Bibliometric Study of Cancer Research in India

Swapan Kumar Patra & Partha Bhattacharya

Abstract

The paper has done a bibliometric analysis of oncology research in India. The data for the study has been downloaded from national centre for Biotechnology (NCBI) *Pub Med*. The study analyses literature growth trends. It also examines research activities in different countries worldwide. Bradford law of scattering was employed to identify the core journal, which published Indian cancer research literature. Lotka's law was employed to study the authors' productivity pattern. The study also identifies the active institutions in India, which published the cancer literature the most.

1. INTRODUCTION

There has been significant growth in the research literature on oncology in India. Searching the literature in this area from the International database gives an insight into the pattern of growth of this literature. The paper intends to make a bibliometric study of cancer-based literature being contributed by Indian authors. Bibliometric study is a simple statistical method of bibliography counting to evaluate and quantify the growth of a subject.[1] The data for the study was downloaded from the National Center for Biotechnology Information (NCBI) *Pub Med*.[2] *Pub Med* (Published Medical Literature) is an online version of *MEDLINE*, available free to anyone with internet access. *MEDLINE* is the National Library of Medicine's bibliographic database covering the fields of medicine, nursing, dentistry, veterinary medicine, health care system and the preclinical sciences. The MEDLINE contains approximately 12 million records dating back to the mid-1960's.

2. OBJECTIVE

Objective of the study is to find:

- ❑ Growth trend of cancer literature in India
- ❑ Research activities of other countries

- ❑ Document types in which cancer literature published

- ❑ Authorship pattern

- ❑ Identify the core journals which publish the articles

- ❑ Indian institutes which publish predominently in cancer research.

3. METHODOLOGY

Data was downloaded from the *Pub Med* database using the software Endnote7. A blank database format was created using the software Endnote.[7] The software has powerful web interface, which can download records into the database. For downloading the data into the database, the search term applied was "Neoplasm" AND Author address= "India". Neoplasm is the MeSH[3] indexed subject heading. As indicated by Lancaster[4], main heading can take care of synonyms, nearly synonyms and homonyms. We expected the maximum retrieval of records. A total of 6484 records were downloaded from *Pub Med* on 25th May 2004. We found 76 records from the year 2004. As 2004 records are incomplete we had removed those records from the database. Now the database has a total of

6408 records. Each record contains English language abstract and bibliographic information (e.g. author, name of journal, author address).

NCBI indexed only the first author's address. We take the address listed in the record as the first author's address for our analysis of author affiliation.

4. DATA ANALYSIS

4.1 Growth of the Literature

Figure 1 shows literature growth trends in Indian cancer research. *Pub Med* has indexed cancer research articles in 1987. In that year 70 articles have been indexed. After that there is a steady growth of literature. Except for the years 1991, 1993, 1995, 1998 and 2003, where the growth of literature decline from the previous year. After observing for the last few years trends it can be concluded that on an average around 500 papers were published yearly.

4.2 Country-Wise Distribution of Literature

A total of 15,37659 records retrieved using the key word "Neoplasm". It has been observed that US is the largest literature producing country with 10% of the total literature published by the US alone. India has a contribution of about 0.4%. Table 1

shows the country-wise distribution of cancer literature. It can be inferred from the table 1 that although in cancer related literature west is dominant, but the cancer related literature is distributed world wide. It means research in this area is carried out worldwide. Unlike other field of research which is mainly concentrated on only in US and UK, worldwide literature distribution shows that the cancer research is getting attention worldwide. It is also important to note that Indian contribution in cancer literature is very less in comparison to other countries.

Table 1. Country-wise distribution of cancer literature

S. No.	Country	No. of publication	%
1.	USA	1,53341	9.97
2.	Japan	79,651	5.18
3.	Italy	34,631	2.25
4.	Germany	32,476	2.11
5.	UK	31,443	2.04
6.	Canada	18,536	2.04
7.	China	10,927	0.71
8.	Australia	10,600	0.68
9.	Spain	10,368	0.67
10.	India	6,48	0.42
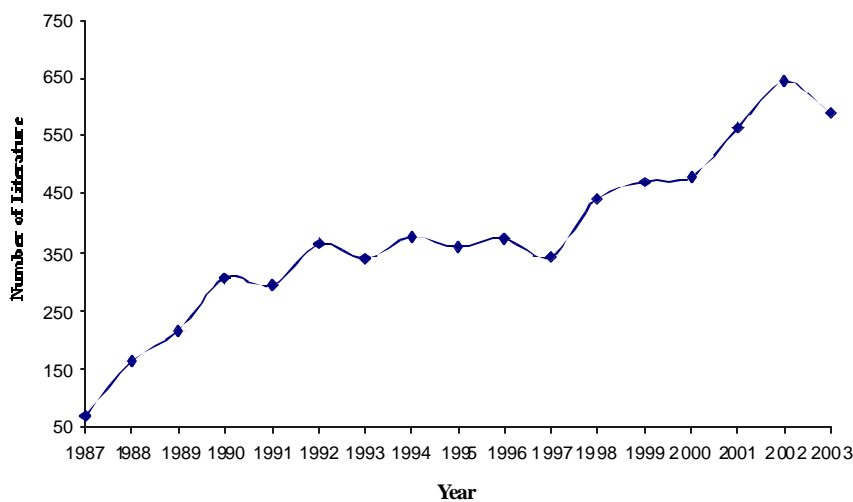
Figure 1. Growth of cancer literature

4.3 Document Type

Although NCBI indexed articles in 54 different categories. For the convenience of the user, we have categorized the literature in five broad type (figure 2) viz journal articles (5643 records), review (279 records), review of reported cases (282 records), randomized controlled trial (81 records) and others (123 records). Other categories include lecture, legal cases, letter to the editor, and so on. Amongst the entire categories, journal article accounts the maximum. About 5643 articles from the journals, which is around 88% of the literature. This means that Indian cancer scientists are heavily dependent on the journal publication. Possibly it can be said that, this is the only mode of communication among the Indian cancer researchers.

It is interesting to note that *Pub Med* has not covered conferences, from India. Proceedings of conference, congress and symposia are most important for a subject like cancer research. This communication channel is more effective and easier. Pre-conference and post-conference proceedings play an important role in communicating the latest development in the field. We did not find any record of conference coverage. It also shows *MEDLINE*s biasness towards Indian conferences and seminars.

4.4 Authorship Pattern

Lotka's law[5, 6,7] describes the frequency of publication by authors in a given field. It states that the number of authors making contributions is about $1/n^2$ of those making one; and the proportion of all contributors, that make a single contribution, is about 60 percent. This means that out of all the authors in a given field, 60 percent will have just one publication, and 15 percent will have two publications ($1/2^2$ times of 60). Seven percent authors will have three publications ($1/3^2$ times of 60), and so on. According to Lotka's law of scientific productivity, only six percent of the authors in a field will produce more than 10 articles. Lotka's law, when applied to large bodies of literature over a fairly long period of time, can be accurate in general, but not statistically exact. The general form of Lotka's law can be expressed as $y=c/x^n$ where y=percentage of authors, x=number of articles published by an author, c=constant and $-n$=slope of the log-log plot. There are altogether 8508 authors who contributed 6408 articles; on an average 1.32 authors per articles. Among 8508 authors, 4985 authors (58.59%) contributed only one article. 1274 authors (14.97%) contributed two articles and 657 (7.72%) authors contributed three articles. So, it can be said
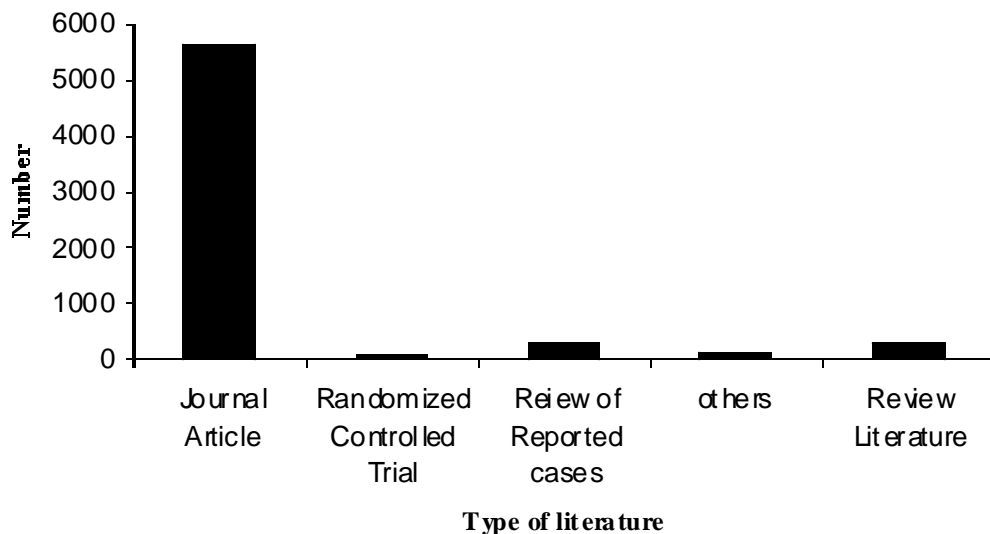

Figure 2. Types of literature where cancer related literature published

that Indian authorship pattern in Indian cancer research is in close conformity to original Lotka's law. Figure 3 shows the graph in which the number of author is plotted with publication. The log-log plot of author number and their contribution gives a straight line which is a good fit to original Lotka's law.

As the authors' list is long, we selected 19 authors (table 2) who published more than 50 articles. M.K. Nair, from Regional Cancer Centre, Trivandrum, is the most productive
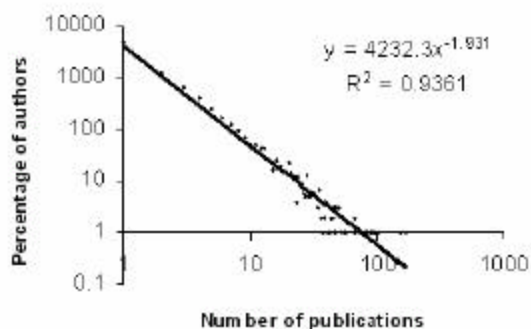
Figure 3. Author's productivity pattern

author who published 167 articles with a productive period of 17 years, followed by S.K. Advani, from Tata Memorial Hospital, Mumbai with 150 publications with same productive years.

Table 2. Authors with more than 50 publications

Sl. No	Name	No. of Publications	Productive years
1	Nair, M.K.	167	1987-2003
2	Advani, S.H.	150	1987-2003
3	Gupta, S.	103	1987-2003
4	Kumar, A	98	1987-2003
5	Kumar, S.	94	1987-2003
6	Goel, A.	92	1989-2003
7	Sharma, S.	88	1987-2003
8	Kumar, R.	83	1988-2003
9	Rath, G.K.	80	1990-2003
10	Sarkar, C.	75	1987-2003
11	Kumar, L.	73	1987-2003
12	Patel, D.D.	67	1987-2003
13	Shukla, N.K.	67	1988-2003
14	Sharma, M.C.	66	1995-2003
15	Pillai, M.R.	65	1987-2003
16	Pandey, M.	57	1994-2003
17	Verma, K.	55	1988-2003
18	Mathur, M.	54	1987-2003
19	Gupta, S.K.	52	1988-2003

4.5 JOURNALS

In total, there are 868 journals, which published 6408 articles. Bradford's law 8,9,10 of scattering can be employed to study journal literature distribution. The Bradford's distribution is used for identifying the 'core' journals. Core journals are central to a subject because they produce most of a subject's content. Other way, it is a Bradford analysis, which gives information about the amount and titles of core journals and about the number of journals needed to cover most of the relevant articles in a specific field. By graphically describing the scattering of articles in a specific field to different journals, it should be possible to divide, which journal should be included in a collection to cover a specific percentage of the relevant articles in the field. Because of increasing cost in today's library and information centers a typical Bradford analysis can suggest the journals to be procured in a library collection, which cover the most articles in a given field.

Figure 4 shows Bradford plot, where cumulative total of publications is plotted against the logarithm of journal's rank. On a Bradford plot, the core journals are those whose points lay on the initial curved part of the 'S' until tangentially becomes a straight line. Here, we observe that the slope of the curve decrease slightly after the nineteenth journal, so it appears that the top nineteen journals are well in their way to form a core.
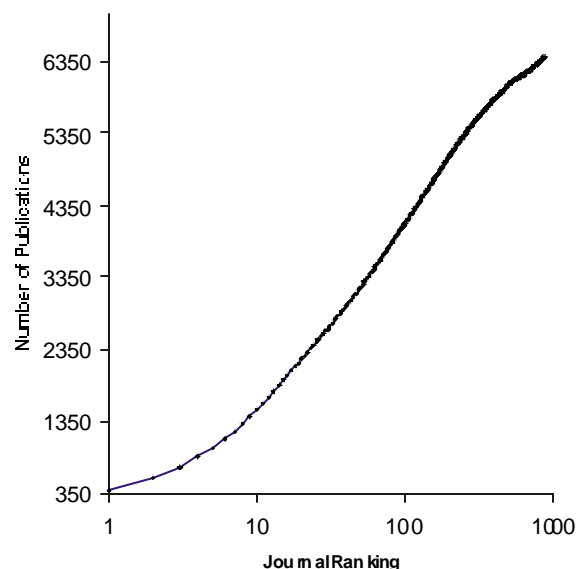
Figure 4. Bradford plot for literature distribution

Table 3. Core journals with the number of publications impact factor and publishers

S. No.	Name of the journal	No. of publications	Impact factor as of 2001	Publisher
1.	*Indian J Cancer*	397	-	Indian Cancer Society, Bombay
2.	*Cancer Lett*	177	1.741	Elsevier Science
3.	*Acta Cytol*	156	1.094	Science Printers and Publishers, St. Louis Mo
4.	*Neurol India*	139	-	Bombay Neurological Society of
5.	*Indian J Exp Biol*	118	-	CSIR, New Delhi
6.	*J Surg Oncol*	117	1.318	NY Wiley-Liss, New York
7.	*Trop Gastroenterol*	109	-	Vikas, Sahibabad, UP
8.	*Neoplasma*	106	0.637	Slovak Academic Press, Ltd, Slovak Republic:
9.	*Diagn Cytopathol*	101	0.956	Igaku-Shoin Medical Publishers, New York
10.	*Indian Pediatr*	96	-	Journal of the Indian Pediatric Society
11.	*Indian J Pathol Microbiol*	94	-	Indian Journal of Pathologists & Microbiologists, Belgaum
12.	*Br J Neurosurg*	86	0.563	Carfax Pub. Co.
13.	*J Assoc Physicians India*	82	-	Bombay Association of Physicians of India
14.	*Australas Radiol*	80	-	Sydney Blackwell Scientific Publications, Australia
15.	*J Laryngol Otol*	73	0.459	Headley Brothers, London
16.	*Int J Cancer*	68	4.233	Wiley-Liss, New York
17.	*Postgrad Med J*	63	0.441	BMJ Publishing Group, London
18.	*J Postgrad Med*	57	-	Journal of Postgraduate Medicine, Bombay
19.	*Cancer*	52(2171)	3.909	BioMed Central, London

Core 19 journals in the order of number of papers published, which published 1/3 of the articles in this area are listed in the table 3. remaining literature is scattered in 849 journals. This shows scattering of the literature. This scattering of information poses a great problem in information retrieval.

Looking deeply into the 19 core journals, 11 journals are from foreign publications and rest 8 are from India. Nine journals do not have any impact factor because they are not indexed in *SCI. Indian Journal of Cancer* published by Indian Cancer Society is the first ranked journal that publishes about 6% of the total articles. International Journal of Cancer, ranked fifteenth has the highest impact factor (4.233) in 2001. It can be concluded that Indian cancer research articles are not published in high impact journals.

4.6 Institutions

It is also important to explore author's affiliation in cancer literature. Authors are affiliated to 435 institutes including private clinics and individual addresses. 289 institutes have only one publication. Table 4 Shows state-wise distribution of literature. Among the metropolitan cities Delhi ranks first. About 50 institutes in Delhi contributed 1640 articles, which is about 25% of total literature. About 57 institutes in Mumbai are in second position with a contribution of 1147 articles, which is 17% of the total literature. 32 institutes in Calcutta published 281 articles (5%) and Chennai published 241 (3.76%) of the total articles. We also analyzed records up to the state level. Delhi is the most productive state with largest number of articles. This is because of high concentration of medical institutes and research centers. From the north east, we found only 12 articles from Assam, 6 articles from Manipur and 16 from Meghalayay. This may be due to the less concentration of research centers, universities or medical colleges in the north east. In the eastern region, the output of research papers is less compared to other regions. As seen, 6 articles from Bihar and 24 articles from Orissa were found. Among the Union territories, Chandigarh ranked first with 503 (7.85%) papers.

Table 4. State-wise distribution of literature

	State	Institution	Articles	%
1.	Delhi	50	1640	25.09
2.	Maharastra	87	1199	18.71
3.	Chandigarh	7	503	7.85
4.	Kerala	32	437	6.82
5.	Uttar Pradesh	34	432	6.74
6.	Karnataka	41	419	6.54
7.	Tamil Nadu	65	369	5.76
8.	West Bengal	44	331	5.16
9.	Andhra Pradesh	29	172	2.68
10.	Gujarat	14	118	1.84

Table 5 lists top 27 institutions according to their number of publications. All India Institute of Medical Sciences (AIIMS), Delhi is in the top with 1037 papers followed by Tata memorial hospital, Mumbai.

5. CONCLUSION

The study is concerned about bibliometric analysis of Indian cancer research as reported in NCBI *Pub Med*. The study shows that cancer research in India is increasing, with a marginal decrease in the year 1991, 1993, 1995, 1997 and 2003. From the last few years' trends, it can be said that Indian cancer researchers published around 500 literatures per year.

Worldwide trend of papers show that the cancer research work is being done on worldwide basis. Still, US is the largest producer of cancer related papers.

Pub Med indexes biomedical literature published in different communication medium, like other field of science and technology. Journal literature (88%) is the single most form of publication among the Indian scientists. Indian conference and symposia on cancer is not covered in MEDLINE.

Authorship pattern shows a close conformity with the original Lotka's law, where 58.59% of Indian authors have single publication, 14.97% publish two articles, 7.72% publish three articles. 19 most productive authors are identified who had published more than 50 articles.

Table 5. Productive institutions according to their number of publications

	Name of the institution	No. of publications
1.	All India Institute of Medical Sciences, Delhi	1037
2.	Tata Memorial Hospital, Bombay	784
3.	Post Graduate Institute of Medical Education & Research, Chandigarh	438
4.	Regional Cancer Center, Trivandrum, Kerala	258
5.	Christian Medical College & Hospital, Vellore	196
6.	Sanjay Gandhi Post Graduate Institute of Medical Sciences, Lucknow	192
7.	Cancer Research Institute and Tata Memorial Hospital, Mumbai	179
8.	Chittaranjan National Cancer Research Centre, Calcutta	151
9.	Maualana Azad Medical College, New Delhi	146
10.	Kastruba Medical College, Manipal, Karnataka	132
11.	King Edward Memorial Hospital, Seth G.S. Medical College, Mumbai	115
12.	Kidwai Memorial Institute of Oncology, Bangalore	110
13.	Banaras Hindu University, & Institute of Medical Sciences, Varanasi	102
14.	Gujarat Cancer & Research Institute, Ahmedabad	88
15.	Cancer Institute (W.I.A), Chennai	87
16.	Bhaba Atomic Research Center, Mumbai	74
17.	Centre for Cellular & Molecular Biology, Hyderabad	65
18.	National Institute of Mental Health & Neuro Sciences, Bangalore	63
19.	Jadavpur University, Calcutta	57
20.	Amala Cancer Hospital & Research Centre, Thrissur	56
21.	Industrial Toxicology Research Centre, Lucknow	56
22.	St. John's Medical College, Bangalore, Karnataka	50
23.	Nizam's Institute of Medical Science, Hyderabad	48
24.	Government Medical College, Chandigarh	46
25.	Jawaharlal Nehru University, Delhi	46
26.	Institute of Nuclear Medicine, Delhi	42
27.	University of Madras, Chennai	42

Bradford's law of scattering is employed to identify core journals. 19 core journals are identified which contains 1/3 of the total articles. Among 19 journals 11 journals are published outside India. Impact factor of the journals shows that Indian oncology related research publish in low impact journals.

All India Institute of Medical Science (AIIMS), Delhi has produced maximum papers on cancer research followed by Tata Memorial Hospital, Mumbai.

REFERENCES

1. Tsay, M.T.; Jou, S.J.; & Ma, S.S. A bibliometric study of semiconductor literature, 1978-1997. *Scientometrics,* 2000, **49** (3), 491-509.

2. http://www.ncbi.nlm.nih.gov (04 April 2004)

3. http://www.nlm.nih.gov/mesh/MBrowser.html (04 April 2004)

4. Lancaster, F.W. Vocabulary control for information retrieval, Ed. 2. Information Resources, Arlington, V.A.1986.

5. Kawamura, M., Thomas, C.D.; Kawaguchi, Y.; & Sasahara, H. Lotka's law and the pattern of scientific productivity in the dental science literature. *Med. Inform. Internet Med,* 1999, **24**(4), 309-15.

6. Lotka, A.J. The frequency distribution of scientific productivity. *Journal of the Washington Academy of Science,* 1926 **16**, 317-23.

7. Sittig, D.F. Identifying a core set of medical informatics serials: An analysis using the MEDLINE database. *Bull. Med. Libr. Assoc.,*1996, **84**(2), 200-04.

8. Brooks, B.C. Bradford's law and the bibliography of science. *Nature,* 1969, **224**, 953-56.

9. Bradford, S.C. Sources of information on specific subjects. *Engineering: An Illustrated Weekly,* 1934, **137**(3550), 85-86.

10. Ungern-Sternberg, S.V. Bradford's law in context of information provision. *Scientometrics,* 2000, **49**(1), 161-86.

Contributors: **Sh. Swapan Kumar Patra**, Department of biochemical Engineering & Biotechnology, Indian Institute of Technology Delhi, New Delhi -110016.

Sh. Partha Bhattacharya, FITT, Indian Institute of Technology Delhi, New Delhi -110016. e-mail: parthab@fitt.iitd.ernet.in Ph: 91-11-26581013