Editorial letter

SCIENTIFIC RESEARCH STATUS IN VIETNAM DURING THE PERIOD 2019-2023 AND DIRECTIONS FOR MEDICAL RESEARCH

Nguyen Minh Tam[1,*], Nguyen Van Nhon[2], Nguyen Van Thanh[2]

[1] *Faculty of Basic Sciences, University of Phan Thiet, Binh Thuan Province, Vietnam*

[2] *Faculty of Health Sciences, Phan Thiet University, Binh Thuan Province, Vietnam*

Abstract: *This study analyzes the trend of scientific publication in Vietnam in the period 2019-2023, focusing on the relationship between the number of publications, the number of citations, and the JNCI (Journal Normalized Citation Impact) index. Data were collected from the InCites database of Web of Science, then processed and analyzed to extract results of important scientific significance. One of the analysis results shows that, although a few research institutions have a low number of publications, their JNCI index is high. On the contrary, most of the units exhibit a higher number of publications, albeit with a lower JNCI index. The research method combines traditional statistical tools and modern clustering algorithms such as K-means clustering and DBSCAN. In addition, word cloud techniques are used for visual illustration, which aids in a deep analysis of each research group's characteristics. The results show that institutions with international collaboration and publications in journals in groups Q1 and Q2 have significantly higher levels of influence, as demonstrated by the number of citations and superior JNCI index. In particular, the fields of medicine and health sciences stand out due to international attention and the urgency of research in the context of the COVID-19 pandemic. These findings emphasize the role of international collaboration strategies and publications in prestigious scientific journals to enhance the quality and influence of Vietnamese scientific research on the world stage.*

Keywords: *Vietnam, scientific research, medicine, health, InCites, WoS, citations*

1. INTRODUCTION

Along with the progress of human science and technology, scientific publication activities are taking place more vigorously than ever. On the basis of previously published scientific works, researchers can pursue and develop research directions, conduct valuable research, and continue to contribute knowledge to the community through the publication of their research. Scientific publication has become an important criterion to evaluate the capacity not only of individual researchers but also of scientific research institutions, including universities and research institutes around the world (Binswanger, 2014).

In addition, the number of scientific publications in each specific field also reflects important factors, such as the size of the field, the interest of the scientific community, social awareness, and the topicality of the research problem. These factors in turn affect scientific management policies from the national level to each research facility. A typical example is the number of scientific publications related to the COVID-19 pandemic. A report by Nature estimates the publication of over 23,500 scientific publications on COVID-19 in the first 6 months of 2020, encompassing research articles, letters, editorials, notes, and reviews (Teixeira da Silva, 2020).

However, by October 2020, this number had increased to more than 87,000 publications and continued to grow rapidly thereafter as scientists around the world

continued to study the coronavirus, the agent causing the pandemic, and its impact on the health of infected people (Grabmeier, 2021). Compared to the number of scientific publications in the field of nanoscience, which was one of the hottest scientific topics in the 1990s, it took more than 19 years for research in the field of nanoscience to increase from 4,000 to 90,000 scientific articles, while the number of research papers on the coronavirus reached that level in just about five months (Teixeira da Silva, 2020). These figures clearly demonstrate the topicality and deep public interest in the COVID-19 pandemic.

In recent years, scientific research in Vietnam has made significant progress in both the number of publications and the level of influence (Nguyễn Minh Quân, 2020; Lương Đình Hải, 2021). Notably, fields related to medicine and health sciences have received considerable attention from the community, especially in the context of the COVID-19 pandemic (Van Luong Nguyen et al., 2022; Bach Xuan Tran et al., 2022). Many research institutions in Vietnam have been increasing their domestic and international cooperation activities with the aim of publishing their research works in prestigious scientific journals with strict peer review processes in two giant databases, Scopus and WoS (Web of Science) (Pranckutė, 2021). Both of these databases offer scientific metrics that evaluate the growth and reputation of scientific articles, the scientists who write them, and the research institutions that carry out the research. Understanding the metrics of scientific publications is very important because, through them, we can evaluate their quality as well as their impact on the research community.

These metrics provide information on the frequency of citations, the reputation of the journal or publishing organization, and the influence of that publication on a specific research field. This helps researchers, regulators, or funding organizations make appropriate decisions based on reliable data and information. However, assessing the capacity of a research institution through the growth in the number of scientific articles, total citations, and other metrics over the years or by each group of interested subjects is still an unsolved problem for research institutions in Vietnam. To answer this question, the research team conducted this study by collecting data and analyzing scientific publication trends in Vietnam from 2019 to 2023. The study specifically examined the correlation between the number of publications, citations, and the JNCI index, both from a broad perspective and within the specific timeframes of each observation group.

2. RESEARCH METHODS

InCites is a product of the Clarivate system, providing a database related to the publications of scientists, research organizations, scientific journals, etc. by country and region in the fields of natural sciences and engineering (according to the SCIE index - Science *Citation Index Expanded*) and social sciences (according to the SSCI index - *Social Sciences Citation Index*), both of which are in the WoS database (Clarivate, 2024). At the same time, it is a support tool for scientists or research planning departments to analyze and evaluate research trends within the unit or in other interested units. Therefore, within the scope of this study, the authors are interested in the scientific research situation of research institutions recorded by InCites, data retrieved on August 20, 2024. The collected data is processed and analyzed using the Python programming language, which is one of the popular tools in data analysis today and is widely used in the fields of software development, data science, artificial intelligence and web development (McKinney, 2017).

Several visualization methods represent the characteristics of the data set during the data description and analysis process. In particular, bar charts were used to compare values between groups; box-plots were used to show the spread of the data, make it easy to see the quartiles, and find outliers; and histograms were used to show the probability distribution of continuous data and give details on skewness and spread. For time series data, line charts show the fluctuations of values over the years to clarify data trends over time. Finally, DePaolo and Wilkinson (2014) apply word clouds to illustrate the frequency of occurrence of topics appearing in published works. These clouds display words that appear more frequently in larger sizes, facilitating easy recognition of important keywords.

At the same time, the group applied two popular clustering methods, K-Means (Jain, 2010) and DBSCAN (Ester et al., 1996). K-means is an unsupervised clustering problem, which aims to divide the data into k groups based on the similarity between the data points, with the idea of dividing the data set into k clusters so that the sum of squared distances between the data points and the center of the cluster is the smallest. The DBSCAN *(Density-Based Spatial Clustering of Applications with Noise)* algorithm, on the other hand, automatically sorts data into clusters based on the number of points in space and can find noise points. However, a drawback of K-means is its sensitivity to the initial centroids *k* because it tends to converge to a local solution, which can lead to sub-optimal results. In contrast, DBSCAN has an advantage in clustering uneven data, but the parameter ε (radius) and the minimum number of points in the cluster affect cluster detection.

3. RESULTS AND DISCUSSION

3.1. Overview of scientific research in Vietnam

From the InCites database retrieved on August 20, 2024, the database collected over 05 years (2019-2023) includes the following elements (InCites Help Center, 2021):

+ Number of published works *(documents)* refers to the total number of scientific articles published by journals in the WoS database.

+ *Times Cited* is the total number of citations of published works.

+ Journal Normalized Citation Impact *(JNCI)* is the ratio of the actual number of citations to the average citation rate of papers published in the same journal in the same year and of the same type of paper. This index shows the performance of a publication relative to how other researchers perform when they publish their work in the journal. If the JNCI numerical value exceeds one, the research paper is considered above average and vice versa.

+ There are three author positions: first author, co-author, and corresponding author.

+ The Journal Impact Factor *(JIF)*: is the ratio of the sum of all citations in the current year to articles published in the previous two years to the total number of publications (including articles, reviews, and proceedings papers) published in the journal in the previous two years.

+ The Journal Citation Indicator *(JCI)*: is the average number of citations received by a journal›s articles compared to the average number of citations received by similar articles in the same field and in the same year. This value represents the average category-normalized citation impact of articles published in the previous three-year period.

The research team observed, analyzed, and made the following comments using data gathered over a five-year period from 75 institutions whose published works the InCites system had recorded:

3.1.1 Research performance

The observed data distribution of research units reveals that each unit employs a unique scientific research development strategy, as evidenced by the diversity of indicators.

Figures 1, 2, and 3, respectively, show the top 40 research units in terms of quantitative indicators, including the number of articles (Figure 1), total citations (Figure 2), and JNCI index (Figure 3).
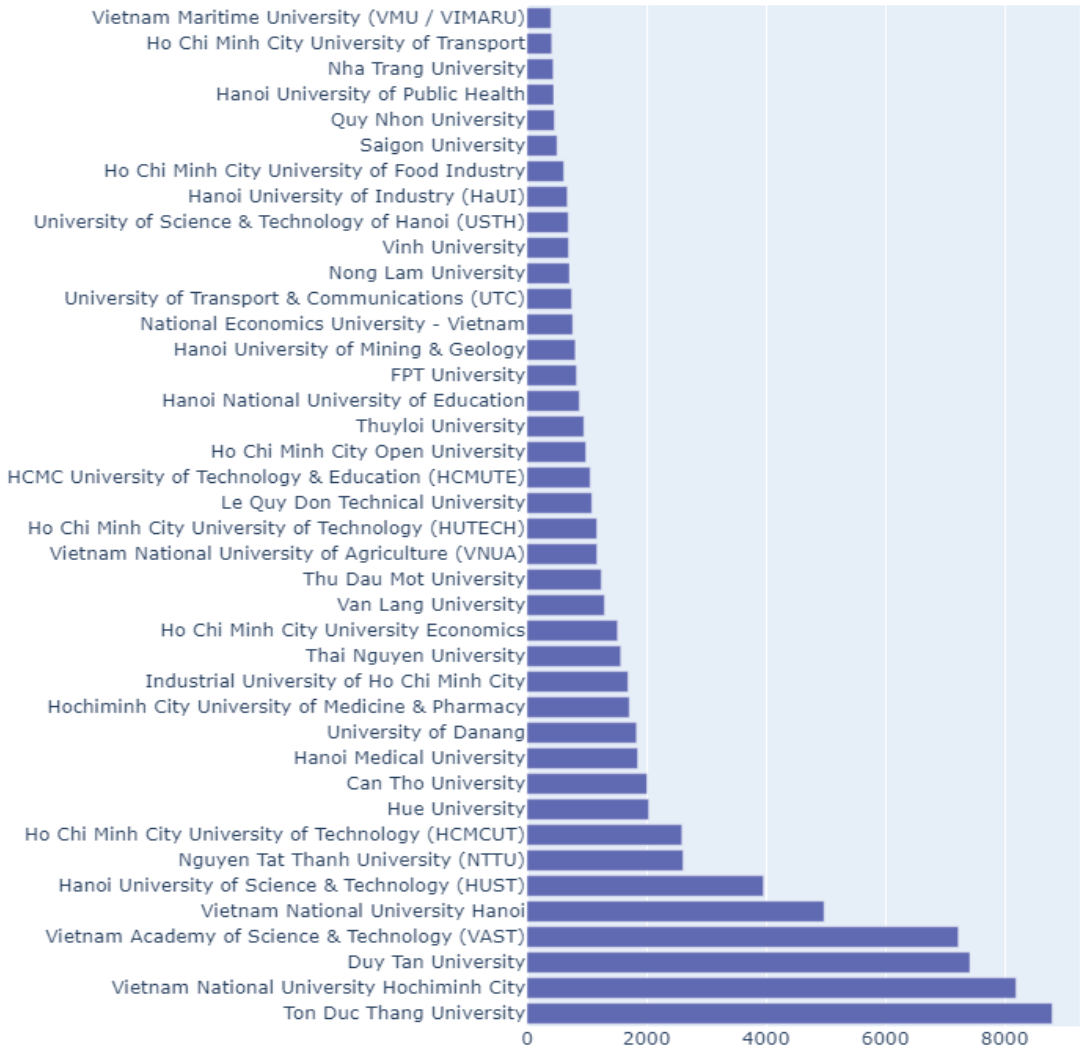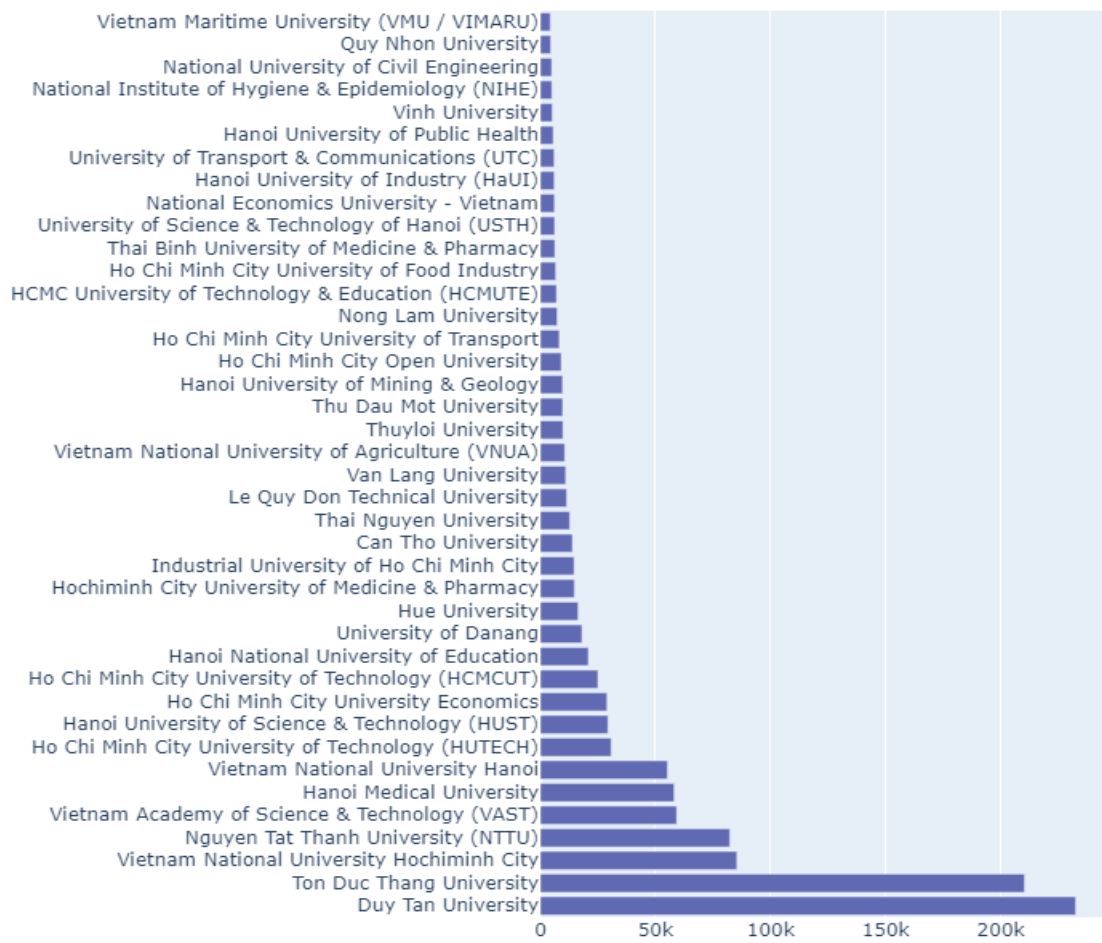

Figure 1. Top 40 units with the highest number of articles in the period 2019-2023

Figure 2. Top 40 units with the highest total number of citations in the period 2019-2023
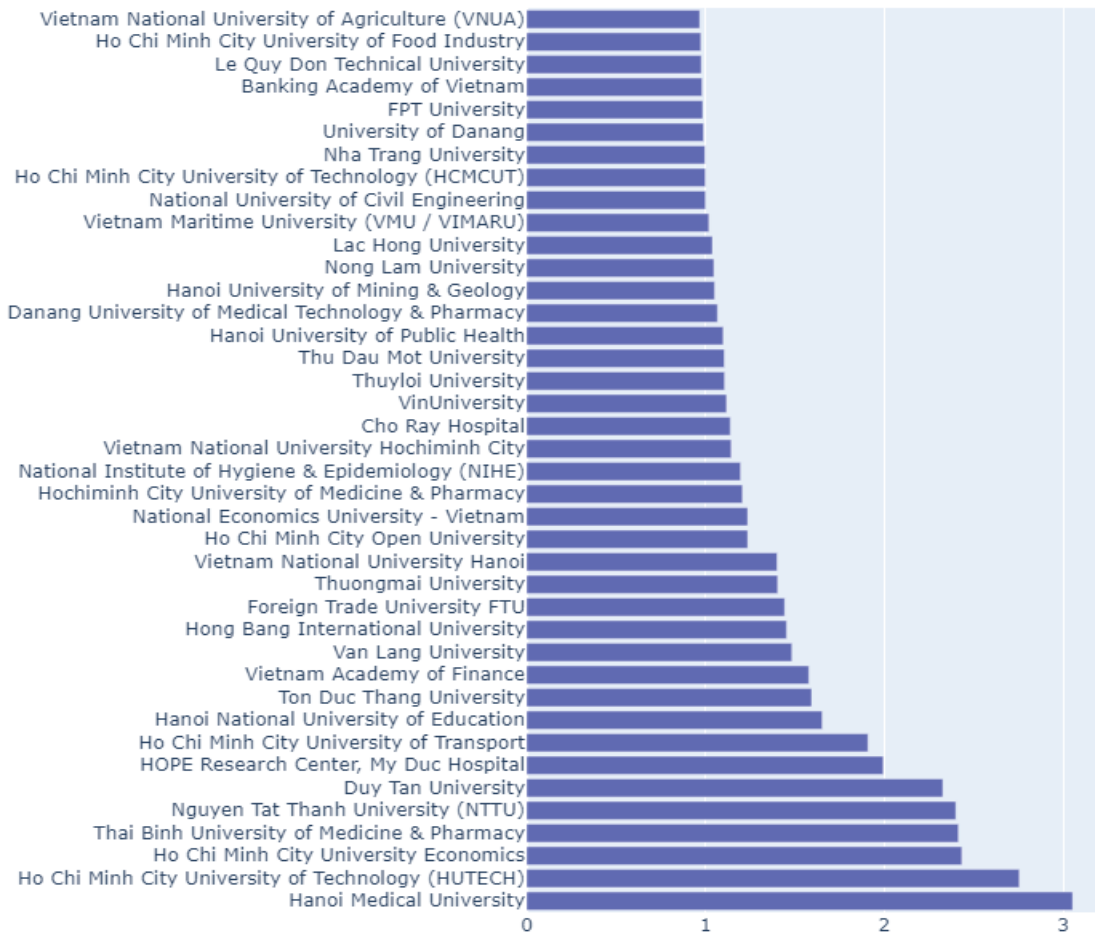
Figure 3. Top 40 units with the highest JNCI index in the period 2019-2023

Figure 4 shows the total number of articles, Figure 5 shows the total number of citations, Figure 6 shows the JNCI index, and Figure 7 shows the percentage of cited articles. The Shapiro-Wilk test with p-value < 0.001 showed that these data sets did not follow a normal distribution. This indicates the presence of outliers in the data or a significant variation between large units. Furthermore, when comparing the differences between research units based on factors such as the number of articles, number of citations, and JNCI index, the Kruskal-Wallis test results showed that there was no significant difference between observation groups at the 5% significance level (p-value > 0.4). Although there were differences in research strategies between units, scientific research factors did not have significant differences between units when tested on the entire sample.
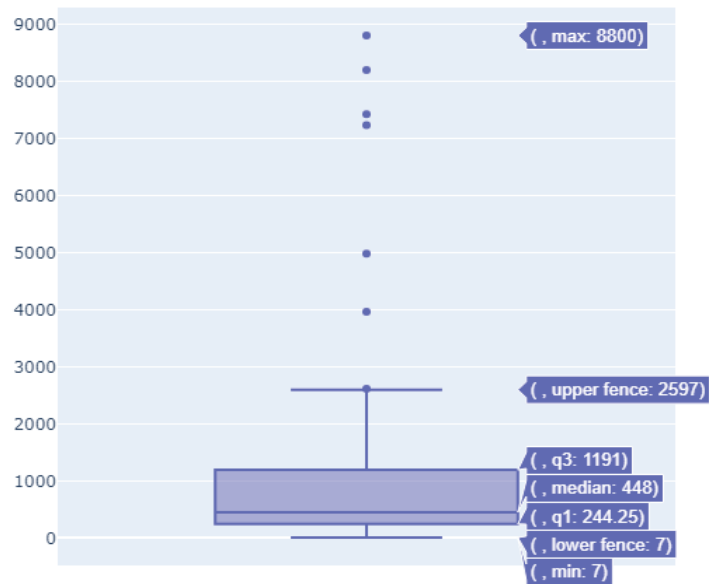
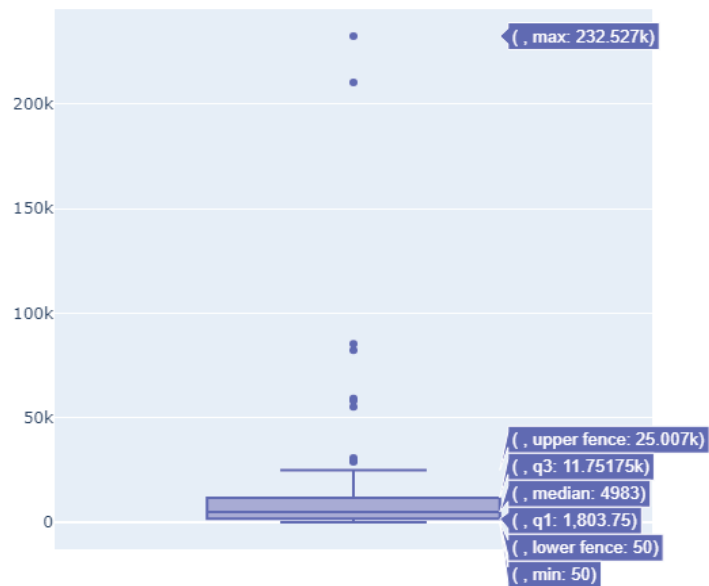**Figure 4. Box-plot of total number of articles in the period 2019-2023**

Figure 5. Boxplot of total citations for the period 2019-2023
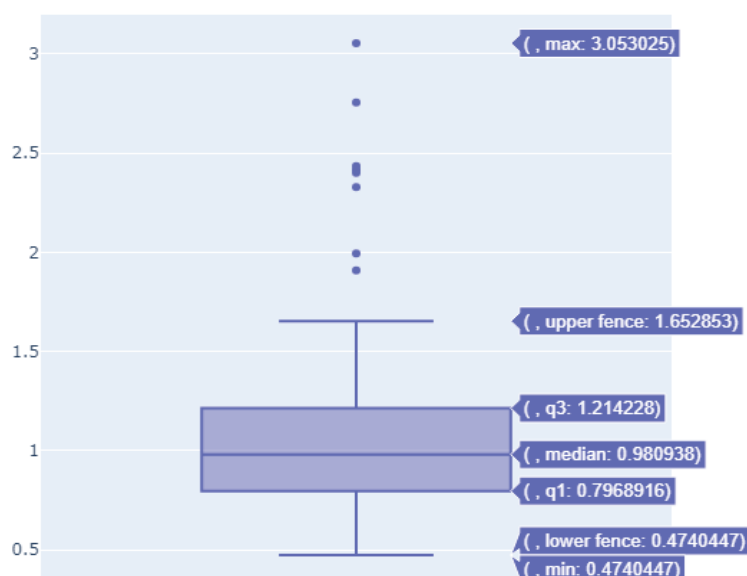
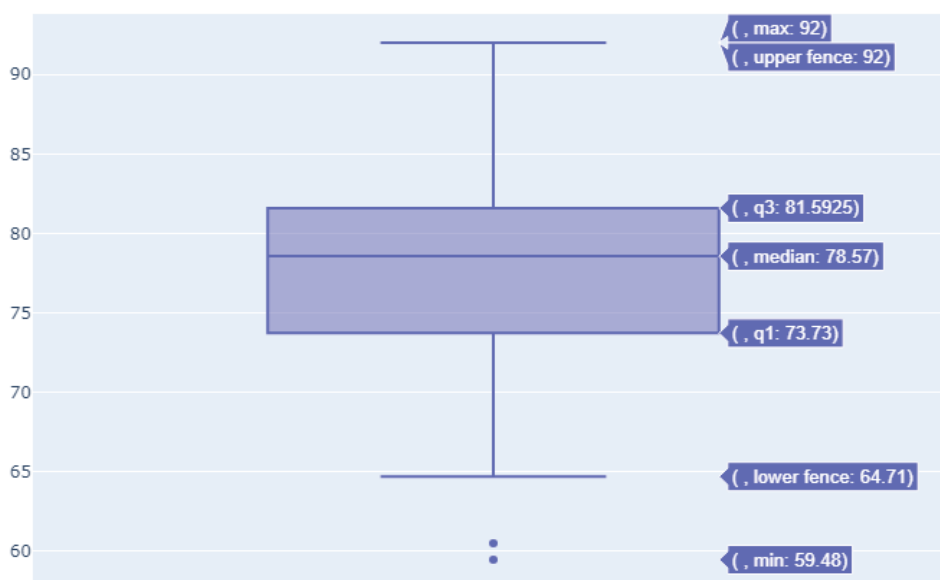**Figure 6. Boxplot of JNCI index for the period 2019-2023**

Figure 7. Box-plot of percentage of cited articles in the period 2019-2023

Ton Duc Thang University and Duy Tan University have the most publications and citations in the last 5 years out of the 75 research institutions that were looked at (Figures 1 and 2). However, Hanoi Medical University has the highest JNCI index (Figure 3), which is the ratio between the actual number of citations and the average citation rate of publications in the same journal in the same year and the same type of document. Notably, the number of publications at Hanoi Medical University is only 21.08% (1855/8800) of the number of publications at Ton Duc Thang University, and the total number of citations in 5 years of this institution is also only 25.0% (58111/232527) compared to Duy Tan University. This result indirectly shows

that although the number of publications and citations is lower, research works in the medical field at Hanoi Medical University still have a significant level of scientific influence, reflecting the important contribution of this field in the academic community.

The next analysis focuses on the research performance of domestic collaboration—articles with two or more authors and separate affiliations but in the same country—and international collaborations— with at least one co-author with an address abroad. Data collected during the period 2019-2023 shows that Vietnam's research trends are inclined toward international collaborations. Whether domestic or foreign, a small

number of institutions account for about 1/4 of research institutions, where a large number of articles reside. The data shows that the number of articles with international collaborations is twice as high as that of domestic articles (51,677 vs. 25,793), but the total number of citations from international works is 6.27 times higher than that of domestic articles (1,022,810 vs. 163,188). The outstanding citation index (JNCI) reflects the higher research impact of works with international cooperation, in addition to their advantage in quantity and quality. Figures 8 and 9 demonstrate that research works incorporating foreign elements have a higher number of citations and citation impact index compared to domestic articles.
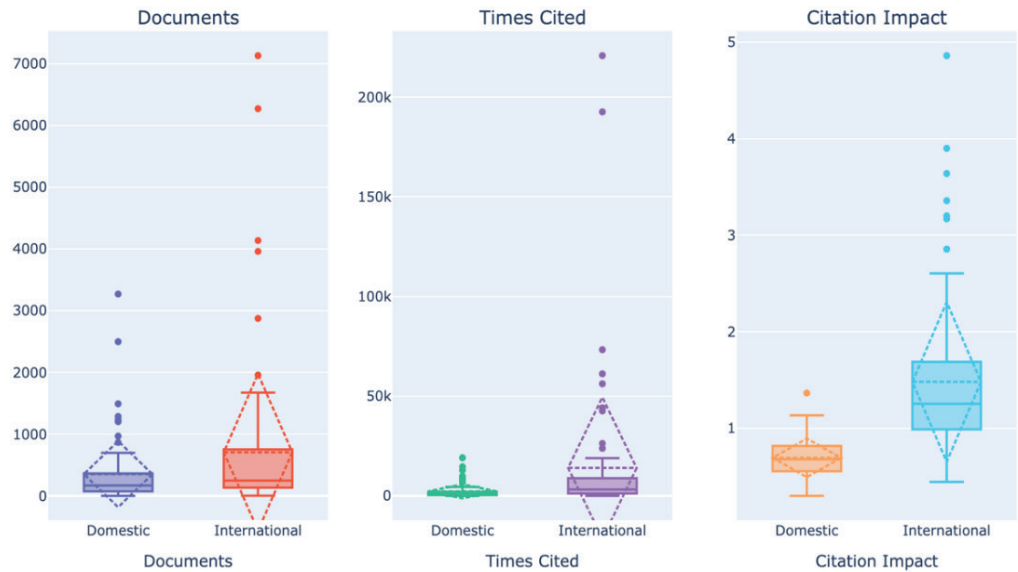

Figure 8. Boxplot of number of articles, number of citations and JNCI index in 5 years by domestic and foreign articles
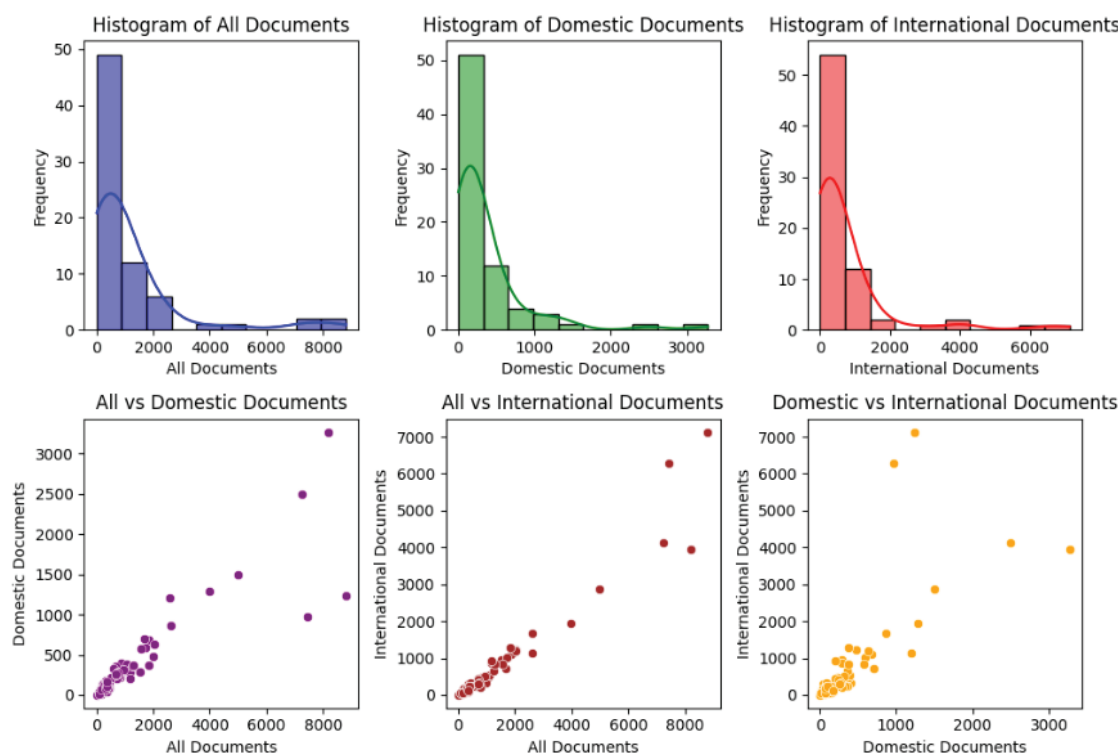
Figure 9. Histogram of the number and classification of domestic and international articles in 5 years (2019-2023)

Both the JIF and JCI measure journal impact based on the number of citations from the WoS database. However, the two indices differ in their calculation methods, scope of application, and ability to normalize by field. The JIF does not normalize by field, instead focusing on the average number of citations to articles over a two-year period. In contrast, the JCI is a newer, field-normalized index based on Category Normalized *Citation Impact (CNCI)*, which uses a three-year time frame and applies to all journals in the Web of Science Core Collection. This allows the JCI to provide a more fair and comprehensive assessment of the citation impact of journals across different scientific fields. Both indices use a Q quartile *to* classify journals, with Q1 comprising journals with high reach and

citation levels. Journals in Q2, Q3, and Q4 have decreasing impact but still play an important role in scientific publication.

In general, Vietnam's scientific research publication performance in the past 5 years has mainly focused on journals in the Q1 and Q2 groups in the WoS database, with a fairly high rate: 71.77% (49,878/69,496) according to the JIF index and 69.78% (55,281/79,224) according to the JCI index. In terms of citations, articles in the Q1 group also recorded the highest number of citations according to both indexes. The distribution according to the JNCI index also reveals that articles in Q1 have the most significant impact. In addition, the number of citations for articles grouped by Q is going down, and there isn't much difference between groups Q2, Q3, and Q4 (Figure 10, Figure 11, Figure 12, Figure 13).
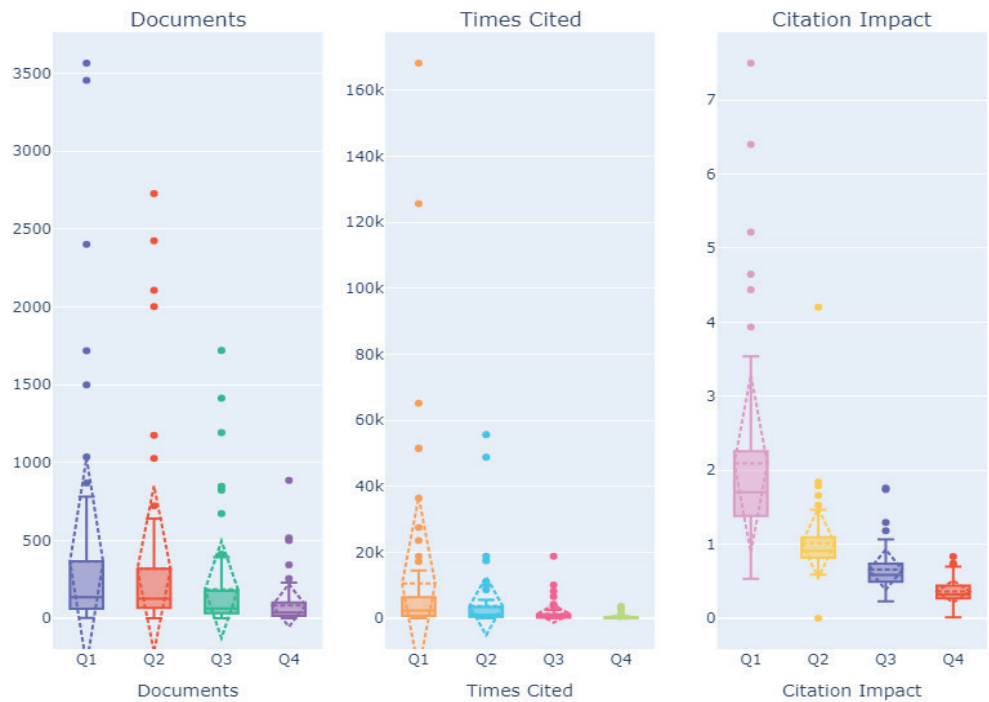
Figure 10. Boxplot of number of articles (documents), number of citations (*times cited*), JNCI index (*citation impact*) in 5 years according to JIF index
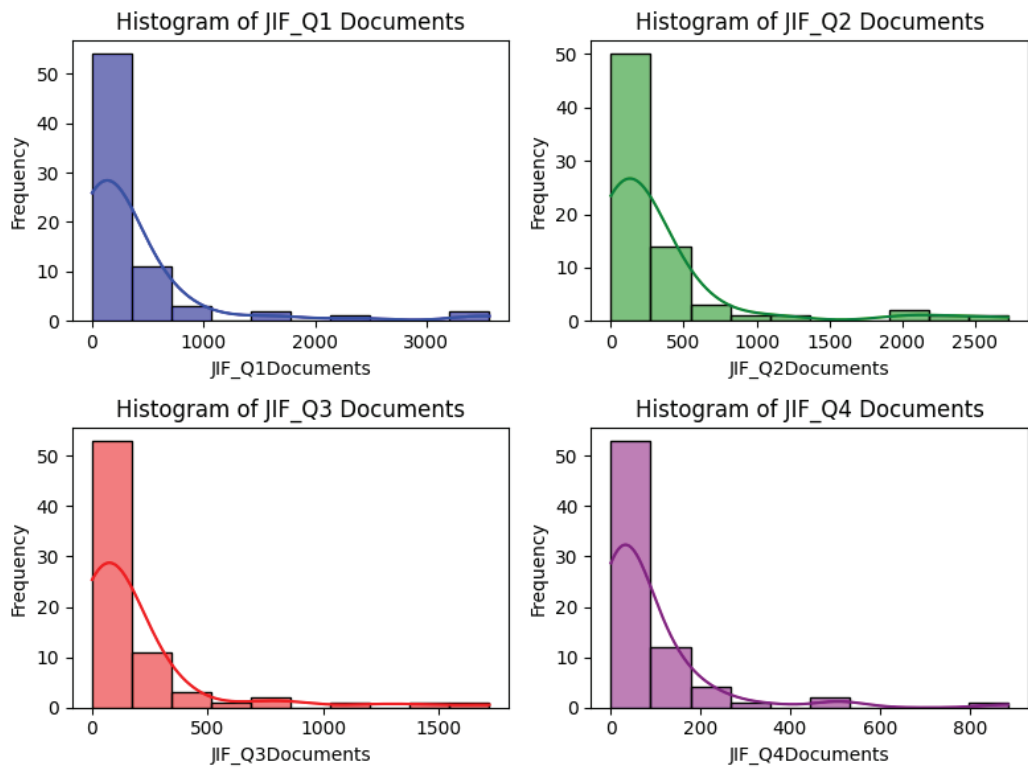

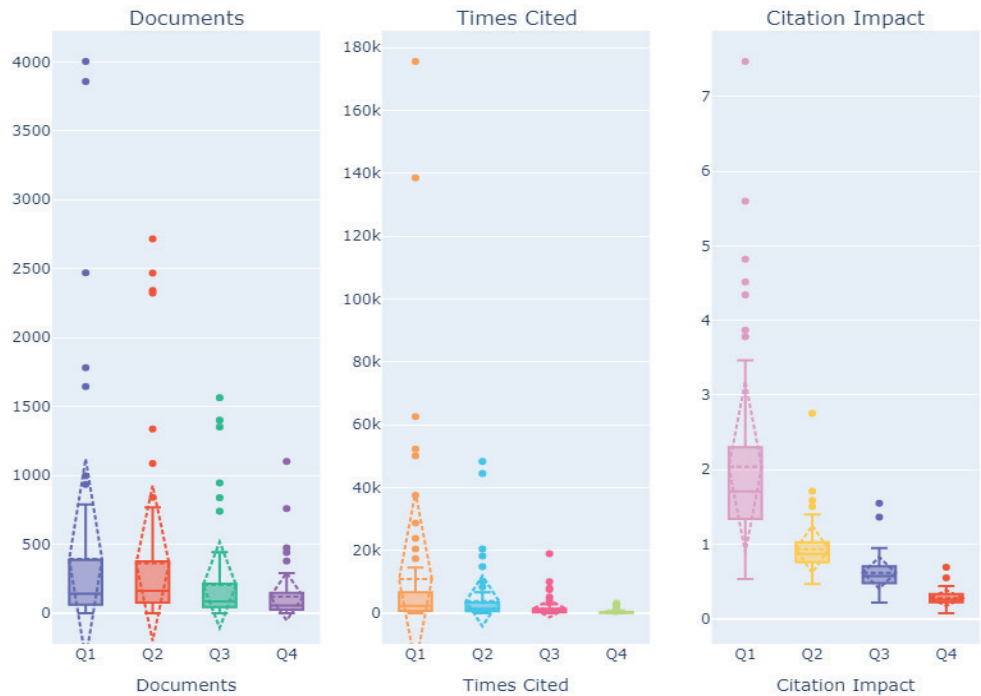**Figure 11. Histogram of JIF index (Q1, Q2, Q3, Q4) by number of articles in 5 years**

Figure 12. Boxplot of number of articles (documents), number of citations (*times cited*), JNCI index (*citation impact*) according to JCI index
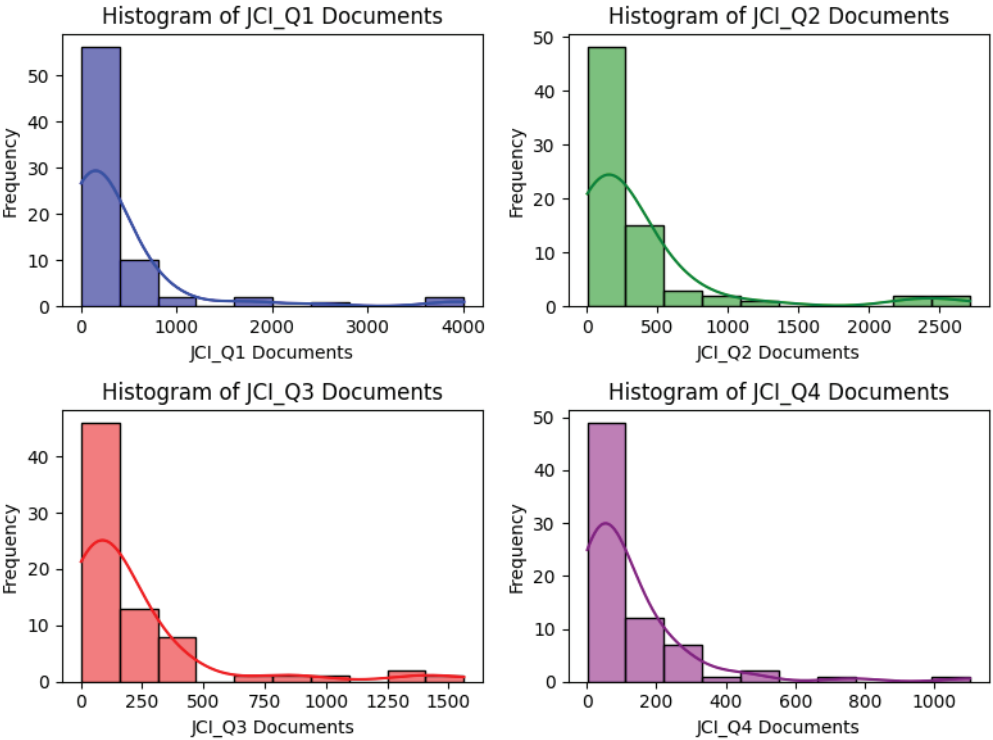

Figure 13. Histogram of JCI index (Q1, Q2, Q3, Q4) by number of articles in 5 years

In a research project, the author can take on different roles such as first author, corresponding author, or last author. *Figure 14 provides a comprehensive view of research performance by author position of 75 research institutions in Vietnam over a 5-year period (2019-2023). The results show that the proportion of works with corresponding authors (*37.02%; 41802/112915) and first authors (35.61%; 40208/112915) is dominant. Despite the high proportion of authors in these positions, the citation impact index (JNCI) for all three positions shows a similar level, with an average JNCI value of 0.84 for first authors, 0.87 for corresponding authors, and 0.89 for non-first authors (Figure 15). This suggests that the influence of authors in all positions in attracting citations is relatively equal.
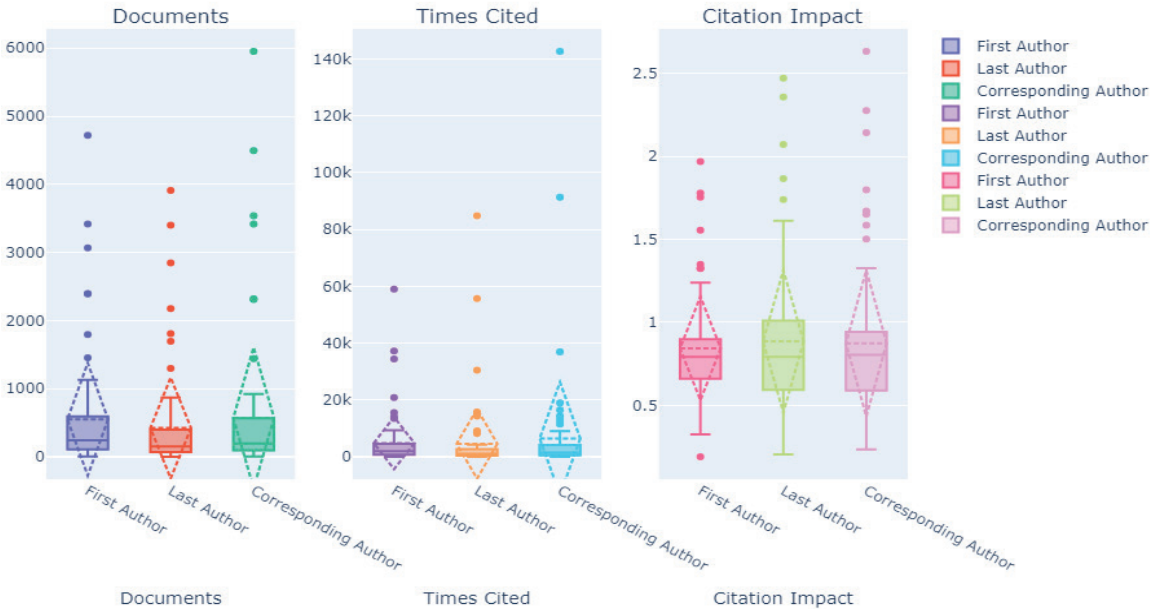

Figure 14. Box-plot of number of articles, number of citations, JNCI index in 5 years by author position
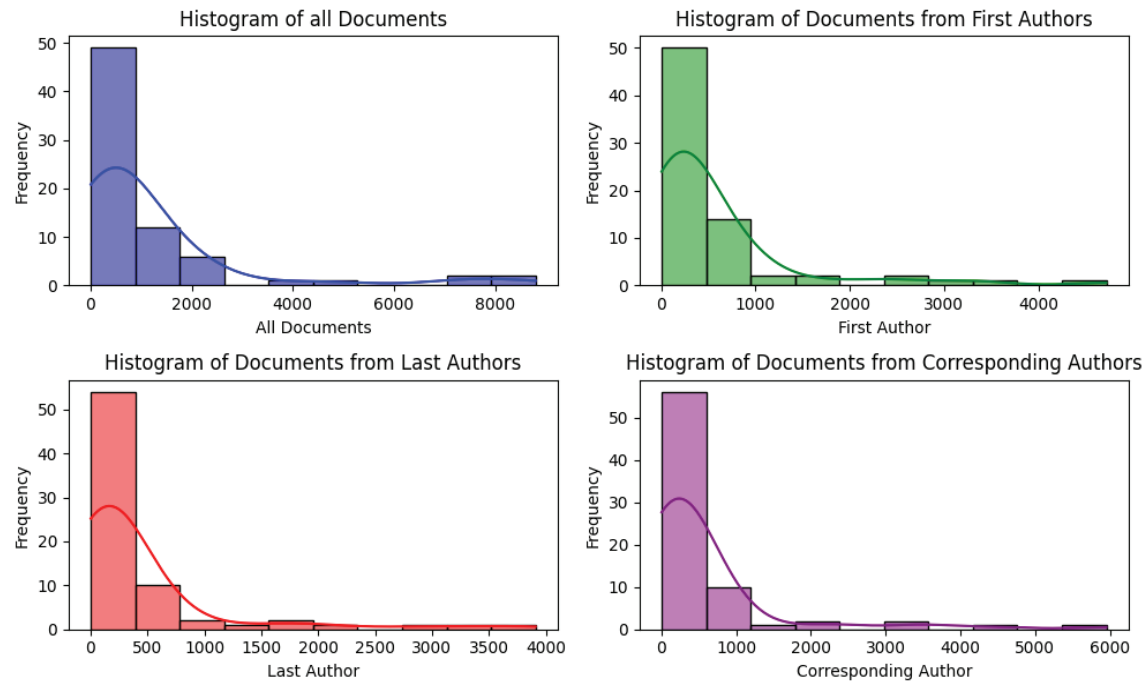
Figure 15. Histogram of number of articles by author position in 5 years

3.1.2. Research performance over the years

Data over the years in the period 2019–2023 of 75 surveyed research institutions show that each university has different research strategies. In general, the number of publications and the number of citations per article tend to be similar between universities. When articles in the same field are published in the same journal at the same time, the JNCI index compares their influence and quality. This shows that the quality and influence of research works depend on the number of citations as well as the level of interest of the scientific community (Figure 16, Figure 17, and Figure 18).
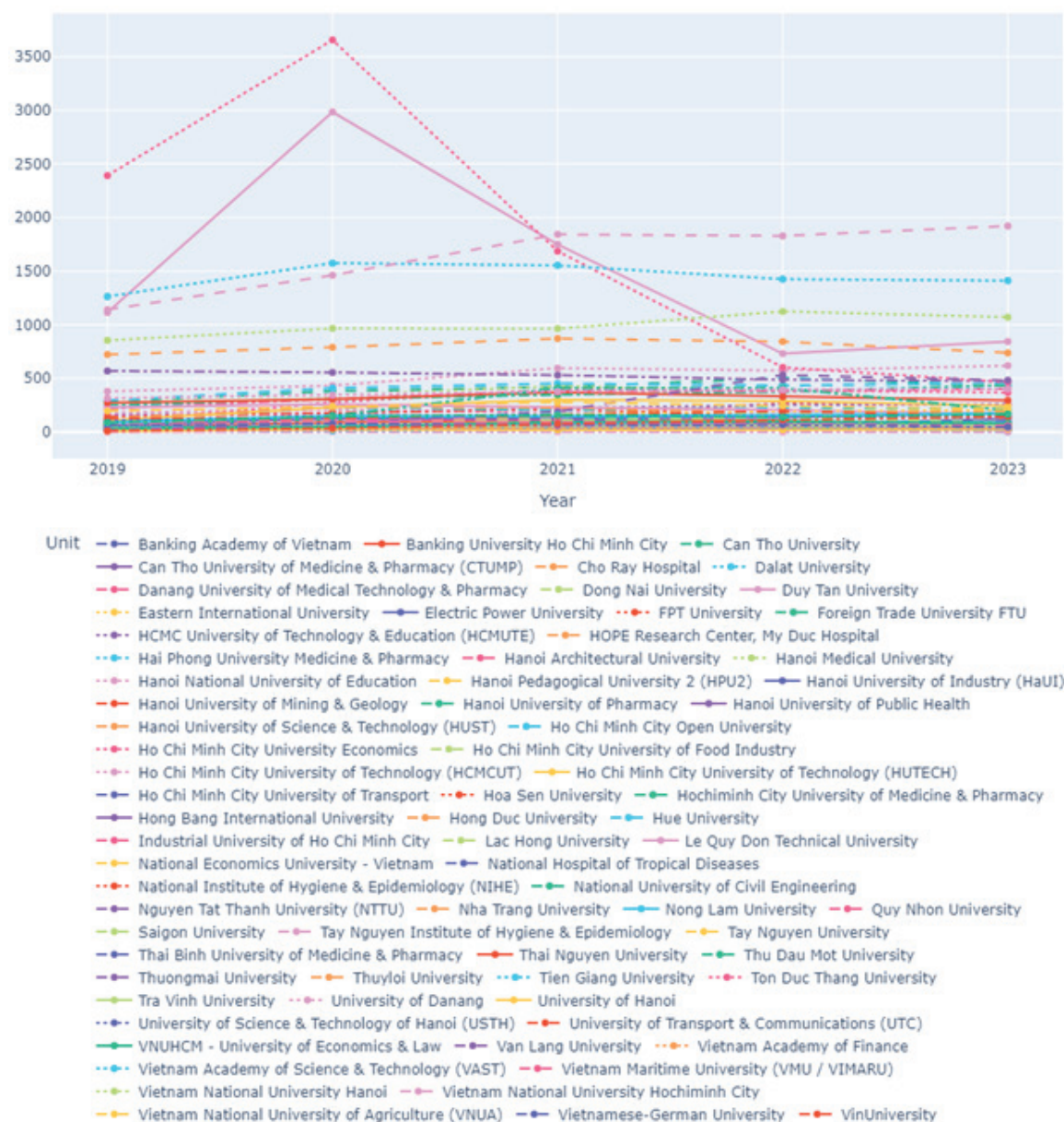
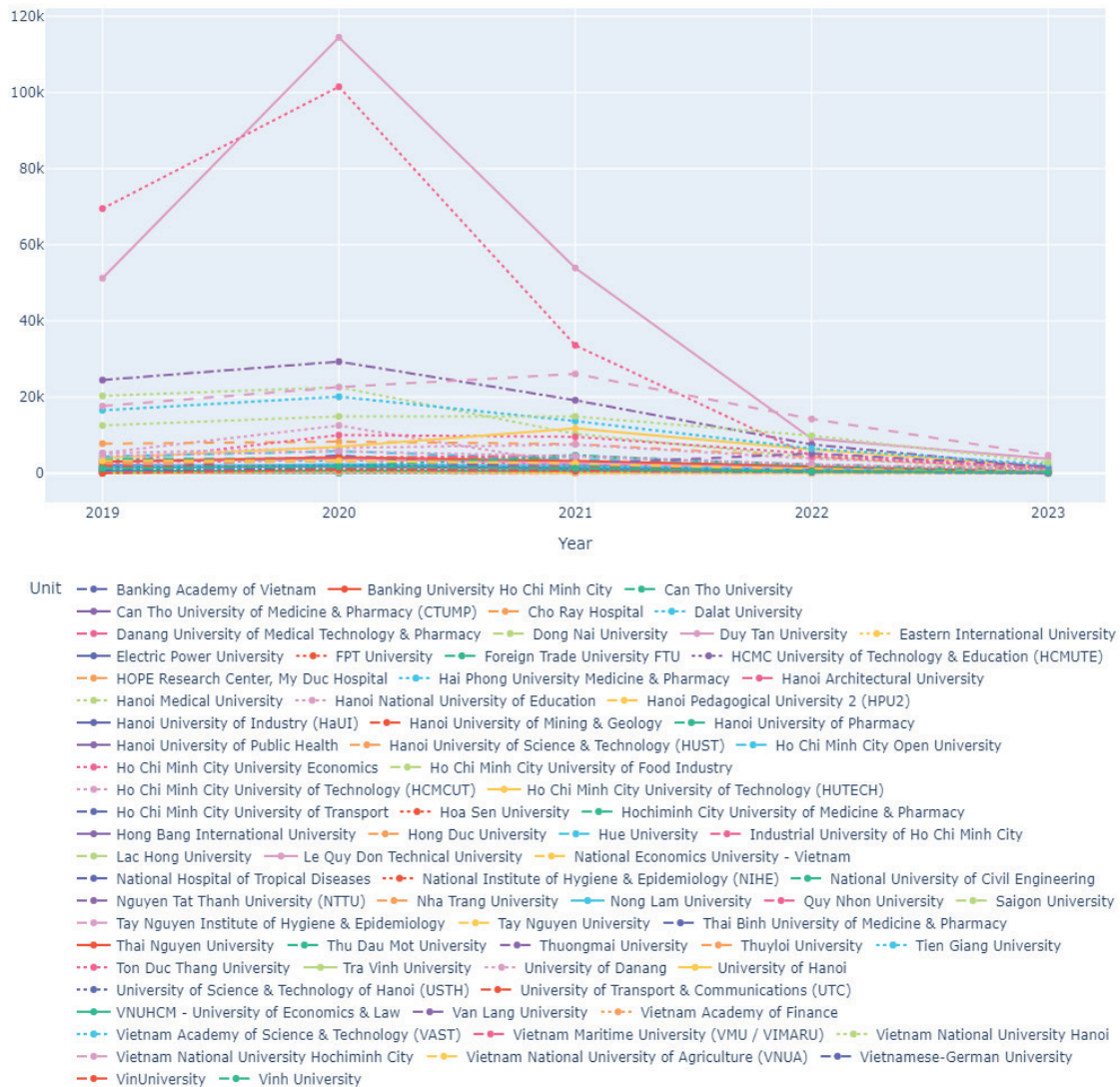**Figure 16. Total number of published works by year 2019-2023**

Figure 17. Total citations by year 2019-2023
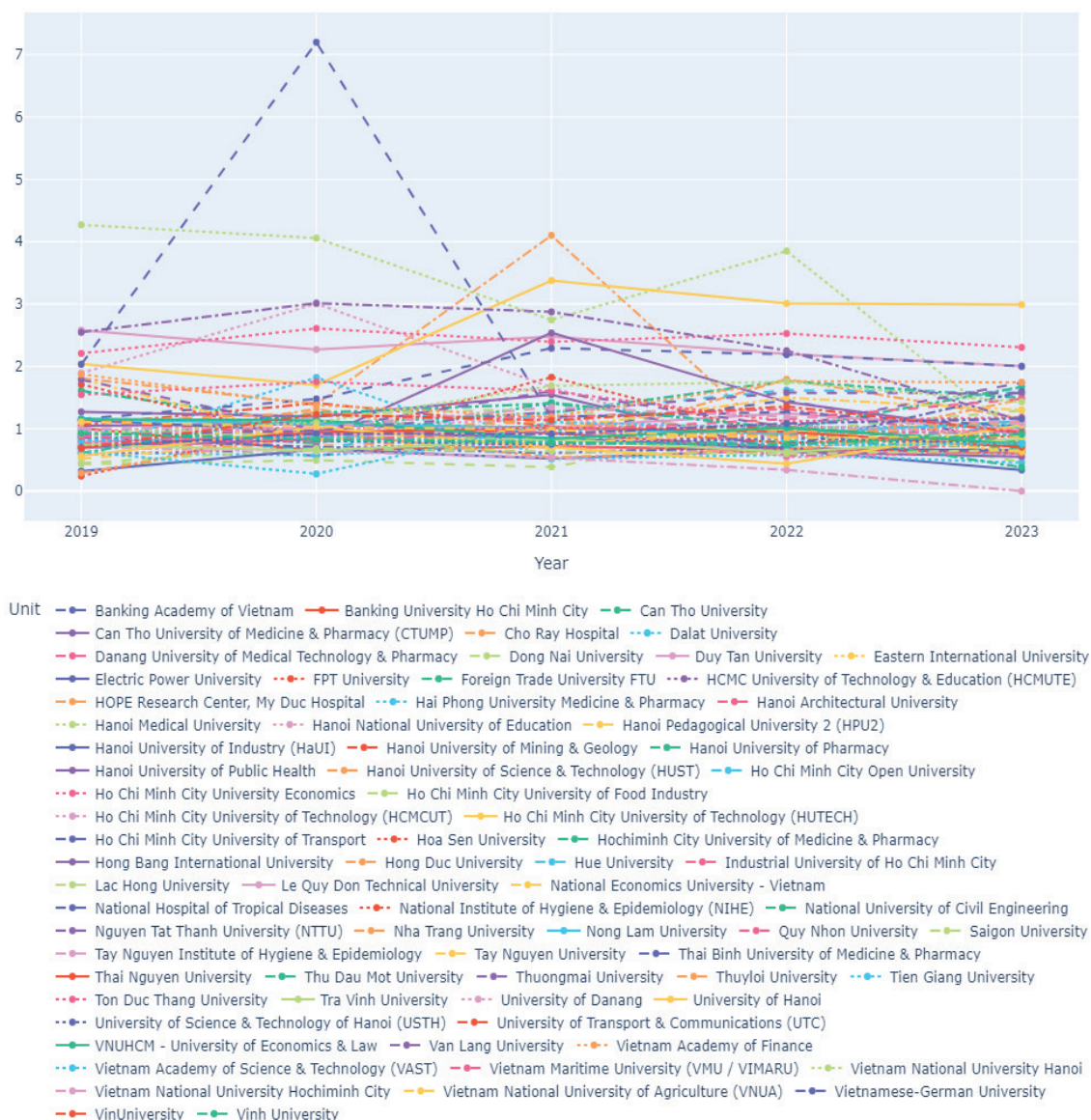
Figure 18. JNCI Index by year 2019-2023

3.1.3. Classification of research institutions based on research performance

Based on three factors, including the total number of articles, the total number of citations, and the JNCI index in the 5-year period (2019-2023) from the institutions whose works are recognized by Incites, two popular clustering methods are applied: *Density-Based Spatial Clustering of Applications with Noise (DBSCAN)* and K-Means. Simultaneously, we combine the random search algorithm to optimize the method of determining epsilon and min_sample values for DBSCAN, and to identify suitable clusters for K-Means. The evaluation indicators in Table 1 yield the following results:

Table 1: Evaluation index for 2 methods, DBSCAN and K-Means

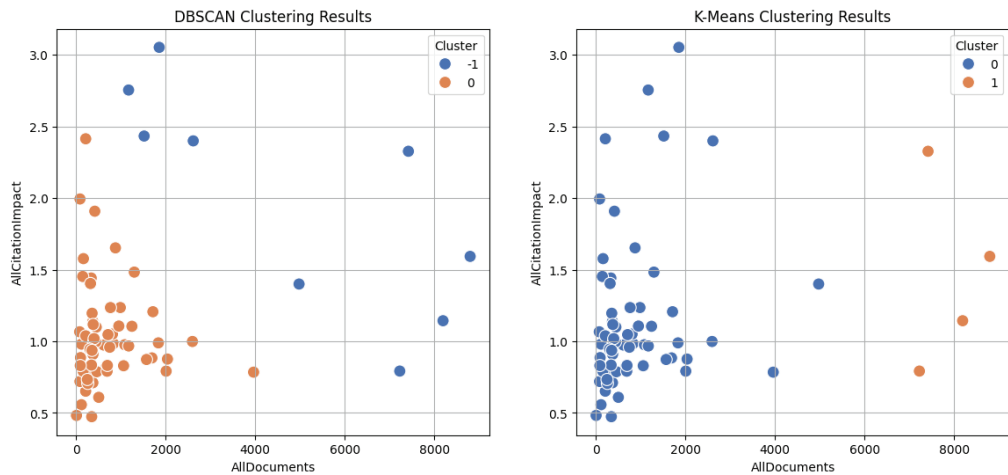| Index	DBSCAN	K-Means	The best option
Silhouette Score	0.7043	0.7512	K-Means
Calinski-Harabasz Score	70.5972	67.0867	DBSCAN
Davies-Bouldin Score	0.8885	0.5803	K-Means

Figure 19. Clustering by DBSCAN and K-Means methods

Based on the results from Table 1 and Figure 19, the DBSCAN method was chosen because of its flexibility in clustering, heterogeneity in shape, and ease of separation. Therefore, in Cluster -1 of the DBSCAN method, the groups with a high number of articles and a high JNCI index were grouped into one cluster for easy comparison with the remaining group. There are 9 research institutions grouped into a high-value cluster by the DBSCAN method, including Duy Tan University, Hanoi Medical University, Ho Chi Minh City University of Economics, Ho Chi Minh City University of Technology (HUTECH), Nguyen Tat Thanh University, Ton Duc Thang University, Vietnam Academy of Science and Technology, Vietnam National University, Hanoi, and Vietnam National University, Ho Chi Minh City. Figure 20, Figure 21, Figure 22, and Figure 23 show the data distribution in both groups as observed by each feature.
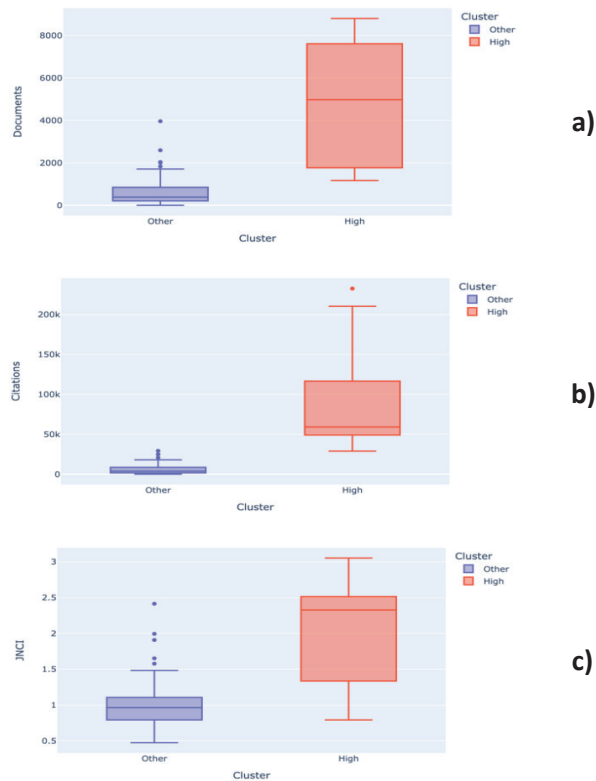
Figure 20. Box-plot of total number of articles (a), total number of citations (b), JNCI index (c) for the period 2019-2023 by 2 groups divided by the DBSCAN method
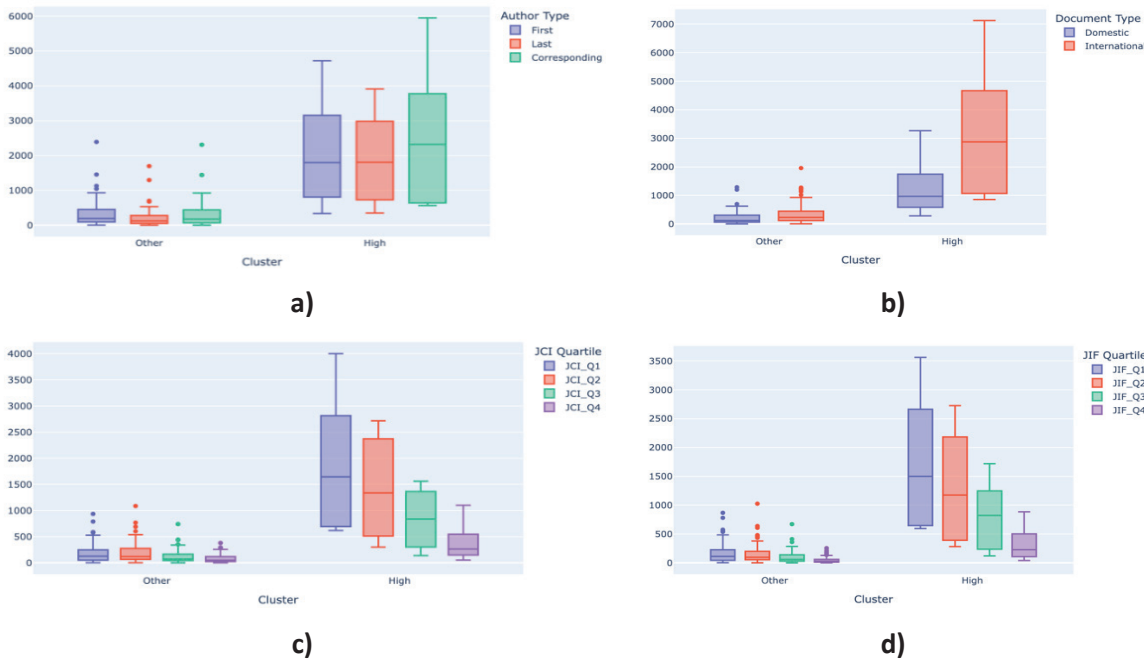
Figure 21. Box-plot of total number of articles by author position (21a) and article type (21b, c, d) in 2 groups divided by the DBSCAN method
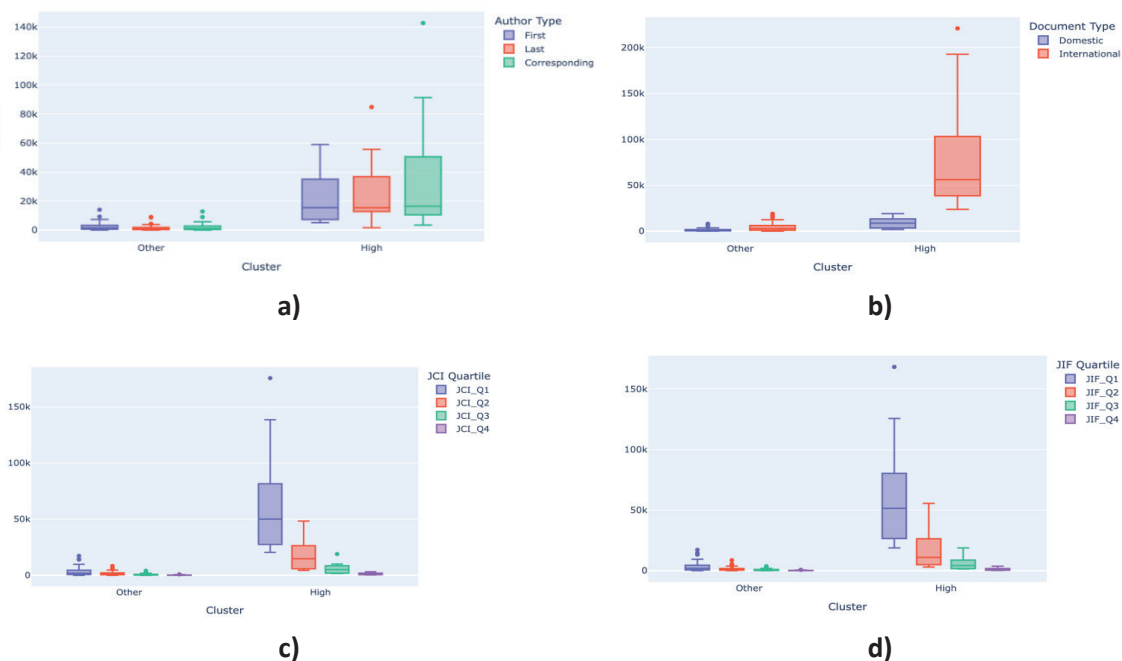
Figure 22. Box-plot of total citations by author position (a), article type (b), JCI by quartile (c), and JIF by quartile (d) in 2 groups divided by the DBSCAN method
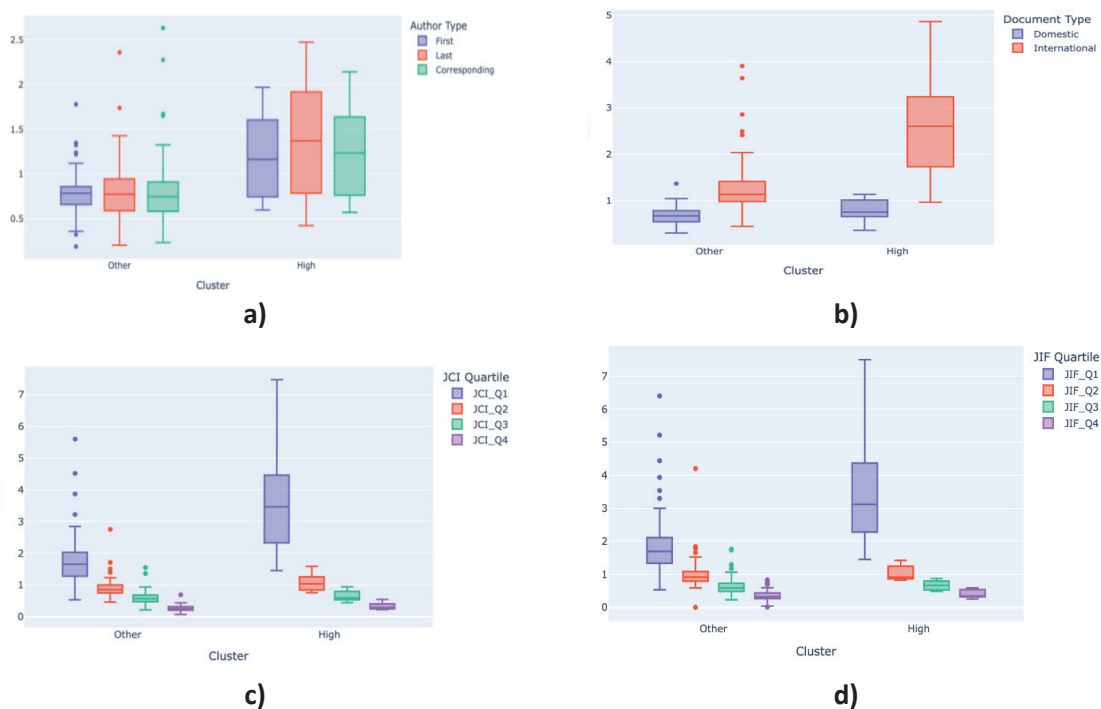
Figure 23. Box-plot of JNCI index by author position (a), article type (b), JCI by quartile (c), and JIF by decile (d) in 2 groups divided by DBSCAN method

While only 9 out of 75 research institutions (12%) are part of the strong research group, their total number of articles stands at 43,780, which is 1.02 times higher than the other group. However, this group receives a larger proportion of citations, accounting for 2.22 times more than the remaining group. In terms of the total number of published works, the strong research institution group tends to focus on international collaborative articles published in prestigious journals, with a high proportion of articles published in Q1 and Q2 journals. The position of the submitting author within the group plays a significant and proactive role in scientific publication. Additionally, the strong research group boasts an impressive number of citations, particularly from articles published in Q1 journals and international collaborative

works, which contribute significantly to the total number of citations. The JNCI index of these institutions is also higher, which indicates that the group's published works have a greater influence and quality than those of the remaining research institutions.

3.2. The state of scientific research in the field of health

From the analysis results 3.1.1, the data is divided into two groups: those that specialize in research on issues in the field of health sciences and the remaining group. The subsequent analysis shows that the results are similar to the previous conclusion. This indicates that the contribution of research in the field of medicine and pharmacy is substantial even though the number of publications and citations may be lower than in other fields (Figure 24).
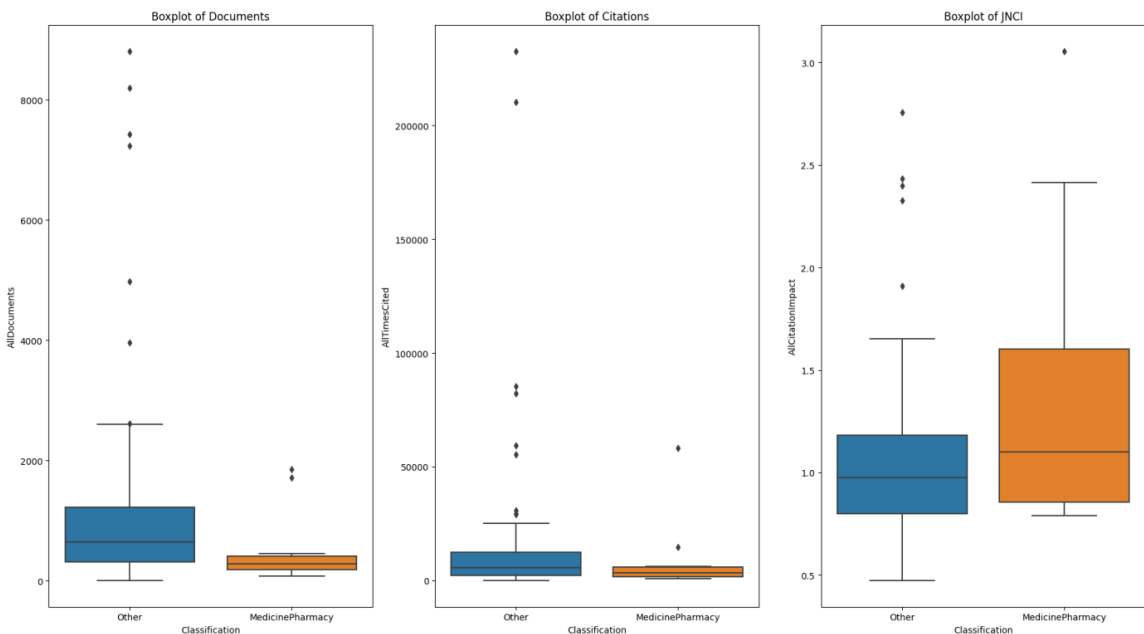
Figure 24. Box-plot of number of articles, number of citations, and JNCI index by medical-pharmaceutical group and the rest

For each observation unit, there will be a data set that represents the research topics of each unit, the purpose of which is to understand the main research direction of each unit. Each topic is recorded with two

indicators: total number of articles and total number of citations. The distribution chart is used to see the distribution of this data. The results show that the data has a clear right skew, meaning that there are topics

with a higher total number of articles or total number of citations than other topics. In other words, some research fields have significantly higher publication and citation levels than the rest. In order to adjust the data to reduce the right skew and make it more even, the analysis applied the logarithmic transformation method to the observed variables (Figure 25, Figure 26, and Figure 27).
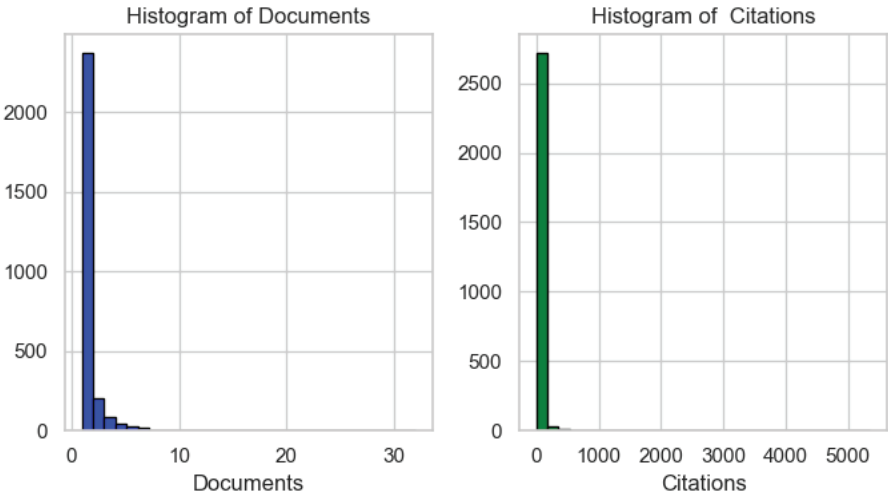
Figure 25. Distribution chart of number of articles and number of citations by the medical-pharmaceutical group in 5 years
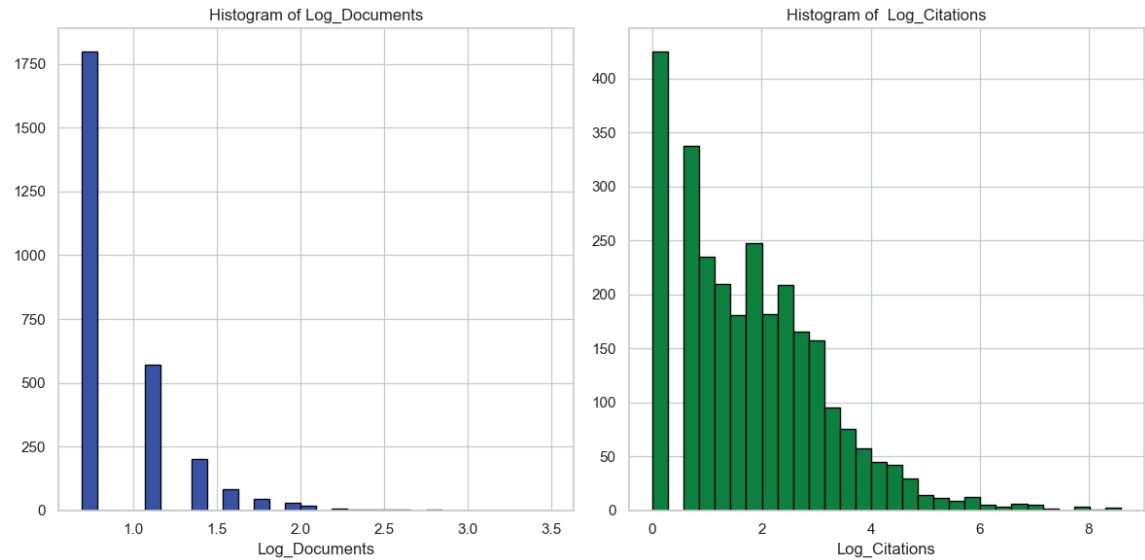
Figure 26. Logarithmic distribution chart of the number of articles and citations by medical-pharmaceutical group in 5 years
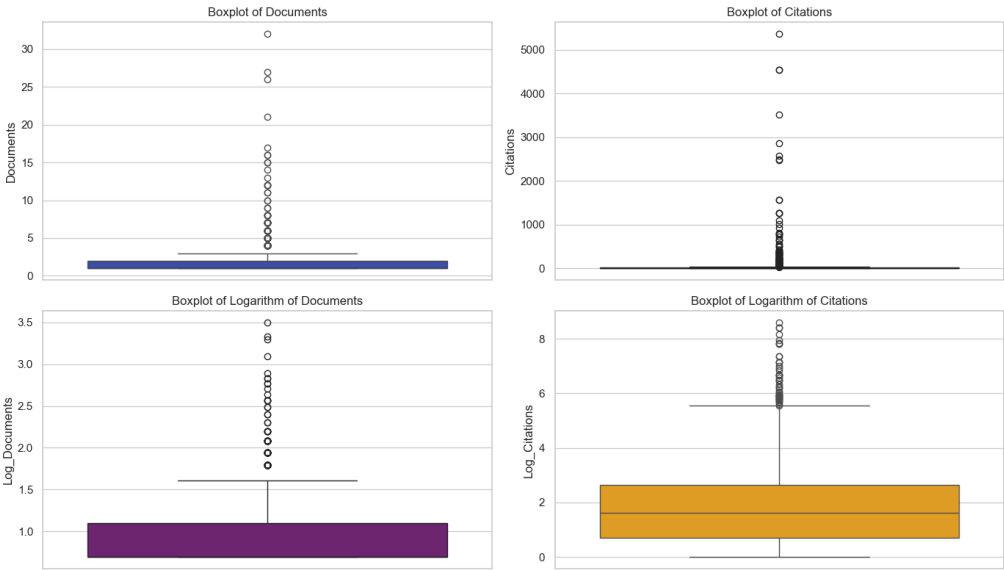
Figure 27. Non-logarithmic and logarithmic Box-plots of the number of articles and citations by medical-pharmaceutical group over 5 years

The Pearson correlation coefficient of 0.49 between the logarithm of the total number of articles and the total citations indicates a linear relationship between the two observed variables, but the correlation is quite low, so the co-variation is not too strong. Theoretically, while an increase in articles on a research topic often leads to an increase in citations, the coefficient indicates the presence of an inverse factor between the two variables. Furthermore, the Pearson correlation coefficient of 0.42 confirms the same trend in the hierarchy between the two variables, indicating the presence of nonlinear factors in addition to the linear factor. There are other factors affecting the distribution of the two observed variables besides the research quality factor, which can be seen in the number of articles and citations (Figure 28). This is shown by the fact that the two coefficients change between 0.4 and 0.49.
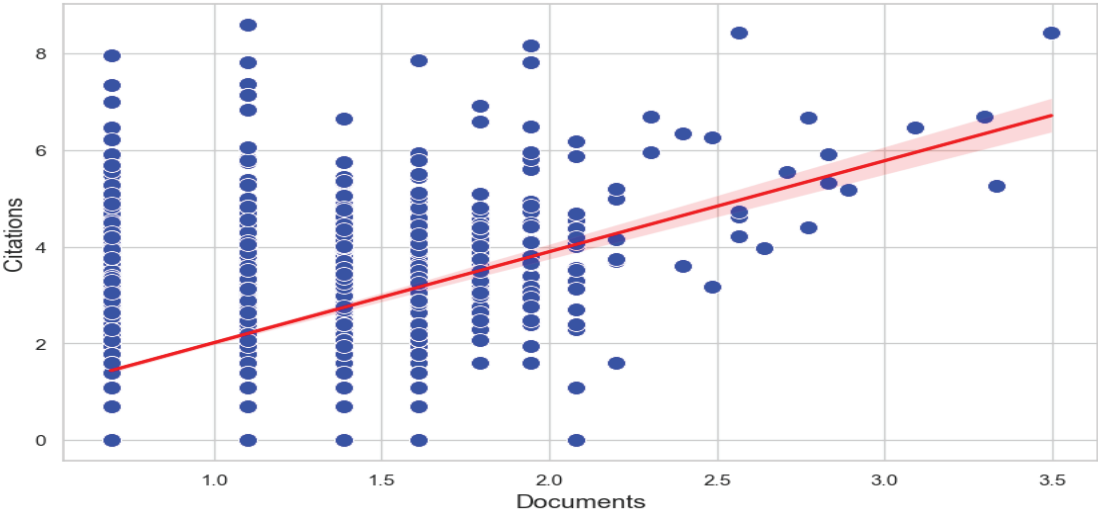
Figure 28. Correlation graph between the logarithm of topic frequency in articles and citations

The word cloud tool is used in Figures 29 and 30 to analyze the phrases of interest in the number of articles and citations collected in the period 2019-2023 of 11 research units in the InCites data, including Cho Ray Hospital, My Duc General Hospital, Central Tropical Hospital, Ho Chi Minh City University of Medicine and Pharmacy, Hanoi University of Pharmacy, Da Nang University of Medical Technology and Pharmacy, Can Tho University of Medicine and Pharmacy, Hai Phong University of Medicine and Pharmacy, Thai Binh University of Medicine and Pharmacy, Hanoi Medical University, and University of Public Health. The word cloud data for the health sector over the past 5 years indicates that research works related to the corona virus have been the most prominent, with a total of 266 articles and 14,174 citations. This clearly demonstrates the urgency and topicality of the research issue, particularly during the pandemic period, with the research peaking in 2020-2022.

Figure 29. Word cloud of articles in the medical research group

Figure 30. Word cloud of citations in the medical-pharmaceutical group

An overview of time series research shows the current research trends in the health sector in Vietnam (Figure 31) as follows:

a) Word cloud of articles in 2019

b) Word cloud of citations in 2019

c) Word cloud of articles in 2020

d) Word cloud of citations in 2020

e) Word cloud of articles in 2021

f) Word cloud of citations in 2021

e) Word cloud of articles in 2022

f) Word cloud of citations in 2022

g) Word cloud of articles in 2023

h) Word cloud of citations in 2023

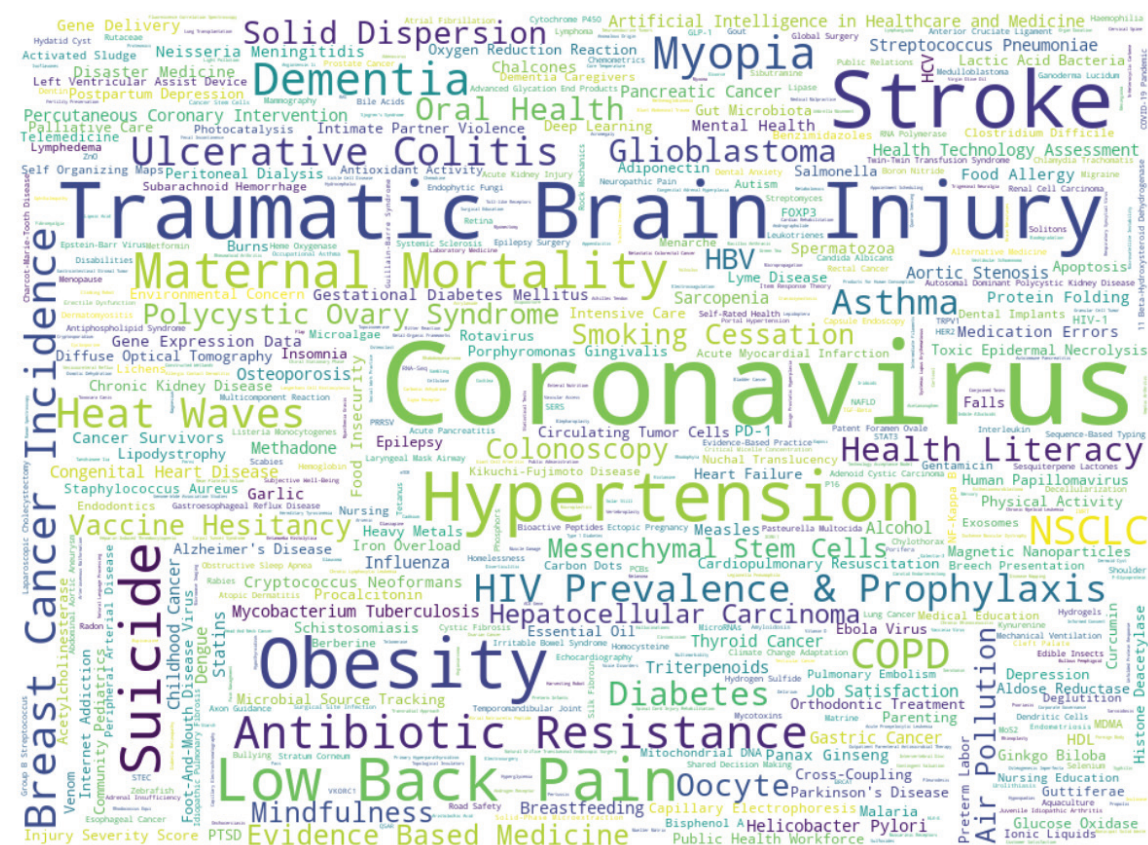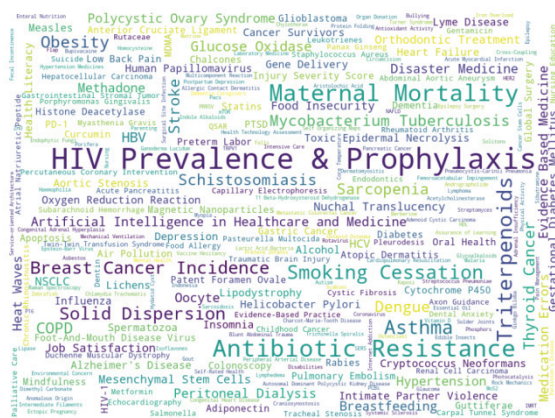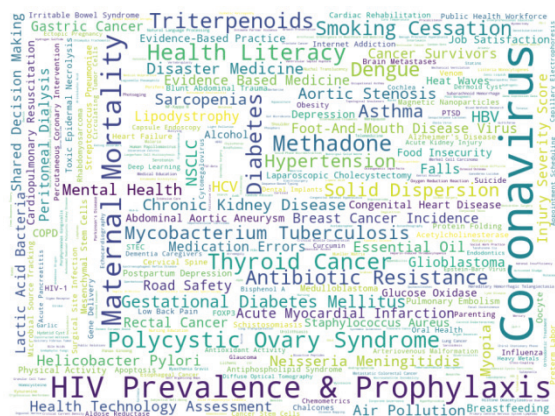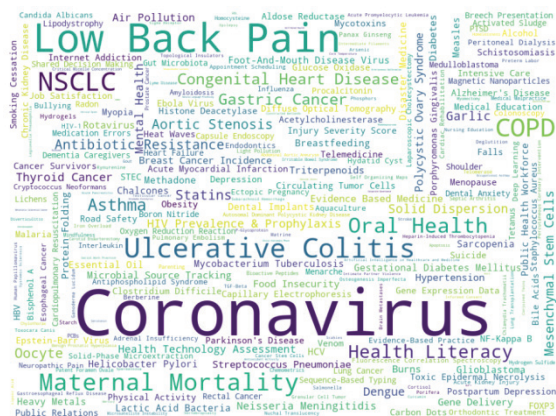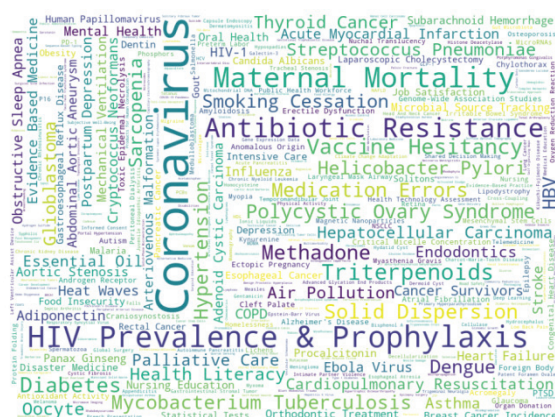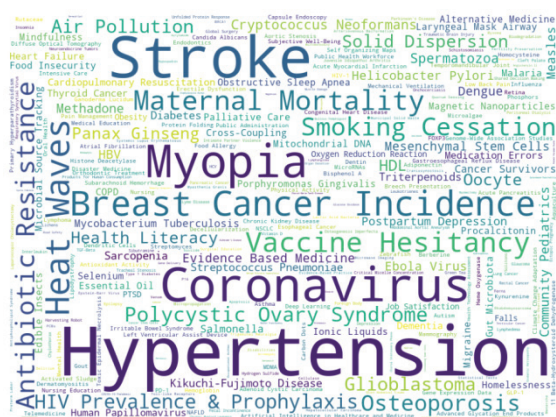**Figure 31. Word cloud of number of articles and number of citations in the period 2019-2023**

By examining word cloud data over time, we can gain a comprehensive understanding of the health research fields in Vietnam. Simultaneously, examining the frequency of citations for various topics offers valuable insights into research trends within this field. For example, based on the data of the most cited articles in 2023, we can see that the community's interest is focused on the following groups of issues:

+ Diabetes, chronic obstructive pulmonary disease (COPD), non-small cell lung cancer (NSCLC), *hepatocellular carcinoma*, *ulcerative colitis*, and circulating tumor *cells* are among the *chronic diseases and cancers.*

+ Research focuses on *infectious diseases and bacteria, such as the* corona virus **and** *Helicobacter pylori,* which can cause stomach infections and ulcers, as well as the issue of bacteria developing resistance to common antibiotics.

+ *Environmental health issues and their impacts* include heat waves*, air* pollution, and *food* insecurity.

+ Artificial Intelligence in Healthcare and Medicine is a focus of *technology and research groups.*

+ *Technology and research groups*, including Artificial Intelligence in Healthcare and Medicine*; Turmeric contains a compound called curcumin,* which may offer health advantages and anti-inflammatory characteristics. *Oocyte –* female sex cells needed for conception.

+ **Other problems**: *iron overload*; *stroke*.

4. CONCLUSION

Scientific research data in Vietnam for the period 2019–2023, extracted from InCites, shows a clear differentiation in the strategies of research units. Only about 12% of the institutions have a high total number of scientific articles, a high total number of citations, or a high JNCI index. Most of the remaining units have a modest research volume. The group of prominent research institutions significantly outperforms the remaining groups in many crucial aspects. Strong research units often have an international cooperation strategy, focusing on publishing in prestigious journals classified as Q1 and Q2.

Notably, the correlation between the number of articles and the number of citations tends to be positive, but its degree is quite weak. This indicates that while some institutions may publish more, there is a higher likelihood of receiving more citations. However, not all cases adhere to this rule. Another important factor is that institutions with high JNCI indexes showing the level of influence and quality of research are mostly concentrated in the field related to public health. Although research units in the field of health sciences do not have a very prominent number of publications or citations, they have quite high JNCI indexes, which show the urgency, importance, and public interest of health-related research. Typically, this trend is reflected in the increase in publications on the topic of epidemics, especially COVID-19, in the past 5 years. The works in the field of health sciences are not only large in volume but also attract much attention from the scientific community, thereby contributing significantly to the increase in the number of publications as well as citations from research institutions.

Author Information:

Dr. Nguyen Minh Tâm *(*Corresponding authors)*, Faculty of Basic Sciences, University of Phan Thiet, Phan Thiet City, Binh Thuan Province, Vietnam

Email: nmtam@upt.edu.vn

Dr. Specialist Level I - Nguyễn Văn Nhơn, Faculty of Health, University of Phan Thiet, Binh Thuan Province, Vietnam.

Email: nvnhon@upt.edu.vn

Dr. Specialist Level II- Nguyễn Văn Thành, Faculty of Health, University of Phan Thiet, Binh Thuan Province, Vietnam.

Email: nvthanh@upt.edu.vn

Note

The authors declare no competing interests.