DESIDOC Journal of Library & Information Technology, Vol. 45, No. 4, July 2025, pp. 378-394, DOI : 10.14429/djlit.21019

Testing the Applicability of Lotka's Law, Bradford's Law, and Zipf's Law on Gastritis Research Output

Amir Amanullah[~,*], S Rajeswari[$] and Sumeer Gul[#]

[~]*Government Degree College Kangan, J&K Government Higher Education Department, Kangan - 191 202, India*
[$]*University College of Engineering BIT Campus, Anna University, Tiruchirappalli - 620 024, India*
[#]*Department of Library and Information Science, University of Kashmir, Srinagar - 190 006, India*
[*]*Email: aman.lis.drass@gmail.com*

ABSTRACT

This study investigates the applicability of the fundamental bibliometric laws-Lotka's Law, Bradford's Law, and Zipf's Law-within the domain of Gastritis research. Utilising data from the Web of Science Core Collection, a total of 19,856 records on Gastritis research were analysed using various scientometric tools and statistical tests. Lotka's Law, which models author productivity, was tested through the Kolmogorov-Smirnov (K-S) test and the Chi-Square test which confirmed significant discrepancies between the observed and expected distributions of publications. Bradford's Law, applied to journal productivity and article scattering, revealed a deviation from the anticipated 1:α:α² ratio. Zipf's Law was validated through an analysis of the most frequently used terms and the results demonstrated an inverse relationship between rank and frequency, supporting the applicability of Zipf's Law to the Gastritis research corpus.

Keywords: Applicability of lotka's law; Applicability of bradford's law; Applicability of zipf's law; Sceintometric analysis; Kolmogorov-smirnov (K-S) test; Chi-square test; Bibliometric analysis; Gastritis

1. INTRODUCTION

Bibliometric or scientometric studies are playing a great role in analysing the scientific literature based on several parameters. Scientometric studies are one of the endeavors to pinpoint the strengths, weaknesses, and opportunities of the research activity of a nation, an organisation, or a discipline. The fundamental laws of bibliometrics include three laws: the law of author productivity, popularly known as Lotka's inverse square law, put forth by Lotka, A.J[1]; the law of scattering or journal productivity, put forth by Bradford, S.C. (1934); and the law of word occurrence or rank-frequency equation, devised by Zipf, G.K[2]. Fairthorne[3], defined bibliometrics as "quantitative treatment of the properties of recorded discourse and behavior appertaining to it". Interestingly, the work of Cole & Eales[4] was regarding the publication output of comparative anatomy from 1550-1860, where they tried to statistically evaluate the output, what we call "scientometrics" or "bibliometrics" today. With the course of time, there came the term "statistical bibliography," which was coined by Hulme[5] during a series of lectures, notes Sengupta in Sengupta[6].

The fundamental laws of bibliometrics are the fundamental parameters for finding out the journal productivity, author productivity, and the keyword occurrences in any field of scientific research. Researchers across the globe use reliable and scientific bibliometric/statistical tools and techniques for quantitative analysis of the research output of n number of entities. The current research is an endeavor to find out the applicability of the fundamental laws of bibliometrics to "Gastritis Research".

2. LITERATURE REVIEW

Review of available literature in the field of study is one of the most important and basic steps to carry out a research. A number of scientific literature has been consulted to understand the research done in the field of scientometrics visa viz use of bibliometric/sceintometric parameters and tests and fundamental scientometric laws. Following are some reviews of the scholarly communications available in the current field of study:

Borgohain, Dhruba Jyoti[7], *et al.* in their research paper explores the theoretical and practical application of Bradford's Law within Information Science literature, analysing 213 source items from the Scopus database (2001-2020). The study assesses both the verbal and algebraic formulations of Bradford's Law and applies the Leimkuhler model to evaluate journal dispersion. While the initial verbal formulation yielded a high error rate, the Leimkuhler model demonstrated a negligible error of 0.0092 %, validating the law's relevance. The

Received : 01 February 2025, Revised : 19 April 2025
Accepted : 21 May 2025, Online published : 15 July 2025

study identified "Scientometrics" and "Bulletin of the Medical Library Association" as top journals. Though Bradford's ideal 1:n:n^2 pattern was not fully observed, the journal distribution aligned more closely through graphical analysis and citation studies. Ram, Shri, and Nitin Paliwal[8] in a study verify Bradford's Law using the Leimkuhler model to examine the distribution of 24,031 psoriasis-related articles across 2,143 journals. While the traditional Bradford model predicted 11 core journals, it resulted in a high error rate. In contrast, the Leimkuhler model identified nine core journals with a minimal error margin of 0.045 %, confirming its greater accuracy. The distribution followed the Bradford pattern of 9:134:2000 across three zones with an average multiplier of 14.91. The British Journal of Dermatology emerged as the most prolific. Graphical analysis via the Bradford-Zipf plot supported these findings. Savanur, Kiran[9], *et al.* in their study investigates the applicability of Lotka's Law to authorship distribution in legal scholarship, focusing on articles published in the Columbia Law Review (1901–2012).

Lotka's Law, a foundational bibliometric principle, was tested using Least Squares, Maximum Likelihood, and Sen's Method, with validation via the Kolmogorov–Smirnov test. Results indicate that authorship distribution aligns with Lotka's predicted pattern, especially with Pao's method yielding values (n = 2.1424, c = 0.6540) closely matching observed data. Over 73 % of contributors authored only one article, with productivity sharply declining beyond that. Amanullah[10] conducted a bibliometric analysis to explore the interdisciplinary nature of Digital Humanities (DH), revealing its broad subject affiliations across more than 95 distinct fields. The top five subject areas include Arts & Humanities, Literature, Computer Science, Information Science & Library Science, and Linguistics, illustrating the wide-ranging scope of DH research. The study found that DH research is primarily published in literature and arts & humanities journals, although journals from computer science and library and information science also play a significant role. Additionally, the analysis highlighted a strong trend of author collaboration, with the largest number of contributors hailing from computer science, followed by arts & humanities and library and information science.

Saxena, Anurag et al[11] in a research endeavor explores the application of Zipf's Law to a randomly selected English-language text from computer science literature, focusing on the influence of ranking methods and text types. It examines three ranking techniques-random rank, maximal rank, and tied rank-by comparing the product of rank and frequency. Findings indicate that while Zipf's Law generally holds, the exponent varies with text type. The maximal and tied rank methods outperform the random rank method. Additionally, the study questions the validity of generalisations in extreme regions of Zipf's curve, particularly regarding Good's data, challenging the linearity suggested by the Mandelbrot-Zipf Law. The study recommends further research on how ranking methods interact with specific text characteristics. Amanullah A. & Rajeswari S.[12,13,14] offered an insightful analysis of gastritis research uncovering growth, authorship and citation pattern in gastritis research. The researchers analysed global research output, research output of G20 nations and Open Access research output on gastritis in three different studies. Khiste and Amanullah, A.[15] examined the global publication trends in "Knowledge Management" and found that the highest number of publications occurred in 2016. The USA, England, and Spain ranked as the top three most productive countries in this field. Hong Kong Polytechnic University emerged as the leading institution for publications on this topic. The majority of the publications were in the form of articles. Suk Wang Ho, Wang & Wang[16] conducted a study on classic scientific literature related to Helicobacter pylori research. Their analysis revealed that the majority of publications were in English, with the USA being the most productive country. Vanderbilt University ranked as the top institution in terms of publication output. Among the 18 classic publications on Helicobacter pylori, M.J. Blaser authored the highest number of publications.

The most-cited paper received 2,964 citations, while the least-cited paper had 1,086 citations. De Souza, Crisomar[17], *et al.* in their study presents a comprehensive bibliometric analysis of ESG (Environmental, Social, and Governance) research using Bradford's, Lotka's, and Zipf's laws. It explores the relationship between corporate social responsibility and access to finance, highlighting reduced agency costs and improved transparency. The findings confirm Lotka's law, with a few prolific authors contributing most publications. Bradford's law identifies finance as the dominant discipline, though key journals belong to environmental studies. Zipf's analysis shows "ESG," "social," and "governance" as dominant keywords. The paper also maps global authorship patterns, revealing strong U.S. output and limited international collaboration. The study encourages future ESG research, especially in emerging areas like innovation, digital transformation, and stakeholder engagement. Batcha, Jahina, and Ahmad[18] analysed the publication patterns of the DESIDOC journal of library and information technology (DJLIT) over a five-year period. They found that the majority of publications were co-authored (48.90 %), followed by single-author papers (30.84 %), with an average productivity of 2.59 publications per author.

Subject analysis showed that the most frequent publications focused on "Scientometrics, Bibliometrics, and Webometrics," followed by topics such as user surveys. The mean doubling time was calculated to be 1.78, with 2014 and 2017 being the most productive years during the study period. Amanullah A. and Khiste GP[19] conducted a comprehensive bibliometric analysis of global publication productivity on "E-learning" between 2003 and 2017. The United Kingdom, United States, and China were identified as the top contributors, with computer science being the dominant subject area. The analysis underscored the significant progress made in E-learning research, particularly in the areas of digital platforms and educational technologies.

3. STATEMENT OF PROBLEM

Research has been conducted on a number of research topics and fields related to medical science. Gastritis has been a topic of interest for gastroenterologists, immunologists, pharmacologists and microbiologists. Given the multidisciplinary nature it remains uncertain whether the fundamental bibliometric laws hold true to this domain. Thus the core problem can be understood by testing the applicability of the three fundamental bibliometric laws to gastritis research. The result of the research may become a base for carrying out further research comparing the applicability of the fundamental bibliometric laws to main subject/basic subject fields and multidisciplinary/interdisciplinary fields.

4. SIGNIFICANCE OF THE STUDY

The theoretical significance of the study lies in the empirical testing of the fundamental bibliometric laws of bibliometrics. The study enhances the understanding of different testing models in the field of scientometrics/bibliometrics. It also adds up to the scholarly communications in the field of gastritis and scientometrics. The practical significance of this study lies in the enhanced valuable insights that can help the practicing librarians, policy makers and the experts involved in collection development especially journal selection in the era of multidisciplinary research. Given the increased importance of scientometric studies in policy making and research evaluation this study provides valuable insights and also encourages further research and comparative analysis across research topics, testing models and databases.

5. OBJECTIVES OF THE STUDY

The objective of this study is to test the applicability of the three fundamental laws of bibliometrics to the gastritis research. A number of statistical tests, scientometrics tools, and parameters have been used in order to attain the goal.

6. METHODOLOGY

6.1 Use of Database for Data Harvesting

Several indexing and abstracting databases are available; using which different studies can be carried out. For the present study, one of the popular databases, Web of Science, has been used. The data is collected from the Web of Science Core Collection by using the search string as TS=Gastritis OR Chronic Gastritis OR Atrophic Gastritis. The resulting output was 19,940 records, and after filtration and cleaning of the data, a total record of 19,856 has been taken for study. The Web of Science database has been used because of its wide range indexing of quality journals in the given field.

6.2 Use of Tools for Analysis

This data was analysed as per the objectives of the study with the help of available software like BibExcel and Microsoft Office, and other aids. For data retrieval, processing, and reporting, this study used a multi-tool strategy. BibExcel has been used to analyze the author productivity, authorship pattern and journal productivity in gastritis research while as MS Excel has been used to tabulate the data and use of required formulae, MS Word is used for writing the manuscript.

6.3 Use of Methods and Statistical Tests

For finding the applicability of Lotka's Law Chi-Square test and K-S test have been applied. In order to find the values of the exponent (n) and the constant (C) the least square method proposed by Pao (1986) and Riemann zeta function have been used and later have been employed to understand the applicability of the fundamental bibliometric laws to the gastritis research.

Estimation of value of n: $$n = \frac{N\sum XY - \sum X \sum Y}{N\sum X^2 - (\sum X)^2}$$

Where,

N = No of sample size (data set)
X = Natural log of contributions (Articles)
Y = Natural log of contributors (Authors)

Estimation of C:

$$C = \frac{1}{\sum_1^{p-1} \frac{1}{x^n} + \frac{1}{(n-1)(p^{n-1})} + \frac{1}{2p^n} + \frac{1}{24(p-1)^{n+1}}}$$

In the above equation
P= Sample size
x= number of contributions
n= exponent

While testing applicability of the Bradford's Law of Scattering on Gastritis the journals are arranged according to their productivity and are distributed in three zones with equal contributions of articles and is then tested with the Bradford ratio.

While testing the applicability of Zipf's Law of Ranking of Word Occurrence the words are ranked according to their frequency of occurrences and the equation rf=c has been used. The value of C is then calculated using the log of r and f.

7. ANALYSIS AND RESULTS

This research is wholly and solely dedicated to understanding the applicability of the fundamental bibliometric laws to gastritis research. The chronological growth, country, organisation, journal, and author metrics have all been excluded. The data harvested and curated for analysis as per the objectives show stunning results. The results in the tables are found using required formulae and equations.

Most of the methods and equations used for calculation of the values are given in the first row of the tables.

7.1 Investigating the Applicability of Lotka's (N Value) Law

7.1.1 Lotka's Law

Popularly known as inverse square law is proposed by Lotka in which he discovered a relationship between authors contributing one publication and those making n number of contributions and this way defined the author productivity in a given field.

The equation thus reads as.

$$Y = \frac{1}{x^2}(\alpha)$$

Here Y represent the number of authors making x contributions and α equals 60 because as per Lotka "the number (of authors) making n 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". In simple words the authors making 2 contributions equal one fourth of those authors making one contribution and those making three contributions equals to one sixth of those making one.

Lotka's general power law can be written as

X^n Y=C

In the above equation Y are the number of authors making X contributions with n exponent and C is a constant. So to find the values of the exponent and the constant Pao (1986) proposed least square method

Equation-1 (Estimation of value of n)

$$n = \frac{N\sum XY - \sum X \sum Y}{N\sum X^2 - (\sum X)^2}$$

Where,

N = No of sample size (data set)

X = Natural Log of Contributions (Articles)

Y = Natural Log of Contributors (Authors)

Equation-2 (Estimation of C)

$$C = \frac{1}{\sum_1^{p-1}\frac{1}{x^n} + \frac{1}{(n-1)(p^{n-1})} + \frac{1}{2p^n} + \frac{1}{24(p-1)^{n+1}}}$$

In the above equation

P= Sample size

x= number of contributions

n= exponent

In Appendix-1 "x" represents the number of contributions, "y" represents the number of authors making "x" contributions. "X" is Log(x), "Y" is Log(y) and thus the data is tabulated accordingly.

$$n = \frac{N\sum XY - \sum X \sum Y}{N\sum X^2 - (\sum X)^2}$$

$$n = \frac{90(96.1188) - (141.7178)(85.6177)}{90(240.7230) - (141.7178)^2}$$

$$n = \frac{8650.512 - 12133.5520}{21665.07 - 20083.9348}$$

$$n = \frac{-3483.04}{1581.1352}$$

$$n = -2.2028$$

To calculate Value of C inversion of Riemann zeta function is used as under

$$C = \frac{1}{\sum_1^{p-1}\frac{1}{x^n} + \frac{1}{(n-1)(p^{n-1})} + \frac{1}{2p^n} + \frac{1}{24(p-1)^{n+1}}}$$

In the above equation

p= sample size= 90

p-1= 89

x= number of contributions = 1, 2, 3.............. in the table

n= -2.2028

Now by putting the values in the equation we get

$$\frac{1}{1.4842 + \frac{1}{(2.2028-1)(90^{2.2028-1})} + \frac{1}{2(90)^{2.2028}} + \frac{1}{24(90-1)^{2.2028+1}}}$$

$$\frac{1}{1.4842 + 0.003708 + 0.00002478 + 0.00000002}$$

$$\frac{1}{1.487933}$$

C = 0.6720

7.1.2 Kolmogorov- Smirnov Testing of Lotka's Law on Gastritis Research

The Lotka's Law is put to test in Kolmogorov-Smirnov Test Model in order to test the goodness of fit while determining the Maximum Deviation (D. Max).

In the Appendix-2 the data is given so as to calculate the value of D. Max to arrive on a conclusion. Around 55943 authors contributed on Gastritis research and the highest number of authors contributing only one article. In order to find out the applicability of Lotka's law on scientific productivity of authors on Gastritis research the K-S test is applied. The n value as already calculated equals -2.2028 and the calculated value of C is 0.6720. In the Appendix-2 the given data is calculated by substituting the values of n and c. the n value (2.2028) is matched with R Rousseau table value to get c value which equals 0.6709 and the same is derived by using the formula also.

The goodness of fit test is applied on the above data with a significance value of 5 % with the condition that p>40

K-S Critical value= $1.36/\sqrt{\sum Y}$

K-S Critical Value= $1.36/\sqrt{55943}$

= 1.36/236.5227

=0.00575

Now

N=2.2028

D (Max) = 0.02672

K-S Critical Value= 0.006892

Since the D (Max) value is greater than the K-S Critical Value thus the present data set is not in conformity with Lotka's Law.

7.1.3 Chi-Square Testing of Lotka's Law on Gastritis Research

Chi-Square model is used to test any discrepancy between the actual and observed results. In the Appendix 3 the Chi-Square test in implemented and the formula used is as under

Chi-Square $\chi^2=\sum(O-E)^2/E$

Where O=F= Observed Value

E=P= Expected Value (Lotka's expected no of authors)

Hence

$\chi^2= \sum (F-P)^2 /P$

$\chi^2= \sum (F-P)^2 /P$= 2879.6535

Now let's calculate the calculated value of χ^2 and tabular value of Chi-Square and for this we have to first find the degree of freedom (df) for goodness of fit

Degree of freedom df=N-1

In table 4.28 the sample size N=90

So df= 90-1=89

By using the Chi-Square distribution table

The tabular value at significance level of 0.05 is **112.022**

The calculated value= **2879.6535**

Hence from the above test the calculated value is greater than the Tabular value, the result is significant.

7.2 Applicability of Bradford's Law of Scattering on Gastritis

In the present study there are 2055 journals that contribute to production of 13930 articles on Gastritis. These articles are scattered in the journals. The scattering of the articles on a given topic is studied by S C Bradford and he found that most of the articles are distributed in few journals and some articles are distributed in higher number of journals. After the study S C Bradford found a general phenomenon in the scattering of the articles and he identified the journals as core zone or Zone-1 Journals, Zone-2 Journals and Zone-3 Journals. In the core zone there are very lesser journals contributing most of the articles, in the second zone there are more journals contributing average number of journals and in the 3rd zone there is large number of journals contributing few articles. Based on his study he found a relationship in each zones in the fashion of 1: α: α^2 : α^3 : α^4 ....

7.2.1 Bradford's Law of Scattering on Gastritis

Bradford SC[20] put forth the law of scattering. The Appendix-4 shows journals ranking as per their publication productivity. The journal with highest number of publication is on the top and the one with least number of publications is on the bottom of the rank table. The data suggest that to produce first 1/3rd of the articles 25 journals are responsible and for the 2nd 1/3rd of articles 209 journals are responsible and for the 3rd 1/3rd of articles 1821 journals are responsible. As per Bradford the first zone is the core.

7.2.2 Journal Distribution in Zones for Gastritis Research

In table-1 the zones and the number of journals and article falling in each zone is given in detail. As per Bradford's distribution the ratio expected is 25:625:15625. But as per the current study shows that the second zone is concentrated and include fewer journals than the Bradford's expected number and the third zone also have showed more concentration where the expected number of articles is produced by lesser number of journals than the Bradford's expected number. The study thus reveals that the core zone is extended to more journals than the expected. The current ratio is not in conformity with the Bradford's ratio.

Table 1. Journal distribution in zones for gastritis research

Zone	Number of journal	Number of articles	Multiplier factors
Zone-1	25	4641	
Zone-2	209	4653	8.36
Zone-3	1821	4636	8.71
Total	2055	**13930**	17.7
Mean value of Bradford Multiplier			**8.85**

7.2.3 Use of Bradford Multiplier to Calculate Percentage of Error

According to Bradford the relationship between the zones is 1:n:n², using the mean value of Bradford Multiplier we can find the percentage of error and is given below:

For the given data in table-5 summation of 1:n:n² equals 2204. The percentage of error can be calculated with the help of the equation (A-B/B) 100 where A is Bradford Expected number of journals and B is the actual number of journals. Putting the values in the equation we get

$$\text{Percentage of error} = \frac{2204-2055}{2044}(100)$$

Percentage of error= 7.26 %

Since the percentage of error is much higher, it can be deduced that the Bradford ration is not confirmed in the current study.

7.3 Testing Applicability of Zipf's Law

Zipf gave the idea of Rank-Frequency table and suggested that in a long textual content, when the words ranked by the frequency of occurrence, an inverse relationship are found where the rank is inversely proportional to its "frequency of occurrence". He propounded an equation r=1/f (c) or rf=c where r is the rank, f is frequency and C is a constant.

In the present study the words are ranked in order of frequency of occurrence and data for the top 50 ranking words are considered for study and found that the word "infection" is the most frequent word with 3393 occurrences with a constant value of 3.53 followed by Gastritis (2217 occurrences) with a constant value of 3.65, Cancer (1630 Occurrences) with a constant value of 3.69 and Helicobacter-Pylori (1538 Occurrences) with a constant value of 3.79.

The data is tabulated in Appendix-5 and from this calculation it can be seen that the C value ranges from 3.53 to 4.17 and hence it can be concluded that the Zipf's Law is valid for the current study.

8. DISCUSSION

The research has been done with the sole objective to understand the applicability of the fundamental bibliometric laws to gastritis research and a number of testing tools and methods have been used. The current study is based on the Web of Science data in the field of Gastritis research. Lotka's law is tested using Kolmogorov-Smirnov Test Model and Chi-Square test Model. The results show that there is a major discrepancy between D-Max value and K-S Critical Value while testing the applicability of the Lotka's law by using Chi-Square test it's found that there is a major deviation between the calculated value and tabular value. Both the tests suggest that the results are not in conformity with the Lotka's law. While testing the applicability of Bradford's law, the journals have been arranged in descending order of productivity and then divided into three groups each containing equal no of articles. The analysis and comparison of the data in Appendix 4 & 5 with the Bradford ratio it's found that there is a major discrepancy suggesting that Gastritis research output doesn't conform the applicability of the Bradford distribution. Zipf's Law has been tested by applying the Zipf's equation to the data given in the Appendix-6. The analysis of the data in the rank-frequency table shows that C value ranges from 3.53 to 4.17 and hence the Zipf's law stays fit for the gastritis research. With the growing prominence of interdisciplinary and multidisciplinary research, the process of journal selection requires increased precision and discernment. In such diverse research environments, the concentration of Bradford's core journals may become more diffuse, challenging the traditional expectations of core journal dominance as observed in more narrowly defined disciplines. Given this shift, it is essential to recognise that the applicability of classical bibliometric laws, such as Lotka's Law and Bradford's Law, may vary across different fields of study. The foundational assumptions of these laws, developed primarily in the context of conventional, discipline-specific research, may not hold uniformly in interdisciplinary or multidisciplinary contexts. To better understand this variation, comparative studies should be conducted to assess the applicability of these laws in basic science fields versus interdisciplinary and multidisciplinary domains. Such investigations would provide valuable insights into how the nature of a research field influences author productivity patterns and the distribution of literature across journals. Ultimately, these findings could inform more effective bibliometric evaluations and journal selection strategies in increasingly complex and cross-cutting research landscapes.

9. CONCLUSION

In conclusion, the present study applied various statistical models to assess the scientific productivity and research patterns in the field of Gastritis. Lotka's Law, tested using the Kolmogorov-Smirnov (K-S) test, and did not fit the observed data as the calculated maximum deviation (D Max) exceeded the K-S critical value. This suggests that the distribution of authorship in Gastritis research deviates from Lotka's expected pattern.

Further analysis using the Chi-Square test confirmed significant discrepancies between the observed and expected distribution of publications, with the calculated Chi-Square value being much higher than the tabular value at a 5 % significance level. This result indicates that the distribution of research output in Gastritis does not fully align with the expected patterns proposed by Lotka's Law.

Bradford's Law, applied to the distribution of journals in Gastritis research, revealed a deviation from the expected $1:\alpha:\alpha^2$ ratio. The study found that the concentration of articles in the core zone (Zone-1) extended to more journals than predicted by Bradford, with fewer journals contributing to the second and third zones than expected. This deviation suggests that the scattering of articles across journals in Gastritis research does not strictly adhere to Bradford's Law.

Lastly, Zipf's Law was validated in this study, as the word frequency analysis revealed an inverse relationship between rank and frequency of occurrence, with a constant value range of 3.53 to 4.17. This supports the validity of Zipf's Law in the context of Gastritis research, demonstrating consistent rank-frequency behavior.

9.1 Future Direction and Recommendations

In future the applicability of the fundamental bibliometric laws can be tested using indexing databases other than Web of Science. Comparative studies using two or three databases can be done in order to evaluate the generalizsability of the Fundamental Bibliometric laws. Comparative studies of different subject fields like multidisciplinary versus basic domains can produce results which can tell us more about the applicability of the laws. A number of methods like Maximum Likelihood and Sen's Method can be used for understanding the applicability of Lotka's law while as Leimkuhler model and Graphical analysis via the Bradford-Zipf plot can be done to evaluate journal dispersion.

10. LIMITATIONS

For the Current study the data has been downloaded from only one of the indexing databases, Web of Science, and the scope is limited to the applicability of the fundamental bibliometric laws to the gastritis research. More indexing databases can be used to download data on different research topics, and a comparative study is possible to understand the applicability of the fundamental bibliometric laws.

REFERENCES

1. Lotka AJ. The frequency distribution of scientific productivity. Journal of the washington academy of sciences. 1926 Jun 19;16(12):317-23.
2. Zipf GK. Human behavior and the principle of least effort: An introduction to human ecology. Ravenio books; 2016 Jan 27.
3. Fairthorne RA. Empirical hyperbolic distributions (Bradford-Zipf-Mandelbrot) for bibliometric description and prediction. J of documentation. 1969 Apr 1;25(4):319-43.
4. Cole FJ, Eales NB. The history of comparative anatomy: Part I -A statistical analysis of the literature. Science Progress (1916-1919). 1917 Jan 1;11(44):578-96.
5. Hulme EW. Statistical bibliography in relation to the growth of modern civilisation. London: Grafton. 1923.
6. Sengupta I. Bibliometrics, informetrics, scientometrics and librametrics: An overview. Libri. 1992;42(2):75-98.
7. Borgohain DJ, Verma MK, Nazim M, Sarkar M. Application of Bradford's law of scattering and Leimkuhler model to information science literature. COLLNET J of Scientometrics and Inf Managem. 2021 Jan 2;15(1):197-212.
8. Ram S, Paliwal N. Assessment of Bradford Law's of Scattering to Psoriasis Literature through bibliometric snapshot. DESIDOC J of Lib & Inf Technol. 2014 Jan;34(1):46-56.
9. Savanur K, Devi SG, Konnur PV. Lotka's law and authorship distribution in the journal of 'Columbia Law Review'. COLLNET J of Scientometrics and Inf Managem. 2014 Jan 2;8(1):193-208. doi: 10.1080/09737766.2014.947840
10. Amanullah A. An Investigation in the interdisciplinary nature of digital humanities: A bibliometric analysis. DESIDOC J of Lib & Inf Technol. 2023 Jul 1;43(4):234-40.
11. Saxena A, Jauhari M, Gupta BM. Zipf's law in a random text from english with a new ranking method. DESIDOC Bulletin of Information Technology. 2007 Jul 1;27(4):51-8.
12. Amanullah A, Rajeswari S. Open access scholarly literature on gastritis (2002-2021) in web of science: A scientometric analysis. Int J of Economic Perspectives. 2022 Jun 30;16(6):76-89.
13. Amanullah A, Rajeswari S. A scientometric analysis of scientific communications on" Gastritis". Int J of Early Childhood Special Education. 2022 Jul 1;14(5):4887-96. doi: 10.9756/INTJECSE/V14I5.592
14. Amanullah A, Rajeswari S. Open access literature on gastritis from G20 countries: A scientometric study. Specialusis Ugdymas. 2022;1(43):8075-88.
15. Khiste G, Amanullah A. Analysis of knowledge management output in web of science during 2007 to 2016. Int Res J of Lib & Inf Sci. 2018;7(4):758-73.
16. Suk FM, Wang MH, Ho YS. A historical review of classic Helicobacter pylori publications in science citation index expanded. COLLNET J of Scientometrics and Inf Managem. 2019 Jul 3;13(2):313-30. doi: 10.1080/09737766.2020.1716645
17. de Souza CL, Kuniyoshi MS, de Gusmão Freitas AB. Mapping The growth of ESG research: A bibliometric analysis using Bradford's, Lotka's, and Zipf's Laws. Revista de Gestão Social e Ambiental. 2024;18(10):1-22. doi: 10.24857/rgsa.v18n10-309
18. Batcha MS, Jahina SR, Ahmad M. Publication Trend in DESIDOC Journal of Library and Information Technology during 2013-2017: A scientometric approach. Inte J of Res in Engineering, IT and Soc Sci. 2018;8(4):76-82.
19. Amanullah A, Khiste GP. Publication productivity of 'E-Learning' by scopus during 2003-2017. In developing smart libraries changes challenges issues & strategies 2018 August 6-8. Srinagar, India. p. 77-82.
20. Bradford SC. Sources of information on specific subjects 1934. J of Inf Sci. 1985;10(4):176–80. doi: 10.1177/016555158501000407

CONTRIBUTORS

Dr. Amir Amanullah is a working library professional working under Department of Higher Education, J&K and is currently posted at Government Degree College Kangan as Libraian. He did his Masters in Library and Information Science from University of Kashmir and has qualified U.G.C State Eligibility Test (SET) in 2008. He also holds a Doctorate in Library and Information Science and has published arround 18 research articles in Journals, Conference proceedings and also authored book Chapter.
He has contributed in data analysis, testing of different models and writing of analysis and results section, discussion section, conclusion section, future direction section, limitation section and abstract.

Dr. S Rajeswari is currently working Assistant Librarian, Department of Library, University College of Engineering BIT Campus, Anna University, Tiruchirappalli-24, Tamil Nadu, India. Formerly She was also Assistant Professor, Department of Library and Information Science, Annamalai University, Annamalainagar, Tamil Nadu, India from 2009 to 2017. Her Qualification is M.com., MLIS, MPhil, PhD.
She has contributed in downloading and tabulation of data. Use of different software for data analysis.

Prof. Sumeer Gul is currently Head, Department of Library and Information Science, University of Kashmir, India. He was also In Charge Librarian, Allama Iqbal Library, University of Kashmir from 2023 to 2024. He did his Bachelors and Masters in Library and Information Science from University of Kashmir and has qualified U.G.C National Eligibility Test (NET) in 2005. He also holds a Doctorate in Library and Information Science.
He has contributed in writing of introduction section, review of literature section, statement of problem, objectives and methodology and reference section.

Appendix 1. Lotka's law on gastritis research output

x	y	X=Log x	Y=Log y	XY	X2
1	39027	0.0000	4.5914	0.0000	0.0000
2	7734	0.3010	3.8884	1.1705	0.0906
3	3207	0.4771	3.5061	1.6728	0.2276
4	1640	0.6021	3.2148	1.9355	0.3625
5	998	0.6990	2.9991	2.0963	0.4886
6	693	0.7782	2.8407	2.2105	0.6055
7	458	0.8451	2.6609	2.2487	0.7142
8	355	0.9031	2.5502	2.3031	0.8156
9	280	0.9542	2.4472	2.3352	0.9106
10	199	1.0000	2.2989	2.2989	1.0000
11	177	1.0414	2.2480	2.3410	1.0845
12	136	1.0792	2.1335	2.3025	1.1646
13	128	1.1139	2.1072	2.3473	1.2409
14	104	1.1461	2.0170	2.3118	1.3136
15	94	1.1761	1.9731	2.3206	1.3832
16	59	1.2041	1.7709	2.1323	1.4499
17	63	1.2304	1.7993	2.2140	1.5140
18	50	1.2553	1.6990	2.1327	1.5757
19	38	1.2788	1.5798	2.0202	1.6352
20	49	1.3010	1.6902	2.1990	1.6927
21	36	1.3222	1.5563	2.0578	1.7483
22	32	1.3424	1.5051	2.0205	1.8021
23	30	1.3617	1.4771	2.0114	1.8543
24	19	1.3802	1.2788	1.7650	1.9050
25	23	1.3979	1.3617	1.9036	1.9542
26	26	1.4150	1.4150	2.0021	2.0021
27	23	1.4314	1.3617	1.9491	2.0488
28	18	1.4472	1.2553	1.8166	2.0943
29	14	1.4624	1.1461	1.6761	2.1386
30	20	1.4771	1.3010	1.9218	2.1819
31	18	1.4914	1.2553	1.8721	2.2242
32	8	1.5051	0.9031	1.3593	2.2655
33	17	1.5185	1.2304	1.8685	2.3059
34	9	1.5315	0.9542	1.4614	2.3454
35	7	1.5441	0.8451	1.3049	2.3841
36	7	1.5563	0.8451	1.3152	2.4221
37	8	1.5682	0.9031	1.4162	2.4593
38	5	1.5798	0.6990	1.1042	2.4957

39	7	1.5911	0.8451	1.3446	2.5315
40	7	1.6021	0.8451	1.3539	2.5666
41	11	1.6128	1.0414	1.6795	2.6011
42	8	1.6232	0.9031	1.4659	2.6349
43	5	1.6335	0.6990	1.1417	2.6682
44	1	1.6435	0.0000	0.0000	2.7009
45	6	1.6532	0.7782	1.2864	2.7331
46	5	1.6628	0.6990	1.1622	2.7648
47	6	1.6721	0.7782	1.3011	2.7959
49	8	1.6902	0.9031	1.5264	2.8568
50	2	1.6990	0.3010	0.5114	2.8865
51	5	1.7076	0.6990	1.1935	2.9158
52	2	1.7160	0.3010	0.5166	2.9447
53	2	1.7243	0.3010	0.5191	2.9731
54	2	1.7324	0.3010	0.5215	3.0012
55	1	1.7404	0.0000	0.0000	3.0289
56	4	1.7482	0.6021	1.0525	3.0562
57	5	1.7559	0.6990	1.2273	3.0831
58	2	1.7634	0.3010	0.5308	3.1097
59	4	1.7709	0.6021	1.0662	3.1359
60	2	1.7782	0.3010	0.5353	3.1618
62	1	1.7924	0.0000	0.0000	3.2127
63	1	1.7993	0.0000	0.0000	3.2376
64	1	1.8062	0.0000	0.0000	3.2623
65	2	1.8129	0.3010	0.5457	3.2867
67	1	1.8261	0.0000	0.0000	3.3345
68	2	1.8325	0.3010	0.5516	3.3581
69	2	1.8388	0.3010	0.5535	3.3814
71	1	1.8513	0.0000	0.0000	3.4272
73	1	1.8633	0.0000	0.0000	3.4720
74	2	1.8692	0.3010	0.5627	3.4940
78	1	1.8921	0.0000	0.0000	3.5800
81	1	1.9085	0.0000	0.0000	3.6423
83	1	1.9191	0.0000	0.0000	3.6829
84	1	1.9243	0.0000	0.0000	3.7029
85	1	1.9294	0.0000	0.0000	3.7227
88	1	1.9445	0.0000	0.0000	3.7810
93	2	1.9685	0.3010	0.5926	3.8749
94	1	1.9731	0.0000	0.0000	3.8932
97	1	1.9868	0.0000	0.0000	3.9473
98	1	1.9912	0.0000	0.0000	3.9650
99	2	1.9956	0.3010	0.6007	3.9826
102	1	2.0086	0.0000	0.0000	4.0345
114	1	2.0569	0.0000	0.0000	4.2309

126	1	2.1004	0.0000	0.0000	4.4116
132	1	2.1206	0.0000	0.0000	4.4968
145	1	2.1614	0.0000	0.0000	4.6715
152	2	2.1818	0.3010	0.6568	4.7604
154	1	2.1875	0.0000	0.0000	4.7852
195	1	2.2900	0.0000	0.0000	5.2443
209	1	2.3201	0.0000	0.0000	5.3831
212	2	2.3263	0.3010	0.7003	5.4118
Total	**55943**	**141.7178**	**85.6177**	**96.1188**	**240.7230**

Appendix 2. Applicability of lotka's law on gastritis using kolmogorov- smirnov test model

x	y	Observed frequency =yx/Σyx	Observed cum. value =Σ(yx/Σyx)	Expected (theoretical) frequency	Expected cum value	Diff.(D)
1	39027	0.69762	0.69762	0.67090	0.67090	**0.02672**
2	7734	0.13825	0.83587	0.14573	0.81663	0.01923
3	3207	0.05733	0.89319	0.05966	0.87629	0.01690
4	1640	0.02932	0.92251	0.03166	0.90795	0.01456
5	998	0.01784	0.94035	0.01936	0.92732	0.01303
6	693	0.01239	0.95274	0.01296	0.94027	0.01246
7	458	0.00819	0.96092	0.00923	0.94950	0.01142
8	355	0.00635	0.96727	0.00688	0.95638	0.01089
9	280	0.00501	0.97227	0.00531	0.96168	0.01059
10	199	0.00356	0.97583	0.00421	0.96589	0.00994
11	177	0.00316	0.97900	0.00341	0.96930	0.00969
12	136	0.00243	0.98143	0.00282	0.97212	0.00931
13	128	0.00229	0.98371	0.00236	0.97448	0.00924
14	104	0.00186	0.98557	0.00200	0.97648	0.00909
15	94	0.00168	0.98725	0.00172	0.97820	0.00905
16	59	0.00105	0.98831	0.00149	0.97970	0.00861
17	63	0.00113	0.98943	0.00131	0.98100	0.00843
18	50	0.00089	0.99033	0.00115	0.98216	0.00817
19	38	0.00068	0.99101	0.00102	0.98318	0.00783
20	49	0.00088	0.99188	0.00091	0.98409	0.00779
21	36	0.00064	0.99253	0.00082	0.98491	0.00761
22	32	0.00057	0.99310	0.00074	0.98565	0.00745
23	30	0.00054	0.99364	0.00067	0.98633	0.00731
24	19	0.00034	0.99398	0.00061	0.98694	0.00704
25	23	0.00041	0.99439	0.00056	0.98750	0.00689
26	26	0.00046	0.99485	0.00051	0.98801	0.00684
27	23	0.00041	0.99526	0.00047	0.98848	0.00678
28	18	0.00032	0.99558	0.00044	0.98892	0.00667
29	14	0.00025	0.99583	0.00040	0.98932	0.00652
30	20	0.00036	0.99619	0.00037	0.98969	0.00650
31	18	0.00032	0.99651	0.00035	0.99004	0.00647

32	8	0.00014	0.99666	0.00032	0.99036	0.00629
33	17	0.00030	0.99696	0.00030	0.99067	0.00629
34	9	0.00016	0.99712	0.00028	0.99095	0.00617
35	7	0.00013	0.99725	0.00027	0.99122	0.00603
36	7	0.00013	0.99737	0.00025	0.99147	0.00590
37	8	0.00014	0.99751	0.00024	0.99170	0.00581
38	5	0.00009	0.99760	0.00022	0.99193	0.00568
39	7	0.00013	0.99773	0.00021	0.99214	0.00559
40	7	0.00013	0.99785	0.00020	0.99234	0.00552
41	11	0.00020	0.99805	0.00019	0.99252	0.00553
42	8	0.00014	0.99819	0.00018	0.99270	0.00549
43	5	0.00009	0.99828	0.00017	0.99287	0.00541
44	1	0.00002	0.99830	0.00016	0.99303	0.00527
45	6	0.00011	0.99841	0.00015	0.99318	0.00522
46	5	0.00009	0.99850	0.00015	0.99333	0.00517
47	6	0.00011	0.99860	0.00014	0.99347	0.00514
49	8	0.00014	0.99875	0.00013	0.99360	0.00515
50	2	0.00004	0.99878	0.00012	0.99372	0.00507
51	5	0.00009	0.99887	0.00012	0.99383	0.00504
52	2	0.00004	0.99891	0.00011	0.99395	0.00496
53	2	0.00004	0.99894	0.00011	0.99405	0.00489
54	2	0.00004	0.99898	0.00010	0.99415	0.00483
55	1	0.00002	0.99900	0.00010	0.99425	0.00475
56	4	0.00007	0.99907	0.00009	0.99435	0.00472
57	5	0.00009	0.99916	0.00009	0.99444	0.00472
58	2	0.00004	0.99919	0.00009	0.99453	0.00467
59	4	0.00007	0.99927	0.00008	0.99461	0.00466
60	2	0.00004	0.99930	0.00008	0.99469	0.00461
62	1	0.00002	0.99932	0.00008	0.99477	0.00455
63	1	0.00002	0.99934	0.00007	0.99484	0.00450
64	1	0.00002	0.99936	0.00007	0.99491	0.00444
65	2	0.00004	0.99939	0.00007	0.99498	0.00441
67	1	0.00002	0.99941	0.00006	0.99504	0.00437
68	2	0.00004	0.99945	0.00006	0.99510	0.00434
69	2	0.00004	0.99948	0.00006	0.99516	0.00432
71	1	0.00002	0.99950	0.00006	0.99522	0.00428
73	1	0.00002	0.99952	0.00005	0.99527	0.00424
74	2	0.00004	0.99955	0.00005	0.99532	0.00423
78	1	0.00002	0.99957	0.00005	0.99537	0.00420
81	1	0.00002	0.99959	0.00004	0.99541	0.00418
83	1	0.00002	0.99961	0.00004	0.99545	0.00415
84	1	0.00002	0.99962	0.00004	0.99549	0.00413
85	1	0.00002	0.99964	0.00004	0.99553	0.00411
88	1	0.00002	0.99966	0.00003	0.99556	0.00410

93	2	0.00004	0.99970	0.00003	0.99559	0.00410
94	1	0.00002	0.99971	0.00003	0.99562	0.00409
97	1	0.00002	0.99973	0.00003	0.99565	0.00408
98	1	0.00002	0.99975	0.00003	0.99568	0.00407
99	2	0.00004	0.99978	0.00003	0.99571	0.00408
102	1	0.00002	0.99980	0.00003	0.99573	0.00407
114	1	0.00002	0.99982	0.00002	0.99575	0.00407
126	1	0.00002	0.99984	0.00002	0.99577	0.00407
132	1	0.00002	0.99986	0.00001	0.99578	0.00407
145	1	0.00002	0.99987	0.00001	0.99579	0.00408
152	2	0.00004	0.99991	0.00001	0.99580	0.00411
154	1	0.00002	0.99993	0.00001	0.99581	0.00411
195	1	0.00002	0.99995	0.00001	0.99582	0.00413
209	1	0.00002	0.99996	0.00001	0.99583	0.00414
212	2	0.00004	1.00000	0.00001	0.99583	0.00417
Total	**55943**	**Present Study D. Max = 0.02672**				

Appendix 3. Applicability of lotka's law on gastritis using chi-square modal

No. of contributions or n	Observed no of authors with n publications or F	Observed percentage of authors $\frac{F}{a1}*100$	Expected number of authors (lotka's value= a1/n²) or P	Expected percentage of authors (Lotka's prediction= 100/n²)	(F-P)² /P
1	39027	100.0000	39027.0000	100.0000	0.0000
2	7734	19.8170	9756.7500	25.0000	419.3525
3	3207	8.2174	4336.3333	11.1111	294.1180
4	1640	4.2022	2439.1875	6.2500	261.8498
5	998	2.5572	1561.0800	4.0000	203.1024
6	693	1.7757	1084.0833	2.7778	141.0834
7	458	1.1735	796.4694	2.0408	143.8367
8	355	0.9096	609.7969	1.5625	106.4641
9	280	0.7175	481.8148	1.2346	84.5329
10	199	0.5099	390.2700	1.0000	93.7408
11	177	0.4535	322.5372	0.8264	65.6702
12	136	0.3485	271.0208	0.6944	67.2665
13	128	0.3280	230.9290	0.5917	45.8772
14	104	0.2665	199.1173	0.5102	45.4371
15	94	0.2409	173.4533	0.4444	36.3950
16	59	0.1512	152.4492	0.3906	57.2831
17	63	0.1614	135.0415	0.3460	38.4325
18	50	0.1281	120.4537	0.3086	41.2086
19	38	0.0974	108.1080	0.2770	45.4650
20	49	0.1256	97.5675	0.2500	24.1761
21	36	0.0922	88.4966	0.2268	31.1412
22	32	0.0820	80.6343	0.2066	29.3336
23	30	0.0769	73.7750	0.1890	25.9743

24	19	0.0487	67.7552	0.1736	35.0832
25	23	0.0589	62.4432	0.1600	24.9149
26	26	0.0666	57.7322	0.1479	17.4415
27	23	0.0589	53.5350	0.1372	17.4164
28	18	0.0461	49.7793	0.1276	20.2881
29	14	0.0359	46.4055	0.1189	22.6291
30	20	0.0512	43.3633	0.1111	12.5877
31	18	0.0461	40.6108	0.1041	12.5890
32	8	0.0205	38.1123	0.0977	23.7916
33	17	0.0436	35.8375	0.0918	9.9017
34	9	0.0231	33.7604	0.0865	18.1596
35	7	0.0179	31.8588	0.0816	19.3968
36	7	0.0179	30.1134	0.0772	17.7406
37	8	0.0205	28.5077	0.0730	14.7527
38	5	0.0128	27.0270	0.0693	17.9520
39	7	0.0179	25.6588	0.0657	13.5685
40	7	0.0179	24.3919	0.0625	12.4007
41	11	0.0282	23.2165	0.0595	6.4283
42	8	0.0205	22.1241	0.0567	9.0169
43	5	0.0128	21.1071	0.0541	12.2915
44	1	0.0026	20.1586	0.0517	18.2082
45	6	0.0154	19.2726	0.0494	9.1405
46	5	0.0128	18.4438	0.0473	9.7992
47	6	0.0154	17.6673	0.0453	7.7049
49	8	0.0205	16.2545	0.0416	4.1919
50	2	0.0051	15.6108	0.0400	11.8670
51	5	0.0128	15.0046	0.0384	6.6708
52	2	0.0051	14.4331	0.0370	10.7102
53	2	0.0051	13.8936	0.0356	10.1815
54	2	0.0051	13.3837	0.0343	9.6826
55	1	0.0026	12.9015	0.0331	10.9790
56	4	0.0102	12.4448	0.0319	5.7305
57	5	0.0128	12.0120	0.0308	4.0933
58	2	0.0051	11.6014	0.0297	7.9462
59	4	0.0102	11.2114	0.0287	4.6385
60	2	0.0051	10.8408	0.0278	7.2098
62	1	0.0026	10.1527	0.0260	8.2512
63	1	0.0026	9.8330	0.0252	7.9347
64	1	0.0026	9.5281	0.0244	7.6330
65	2	0.0051	9.2372	0.0237	5.6702
67	1	0.0026	8.6939	0.0223	6.8089
68	2	0.0051	8.4401	0.0216	4.9140
69	2	0.0051	8.1972	0.0210	4.6852
71	1	0.0026	7.7419	0.0198	5.8711

73	1	0.0026	7.3235	0.0188	5.4601
74	2	0.0051	7.1269	0.0183	3.6882
78	1	0.0026	6.4147	0.0164	4.5706
81	1	0.0026	5.9483	0.0152	4.1164
83	1	0.0026	5.6651	0.0145	3.8416
84	1	0.0026	5.5310	0.0142	3.7118
85	1	0.0026	5.4017	0.0138	3.5868
88	1	0.0026	5.0396	0.0129	3.2381
93	2	0.0051	4.5123	0.0116	1.3988
94	1	0.0026	4.4168	0.0113	2.6432
97	1	0.0026	4.1478	0.0106	2.3889
98	1	0.0026	4.0636	0.0104	2.3097
99	2	0.0051	3.9819	0.0102	0.9865
102	1	0.0026	3.7512	0.0096	2.0177
114	1	0.0026	3.0030	0.0077	1.3360
126	1	0.0026	2.4582	0.0063	0.8650
132	1	0.0026	2.2398	0.0057	0.6863
145	1	0.0026	1.8562	0.0048	0.3949
152	2	0.0051	1.6892	0.0043	0.0572
154	1	0.0026	1.6456	0.0042	0.2533
195	1	0.0026	1.0264	0.0026	0.0007
209	1	0.0026	0.8935	0.0023	0.0127
212	2	0.0051	0.8683	0.0022	1.4748
Total	**55943**			**$\chi^2 = \sum(F-P)^2/P$**	**2879.6535**

Appendix 4. Bradford's law of scattering on gastritis

No. of journals	No. of articles	Total articles	Cum. articles
1	440	440	440
1	397	397	837
1	376	376	1213
1	343	343	1556
1	264	264	1820
1	236	236	2056
1	217	217	2273
1	215	215	2488
1	193	193	2681
1	191	191	2872
1	189	189	3061
1	173	173	3234
1	171	171	3405
1	136	136	3541
1	135	135	3676
1	127	127	3803
1	110	110	3913

1	109	109	4022
1	103	103	4125
1	96	96	4221
1	94	94	4315
1	92	92	4407
1	86	86	4493
1	75	75	4568
1	**73**	**73**	**4641**
1	68	68	4709
1	65	65	4774
2	61	122	4896
1	60	60	4956
1	58	58	5014
2	56	112	5126
1	54	54	5180
1	53	53	5233
1	50	50	5283
1	49	49	5332
2	48	96	5428
1	47	47	5475
1	44	44	5519
1	43	43	5562
3	42	126	5688
2	41	82	5770
2	39	78	5848
2	38	76	5924
6	37	222	6146
3	35	105	6251
3	34	102	6353
3	33	99	6452
3	32	96	6548
1	31	31	6579
5	30	150	6729
6	29	174	6903
2	28	56	6959
1	27	27	6986
3	26	78	7064
6	25	150	7214
2	24	48	7262
2	23	46	7308
8	22	176	7484
6	21	126	7610
5	20	100	7710
5	19	95	7805

6	18	108	7913
9	17	153	8066
7	16	112	8178
8	15	120	8298
7	14	98	8396
16	13	208	8604
30	12	360	8964
30	**11**	**330**	**9294**
21	10	210	9504
31	9	279	9783
43	8	344	10127
45	7	315	10442
59	6	354	10796
92	5	460	11256
144	4	576	11832
185	3	555	12387
342	2	684	13071
859	1	859	13930
2055		**13930**	

Appendix 5. Ranking of word occurrence in zipf's law

S. No	Words	Frequency	Rank	Log f	Log r	C
1	Infection	3393	1	3.53058	0.00000	3.53058
2	Gastritis	2217	2	3.34577	0.30103	3.64680
3	Cancer	1630	3	3.21219	0.47712	3.68931
4	Helicobacter-pylori infection	1538	4	3.18696	0.60206	3.78902
5	Eradication	1430	5	3.15534	0.69897	3.85431
6	Expression	1366	6	3.13545	0.77815	3.91360
7	Disease	1345	7	3.12872	0.84510	3.97382
8	Risk	1266	8	3.10243	0.90309	4.00552
9	Prevalence	1264	9	3.10175	0.95424	4.05599
10	Atrophic gastritis	1252	10	3.09760	1.00000	4.09760
11	Campylobacter-pylori	1120	11	3.04922	1.04139	4.09061
12	Helicobacter-pylori	1057	12	3.02407	1.07918	4.10326
13	Stomach	1027	13	3.01157	1.11394	4.12551
14	Duodenal-ulcer	999	14	2.99957	1.14613	4.14569
15	Carcinoma	929	15	2.96802	1.17609	4.14411
16	Diagnosis	911	16	2.95952	1.20412	4.16364
17	Intestinal metaplasia	892	17	2.95036	1.23045	4.18081
18	Mucosa	877	18	2.94300	1.25527	4.19827
19	Association	875	19	2.94201	1.27875	4.22076
20	Population	597	20	2.77597	1.30103	4.07700
21	Children	585	21	2.76716	1.32222	4.08938
22	Epidemiology	559	22	2.74741	1.34242	4.08983

23	Therapy	556	23	2.74507	1.36173	4.10680
24	Inflammation	550	24	2.74036	1.38021	4.12057
25	Peptic-ulcer	545	25	2.73640	1.39794	4.13434
26	Follow-up	529	26	2.72346	1.41497	4.13843
27	Cells	528	27	2.72263	1.43136	4.15400
28	Classification	526	28	2.72099	1.44716	4.16814
29	Management	524	29	2.71933	1.46240	4.18173
30	Gastric-cancer	508	30	2.70586	1.47712	4.18298
31	Strains	485	31	2.68574	1.49136	4.17710
32	Ulcer	462	32	2.66464	1.50515	4.16979
33	Omeprazole	444	33	2.64738	1.51851	4.16590
34	Peptic-ulcer disease	432	34	2.63548	1.53148	4.16696
35	Gene	427	35	2.63043	1.54407	4.17450
36	Pathogenesis	423	36	2.62634	1.55630	4.18264
37	Pernicious-anemia	421	37	2.62428	1.56820	4.19248
38	Activation	413	38	2.61595	1.57978	4.19573
39	Epithelial-cells	411	39	2.61384	1.59106	4.20491
40	Caga	410	40	2.61278	1.60206	4.21484
41	Adenocarcinoma	379	41	2.57864	1.61278	4.19142
42	Protein	371	42	2.56937	1.62325	4.19262
43	Identification	351	43	2.54531	1.63347	4.17878
44	Carcinogenesis	348	44	2.54158	1.64345	4.18503
45	Mice	345	45	2.53782	1.65321	4.19103
46	In-vitro	328	46	2.51587	1.66276	4.17863
47	Lesions	320	47	2.50515	1.67210	4.17725
48	Risk-factors	303	48	2.48144	1.68124	4.16268
49	Antibodies	303	49	2.48144	1.69020	4.17164
50	Colonization	302	50	2.48001	1.69897	4.17898