Bibliometric Analysis of References Selection that Influence Citations among Articles of Thai Multidisciplinary Journals

Sutthisak Srisawad[1],*, Kullacha Lertsittiphan[2]

[1]Faculty of Nursing, Mahidol University, Nakhon Pathom, THAILAND.
[2]Mahidol University International College, Mahidol University, Nakhon Pathom, THAILAND.

ABSTRACT

The purpose of this study is to examine association between reference list factors and citation count among articles published in Thai multidisciplinary journals based on the Scopus database. Spearman's rank correlation and Negative binomial regression models were used for univariate and multivariate analysis, respectively. The results from 900 articles revealed that many reference list factors such as number of references, proportion of past 10 years references, source items references and impact of references had significantly positive correlation with citation counts both in univariate and multivariate analysis. The number of references is still a good quantitative basic indicator, which is easy to access information. For impact of references list, this study measured by average citation and *h*-Index of references list. Despite being a factor that reflects the quality of the research as well but for the evaluation of the unpublished research articles in the submission process, the indicators still have a long memory retention period. Such indicators require a long time to collect data and are not flexible.

Keywords: Bibliometric Analysis, Citation, List of References, Multidisciplinary Journal, Correlation, Negative Binomial Regression.

INTRODUCTION

Currently, researchers and academic staff emphasize more the importance of the citation count in their research publications because citations are an accepted indicator of research impact. Articles that use numerous citations immediately reflect the desirable quality of the research. Therefore, citation counts are used as an indicator in evaluating research quality at all levels, including individual level, faculty or academic institution level and nationality, and are also used to measure the quality of published journals, called journal quality indicators or journal metrics. The previous three-year citation counts, excluding self-citations, were used as part of estimating the journal quality indicators on the Scopus database such as Source Normalized Impact per Paper (SNIP) and the SCImago Journal Rank (SJR);[1] another reputable indicator is Journal Impact Factors, indicators on the ISI Web of Knowledge database, which were calculated from all previous two-year citation counts.[1,2]

Thai educational institutions paid more attention to international publications and citations both on the Scopus and ISI Web of Knowledge databases to push the concept of becoming world class universities and international research universities in accordance with Thailand 4.0 policy. This policy engenders potential development of researchers and academic staff, quality assessment of researchers and articles, including improving the quality of Thai journals from the Thai national database to become international journals based on the Scopus database as part of the push for research articles to be published in a national database to a more international level. At present, about 40 Thai journals have reached international level on the Scopus database, and there were at least 20 journals from the project to improve the quality of Thai journals on the Thai-Journal Citation Index Centre (TCI). One of the guidelines for the development of national journals and the TCI database is the study of bibliometric analysis.[3]

Bibliometric analysis is quantitative research by applying statistical methods to analyze bibliographic data or information of publication contained in the research online database. It is useful for evaluating and comparing research performance, for study of research directions or trends in each subject and is also an important tool in indicating the characteristics and values of research published in journals.[4] Citations analysis and their predictive variables studies are part of bibliometric analysis research. It is accepted that the citation of the research

article does not solely depend on the research quality, research methodology or scientific merit, but also on various factors that can be found from the analyses of bibliometric data.[5] The number of citations is mainly influenced by the quality of the published journals. In addition, there are other variables that need to be considered such as fields, institutions, authors, author's *h*-index and references list. Liu, Yang and Chen (2021)[6] have classified citation into three forms: direct citation, co-citation and bibliographic coupling, which is a study of citation in the literature. This study reflects the reference literature influencing indirect citation. It also found that the use of references in research articles plays a different role in each part of article writing (Zhang, Liu and Wang, 2021).[7] For editors of journals, according to the citation analysis, there are many aspects to assess the quality of the research that has been published in journals, which can compare the characteristics or values of the research articles with other journals in the same field, and it can also be used to define the indicators or criteria to consider research published in the journals. These assessment indicators are quantifiable and tangible.

The purposes of this study emphasize the citation analysis and the association with reference list factors in Thai multidisciplinary journals based on the Scopus database. We used articles from same subject field of journals for controlling some confounding factors. For the association factors, this study focused on quantitative and impact of references list of the articles. This is one of the issues to be considered for publication in research journals.[8] In addition, the initial concept of the study needs to incorporate guidelines for improving the criteria for considering and evaluating the quality of research articles published in Thai journals on the Scopus database. Analysis of characteristics of reference list of research articles will be detailed in more depth. For data collection plan, we collected data about the number of references and other relevant details such as number of references from open access journals, number of references from source items (Scopus), number of references published not more than 10 years, and the impact of the references list in term of average citation and *h*-Index. The main objective is to indicate the association of these two factors, the citation counts and references lists by using bibliometric data from the Scopus database. Spearman's rho correlation was used for univariate analysis between all independent variables and citation counts. For multivariate analysis, regression models for count data were used for analyzing reference list' influence on citation counts when controlling for other significant factors. Further detail about methodology will be presented in research methodology.

LITERATURE REVIEW

Bibliometric Data and Citation Analysis

Bibliometric data are characteristic data of academic articles, books and research articles publication and appear on online databases. These online databases are considered a source of big data for assessing the quality, value or performance about research publication at the international level. Initial bibliometric data analysis is statistical analysis to study the distribution of research publication in each year, as well as to study the growth and trends of research articles published in various fields. Data analysis will involve collecting data in times series. Later, more studies have been conducted on the relationship between various bibliometric factors by using cross-sectional data or individual data. This has become an important research topic in the field of library and information science.[9] Agarwel *et al*. (2016)[10] explained about bibliometrics as statistical analysis for extracting measurable data of research publication and knowledge or scientific merit as well as quantity (productivity) and quality (performance) within a publication. Bibliometrics was also a key method for measurement of scholarly publication impact. Brown and Gutman (2019)[1] claimed that bibliometrics could also provide helpful information about publication profile, identifying weak points, and providing research network and research collaboration, which were related to research funding.

Citation count of scholarly publications is the part of bibliometric data, which is an acceptable indicator for evaluating quality of articles. The quality assessment of articles is difficult to determine with clearly concrete measurements. However, citation count is one of the most widely used indicators for evaluating the quality of research publication both of academic institutions and journals.[5,11,12] However, the citation analysis still needs to be taken into consideration about self-citation that has contributed to the quality assessment being inaccurate.[13,14] Another factor to consider when using the citation count as a measure of the research quality is citation speed or first time citation.[15] The study is usually done in the form of period time limit for collecting cited data since publication.[16] This is consistent with the estimating of journal quality indicator, namely SCImago Journal Rank (SJR) and Source Normalized Impact per Paper (SNIP), which used the citation counts of the articles within three years published for calculation and excluding Self-Citation.[17] Another example of a widely used indicator is journal impact factor, which uses all citation counts within two years of publication and was introduced by Garfield (1972).[5]

References list Factors

Garfield (1972)[5] investigated citation analysis and argued that citation count was a function with many variables besides scientific merit, for example, author's reputation, controversial nature of subject matter, circulation and others. Consistent

with the conclusions in the study of Bornmann *et al.* (2008)[18] that bibliometric studies have revealed factors influenced by citation counts, factors should be considered in statistical analysis of bibliometric data including: 1) Field-dependent factors; 2) Journal-dependent factors (journal accessibility, visibility, internationality, journal quality indicators, etc.); 3) Article-dependent factors (articles type, number of authors, number of references, article length, etc.); and 4) Author/reader-dependent factors (article language, social network effect, etc.). In addition, the analysis of the factors that influence the citation counts must also be careful about the distribution of citation counts that are countable data with non-normality distribution. The appropriate statistics for regression model were Poisson regression, negative binomial regression, and zero-inflate Poisson or negative regression model. Depending on the model suitability test and the over-dispersed data problem, the Poisson Model cannot be used (Zong *et al.*, 2020).[19]

The references list within articles is the number of scholarly publications cited by new articles, which often cited basic knowledge, theory, concepts and results of these publications. Ucar *et al.* (2014)[20] showed the growth in the number of references used per research article in engineering journals. In addition, the editorial perspective showed the direction of the increase in the number of references, the association with journal impact factors and emphasized the importance of using references for publication community.[21] When the number of references affected the evaluation of scholarly performance, the increase in references per publication was statistically significant. However, there are still researchers who are interested in the quality of reference documents, especially in the accuracy of references. Spivey and Wilks (2004)[8] revealed about percentage of references accuracy in social network journals. They had suggested that the solution to the problem of inaccuracies in the reference list, which is introduced by Fenton *et al.* (2000)[22] such as the sharing of responsibility between the authors, editors and reviewers, was that the references list should be part of the acceptance process, limiting for number of references, presenting first page of references list when submitting articles and using references management software such as Endnote. Mitchell-Williams *et al.* (2017)[23] studied the same concept to give importance to the accuracy and completeness of the references list, which affected the quality assessment of articles.

For association analysis with citation counts, many articles found significant positive correlation between the number of references and citation counts. Moreover, the number of references could still predict citation counts when controlling other bibliometric factors.[24-26] There were also other research articles that investigated references list factors on different issues. Didegah and Thelwall[27] found factors that predicted citation counts by using zero-inflate negative binomial regression. The results showed that number of references and impact of references positively correlated with citation counts in nanoscience and nanotechnology. The impact of references was measured by an average of number of citations to the cited references. Bornmann *et al.*[28] revealed the association between the references list factor and the citation as well when analyzing using the negative regression model. Single publication *h* index for the cited references was used as a variable in references list factor. Jiang *et al.*[29] investigated the correlation between article citations and references factors. This study focuses only on variables in the references and their impact measures in order to answer two research questions including the effect of conducting the research process to the quality of research output and selection for the most useful reference to scholars and contribution to their research. The results indicated that variables had a positive relationship with citation counts such as citation count of the references, impact factor of references' journal, and *h*-index of authors in the references list.

DATA AND METHODOLOGY

Population and Sample Size

The population in this study is research articles published in Thai international multidisciplinary journals based on the Scopus database found by searching the list of journals from Scimago Journal and Country Rank (SCImago, 2020).[30] We found five multidisciplinary journals including Chiang Mai University Journal of Natural Sciences, Maejo International Journal of Science and Technology, ScienceAsia, Songklanakarin Journal of Science and Technology, and Walailak Journal of Science and Technology. All five journals have similar characteristics, and are mainly published in applied science and natural science, with four journals being published by Thai universities. Another journal is ScienceAsia, published by The Science Society of Thailand and The National Research Council of Thailand. The citation analysis study within the same field will help control the impact of the field on the citation impact. For sample size calculation, we calculated with sample size estimating for correlation analysis formula (Montgomery, 2012)[31] as follows:

$$N = \left[\frac{z_\alpha + z_\beta}{C} \right]^2 + 3, \text{, which } one-side \ \alpha = 0.05, \ \beta = 0.2 \text{ and}$$

$$C = 0.5 \ln \left[\frac{1+r}{r} \right].$$

The correlation coefficient was cited from study of So *et al.* (2015)[25] who studied factors affecting citation network in science and technology. The results revealed that the number of references had positive significant correlation with citation

counts (r = 0.085; $p < 0.001$). When replacing the correlation coefficient from this study, it was found that the minimum sample size that should be used in this study was 855 articles. For data sampling procedure, we used purposive sampling with inclusion criteria to be only research article and number of page of not fewer than three pages. Bibliometric data was gathered from all research articles of five journals and starting from 2016 going back until the sample size was similar to the calculation. Therefore, this study selected research articles published from the years 2014-2016 with a total of 902 articles, excluding two articles because they were found to be message articles from the editorial board.

Variables

This study collected citation counts, characteristics of references lists and other bibliometric data, including publication year, journal quartile ranking of publication year, number of pages, number of authors, number of institutions of authors, number of countries and the number of keywords, which were collated from the Scopus and Scimago Journal and Country Rank database. For citation counts data, we collated number of citations within three years after publication and not excluding self-citation. All citation count data was collected until 31 January 2020. For the part of references list factors, we emphasized number of references and their characteristics that may affect citation count such as number of references from open access journals, number of references from same database (Scopus), number of new references (not more than 10 publication years) and impact of references in two variables including *h*-Index value and average citations of references lists. For impact indicators, we collected data in two parts: 1. Data of references lists from the Scopus database; and 2. Data of references lists from the Scopus database not more than 10 years old. As shown in Table 1, all characteristics variables, when used to analyze relationships, were transformed in terms of proportions by the number of references used as the basis for calculations.

Statistical Analysis

The univariate association between the references list variables shown in Table 1 and three-year citation counts was analyzed by using Spearman's rho correlation statistics, which was the appropriate statistic for citation data that was countable data, non-normal distribution and right skewed. Spearman's rho correlation was also used to analyze the relationship between citation and other bibliometric or some social variables (Rahimi, Soheili and Amini Nia, 2020).[32] Those variables that were statistically significant at the 95% confidence level were used as control variables for the multivariate models.

The aim of multivariate analysis was to indicate the influencing of references list factors to citation counts when controlling for significant bilbliometric variables. We preferred negative

Table 1: List of Dependent Variable and Independent Variables in This Study.

	Factors	Acronym
	Dependent Variable	
1.	3-Years citation counts (exclude self-citations)	CITE
	Independent Variables	
1.	Number of references	NREF
2.	Proportion of open access references	OA_REF
3.	Proportion of source items references	SC_REF
4.	*h*-index for Scopus reference list / Number of Scopus references	HI_SC_REF
5.	Average citations for Scopus reference	AC_SC_REF
6.	Ratio of last 10 years Scopus references	10y_REF
7.	*h*-index for last 10 years Scopus references / Number of last 10 years Scopus references	HI_10y_REF
8.	Average citations for last 10 years Scopus reference	AC_10y_REF

binomial regression statistics to be consistent with the characteristics of dependent variables, which are count data with over dispersion problem. Regarding the impact of references factors such as "AC_SC_REF" and "AC_10y_REF" collected data about all number of citations of references lists per number of references they will have very wide range. When we analyzed citation predicted by regression model for these variables, the beta coefficient will be close to 0 although it is statistically significant and cannot be interpreted clearly. Categorical variable transformation was used for solving this problem by using 1st Quartile, 2nd Quartile and 3rd Quartile values that were cut-off points for categorizing into four groups.

RESULTS

Table 2 shows a total of 900 articles used in data analysis of this study, 263 from Songklanakarin Journal of Science and Technology, representing 29.22%. The research articles from Maejo International Journal of Science and Technology had a minimum number of 69 articles (7.67%) and each year, there were approximately 300 articles from five journals. The number of research papers each year do not have a clear direction to increase or decrease the number of exemptions, Chiang Mai University Journal of Natural Sciences and ScienceAsia promising research article published in 2015, a decrease compared to 2014, and Maejo International Journal of Science and Technology, which is likely to publish more research articles in 2015 compared to 2014.

The quartile ranking of each multidisciplinary journal, which is collected on the Scimago Journal and Country Rank website, will use the data differently in each publication year from the calculated SJR indicators that change every year. Table 2 shows

Table 2: Percentage of Articles Publication and Journal Quartile Ranking of each Thai International Multidisciplinary Journals (2014-2016).

	2014 (*n*=314)		2015 (*n*=295)		2016 (*n*=291)		Total (*n*=900)
	Quartile*	%	Quartile	%	Quartile	%	%
1. Chiang Mai University Journal of Natural Sciences	4	16.88	4	10.17	4	11.00	12.78
2. Maejo International Journal of Science and Technology	3	3.18	2	10.51	4	9.62	7.67
3. ScienceAsia	3	25.80	2	20.00	3	19.59	21.89
4. Songklanakarin Journal of Science and Technology	2	27.07	3	29.83	2	30.93	29.22
5. Walailak Journal of Science and Technology	3	27.07	3	29.49	3	28.87	28.44

* Publication Year Journal Quartile Ranking in Multidisciplinary Area

Table 3: Descriptive Statistics of Citation Counts.

Cite	n	%
0 time	487	54.11
1 time	207	23.00
2 time	94	10.44
3 time	46	5.11
4 time and above	66	7.33
Range, Mean (SD)	0 - 10	0.97 (1.52)

Table 4: Descriptive Statistics of Other Bibliometric Data and Spearman's Rho Correlation Analysis with Citation Counts.

Variables	Range	Mean (SD)	Median [IQR]	r_s with CITE
Publication Year (Ordinal) 1 (2014) > 3 (2016)	2014 - 2016	-	-	-0.040
Journal Quartile (Ordinal) 2 (Q2) > 4 (Q4)	2 – 4	-	-	-0.122 *
# Pages	3 - 26	9.30 (3.72)	8.00 [7-11]	-0.004
# Authors	1 - 14	3.20 (1.51)	3.00 [2-4]	0.066 *
# Institution of Authors	1 - 9	1.70 (0.91)	1.00 [1-2]	0.053
# Country of Authors	1 - 4	1.17 (0.47)	1.00 [1-1]	0.015
# Keywords	0 - 9	4.37 (1.14)	5.00 [5-5]	0.012

* Significant at 0.05 level

the quartile ranking of Thai multidisciplinary journals at the level of Q2-Q4 in year 2014-2016. Only two journals have a stable quartile ranking in this three-year period, including Walailak Journal of Science and Technology and Chiang Mai University Journal of Natural Sciences. Quartile ranking is a variable that reflects the quality of journals and affecting the citation, therefore data is collected for use as a control variable in multivariate analysis.

Table 3 shows descriptive analysis for citation counts within the first three publication years excluding self-citations. The results find that 487 articles, or 54.1%, were research papers that have not been cited yet. Followed by 207 articles were cited 1 time, or 23.00%. The maximum number of citations per article is 10, with only 1 article and the average citation counts with standard deviation being 0.97 ± 1.52 times. Overall, more than 90% of the articles published in these five journals have not more than three times of citation counts.

Table 4 shows the descriptive statistics such as maximum, minimum, mean, standard deviation, median and inter quartile range for other bibliometric data (number of pages, number of authors, number of institution of authors, number of country of authors, and number of keywords) and showing the Spearman's rho correlation analysis with citation counts. For the descriptive statistics, the mean and standard deviation of number of pages was 9.30 ± 3.72; that of number of authors, 3.20 ± 1.51; that of number of institution of authors, 1.70 ± 0.91; and that of number of country of authors, 1.17 ± 0.47. The results found that 51.4% of all articles were investigated from a single university, and 83.0% were investigated from a single

country, most of the articles being from Thailand. For the number of keywords, mean and standard deviation was 4.37 ± 1.14, which contained three articles without keywords.

For the Spearman's rho correlation analysis, Table 4 shows two factors that had significant association with citation counts at a significant level 95% including journal quality indicator using journal quartile ranking, which had negative correlation with citation. When the quartile ranking is higher (the quality index of the journal decreases), it will reduce the citation counts of articles (r_s = -0.122; $p < 0.001$). In addition, the number of authors had positive correlation with citation counts (r_s = 0.066; $p = 0.048$). The journal quartile ranking and the number of authors will be used as a control variable to analyze the relationship between reference list factors and citation counts by using negative binomial regression.

Table 5 shows the results for descriptive statistics (range, mean, standard deviation, median and inter-quartile range) of references list factor and their correlation with citation counts using Spearman's rho correlation. The descriptive analysis revealed that a number of references were cited in articles published in these five journals that had minimum and

Table 5: Descriptive Statistics of References list Factors and Spearman's Rho Correlation Analysis with Citation Counts.

References Factors	Range	Mean (SD)	Median [IQR]	r_s with CITE
NREF	3 - 83	25.89 (11.59)	25.00 [18-32]	0.167 [*]
OA_REF	0 - 0.75	0.11 (0.12)	0.07 [0-0.15]	-0.016
SC_REF	0 - 1	0.70 (0.21)	0.74 [0.58-0.86]	0.067 [*]
HI_SC_REF	0 - 1	0.83 (0.14)	0.84 [0.75-0.93]	-0.001
AC_SC_REF	0 – 5,042	212.17 (421.49)	60.46 [29-166]	0.118 [*]
10y_REF	0 - 1	0.40 (0.20)	0.41 [0.25-0.55]	0.066 [*]
HI_10y_REF	0 - 1	0.84 (0.21)	0.88 [0.78-1.00]	0.032
AC_10y_REF	0 – 2,777	80.89 (213.92)	31.42 [15-63]	0.130 [*]

[*] Significant at 0.05 level

maximum values of 3-83, mean and standard deviation was 25.89 ± 11.59 articles. The mean and standard deviation of proportion of references from open access journal was 0.11 ± 0.12 or 11% of number of references. In which the proportion of article references from open access journals was the only factor that did not find any relationship with citation counts in both univariate and multivariate analysis. The mean and standard deviation of proportion of references from journal based on Scopus database was 0.70 ± 0.21, and 0.40 ± 0.20 for proportion of references from journals based on Scopus database and not more than 10 publication years. For the impact of references list factors, we studied two indicators such as *h*-Index and average citations. Mean and standard deviation of *h*-Index factors was 0.83 ± 0.14 based on all Scopus references and 0.84 ± 0.21 based on Scopus references and not more than 10 publication years, which had similar values. It can be interpreted that the *h*-Index of all reference articles will be close to 80% of number of references. As for the average citation of references list which had been gathered for past five years, the results found that mean and standard deviation based on Scopus database journal was 212.17 ± 421.49, and 80.89 ± 213.92 for not more than 10 years references.

Results of correlation analysis using Spearman's rho correlation revealed that many references list factors had a significantly positive correlation with citation counts of articles at a 95% confidence level, including number of references (r_s = 0.167; p < 0.001), proportion of references from source items (r_s = 0.067; p = 0.046), proportion of references from Scopus within last 10 years (r_s = 0.066; p = 0.048) and average citation counts of references list both in same database list and new list (r_s = 0.118; p < 0.001 and r_s = 0.130; p < 0.001, respectively).

Negative binomial regression was used to find influencing level of references list factors to citation counts when controlling other bibliometric factors such as journal quartile ranking and number of authors; all results are shown in Table 6. Before analyzing, the Vuong Test was used to find

an appropriate model compared with NB Regression and Zero-Inflate NB Regression. The results show that every model had *p* value > 0.05; therefore, it can be concluded that NB Regression is more appropriate for data analysis than ZINB Regression. In addition, the LR Test was also used to analyze the appropriateness between the NB Regression and Poisson Regression, in which the NB Regression was more appropriate.

The results confirmed references list factors can positively influence citation counts of articles, even though journal quartile and number of authors have been controlled. Variables that were statistically significant influencing to citation were NREF (Coef. = 0.015; 95% CI = 0.006, 0.024), SC_REF (Coef. = 0.718; 95% CI = 0.197, 1.238), dummy variables of AC_SC_REF (Coef. = 0.460, 0.351 and 0.387), 10y_REF (Coef. = 0.637; 95% CI = 0.065, 1.209) and dummy variables of AC_10y_REF (Coef. = 0.310, 0.508 and 0.400), which were consistent with the study results of univariate analysis in Table 5. In addition, the results which showed differences from univariate analysis and are more interesting are proportion of *h*-Index of references from source items within a period of 10 years can affect the citation counts when controlling confounding factors (Coef. = 0.637; 95% CI = 0.065, 1.209).

DISCUSSION AND CONCLUSION

Currently, bibliometric data analysis is a field that has seen more article publications. This research is useful in big data analytic tools for online research databases as well as to propose policies for the development of research potential both at individual and institutional levels and for evaluating the quality of research published in various journals. Many studies show the relationship of citation counts, which is an indicator of research article quality and various bibliometrics data, not merely the quality of the research in terms of scientific merit. This study focuses mainly on references list factors of articles which are considered as one of the indicators in the publication submission process of various journals, each with different evaluation criteria both in terms of quantity and quality.

The quantitative consideration of a references list as the number of references used in a research article is still a factor that can affect the quality indicators of the research in terms of citation counts by having positive association. The results reflect two points: first, the number of references is one of the indicators for quality assessment of scholarly publications. Another issue that can be seen is that current researchers' References may not only refer to the content, knowledge, or the results of the cited article, but that researchers still use the references of a cited articles as a source of instruction, theory or knowledge. As a result, articles in the References list are often cited as well. However, when researchers are aware of such

Table 6: Analysis of Negative Binomial Regression Predicting Citation Counts by References list Factors.

Model	Adjusted NB Regression [a]	Coef.	Std. Err.	95% CI	Exp (Coef.)	P Value	LR Test [I]
1	NREF	0.015	0.004	0.006, 0.024	1.015	0.001[*]	<0.001[*]
2	OA_REF	-0.357	0.185	-1.228, 0.514	0.700	0.422	<0.001[*]
3	SC_REF	0.718	0.265	0.197, 1.238	2.049	0.007[*]	<0.001[*]
4	HI_SC_REF	0.593	0.406	-0.202, 1.389	1.910	0.144	<0.001[*]
5	AC_SC_REF						<0.001[*]
	- Less than 1st Quartile	Ref.	-	-	-	-	
	- 1st – 2nd Quartile	0.460	0.149	0.168, 0.752	1.584	0.002[*]	
	- 2nd – 3rd Quartile	0.351	0.150	0.056, 0.646	1.421	0.020[*]	
	- 3rd Quartile and above	0.387	0.150	0.093, 0.681	1.472	0.010[*]	
6	10y_REF	0.608	0.265	0.089, 1.128	1.837	0.022[*]	<0.001[*]
7	HI_10y_REF	0.637	0.291	0.065, 1.209	1.891	0.029[*]	<0.001[*]
8	AC_10y_REF						<0.001[*]
	- Less than 1st Quartile	Ref.	-	-	-	-	
	- 1st – 2nd Quartile	0.310	0.151	0.014, 0.608	1.364	0.040[*]	
	- 2nd – 3rd Quartile	0.508	0.149	0.216, 0.800	1.663	0.001[*]	
	- 3rd Quartile and above	0.400	0.150	0.105, 0.694	1.492	0.008[*]	

[a] Controlled by Journal Quartile and Number of Authors, [I] LR Test for NB Regression vs Poisson Regression, [*] Significant at 0.05 level

effects, there may be issues with adding more references than necessary. Therefore, there is a question to journal editorial boards whether the consideration of the number of reference documents should specify the number of references clearly or not according to the concept of Fenton *et al.* (2000)[22] as it appears in articles of Spivey and Wilks (2004),[8] limiting the number of references should be part of the submission process of published research articles for check the accuracy of the references list.

In terms of qualitative consideration of a references list, when we talk about defining metrics in terms of quality, there are often problems in creating concrete indicators. Because the quality is interpreted in various approaches, and also depends on the individual. This study has investigated collecting more detailed information of references lists in order to determine which elements are interesting and which can be used as a quality assessment indicator. We would like to emphasize two issues for consideration in the quality of the references lists found from this study, namely, the proportion of the references that are on the same database as the research article, and the proportion of references that are on the same database as the research article and are new references. Both of these factors, although having a low correlation with the citation counts, can still have significantly positive influencing both in univariate and multivariate analysis. During the collection of data, we found that some articles have a high proportion of inaccessible reference documents or from non–source items (Chi, 2014). [33] Such an issue may disrupt the editorial boards of journals in checking the authenticity of the document references. It also affects the citations because the reader will not be able to take advantage of all references. Another important issue that we are interested in is the use of local language in research articles, reference documents that are outside the Scopus but in a local database and usually research articles in the local language, which will prevent readers from making full use of the references list as well. Regarding the newness of the references list, both in consideration of research funding or publication submission process, researchers are often examined as to their references to previous studies or literature reviews. The reference documents should have a study result that matches the current global situation in that research field. The period of scholarly publication is therefore important in consideration, especially in references with the data analysis results and outcomes of that research. In addition, a list of reference documents that have a long time publication is usually a list of references cited for basic theory or background information. Hou *et al.* (2011)[34] proposed a new reference list counting method, with more emphasis on detailed content than counting from the references list in the last section of the articles. This is due to the problem that some references lists are just for background information or are incidentally mentioned. As for the proportion of OA references, this study did not find any correlation with the citation counts. Although in other studies has been found that the OA policy will benefit the article published in that journal and found to have a positive correlation on citations (Fraser, Momeni, Mayr and Peters, 2020;[35] Oh *et al.* 2017).[36] However, in this study, OA Policy had no or little effect on citaitons, possibly because this study examined the impact of OA policy as an indirect

effect through reference selection. The study also found that the proportion of OA references was relatively low, so the results may not be conclusively confirmed (Mean of P_{OA_Ref} = 0.11; 95%CI = 0.098, 0.114).

Finally, another indicator of the quality assessment concerns the impact of references lists whether they measure qualitatively better or not. This study collected two variables that are h-Index and average citation estimating based on number of references in the Scopus database comparing with all references and newest references (10 publication years). The study found that the impact of reference list can affect citation count and will yield better results if the focus is on the citing of articles that are high impact and published within the past 10 years, when measured using the h-Index of references list. The impact of references study is consistent with the results of many articles.[27-29] Although the citation and h-Index values are highly accepted indicators in terms of evaluating the quality of research and the use of research results, we see indications that assessment from citation or h-Index is a measure that is useful in terms of quality assessment of research after publication.[37] But in terms of consideration of the journal editorial boards in the submission process, they may not be appropriate indicators because there are significant disadvantages in the evaluation period and are not flexible. The papers in the submission process are still unable to take full advantage of bibliometric analysis for describing all the references lists on the online research article database.

This paper highlights the benefits of bibliometric data analysis, which is bringing big data to statistical analysis and can contribute to the development of researcher performance, journal quality and research of online databases. Bibliometric analysis is still necessary for research potential improvement at the individual, institutional and national level. The authors of this study have received the question, "How can we improve the research potential of researchers and our institution so that research articles will be published internationally and will be cited more?" This issue led to the research question and the beginning of the bibliometric analysis study. However, this study may not be consistent with the size of the Scopus database, which is considered as one of the largest research database. The database contains 23,452 active journals, 120,000 conferences and 206,000 books with approximately 3 million records added every year.[38] From this point on, the sample size needs to be enlarged so that it can be a good representation. In terms of macro-level analysis, we still need to focus on the analysis of other issues as well, not just in the references list and cover other research fields. Important issues that must be considered are, "Can we evaluate the quality of the research request from other indicators in concrete terms, in addition to citation of articles?"

LIMITATIONS

This study has limitation in terms of the size of study as it was a small and the research articles were selected for analysis within a single subject and collected at least sufficient data for minimum appropriate sample size for statistical analysis. That results in limitations on generalization. Therefore, it is necessary to expand the scale of research to be more macro-scale as mentioned above. For the next study, consideration should be given to the study of research articles in various fields, various factors, increasing the sample size and the study of articles published in journals from other countries in order to apply the results to other journals that are in the process of developing quality to internationality level.

CONFLICT OF INTEREST

The authors declare that there is no conflict of interest.

REFERENCES

1. Brown T, Gutman SA. Impact factor, eigenfactor, article influence, Scopus SNIP, and ScImage journal rank of occupational therapy journals. Scandinavian Journal of Occupational Therapy. 2019;26(7):475-83. doi: 10.1080/11038128.2018.1473489, PMID 29772931.

2. Haddawy P, Hassan SU, Asghar A, Amin S. A comprehensive examination of the relation of three citation-based journal metrics to expert judgment of journal quality Journal of Informetrics. 2016;10(1):162-73. doi: 10.1016/j.joi.2015.12.005.

3. Sombatsompop N, Chancheewa S, Markpin T, Premkamolnetr N, Ittiritmeechai S, Wongkaew C, et al. (TCI) centre: 10 years of experiences, lessons learned, and ongoing development. Malaysian Journal of Library and Information Science. 2012. Thai-Journal citation index;17(3):17-33.

4. Lovakov A, Agadullina E. Bibliometric analysis of publications from post-Soviet countries in psychological journals in 1992-2017. Scientometrics. 2019;119(2):1157-71. doi: 10.1007/s11192-019-03087-y.

5. Garfield E. Citation analysis as a tool in journal evaluation. Science. 1972;178(4060):471-9. doi: 10.1126/science.178.4060.471, PMID 5079701.

6. Liu Y, Yang L, Chen M. A new citation concept: Triangular citation in the literature. Journal of Informetrics. 2021;15(2). doi: 10.1016/j.joi.2021.101141.

7. Zhang C, Liu L, Wang Y. Characterizing references from different disciplines: A perspective of citation content analysis. Journal of Informetrics. 2021;15(2). doi: 10.1016/j.joi.2021.101134.

8. Spivey CA, Wilks SE. Reference list accuracy in social work journals. Research on Social Work Practic. 2004;14(4):281-6. doi: 10.1177/1049731503262131.

9. Peritz BC. On the careers of terminologies; The case of bibliometrics. Libri. 1984;34(1):233. doi: 10.1515/libr.1984.34.1.233.

10. Agarwal A, Durairajanayagam D, Tatagari S, Esteves SC, Harlev A, Henkel R, et al. Bibliometrics: Tracking research impact by selecting the appropriate metrics. Asian Journal of Andrology. 2016;18(2):296-309. doi: 10.4103/1008-682X.171582, PMID 26806079.

11. Bornmann L, Daniel HD. What do citation counts measure? A review of studies on citing behavior. Journal of Documentation. 2008;64(1):45-80. doi: 10.1108/00220410810844150.

12. Cole JR, Cole S. The ortega hypothesis: Citation analysis suggests that only a few scientists contribute to scientific progress. Science. 1972;178(4059):368-75. doi: 10.1126/science.178.4059.368, PMID 17815351.

13. Lawani SM. Some bibliometric correlates of quality in scientific research. Scientometrics. 1986;9(1-2):13-25. doi: 10.1007/BF02016604.

14. Kacem A, Flatt JW, Mayr P. Tracking self-citations in academic publishing. Scientometrics. 2020;123(2):1157-65. doi: 10.1007/s11192-020-03413-9.

15. Kumari R, Uddin A, Lee BH, Choi K. Analyzing the factors influencing the waiting time to first citation and long-term impact of publications. Journal of Scientometric Research. 2020;9(2):127-35. doi: 10.5530/JSCIRES.9.2.16.

16. Schubert A, Glänzel W. Mean response time - A new indicator of journal citation speed with application to physics journals. Czechoslovak Journal of Physics. 1986;36(1):121-5. doi: 10.1007/BF01599743.

17. González-Pereira B, Guerrero-Bote VP, Moya-Anegón F. A new approach to the metric of journals' scientific prestige: The SJR indicator. Journal of Informetrics. 2010;4(3):379-91. doi: 10.1016/j.joi.2010.03.002.

18. Bornmann L, Mutz R, Neuhaus C, Daniel HD. Citation counts for research evaluation: Standards of good practice for analyzing bibliometric data and presenting and interpreting results. Ethics in Science and Environmental Politics. 2008;8(1):93-102. doi: 10.3354/esep00084.

19. Zong Q, Fan L, Xie Y, Huang J. The relationship of polarity of post-publication peer review to citation count: Evidence from Publons. Online Information Review. 2020;44(3):583-602. doi: 10.1108/OIR-01-2019-0027.

20. Ucar I, López-Fernandino F, Rodriguez-Ulibarri P, Sesma-Sanchez L, Urrea-Micó V, Sevilla J. Growth in the number of references in engineering journal papers during the 1972-2013 period. Scientometrics. 2014;98(3):1855-64. doi: 10.1007/s11192-013-1113-6.

21. Jaunich M. Editorial: The average number of references per article published in Polymer Testing over the years. Polymer Testing. 2018 [editorial];68:A1-2. doi: 10.1016/j.polymertesting.2018.05.034.

22. Fenton JE, Brazier H, De Souza A, Hughes JP, McShane DP. The accuracy of citation and quotation in otolaryngology/head and neck surgery journals. Clinical Otolaryngology and Allied Sciences. 2000;25(1):40-4. doi: 10.1046/j.1365-2273.2000.00322.x, PMID 10764235.

23. Mitchell-Williams MT, Skipper AD, Alexander MC, Wilks SE. Reference list accuracy in social work journals: A follow-up analysis. Research on Social Work Practice. 2017;27(3):348-52. doi: 10.1177/1049731515578536.

24. Fox CW, Paine CET, Sauterey B. Citations increase with manuscript length, author number, and references cited in ecology journals. Ecology and Evolution. 2016;6(21):7717-26. doi: 10.1002/ece3.2505, PMID 30128123.

25. So M, Kim J, Choi S, Park HW. Factors affecting citation networks in science and technology: Focused on non-quality factors. Quality and Quantity. 2015;49(4):1513-30. doi: 10.1007/s11135-014-0110-z.

26. Yu T, Yu G, Li PY, Wang L. Citation impact prediction for scientific papers using stepwise regression analysis. Scientometrics. 2014;101(2):1233-52. doi: 10.1007/s11192-014-1279-6.

27. Didegah F, Thelwall M. Determinants of research citation impact in Nanoscience and nanotechnology. Journal of the American Society for Information Science and Technology. 2013;64(5):1055-64. doi: 10.1002/asi.22806.

28. Bornmann L, Schier H, Marx W, Daniel HD. What factors determine citation counts of publications in chemistry besides their quality? Journal of Informetrics. 2012;6(1):11-8. doi: 10.1016/j.joi.2011.08.004.

29. Jiang J, He D, Ni C. The correlations between article citation and references' impact measures: What can we learn? Proceedings of the ASIST Annual Meeting. 2013;50(1):1-4. doi: 10.1002/meet.14505001162.

30. SCImago. 2020. SJR — SCImago Journal and Country Rank [Portal].

31. Montgomery DC. Design and analysis of experiments. 8th ed. John Wiley and Sons, Incorporated; 2012.

32. Rahimi S, Soheili F, Nia YA. Social influence, research productivity and performance in the social network Co-authorship: A structural equation modelling. Journal of Scientometric Research. 2020;9(3):326-34. doi: 10.5530/jscires.9.3.40.

33. Chi P-S. Which role do non-source items play in the social sciences? A case study in political science in Germany. Scientometrics. 2014;101(2):1195-213. doi: 10.1007/s11192-014-1433-1.

34. Hou WR, Li M, Niu DK. Counting citations in texts rather than reference lists to improve the accuracy of assessing scientific contribution: Citation frequency of individual articles in other papers more fairly measures their scientific contribution than mere presence in reference lists. BioEssays. 2011;33(10):724-7. doi: 10.1002/bies.201100067, PMID 21826692.

35. Fraser N, Momeni F, Mayr P, Peters I. The relationship between bioRxiv preprints, citations and altmetrics. Quantitative Science Studies. 2020;1(2):1-21. doi: 10.1162/qss_a_00043.

36. Oh J, Chang H, Kim JA, Choi M, Park Z, Cho Y, et al. Citation analysis for biomedical and health sciences journals published in Korea. Healthcare Informatics Research. 2017;23(3):218-25. doi: 10.4258/hir.2017.23.3.218, PMID 28875057.

37. González-Betancor SM, Dorta-González P. An indicator of the impact of journals based on the percentage of their highly cited publications. Online Information Review. 2017;41(3):398-411. doi: 10.1108/OIR-01-2016-0008.

38. Singh VK, Singh P, Karmakar M, Leta J, Mayr P. The journal coverage of Web of Science, Scopus and Dimensions: A comparative analysis. Scientometrics. 2021;126(6):5113-42. doi: 10.1007/s11192-021-03948-5.