Growth Analysis of Cancer Biology Research, 2000-2011

Keshava *

Dept. of Studies and Research in
Library and Information Science
Tumkur University, India
keshtut@gmail.com

B. N. Thimmaiah

Dept. of Studies and Research in
Library and Information Science
Tumkur University, India
thimmegowda55@gmail.com

K. B. Agadi

Gujarath Central University
Library, India
kbagadi@gmail.com

ABSTRACT

Methods and Material: The PubMed database was used for retrieving data on 'cancer biology.' Articles were downloaded from the years 2000 to 2011. The articles were classified chronologically and transferred to a spreadsheet application for analysis of the data as per the objectives of the study.

Statistical Method: To investigate the nature of growth of articles via exponential, linear, and logistics tests.

Result: The year wise analysis of the growth of articles output shows that for the years 2000 to 2005 and later there is a sudden increase in output, during the years 2006 to 2007 and 2008 to 2011. The high productivity of articles during these years may be due to their significance in cancer biology literature, having received prominence in research.

Conclusion: There is an obvious need for better compilations of statistics on numbers of publications in the years from 2000 to 2011 on various disciplines on a worldwide scale, for informed critical assessments of the amount of new knowledge contributed by these publications, and for enhancements and refinements of present Scientometric techniques (citation and publication counts), so that valid measures of knowledge growth may be obtained. Only then will Scientometrics be able to provide accurate, useful descriptions and predictions of knowledge growth.

Keywords: Scientometric, Cancer biology, PubMed, Relative growth rate, Exponential trend

1. INTRODUCTION

Cancer can be defined as a disease in which a group of abnormal cells grow uncontrollably by disregarding the normal rules of cell division. Normal cells are constantly subject to signals that dictate whether the cell should divide, differentiate into another cell, or die. Cancer cells develop a degree of autonomy from these signals, resulting in uncontrolled growth and proliferation. If this proliferation is allowed to continue and spread, it can be fatal; almost 90% of cancer-related deaths are to tumour spreading. The foundation of modern cancer biology rests on a simple principle that virtually all mammalian cells share similar molecular networks that control proliferation, differentiation, and cell death.

Phenomenal advances in cancer research have given an insight into how cancer cells develop (Hejmadi, 2010). The growth of literature is a key work for all scholars and students of comparative literature (Chadwick, 1986). The changes in the size of literature over a specific period are termed as growth literature. Gilbert (1978) has reviewed the existing literature on the indicators of growth of knowledge in scientific specialties, and has listed many ways of measuring it, noting their strengths and limitations and commenting, at the same time, on their uses. Gupta, Sharma, and Karisiddappa (1977) suggested two approaches that have normally been considered in understanding knowledge growth: (i) Qualitative and (ii) Quantitative. A qualitative approach suggests structural or descriptive models of knowledge growth, while a descriptive model uses social phenomenon to explain diffusion and creation of knowledge. A quantitative approach employs summarization of statistics to describe the observed behaviour, while applying growth and technology diffusion models and bibliometric/Scientometric techniques. Many studies have been made on the growth of literature in the field of science literature but only very few studies have been conducted on cancer biology research. Hence an attempt has been made to analyse the growth of cancer biology research literature. The authors studied only growth indicators as a pilot study for further study, viz. collaborative research trend; cross national analysis; obsolescence; etc.

2. OBJECTIVES

- To study and explore the applicability of selected growth models in the world in cancer biology literature;
- To determine the rate of growth of cancer biology literature by calculating relative growth rates and doubling time for publications;
- To fit both a modified exponential curve and logistic curve for the original publications data, studying actual growth.

3. REVIEW LITERATURE

Literature growth studies have become very common in the field of Bibliometrics, Informatrics, and Scientometrics. Studying growth patterns in the NLM's serials collection and in Index Medicus journals between 1966 and 1985, Leeds (1964) "Law of Exponential Growth" has been further dealt with by Tague, Beheshti, and Rees-Potter (1981), and by Ravichandra Rao and Meera (1992). The exponential growth of the literature is described mathematically by the exponential function YT =a.ebt where YT represents the size at time t, a is the initial size, and b is the continuous growth rate which is related to the annual percentage growth rate r, as: r =100(eb-1). Egghe and Ravichandra Rao (1992) concluded, however, that the power model (with exponent >1) is the best growth model for sciences and technology fields, while the Gompertz S-shaped distribution better fits databases of the social sciences and the humanities.

4. SCOPE AND LIMITATION

The present study is confined to cancer biology literature as reflected in the PubMed database from the years 2000 to 2011. $R_t(P) = \frac{1}{t}\left[\log_e p(t) - \log_e p(0)\right]$

5. METHODOLOGY

The PubMed database was used for retrieving data on 'cancer biology'. Articles were downloaded cover-

ing the years 2000 to 2011. The articles were classified chronologically and transferred to a spreadsheet application, and the data is analysed as per the objectives of the study.

6. STATISTICAL METHOD

To investigate the nature of growth of articles via exponential, linear, and logistic tests. The exponential growth is defined as

$$F(t) = a\ e^{bt}$$

Where

a = the initial size of literature, i.e. at time t=o and b, the continuous growth rate is related to the percentage by which the size increases each year.

The logistic has a lower limit and an upper limit or a ceiling beyond which the size cannot grow, and can be represented mathematically as

$$U_t = \frac{K}{1+\mu}$$

Where,

U = expected size of literature

K and μ = constants and t= time.

Similarly, the linear growth is represented as

$$U_e = a + bt$$

Relative Growth Rate (RGR) and Doubling Time (Dt) has been applied. RGR means the increase in the number of articles per unit of time. The mean RGR of articles over the specific period of interval is represented as

$$R_t(P) = \frac{1}{t}\left[\log_e p(t) - \log_e p(0)\right]$$

R_t = Relative Growth Rate of articles over the specific period of time

$\log_e p(0)$= Logarithm of initial number of articles

$\log_e p(t)$= Logarithm of final number of articles

Similarly, RGR of subject articles has increased in number of articles per unit of time. The mean RGR of subject articles $R_t(SA)$ over specific period of time is determined as

$$R_t(SA) = \frac{1}{t}\left[\log_e p(t) - \log_e p(0)\right]$$

$R_t(SA)$ = Relative Growth Rate of articles over the specific period of time

$\log_e p(0)$= Logarithm of initial number of articles

$\log_e p(t)$= Logarithm of final number of articles

D_t (Doubling Time) has been calculated using the following formula:

$$\text{Doubling Time } D_t = 0.693/R$$

D_t (Doubling Time) is directly related to RGR and is defined as the time required for the articles to double in number from the existing amount. If the number of articles in a subject doubles during a given period, then the difference between logarithms of number at the beginning and at the end of this period must be the logarithm of the number 2. We used a Napier logarithm and so the taken value of $\log_e 2$ is 0.693. Hence, as per this (0.693) and an average growth rate we calculated by what time interval does the Napier logarithm of numbers increase by 0.693. So the Doubling Time is calculated as

$$D_t(SA) = \frac{\log_e 2}{R_t(SA)} = \frac{0.693}{R_t(SA)}$$

Here, $D_t(SA)$ = average doubling time of subject articles

7. ANALYSIS AND DISCUSSION

As Table 1 clearly indicates, the value of an average RGR of articles $R_t(P)$ increased gradually from 0.57 to 2.05. Correspondingly, the values of Doubling Time of the publication of articles $D_t(P)$ increased gradually from 2002 (1.84 years) to 2011 (4.10 years). The mean relative growth $R_t(P)$ for the first 6 years (from 2000 to 2005) indicates a growth rate of 1.00, whereas for the latter 6 years (from 2006 to 2011) it increased to 1.92. The linear and exponential growth trend is fitted to number of articles for the years 2000 to 2011. Table 1 and Figs. 2 and 3 reveals that the R^2 value for the exponential trend (0.9427) is more than that of the linear trend (0.9339); this indicates that the exponen-

Table 1. Relative Growth-rate (RGR) and Doubling Time (Dt) of Articles in Cancer Biology from 2000 to 2011

Year	No. of Articles	Cumulative	$\log_e 1^P$	$\log_e 2^P$	$R_t(P)$	Mean $R_t(P)$	$D_t(P)$	Mean $D_t(P)$
2000	341	341	5.83	5.83	0		0	
2001	440	781	6.08	6.66	0.57		1.15	
2002	517	1298	6.24	7.16	0.92		1.84	
2003	450	1748	6.10	7.46	1.36		2.71	
2004	525	2273	6.26	7.72	1.47		2.93	
2005	529	2802	6.27	7.93	1.67	1.00	3.33	1.99
2006	569	3371	6.34	8.12	1.78		3.56	
2007	681	4052	6.52	8.30	1.78		3.57	
2008	728	4780	6.59	8.47	1.88		3.76	
2009	750	5530	6.62	8.61	2.00		4.00	
2010	869	6399	6.76	8.76	2.00		3.99	
2011	944	7343	6.85	8.90	2.05	1.92	4.10	3.83
Total	7343	40718						

R^2 (Linear trend for no. of articles) = **0.9339**
R^2 (Exponential trend for no. of articles) = **0.9427**
R^2 (Exponential trend for cumulative no. of articles) = **0.9138**

tial trend is more suitably fitted as compared to the linear trend. Further, the exponential trend is fitted to cumulative number of articles from 2000 to 2011. The R^2 value for this trend is 0.9138; this shows 91.38% variation observed from the cumulative number of articles.

Furthermore, mean $D_t(P)$ for the first six years was 1.99 and increased to 3.83 in the latter six years, i.e. from 2006 to 2011. It shows that the mean relative growth of cancer biology literature has shown an increasing trend. This may be due to the interdisciplinary and multidisciplinary nature of research and the communication patterns of medical researchers.

The year wise analysis of the growth of articles output shows that growth is poor in the year 2000 to 2005 and then there is a sudden increases in productivity during the years 2006 to 2007 and 2008 to 2011. The high productivity of articles during these years may be due to their significance in cancer biology literature, having received prominence in research. Another

reason is the research area may benefit from good infrastructure facilities in R&D institutions.

8. SUGGESTIONS

a) Since the growth of literature is on an exponential trend, medical libraries may think of resource sharing and networking options in order to avoid financial constraints;

b) Medical libraries may allocate budget for various resources scientifically.

9. CONCLUSION

The many papers have tried to estimate the growth of knowledge in various ways, as many questions have been raised about the validity and reliability of Scientometrics measures for this process. It appears
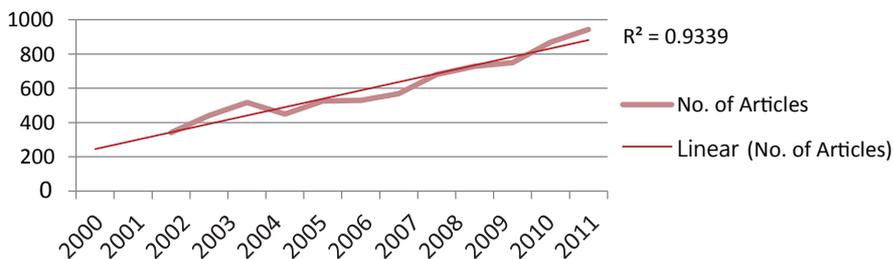
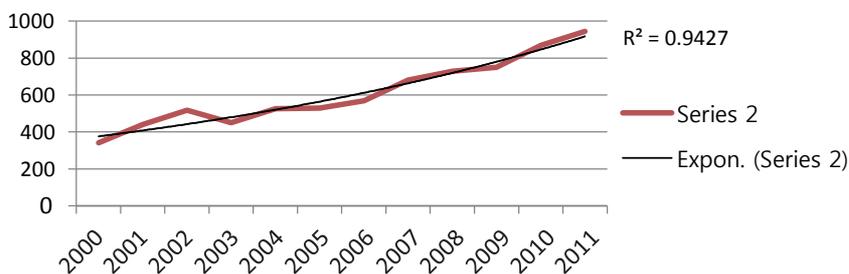**Fig. 1** Linear trend for no. of articles

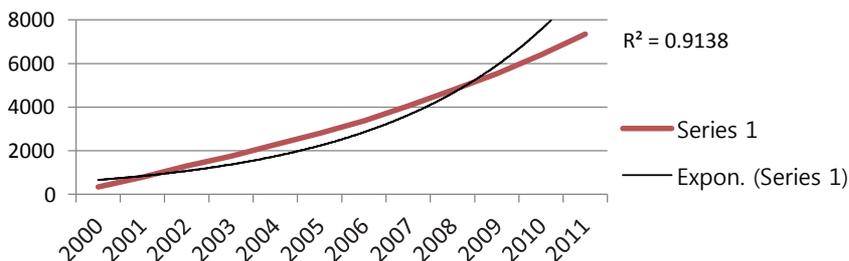**Fig. 2** Exponential trend for no. of articles

Fig. 3 Exponential trend for cumulative no. of articles

that, for the "growth of knowledge" subfield, the time is not yet ripe for a logarithmic decline in the number of first-rate papers. There is an obvious need for better compilation of statistics on numbers of publications in the years from 2000 to 2011 in various disciplines on a worldwide scale, for informed critical assessments of the amount of new knowledge contributed by these

publications, and for enhancement and refinement of the present Scientometrics techniques (citation and publication counts), so that valid measures of knowledge growth may be obtained. Also, studies of literature growth need to become more exact in the description of their models and more rigorous in the application of statistical tests to determine how well these models fit

reality. Only then will Scientometrics be able to provide accurate, useful descriptions and predictions of knowledge growth.

REFERENCES

Chadwick, H.M., & Chadwick, N.K. (1986). *Growth of literature*. Cambridge: Cambridge University Press. (Original work published 1932).

Egghe, L., & Ravichandra Rao, I.K. (1992). Classification of growth models based on growth rates and its applications. *Scientometrics, 25*(1), 5-46.

Gilbert, G.N. (1978). Measuring the growth of science: A review of indices. *Scientometrics, 1*(1), 9-34.

Gupta, B.M., Sharma, P., & Karisiddappa, C.R. (1977). Growth of research literature in scientific specialties: A modeling perspective. *Scientometrics, 40*(3), 507-528.

Hejmadi, M. (2010). *Introduction to cancer biology*, 2nd ed. London: Bookboon.

Keshava. (2004). *Scientometric analysis of social science research in India* (Ph.D. thesis). Dept. of Library and Information Science, Karnataka University, Dharwad.

Line, M.B., & Roberts, S. (1976). Growth and composition of social science literature. *International Social Science Journal, 28*(1), 122-159.

Orr, R.H., & Leeds, A.A. (1964). Biomedical literature: Volume, growth, and other characteristics. *Federation Proceedings, 23*(6), 1310-1331.

Ravichandra Rao, I.K., & Meera, B.M. (1991). Growth and obsolescence of literature: An empirical study. In I.K. Ravichandra Rao (Ed.), *INFORMETRICS '91: Selected Papers from the Third International Conference on Informetrics, 9-12 August 1991* (pp. 377-394). Bangalore, 1992.

Roberts, S.A., & Chak, M. (1981). Size growth and characteristics of the serial literature of geography. *Social Science Information Studies, 1*(5), 317-388.