) Goh, K.L.; Farrell, G.C. Publications from China: the sleeping giant awakens. *J. Gastroenterol. Hepatol.* **2008**, *23*, 341-343.
- (13) Klaewsongkram, J.; Reantragoon, R. Asthma Research Performance in Asia-Pacific: A Bibliometric Analysis by Searching PubMed Database. *J. Asthma* **2009**, *46*, 1013-1020.
- (14) He, T.W. International scientific collaboration of China with the G7 countries. *Scientometrics* **2009**, *80*, 571-582.
- (15) Klar, M.; Foldi, M.; Denschlag, D.; Stickeler, E.; Gitsch, G. Estimates of Global Research Productivity in Gynecologic Oncology. *Int. J. Gynecol. Cancer* **2009**, *19*, 489-493.
- (16) Leydesdorff, L.; Wagner, C. Is the United States losing ground in science? A global perspective on the world science system. *Scientometrics* **2009**, *78*, 23-36.
- (17) Kostoff, R.N.; Barth, R.B.; Lau, C.G.Y. Quality vs. quantity of publications in nanotechnology field from the People's Republic of China. *Chin. Sci. Bull.* **2008**, *53*, 1272-1280.
- (18) Uthman, O.A.; Uthman, M.B. Geography of Africa biomedical publications: An analysis of 1996-2005 PubMed papers. *Int. J. Health Geogr.* **2007**, *6*, 46.
- (19) Karlsson, S.; Srebotnjak, T.; Gonzales, P. Understanding the North-South knowledge divide and its implications for policy: a quantitative analysis of the generation of scientific knowledge in the environmental sciences. *Environ. Sci. Policy* **2007**, *10*, 668-684.
- (20) Packer, A.L.; Meneghini, R. Learning to communicate science in developing countries. *Interciencia* **2007**, *32*, 643-647.
- (21) Boldiš, P.; Landová, H. Comparison of citation databases SCOPUS and Web of Science: Czech and Slovak agricultural and related disciplines. *Plant Soil Environ.* **2006**, *52*, 481-484.