Asian Journal of Information Science and Technology
ISSN: 2231-6108 (P) Vol.8 No.1, 2018 pp.44-51

DOI: https://doi.org/10.51983/ajist-2018.8.1.166

Ambiguity in Rendering of Personal Names Lead to Vagueness in the Results of Scientometric Studies: A Study Using Scientometric Analysis of Arrhythmia Research

Prasanna Raj
Head Librarian, University of Fujairah, Fujairah, UAE
E-Mail: rpraj05@gmail.com
(Received 23 February 2018; Revised 14 March 2018; Accepted 4 April 2018; Available online 15 April 2018)

***Abstract* - Most of the bibliographic databases like Web of Science, Scopus etc. use author name and full author name as two separate fields in the citation database. The analysis part uses short form of author name to list the high prolific authors. There are discrepancies in this listing since it is possible to render two different authors like Zang, Jio and Zang, James under one caption Zang, J.While listing the high productive authors in the analysis part, Zang, J will have two publications though the two authors are different. This study illustrates this ambiguity in rendering of author names taking scientometric analysis of Arrhythmia research as case study.**
Keywords: **Scientometric Studies, Arrhythmia Research**

I. INTRODUCTION

Heart diseases have been one of the leading causes of death in many countries. In today's world death rate due to heart diseases is a major threat to human beings. Arrhythmia is such a type of heart disease which is increasing rapidly and major deaths in cardiac patients are due to arrhythmia. An arrhythmia is any disorder of heart rate. The time interval between successive heart beats is called as rhythm and when the rhythm is not regular it is called as arrhythmia (a+rhythm=arrhythmia).Symptoms of arrhythmia are fast or slow heartbeat, skipping beats, chest pain etc. that can be seen in the early stage. Due to the complex nature and various treatment methods, early diagnoses and treatment of such diseases have attracted great interest among researchers.

One of the most common heart diseases is Cardiac Arrhythmia which is caused by disorders in electrical conduction system of the heart. The arrhythmias are identified according to their occurrence area within heart (atrial or ventricle) and their effects on heartbeat[3]. Arrhythmias started at the atrial are termed as atrial arrhythmia and those originated from ventricle, ventricle arrhythmia.

Quantitative study of science, and particularly bibliometrics, is a well-developed field of research with its own international community, international journals, conferences, institutes and research groups within universities and national research organizations. Several of the institutes and research groups have long-standing, extensive experience in the application of bibliometric methods for evaluation purposes. According to Francis Narin[7], publication and citation count is used for assessing the scientific activity. This is being adopted for nearly a century. While publication count is used as a quantitative measure, citation count is used as a qualitative tool. . One of the quantitative measures in the field of bibliometrics is author productivity and identification of high productive authors in a field.

Narin says "that scientific talent is highly concentrated on limited number of individuals" Zukerman (1967)[11] found that the most prolific laureates publish 10 papers annually – one for every five weeks for more than 20 years. Lotka, investigated the literature output of a sample of chemists, and found that, "… the number (of authors) making n contributions is about 1/n2 of those making one; and the proportion of all contributors, that make a single contribution, is about 60 %." This can be mathematically expressed as $X*Y^2 = k$ (Constant characteristic of a particular subject area) where X is the number of authors making Y contributions.

Large scale bibliometric research was made possible by the creation and development of the Science Citation Index (SCI), which is now part of Web of Science (WoS) along with two other indexes: the Social Science Citation Index (SSCI) and the Arts and Humanities Citation Index (A&HCI), SCOPUS, PUBMED etc. All these databases have their own rendering of the bibliographic elements like author, title, publication type, abstract, author address, references etc. The present investigation is an attempt to reveal the anomalies or ambiguities in the rendering of author names in these databases by conducting a bibliometric analysis of arrhythmia disease.

II. PREVIOUS STUDIES

Farhat, *et al.*, (2013)[2] conducted a comparative study on Research in congenital heart disease (CHD) between developing and developed countries. This study was based on a systematic search on MEDLINE, PubMed, and Scopus. The final findings of the study is the Arab world research

output in the field of CHD per capita is substantially low, estimated to be 29 times less than in developed countries. Despite the minimal increase in published research articles in global periodicals, most of the research relating to CHD continues to be far from innovative. Regional collaborations with international linkage are starting to evolve. The research facilities in the Arab countries need to increase substantially in research and infrastructure funding to keep up with the pace of research in developing countries.

Ugolini, *et al.,* (2013)[10] made a study on Bibliometric analysis of literature in cerebrovascular and cardiovascular diseases (CCD) rehabilitation. Citations from 1967 to 2008 were downloaded from the PubMed database. Based on the data analysis they conclude as the overall scientific production in the field of CCD rehabilitation showed a steep growth in the last decade, especially because of cerebrovascular research. In the same period, a decrease in the overall IF was observed. The European Union and the United States contributed 3 of every 4 articles in the field, although some Asian countries showed promising performance.

Most of the bibliometric studies were made on general cardiology related diseases (Miguel-Dasit et al (2004)[6], Bordons and Zulueta (2002)[1], Pagel and Hudetz (2011)[8], Pennell, et al (2012)[9] etc.). The bibliometric studies based on the citation data available from the bibliographic databases like Pubmed, WoS, Scopus and the like are either qualitative or quantitative in nature. None of the studies interpret the validity and vulnerability of the data is available.

III. OBJECTIVES

Though the prime aim of this aim is to reveal the results of the ambiguity in rendering author names in the bibliographic databases, the other objectives include

1. Trend of research in Arrhythmia disease
2. Authorship pattern in Arrhythmia research
3. High productive authors in Arrhythmia research
4. Identification of variation in the rendering of author names and thereby change in the list of high productive authors
5. Variation in the collaboration index due to the rendering of author names

IV. METHODS

The data for analysis have been downloaded from Pubmed database using "Arrhythmia" as search term in the Mesh Headings. The available data covers a period from 1945 to 2016. The total records downloaded were 1, 83,033 and these records which are in text form with corresponding tags are converted into MS Access database. Using SQL Query language necessary information is extracted from the database for analysis.

V. DISCUSSION

TABLE I TREND OF RESEARCH IN ARRHYTHMIA (DECADE-WISE)

Period	Publications	Percent	Relative Growth Rate
1940's	1025	0.56	1.04
1950's	4203	2.30	0.01
1960's	14039	7.67	0.20
1970's	23119	12.63	-0.01
1980's	28073	15.34	0.03
1990's	30492	16.66	0.01
After Millennium	82082	44.85	0.02
	183033	100.00	0.19

Research publications in the field of Arrhythmia are available in Pubmed from 1945 onwards. A close look at the year wise publication from 1945 to 2016 shows a growing trend though there is decline in certain years (1949, 1952, 1959, 1960, 1965, 1970, 1974, 1076, 1978, 1979, 1980, 1986, 1987, 1989, 1993, 1994, 1995, 1997, 2014, and 1016). Table 1 shows the trend of research in Arrhythmia in the seven decades from 1945. Except 1970's, the other periods show growth and the maximum growth is in 1940's, the nascent period.

TABLE II RESEARCHES IN ARRHYTHMIA AFTER MILLENNIUM

Year	Publications	Percent	Growth rate
2001	3450	4.20	
2002	3607	4.39	0.05
2003	4073	4.96	0.13
2004	4311	5.25	0.06
2005	4500	5.48	0.04
2006	4734	5.77	0.05
2007	5048	6.15	0.07
2008	5201	6.34	0.03
2009	5285	6.44	0.02
2010	5421	6.60	0.03
2011	5620	6.85	0.04
2012	6079	7.41	0.08
2013	6808	8.29	0.12
2014	6578	8.01	-0.03
2015	6857	8.35	0.04
2016	4510	5.49	-0.34
	82082	100.00	0.02

The research in arrhythmia has gained momentum after the millennium. But, the growth is not uniform. The average growth rate is 0.02.

TABLE III AUTHORSHIP PATTERN IN RESEARCH

No of Authors	Publications	Percent
Solo research	30876	16.87
Joint Authorship	31162	17.03
Three Authors	28712	15.69
Four Authors	24396	13.33
Five Authors	18827	10.29
Six Authors	14615	7.98
Seven Authors	9998	5.46
Eight Authors	7324	4.00
Nine Authors	5105	2.79
Ten Authors	3771	2.06
More than 10 authors	8247	4.51
	183033	100.00

Table III shows that the authorship pattern ranges from solo research to as many as more than 10 authors. In this case, joint authored publications are the highest. The next highest numbers of publications are the result of solo research (16.87%). Here it is to be noted that as the number of authors increases from 2 to more than 10, the number of publications decreases. Hence it can be presumed that team research is more in Arrhythmia and the optimum number of members in team research is 2 to 3.

In 1980, Lawani5 introduced collaboration index (CI) as the average number of authors per article. From an analysis of the collaboration index during the decades from 1940's onwards it is found that collaboration index varies from 0.93 to 1.98 when the short form of the authors are considered. At the same time the collaboration index varies from 0.91 to 2.01 when the full names of the authors are considered.

TABLE IV COLLABORATION INDEX

Period	No of Publications	No of Authors (Short Form)	Collaboration Index	No of Authors (Full Name)	Collaboration Index
1940's	1025	957	0.93	936	0.91
1950's	4203	3939	0.94	3941	0.94
1960's	14039	14370	1.02	14371	1.02
1970's	23119	26269	1.14	26271	1.14
1980's	28073	38353	1.37	38286	1.36
1990's	30492	46776	1.53	46724	1.53
After Millennium	82082	129883	1.58	165142	2.01

TABLE V COLLABORATION INDEX AFTER MILLENNIUM

Year	No of Publications	No of Unique Authors (Short Form)	No of Publications	Collaboration Index	No of Unique Authors (Full Form)	Collaboration Index
2001	3450	9352	3450	2.71	9360	2.71
2002	3607	9528	3607	2.64	10258	2.84
2003	4073	11146	4073	2.74	11973	2.94
2004	4311	12231	4311	2.84	13056	3.03
2005	4500	12786	4500	2.84	13679	3.04
2006	4734	13635	4734	2.88	14583	3.08
2007	5048	14533	5048	2.88	15445	3.06
2008	5201	15323	5201	2.95	16390	3.15
2009	5285	15667	5285	2.96	16834	3.19
2010	5421	16807	5421	3.10	18070	3.33
2011	5620	17827	5620	3.17	19267	3.43
2012	6079	19058	6079	3.14	20521	3.38
2013	6808	21599	6808	3.17	23590	3.47
2014	6578	21573	6578	3.28	23685	3.60
2015	6857	22758	6857	3.32	25134	3.67
2016	4510	17471	4510	3.87	19052	4.22

The variation in the collaboration index calculated for short form of author names and full author names lead to suggest that standardization must be followed in rendering of the name of the authors. Table V shows the variation of collaboration index in various years after the millennium. Though the collaboration index is increasing from 2001 to 2016, here too it is observed that the collaboration index is 3.87 in 2016 for short form of the authors while it is 4.22 for full name of the authors.

TABLE VI RANKING OF AUTHORS (HIGH PRODUCTIVE AUTHORS)

	Authors with Short Name		Full Name of Authors	
Rank	**Author Name**	**Publications**	**Author Name**	**Publications**
1	Lip GY	606	Lip, Gregory Y H	533
2	Zhang Y	335	Jais, Pierre	293
3	Jais P	316	Hocini, Meleze	256
4	Wilde AA	300	Morady, Fred	251
5	Hindricks G	298	Haissaguerre, Michel	248
6	Hocini M	279	Hindricks, Gerhard	243
7	Breithardt G	278	Calkins, Hugh	238
8	Haissaguerre M	275	Zareba, Wojciech	238
9	Morady F	258	Wilde, Arthur A M	234
10	Zareba W	255	Di Biase, Luigi	218
11	Calkins H	247	Sacher, Frederic	218
12	Wang Y	237	Breithardt, Gunter	215
13	Brugada J	235	Lin, Yenn-Jiang	214
14	Sacher F	229	Brugada, Josep	210
15	Di Biase L	223	Moss, Arthur J	207
16	Li Y	222	Chang, Shih-Lin	196
17	Lin YJ	220	Nattel, Stanley	192
18	Kuck KH	217	Callans, David J	182
19	Moss AJ	211	Shimizu, Wataru	176
20	Yamada T	210	Ackerman, Michael J	175
21	Wang J	206	Kirchhof, Paulus	175
22	Camm AJ	204	Clementy, Jacques	173
23	Kirchhof P	199	Kuck, Karl-Heinz	173
24	Nattel S	199	Natale, Andrea	173
25	Chang SL	198	Stevenson, William G	172

The first ranked author according to the short form is Lip GY with 606 publications. At the same time, the first ranked author according to the full form is Lip, Gregory Y H with 533 publications. The second ranked author with short name Zhang Y has different variations in full form like Zhang, Youhua, Zhang, Yu, Zhang, Y, Zhang, Yan, Zhang, Ying, Zhang, Yong etc. with various number of publications. This pattern exists in most of the author names. Hence it can be deduced that ranking of authors according to short form of the name of the authors will not give accurate results. The same principle exists in highly cited authors also.

Author productivity is a measure for ranking the authors according to their publication output. The most common methods for ranking authors are 1. Publication count and 2. Equal share. In a collaborative publication it is not necessarily that all the co-authors contribute equal share in the research. In General it is presumed that the author named first might have put maximum effort. As the position of the author name moves from the first to the last position, the effort of the co-authors may decrease. There may be some authors whose name may be included just because of the mantra "Publish or Perish". Hence a new method for ranking of authors is found to be appropriate if each author may be given an ordinal value in the decreasing order according to their position. Hence the authors can be ranked on the basis of weighted share by their position in the author list.

TABLE VII AUTHOR SPECIALIZATION (PREPOTENCY INDEX)

Author	Count	Potency	Rank	Specialization(PI)
Lip, Gregory Y H	533	114.63	1	0.22
Madias, John E	117	110.48	79	0.94
Tanel, Ronn E	121	98.22	74	0.81
Nattel, Stanley	192	70.69	17	0.37
Barold, S Serge	120	55.52	78	0.46
Stollberger, Claudia	125	51.96	69	0.42
Yamada, Takumi	159	49.91	32	0.31
Moss, Arthur J	207	49.47	15	0.24
Asirvatham, Samuel J	141	45.04	45	0.32
Antzelevitch, Charles	136	43.61	49	0.32
Calkins, Hugh	238	42.79	7	0.18
Jastrzebski, Marek	93	42.24	125	0.45
Wilde, Arthur A M	234	39.75	9	0.17
Callans, David J	182	38.11	18	0.21
Israel, Carsten W	80	37.39	187	0.47
Camm, A John	159	37.36	31	0.23
Roden, Dan M	164	36.09	27	0.22
Liu, Tong	117	35.47	80	0.30
Miyazaki, Shinsuke	139	34.23	48	0.25
Shimizu, Wataru	176	33.96	19	0.19
Schwartz, Peter J	134	32.60	55	0.24
Zareba, Wojciech	238	32.51	8	0.14
Di Biase, Luigi	218	31.69	10	0.15
Arias, Miguel A	86	31.54	156	0.37
Boriani, Giuseppe	148	31.19	38	0.21

Dr.S.R.Ranganathan's canon of Prepotence supports this method. The canon says that "The potency of an author is concentrated more on the first author who is also called prime author". According to Kumaravel[4], each author named in a publication is given a value according to his/her position in the authors place and this value is termed as potence value (PV). Therefore, prime PV is accorded to the first author and then PV goes decreasing to the second, third and so on.

For example, if there are n authors for a publication, the potency value (PV) of an author in p^{th} position ($p \leq n$) for that publication can be calculated as

$PV = (n - p + 1) / n\sum$ where $n\sum = 1+2+3+ \ldots n$ and $PV \leq 1$

For example, the potency of each author in a work by 4 authors, can be calculated as

1st Position = $(4 -1 +1) / 4\sum = 4 / (1+2+3+4) = 4 / 10$ i.e. 0,4

2nd position = $(4 -2 +1)/4\sum = 3/10$

i.e. 0.3

3rd position = $(4 -3 +1)/4\sum = 2/10$

i.e. 0.2

4th position = $(4 -4 +1)/4\sum = 1/10$

i.e. 0.1

A. Prepotence Index (PI) – a measure to evaluate Authors Specialization

Kumaravel proposed that the prepotence index of an author can be measured by arriving the potence value of the author. The formula for PI is PV/N where N is the total number of publications by the author.

The value of PI ranges from 0 to 1. The PI value nearer to 1 indicates the higher involvement of the author in most of his collaborative publications. The PI value nearer to zero indicates that the author has been involved in majority of his collaborative publications for name sake. From this index, the potential or specialization of an author in a subject can be measured.

A close look at the table 7 shows the number of publications by an author cannot be a measure to designate an author to be a specialist in the field. The specialization of an author in a field can be measured by PI.

TABLE VIII VERIFICATION OF LOTKA'S LAW (SHORT NAME)

No of Papers (x) (1)	No of Authors (y) (2)	$X^n*y = k$ (3)	% of authors making one contribution (4)	$1/n^2$ of number of authors making one publication (5)
1 Paper	91267	91267		
2 Papers	40241	160964	44.09	22816.75
3 Papers	10720	96480	11.75	10140.78
4 Papers	6566	105056	7.19	5704.188
5 Papers	3549	88725	3.89	3650.68
6 Papers	2528	91008	2.77	2535.194
7 Papers	1705	83545	1.87	1862.592
8 Papers	1318	84352	1.44	1426.047
9 Papers	950	76950	1.04	1126.753
10 Papers	805	80500	0.88	912.67
11 Papers	675	81675	0.74	754.2727
12 Papers	531	76464	0.58	633.7986
13 Papers	442	74698	0.48	540.0414
14 Papers	360	70560	0.39	465.648
15 Papers	323	72675	0.35	405.6311
16 Papers	270	69120	0.30	356.5117
17 Papers	209	60401	0.23	315.8028
18 Papers	206	66744	0.23	281.6883
19 Papers	194	70034	0.21	252.8172
20 Papers	158	63200	0.17	228.1675

Lotka's Law states that "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 (Lotka 1926,cited in Potter 1988). This means that out of all the authors given in a field, 60 percent will have just one publication, and 15 percent will have two publications ($1/2^2$ times of 60), 7 percent of authors will have three publications ($1/3^2$ times of 60), and so on. According to Lotka's Law of scientific productivity, only 6% 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. It is often used to estimate the frequency with which authors will appear in an online catalog (Potter 1988).

The total number of unique authors (identified by short form) who have contributed to arrhythmia research is 165142 of which 55.27% have contributed only one paper and 3.33% of the total authors have contributed more than 10 papers. These two figures do not coincide with lotka's findings. Also the mathematical calculations for X^n*y do not result in a constant value k (Column 3). Hence the present study deviates Lotka's law.

The total number of unique authors (identified by short form) who have contributed to arrhythmia research is 129885 of which 48.85% have contributed only one paper and 6.03% of the total authors have contributed more than 10 papers. Among these two figures, the percentage of authors making one publication does not match with Lotka's findings. But the percentage of the number of authors making more than 10 publications coincide with Lotka's findings. The mathematical calculations for X^n*y do not result in a constant value k (Column 3) for the number of authors making one two and three publications. Hence the present study coincides with Lotka's law with limitations.

The results of arrhythmia research are published in 5288 journals. These journals are categorized into three zones each containing equal number of publications. The three zones are in the ratio 25:201:5089 == 1:8:203.6 which is not in the form $1:n:n^2$ hence violating Bradford's law. Here it is to be noted that most of the present studies do not

validate Bradford's law. The reason is that when Bradford proposed the law in 1926, the number of journals in a subject is less in number. But, now there is a mushrooming growth of the number of journals of which most of them are predatory journals.

TABLE IX VERIFICATION OF LOTKA'S LAW (FULL NAME)

No of Papers (x)	No of Authors (y)	$X^n*y = k$	% of authors making one contribution	$1/n^2$ of number of authors making one publication
1 Paper	63444	63444		
2 Papers	31158	108498.5	49.11	15861.00
3 Papers	9837	88533	15.51	7049.33
4 Papers	6233	99728	9.82	3965.25
5 Papers	3646	91150	5.75	2537.76
6 Papers	2567	92412	4.05	1762.33
7 Papers	1771	86779	2.79	1294.78
8 Papers	1394	89216	2.20	991.31
9 Papers	1075	87075	1.69	783.26
10 Papers	926	92600	1.46	634.44
11 Papers	707	85547	1.11	524.33
12 Papers	610	87840	0.96	440.58
13 Papers	544	91936	0.86	375.41
14 Papers	459	89964	0.72	323.69
15 Papers	386	86850	0.61	281.97
16 Papers	310	79360	0.49	247.83
17 Papers	320	92480	0.50	219.53
18 Papers	228	73872	0.36	195.81
19 Papers	222	80142	0.35	175.75
20 Papers	196	78400	0.31	158.61

TABLE X VALIDATION OF BRADFORD'S LAW OF SCATTERING

Zone	No. of Journals	No. of Publications
Zone 1	25	60803
Zone 2	201	61125
Zone 3	5089	61105

VI. CONCLUSION

Changes in the normal rhythm of a human heart may result in different cardiac arrhythmias, which may be immediately fatal or cause irreparable damage to the heart sustained over long periods of time. Though many automated techniques have been developed to identify and cure arrhythmia, still research is going on to cure the disease. This bibliometric study will enable the researchers in this field to identify the trend of research in this field and move in the right direction. This study has shown that the results of author productivity studies and the validation of Lotka's law vary due to the ambiguity of rendering of personal name of the authors. Also this can be proved for the other qualitative indicators like h-index since Thomsons Reuters calculates the H-index by selecting the name of the author in the dropdown menu of the search box. Though there are solutions like Orcid, Researcher id etc. for this, research can be carried out to direct the authors to convert the names into unique identifies.

REFERENCES

[1] H.A. Zuckerman, "Nobel Laureates in Science: Patterns of Productivity, Collaboration, and Authorship", *American Sociological Review*, Vol. 37, pp. 391-403, 1967.

[2] S.M. Lawani, "Quality, collaboration, and citations in cancer research: a bliometric study", PhD thesis, Florida: Florida State University, 1980.

[3] L Goldberger, E. Goldberger, "*Clinical electrocardiography: A simplified approach*", 5th edition, Saint Louis, The Mosby Company 1994.

[4] Narin,Francis, "Evaluative bibliometrics: The use of publication and citation analysis In the evaluation of scientific activity". Contract report to the National Science Foundation. USA, 1976.

[5] Bordons María and Zulueta M. Ángeles, "Inter disciplinarity of Spanish Cardiovascular Research Teams", *Rev Esp Cardiol*, Vol. 55, No. 9, pp. 900-12, 2002.

[6] Miguel-Dasit, Alberto *et al,*. "Spanish Production of Research Articles on Diagnostic Imaging in Cardiology and Radiology," *Rev Esp Cardiol*. Vol. 57, No. 9. pp 806-14, 2004.

[7] Paul S Pagel and Judith A Hudetz, "Scholarly productivity of United States academic cardiothoracic anesthesiologists: influence of fellowship accreditation and transesophageal echocardiographic credentials on h-index and other citation bibliometrics", *J Cardiothorac Vasc Anesth.* Vol. 25, No. 5, pp. 761-5, 2011

[8] J.P.S. Kumaravel, A. Manoharan, B. Kanagavel, M. Karthi Rajan and L. Mohamed Idhris, "Dr. S. R. Ranganathan's Canon of Prepotence applied to Bibliometrics leading to a new indicator - Prepotency index (PI)," Paper presented at *8th International Conference on Webometrics, Informetrics and Scientometrics (WIS) and 13th COLLNET Meeting in Seoul, Korea*, October 23-26, 2012

[9] Dudley J Pennell, *et al.*, "Review of Journal of Cardiovascular Magnetic Resonance 2011", *Cardiovasc Magn Reson*, Vol. 14, pp.78, 2012.

[10] Theresa Farhat, *et al.*, "Research in congenital heart disease: a comparative bibliometric analysis between developing and developed countries", *Pediatr Cardiol*. Vol. 34. No. 2, pp.375-82, 2013

[11] Donatella Ugolini, *et al.*, "Bibliometric analysis of literature in cerebrovascular and cardiovascular diseases rehabilitation: growing numbers, reducing impact factor". *Arch Phys Med Rehabil*. Vol. 94, No. 2, pp. 324-31, 2013.