Systematic Review and Meta-Analysis

Medicine

OPEN

The model of descriptive, diagnostic, predictive, and prescriptive analytics on 100 top-cited articles of nasopharyngeal carcinoma from 2013 to 2022

Bibliometric analysis

Chung-Chia Hung, MD[a], Mei-Yu Tu, MS[b,c], Tsair-Wei Chien, MBA[d], Cheng-Yao Lin, MD[e,f,g], Julie Chi Chow, MD[a,h], Willy Chou, MD[i,j,*]

Abstract

Background: Nasopharyngeal carcinomas (NPCs) are prevalent in southeast Asia. There is a need to systematically review the current trend and status of NPC research. However, most bibliometric analyses have tended to focus on descriptive and diagnostic analytics rather than predictive and prescriptive analyses. Thus, it is necessary to use the model of the 4 (called the descriptive, diagnostic, predictive, and prescriptive analytics [DDPP]) to derive insights from the data. This study aimed to apply the DDPP model to classify article themes and illustrate the characteristics of NPCs; compare NPC researcher achievements across countries, institutes, departments, and authors; determine whether the mean citations of keywords can be used to predict article citations; and highlight articles that are worthy of reading.

Methods: The Web of Science Core Collection was searched for 100 top-cited articles and reviews related to NPCs published between 2013 and 2022. As part of Microsoft Office Excel 2019, Visual Basic for Applications was used to illustrate the number of publications and scientific productivity of authors over time and to generate network/temporal heatmaps, chord/Sankey diagrams, radar/impact beam plots, and scatter/pyramid charts about collaborations among countries. The DDPP model identifies institutions, authors, and hotspots of NPC research. The category, journal, authorship, and L-index (CJAL) score was applied to evaluate individual research achievements.

Results: A total of 10,564 publications were extracted from Web of Science Core Collection and screened for 100 top-cited articles and reviews related to NPCs. Despite having the highest number of publications (36%), China lags slightly behind the US in CJAL scores. CJAL was higher at Sun Yat-Sen University, Radiat Oncol department, and author Jun Ma from China. The number of article citations was significantly correlated with the number of weighted keywords ($F = 1791.17$; $P < .0001$). Six articles with significantly increasing citations over the last 4 years were recommended.

Conclusion: This bibliometric study utilizes the DDPP model to analyze the scientific progress of NPC over the past decade. The whole genome is a hot topic that may prove to be a promising research area in the future. A temporal heatmap may serve as a tool for providing readers with articles that are worth reading, which could lead to additional research in bibliometrics.

Abbreviations: AAC = advantage coefficient, CC = correlation coefficient, CIDA = country, institute, department, and author, CJAL = category, journal, authorship, and L-index, DDPP = descriptive, diagnostic, predictive, and prescriptive analytics, EBV = Epstein–Barr virus, IBP = impact beam plot, NPC = nasopharyngeal carcinomas, PD-L1 = programmed death ligand-1, RA = research achievement, THM = temporal heatmap, T100NPC = 100 top-cited articles and reviews related to NPCs, WoSCC = Web of Science Core Collection.

Keywords: CJAL score, heatmap, nasopharyngeal carcinoma, Web of Science, whole genome

The authors have no funding to disclose.

The authors have no conflicts of interest to disclose.

The datasets generated during and/or analyzed during the current study are publicly available.

Supplemental Digital Content is available for this article.

[a] *Department of Pediatrics, Chi Mei Medical Center, Tainan, Taiwan,* [b] *Department of Nutrition, Chi Mei Medical Center, Tainan, Taiwan,* [c] *Department of Food Nutrition, Chung Hwa University of Medical Technology, Tainan, Taiwan,* [d] *Department of Medical Research, Chi-Mei Medical Center, Tainan, Taiwan,* [e] *Division of Hematology-Oncology, Department of Internal Medicine, Chi Mei Medical Center, Liouying, Tainan, Taiwan,* [f] *Department of Senior Welfare and Services, Southern Taiwan University of Science and Technology, Tainan, Taiwan,* [g] *Department of Environmental and Occupational Health, National Cheng Kung University, Tainan, Taiwan,* [h] *Department of Pediatrics, School of Medicine, College of Medicine, Taipei Medical University, Taipei, Taiwan,* [i] *Department of Physical Medicine and Rehabilitation, Chiali Chi-Mei Hospital, Tainan, Taiwan,* [j] *Department of Physical Medicine and Rehabilitation, Chung San Medical University Hospital, Taichung, Taiwan.*

** Correspondence: Willy Chou, Department of Physical Medicine and Rehabilitation, Chiali Chi-Mei Hospital, Tainan 710, Taiwan (e-mail: smilewilly@mail.chimei.org.tw).*

How to cite this article: Hung C-C, Tu M-Y, Chien T-W, Lin C-Y, Chow JC, Chou W. The model of descriptive, diagnostic, predictive, and prescriptive analytics on 100 top-cited articles of nasopharyngeal carcinoma from 2013 to 2022: Bibliometric analysis. Medicine 2023;102:6(e32824).

Received: 4 December 2022 / Received in final form: 10 January 2023 / Accepted: 11 January 2023

http://dx.doi.org/10.1097/MD.0000000000032824

Key points:

- 100 Top-cited articles and reviews related to nasopharyngeal carcinomas was classified by thematic analysis, visualized by the chord diagram, and highlighted by heatmaps for articles worth reading, which is a novel and modern approach.
- We developed a heatmap based on 4 distinct types of citation trends over the last 4 years to identify 100 top-cited articles and reviews related to nasopharyngeal carcinomas articles worth reading.
- Descriptive, diagnostic, predictive, and prescriptive analytics is a comprehensive bibliometric analysis and is worthy of recommendation for bibliographical studies in the near future.

1. Introduction

Nasopharyngeal carcinomas (NPCs) are uncommon cancers in Western countries but are common in Southeast Asia, the Middle East, and North Africa.[1,2] The annual incidence rate of NPC is the highest in the world in China (approximately 60.6 cases per 100,000 people), with nearly half of new cases occurring in China.[2,3] Considering that most of the high-risk population of NPC is composed of males between the ages of 40 and 50, both the social and economic costs as well as the medical burden are substantial.[3] NPC has been classified by the World Health Organization into 3 pathological subtypes, namely, keratinizing squamous, nonkeratinizing squamous, and basaloid squamous.[1,4] The incidence, prevalence, and mortality of head and neck tumors are on the rise, and they have become an important public health issue. Globally, approximately 550,000 people die from head and neck cancer each year, and 380,000 people die from diseases related to head and neck cancer.[5]

1.1. Identifying important papers in the NPC literature is challenging

The mortality rate associated with NPC has decreased over the past decades,[6,7] resulting from years of research aimed at developing innovative methods for controlling and managing the disease. Nevertheless, there has been a proliferation of NPC literature involving a variety of researchers, countries, specialties, and scientific journals,[8] so it is difficult to identify the papers of greatest significance.

Citation analysis is a bibliometric analysis method[9] that evaluates the influence and importance of an article in a certain field through the analysis of citation counts.[10] Affirmations are also given to authors, institutions, and countries that have made significant contributions to the field. Many disciplines have utilized this method, including urology and nephrology,[11–14] dermatology,[14] spine surgery,[15] orthopedics,[16] neurosurgery,[17] ophthalmology,[18] and otolaryngology.[19,20] A previous study[9] examined highly cited NPC papers, but only simple descriptive analytics were applied to conduct the bibliometric analysis, such as identifying the leading countries, institutes, authors, and journals.

The major disadvantage of these articles[9,11–20] is that they provide a list of all highly cited articles that may have become obsolete with a few citations in recent years. Therefore, it is important and meaningful for readers to read articles that have received increasing numbers of citations in recent years. To date, no such articles have been found in reporting articles worthy of reading on a specific topic. The reason is that the selection of articles that are worth reading requires the use of a robust algorithm.

1.2. Descriptive analytics used in bibliometrics are quite inadequate

Data can be analyzed in a variety of ways. The choice of method depends on the questions we are asking and the information we are seeking from our dataset. It is sufficient to use descriptive analytics to explain what has happened.[21] On the other hand, diagnostic analytics can be useful in determining why it happened [e.g., by classifications of author collaborations using coword analysis to understand their relationship in CiteSpace (College of Computing and Informatics, Drexel University)[22,23] and VOS Viewer (Universiteit Leiden)[24]]. Furthermore, for questions relating to what will happen in the future and what should be done next for readers, we could use predictive and prescriptive analytics[21] in bibliometrics.

Therefore, the research question of this study aims to examine whether the 4 types of data analysis can be applied to 100 top-cited articles and reviews related to NPC (T100NPC) utilizing the descriptive, diagnostic, predictive, and prescriptive analytics (DDPP) model consisting of descriptive, diagnostic, predictive, and prescriptive analyses, which has not been previously mentioned and demonstrated in bibliometrics.

1.3. Temporal heatmaps (THMs) used to highlight articles worth reading

Recent publications may not have sufficient citations due to the time-dependent citation analysis.[25] As an example, none of the clinical articles published from 2013 to 2018 were ranked among the top 100 most-cited NPC articles.[7] A corresponding analysis was conducted by the authors[7] to identify the top 10 clinical research articles from 2013 to 2018, with the objective of more comprehensively revealing the NPC development trend, research focus and/or articles worth reading.

Nevertheless, articles with potentially increasing citations from citing papers in recent years should be focused on. Even in the Web of Science core collection (WoSCC),[26] citation trends for articles worth reading are not particularly presented, but the website does provide recent 5-year citation counts in a contingency table. Thus, a THM, as developed in this study, is required for the visual representation of NPC articles worthy of reading based on prescriptive analytics.

1.4. Study aims

In this study, we verified the hypotheses that the DDPP model could be used to analyze bibliometric data, and the THM would be useful and viable for selecting articles that merit additional attention.

2. Methods

2.1. Data sources

We searched "TS = ((nasopharyngea carcinoma) OR (nasopharyngeal carcinomas) OR (carcinoma, nasopharyngeal) OR (carcinomas, nasopharyngeal))" on WoSCC for T100NPC since 2013. The document types were original articles and reviews. The retrieved articles were ranked from highest to lowest based on the number of citations. Six researchers reviewed and screened the titles and abstracts of the articles. If necessary, some studies without a PubMed unique identifier number or with complexly mixed NPC factors, such as head and neck cancer and Burkitt lymphoma, were excluded. The T100NPC articles are listed at the link[27] and deposited in Supplemental Digital Content S1, http://links.lww.com/MD/I419.

Because no human subjects were enrolled, ethics approval was not required for this study.

2.2. Four analytics applied to this study

2.2.1. Descriptive analytics in T100NPC. A 4-quadrant radar plot[28] and a pyramid plot in R[29] were used to visualize the productive entities and journals in T100NPC articles.

The absolute advantage coefficient (AAC) (see Equations 1–3)[14,30–32] was applied to evaluate the dominance extent for the most influential country, institute, department, and author (CIDA) in category, journal, authorship, and L-index (CJAL) scores[28] (based on the CJA score[33] and the L-index[34] to evaluate research achievements [RAs]). The Y-index[35,36] was applied to locate their coordinates on the 4-quadrant radar plot.[28]

$$\mathrm{AAC} = (\mathrm{R}_{12}/\mathrm{R}_{23}) / (1 + (\mathrm{R}_{12}/\mathrm{R}_{23})), \quad (1)$$

$$\mathrm{R}_{12} = \mathrm{A1}/\mathrm{A2}, \quad (2)$$

$$\mathrm{R}_{23} = \mathrm{A2}/\mathrm{A3}, \quad (3)$$

where the AAC ratio is determined by the 3 consecutive numbers of values (e.g., top 3 CJAL scores in descending order denoted by A1, A2, and A3 in Eqs. 2 and 3). The ACC ranged from 0 to 1.0, representing the strength of dominance for the top member when compared to the next 2 members. Through the computation of AAC, the dominance strength in a variable (i.e., country, journal, category, or CIDA) can be measured and judged by the effect size, with criteria of <0.5, between 0.5 and 0.7, and not less than 0.7 as the small, medium, and large effect sizes, respectively.[14]

2.2.2. Diagnostic analytics in T100NPC. Three visual representations were used to display the relationship between variables using diagnostic analytics, including a network chart,[37,38] a chord diagram,[39,40] and a Sankey diagram.[14,41,42]

Coword analysis[37,38] was performed to extract the chief keywords in clusters as themes (or leaders) represented by Keywords Plus that were retrieved from the WoSCC. The themes denoted by the chief keywords in clusters were assigned to T100NPC and countries using Equation 4.[40,43]

$$Theme\# = At\left[\max_{0\le x\le 1}\sum_{i=1}^{\bar{L}}\sum_{j=1}^{n}\left(m = \underset{k\in F}{m} + 1\right)\right], \quad (4)$$

where L is the number of keywords in article i. n is the number of keywords denoted by keyword k belonging to the theme defined in coword analysis.[37,38] Accordingly, theme# is redirected to the maximal number of keywords (=m) involved in theme# via equation 4. The top five keywords in each cluster were displayed on a network chart.[37,38]

In the country-based author collaboration networks, the cluster names can be assigned using equation 5 based on the themes in articles.

$$Theme_{rj} = \max_{r\,in\,j}\left(\sum_{n=1}^{N}\sum_{l=1,a\in D,a\in r}^{L}\sum_{j=1,t\in c}^{J} term_{rj}\,(count < -count + 1/L)\right), \quad (5)$$

where L is the number of terms (e.g., country names) in an article. A contingent table with clusters in row (r) and themes in column (j) was built to record the summed counts. The term was matched with the cluster number corresponding to the theme defined in an article (e.g.,$a \in D$ means the article belongs to a theme) via equation 4. The total weighted scores were summed and selected according to the maximum likelihood in equation 5.

Themes were visually assigned to T100NPC, and countries were displayed on chord diagrams.[39,40]

2.2.3. Predictive analytics in T100NPC. Two visualizations were applied to this section, including a forest plot[44] used for comparing the difference in proportional counts for themes between the 2 early and recent stages (i.e., from 2013 to 2017 and 2018 to 2022), and a scatter plot[14,40] described below:

The number of connections for a specific Keyword plus was computed based on an equal weight in an article.[14,40] The weighted mean citation for each term was applied to predict article citations based on the correlation coefficient (CC) in determining the predictive power related to original article citations. The CC t value was calculated using the following formula $\left(=\mathrm{CC}\times\sqrt{\frac{n-2}{1-\mathrm{CC}\times\mathrm{CC}}}\right)$ [14,45–47] A prediction equation was produced through the simple regression analysis using MedCalc statistical software, version 9.5.0.0 (MedCalc, NY). A scatter plot[14,40] was used to display the relationship between weighted mean citations and article citations in T100NPC.

2.2.4. Prescriptive analytics in T100NPC. The THM[48] contains 4 dimensions, namely, articles on the row, years on the column, colors darked by article citation, and the burst points start at the beginning of red font and ends at citations smaller than the count at the burst point.[49–51] With the THM, the most worth-reading articles with higher growing citation trends in T100NPC were highlighted by the respective burst strength.[49–51] The trend signals include 5 possible outcomes (e.g., ready to rise, increasing, slow down, declining, and stationarity). The definitions of these terms have been described in the previous study.[52]

Two types of articles on T100NPC were displayed on an impact beam plot (IBP),[53] including the top 20 articles with an increasing trend over the last 4 years and 80 others in T100NPC articles. All those articles could be linked to PubMed once the dot representing an article in IBP was selected and clicked.

2.3. Creating dashboards on Google Maps

All graphs were drawn by author-made modules in Excel (Microsoft Corp at https://office.microsoft.com/excel). We created pages of HTML used for Google Maps. All graphs displayed with a dashboard type can be presented on Google Maps. The way to draw the visualization involved in this study is deposited in Supplemental Digital Content S2, http://links.lww.com/MD/I420.

3. Results

3.1. Descriptive analytics in T100NPC

China, Sun Yat-Sen University, the Radiat Oncol department, and author Jun Ma are the most prolific and influential CIDAs. The CJAL score indicates that only Sun Yat-Sen University has a medium effect of dominance (0.62); see Figure 1. Despite having the greatest number of publications (36%), China lags slightly behind the US in terms of CJAL scores.

Lancet Oncol and *J Clin Oncol* contributed the most to the number of T100NPC articles with higher mean citations (=13.6 and 13.5, respectively), as shown in Figure 2.

3.2. Diagnostic analytics in T100NPC

In T100NPC, 6 themes were identified using coword analysis, as shown in Figure 3, including Epstein–Barr virus (EBV), gene expression, whole genome, RNA-guided endonuclease, serum, and chronic lymphocytic leukemia as chief keywords in each cluster. Figure 4 shows a chord diagram[39,40] to present the relationship between themes and articles (in top panel) and counties (in bottom panel). The chord diagram enables us to understand the relationship between themes (or countries) through color-coded curves.

The Sankey diagram in Figure 5 is another useful visual representation. The majority of the publications were published in 2016 (22%), EBV (60%) under themes, oncology (47%) under

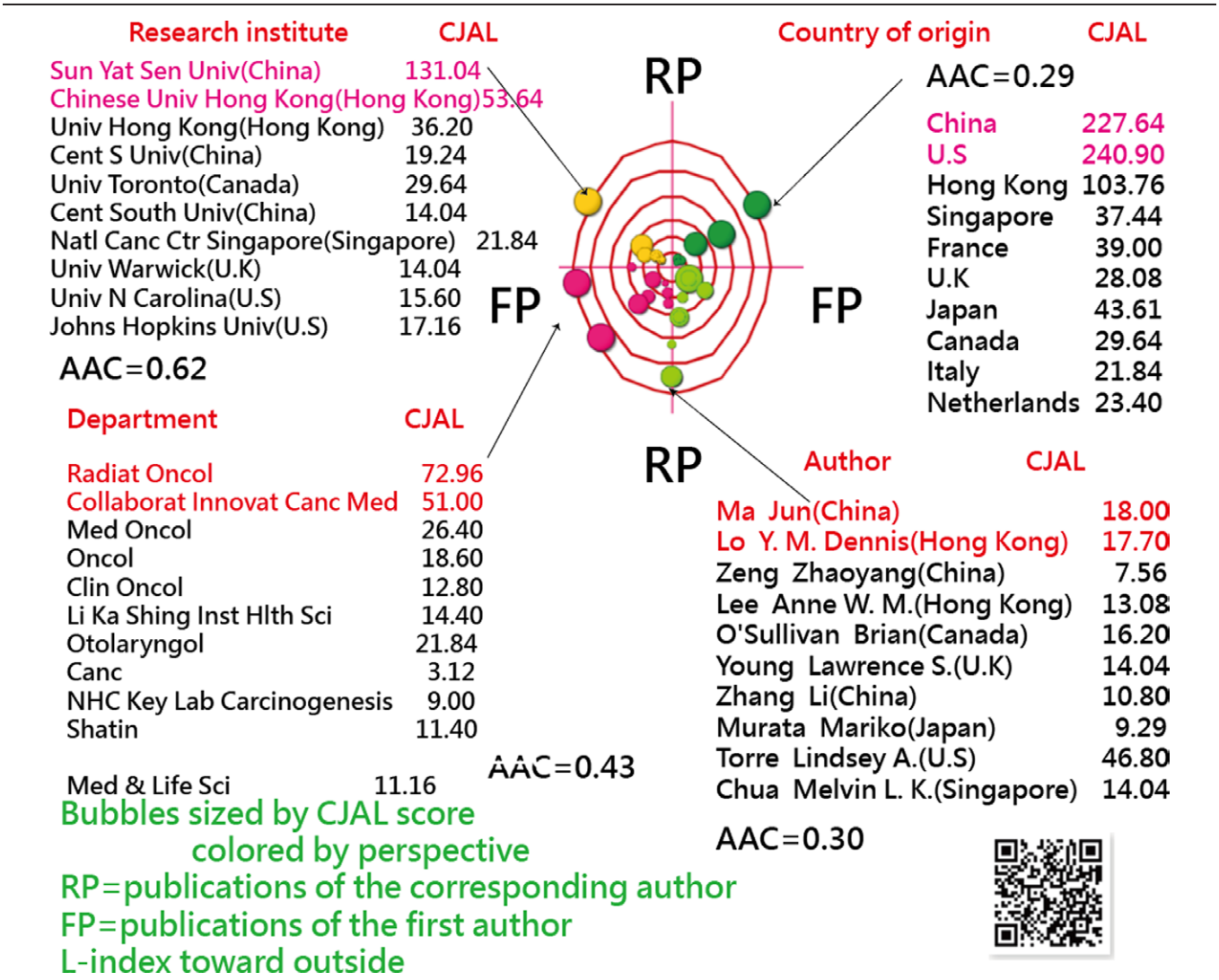

Figure 1. Comparison of RAs on CJAL scores for CIDA entities using the 4-quadrant radar plot. CIDA = country, institute, department, and author, CJAL = category, journal, authorship, and L-index, RAs = research achievements.

subject categories, the journal of *Lancet Oncology* (6%), China (36%), and Sun Yat-Sen University (17%) in China. There is a closer relationship between the 2 adjacent members of the article entity when the curve is wider. A larger block indicates that there are more publications in T100NPC.

3.3. Predictive analytics in T100NPC

Using the forest plot shown in Figure 6, there was no difference in proportional counts across 6 themes between the early and recent states.

To utilize the keyword weights to predict article citations, according to our results, keywords plus in WoS were evident in prediction power on the number of article citations (CC = 0.97, $t = 42.32$), as shown in Figure 7. The regression equation is defined as article citation (y) = −461.3929 + 2.0835 × weight (x) of the keyword plus. The slope coefficient presented statistical significance ($F = 1291.17$, $P < .0001$). As soon as the QR code in Figure 7 is scanned, the article appears instantly on PubMed when the bobble of interest is clicked.

3.4. Prescriptive analytics in T100NPC

Figure 8 illustrates the top 20 highly cited articles in T100NPC worth reading. Six articles with significantly increasing citations over the last 4 years were recommended. Based on the THM (Fig. 8), 3 articles[54–56] will be abstracted in the Discussion section due to receiving increasing citations since 2019.

One of the most highly cited articles[57] with 15,768 citations was published in 2015 and was not included in any article worth reading of this study. The reason is that the citation trend of {3165,1570,1160,762] has decreased over the last 4 years.

The IBP shown in Figure 9, which includes 20 of the most highly cited articles in the increasing citation trend, indicates that no difference in mean citation was found between these 2 groups (e.g., 20 articles worth reading vs 80 others). The article appears on PubMed immediately after clicking the dot of interest.

3.5. Online dashboards shown on Google Maps

All the QR codes in Figures are linked to the dashboards if the QR code is scanned. Readers are suggested to examine the details about article information laid on Google Maps.

4. Discussion

We observed that despite having the highest number of publications (36%), China lags slightly behind the US in CJAL scores.

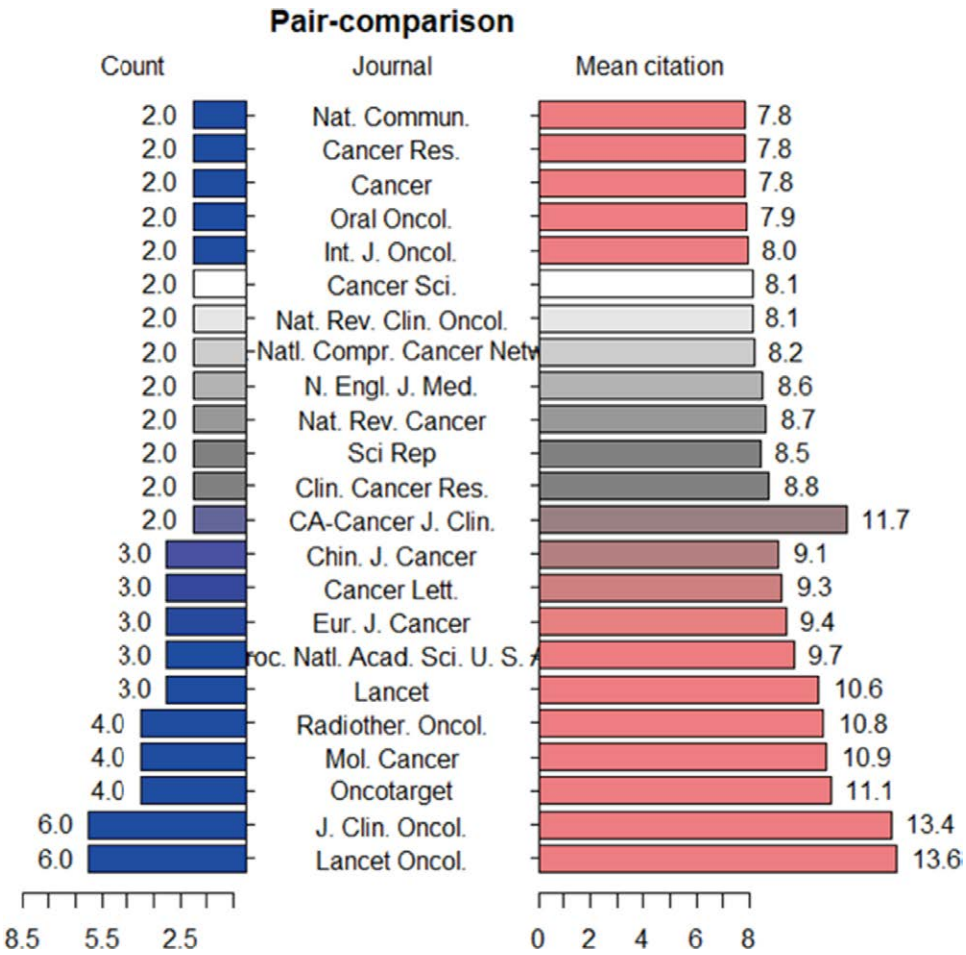

Figure 2. Comparison of productive journals in T100NPC by count and mean citation. T100NPC = 100 top-cited articles and reviews related to nasopharyngeal carcinomas.

CJAL was higher at Sun Yat-Sen University, Radiat Oncol department, and author Jun Ma from China. The number of article citations was significantly correlated with the number of weighted keywords ($F = 1791.17$; $P < .0001$). Six articles with significantly increasing citations over the last 4 years were recommended.

Accordingly, the 2 hypotheses that the DDPP model could be used to analyze bibliometric data and the THM would be useful and viable for selecting articles that merit additional attention are confirmed in this study.

4.1. Additional information

We applied the DDPP model to analyze bibliometric data, which had not yet been illustrated and discussed in the literature. Since the 4 facets of the DDPP model are comprehensive and essential for readers to better understand article characteristics, it is obvious and definite that the model is important.

A bibliographic study should begin with a description of the studies under investigation (i.e., a map of the field is essential for any bibliographic study),[58] as illustrated in Figures 1 and 2. The map of the field is a starting point to dig deeper into bibliographic data using bibliometric methods (e.g., diagnostic, predictive, and prescriptive analytics) because the knowledge in the field has not been structured merely by descriptive analysis and does not lead to a discussion of what we know and where we are and does not provide an agenda for further research.[58] Bibliographic studies must therefore go beyond a simple summary of prior literature (e.g., descriptive analytics only). It requires the authors to interpret and discuss the development and state of the field and to suggest meaningful directions for further research, as we did in this study by proposing the DDPP model.

Second, a THM can be used to select articles that are worthy of reading based on their increasing citations over the last 4 years. CiteSpace[23] only shows keyword citations using timeline graphs, but the disadvantages include less information about a series of data and growth trends on the temporal bar graph[59]; see the timeline graph was produced by CiteSpace[23] in a previous study.[60]

In both more and less developed countries, NPC incidence is approximately 2 to 3 times higher in males than in females. An estimated 92% of new cases of NPC occur in economically less developed countries as a result of geographical disparities in resource availability. Southeastern Asia has the highest incidence rate (at least twice that of any other region). In this region, NPC is the sixth most common type of cancer among men. Chinese and Malay populations have high incidence rates of NPC.[57,61] Other populations with relatively high rates of NPC include the Inuits of Alaska, Greenland, and North Canada, as well as the Chinese and Filipinos living in the US.[62]

NPC appears to be caused by a combination of viral, environmental, and genetic factors.[63,64] Incidence rates are intermediate between migrants from high-risk and low-risk countries,[65] suggesting that both environmental and genetic factors may play a

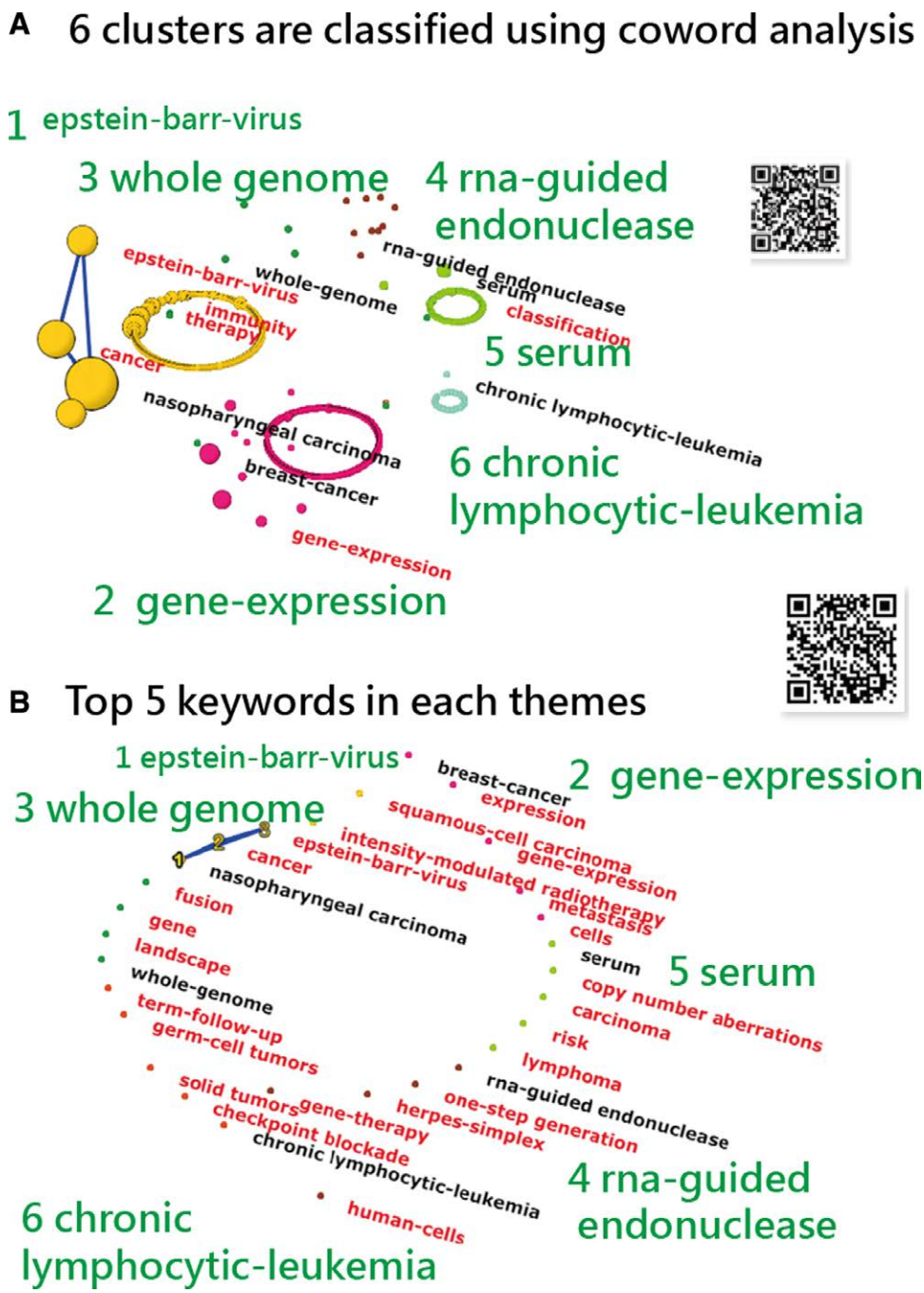

Figure 3. Coword analysis of keywords to classify articles on network charts.

role, as well as possible interactions with EBV. The EBV virus is considered to be an important step in the progression of NPC. However, only a small proportion of those infected with the virus develop the disease. As a result, infection with the virus is considered both lifelong and ubiquitous in most parts of the world.[66]

4.2. Three worth-reading articles with a higher growth rate of citations in T100NPC

A study written by Chen et al and published in 2019 was the most noteworthy article.[67] The article received 940 citations. The authors addressed that NPC incidence has declined gradually but progressively in the past decade, with improved survival due to lifestyle and environmental changes, enhanced understanding of the pathogenesis and risk factors, population screening, advancements in imaging techniques, and individualized comprehensive chemoradiotherapy strategies.

The second article[68] written by Sun et al and published in 2016 was another noteworthy article with 523 citations. An open-label, phase 3, multicenter, randomized controlled trial at ten institutions in China was performed to determine the efficacy of induction chemotherapy plus concurrent chemoradiotherapy in treating stage III to IVB NPC. Three-year survival results and acute toxic effects were followed up. After a median follow-up of 45 months, patients who received induction chemotherapy plus

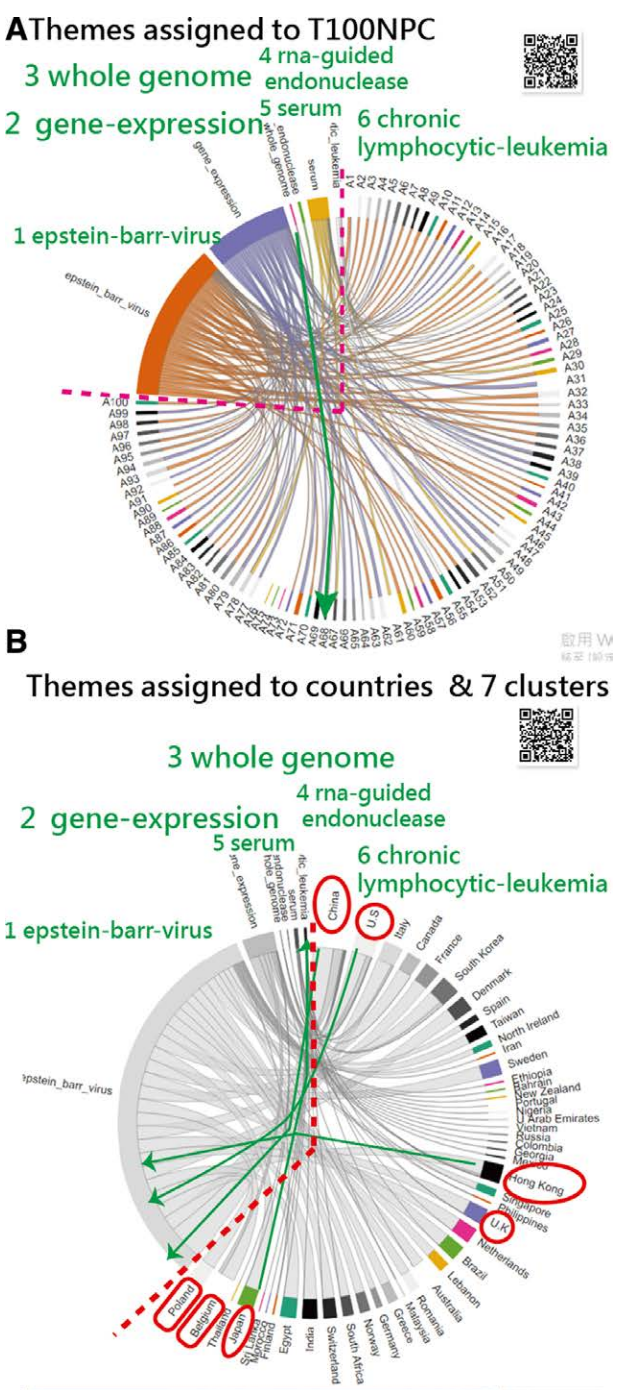

Figure 4. Themes assigned to T100NPC and countries using chord diagrams. T100NPC = 100 top-cited articles and reviews related to nasopharyngeal carcinomas.

concurrent chemoradiotherapy had a 3-year failure-free survival rate of 80%, compared to 72% in the concurrent chemoradiotherapy alone group.

The third article[69] written by Jiang et al and published in 2019 was also worth reading, with 471 citations. Tumor immune escape is an important strategy of tumor survival. The programmed death ligand-1 (PD-L1)/programmed death-1 signaling pathway is an important component of tumor immunosuppression, and the induction of PD-L1 by inflammatory factors in the tumor microenvironment may affect the therapeutic efficiency of PD-L1/PD-1 blockade.

4.3. Implications and changes

Top keywords with the strongest frequency bursts are frequently displayed on temporal bar graphs[49,50,70] and timeline bar graphs in CiteSpace[22,23] (e.g., notes described by authors on the timeline bar graph that the red bars represent frequently cited spots and the green bars represent infrequently cited periods[71] in CiteSpace[22,23]). Of these, they do not contain such features as values displayed on the heatmap and citation trend over the past years as we did on the THM in Figure 8: a strong frequency burst indicates that a variable has undergone a great change in a short period of time with the red fonts, implying the durations of the bursts in citations.

An article[57] with 15,768 citations was published in 2015 and was not included in any article worth reading of this study because the citation trend of {3165,1570,1160,762] has decreased over the last 4 years. If the citation trend was observed using the Newton–Raphson iteration method[72–74] and the link,[75] the decreasing citation trend appears. It is thus necessary to present articles worth reading for readers using the THM of prescriptive analytics in the future.

PubMed contains over 410 publications with the titles of 100 top-cited articles.[76] Most of the articles included descriptive and diagnostic analytics without using predictive or prescriptive analytics. As a potential future application of bibliometrics, the DDPP model is promising.

Moreover, CiteSpace[22,23] was used to group and label article themes. There are, however, no such methods for assigning themes to articles using Equations 4 and 5.[40,43] Human classification of themes assigned to articles[77] is a laborious and time-consuming process.

4.4. Limitations and suggestions

A number of issues must be addressed in detail in further research. As a first concern, only T100NPC articles were included. The CJAL score should be used for future studies on NPC rather than being limited to T100NPC alone for analyzing their RAs.

The second point is that although the Y-index[35,36] and the CJAL score[28] have been considered to be fair measures of RAs for CIDA, it is assumed that the co-first authors contribute equally to the articles. If authorship does not follow the rule as designed using the author weighted scheme,[78–81] the results regarding the RAs for CIDA will be biased.

Third, it takes some time to draw the THM. The advancement in hardware has made this task trivial with a dedicated software program, equally easy and quick, as demonstrated in Supplemental Digital Content S2, http://links.lww.com/MD/I420.

Fourth, the DDPP model was proposed for future bibliometrics, but its usefulness must be validated further in future bibliometrics.

Fifth, according to Figure 1, only regions with higher CJAL scores are compared. Readers may also be interested in the list of productive regions with the Y-index[35,36] shown on the radar plot. Using the 4-quadrant radar plot[28] to display the productive regions should be involved in a future study (see how to draw the radar plot in Supplemental Digital Content S2, http://links.lww.com/MD/I420).

Finally, although the THM is considered useful and applicable, some basic R expertise is required to draw them effectively in R for use in the future.

5. Conclusion

This bibliometric study utilizes the DDPP model to analyze the scientific progress of NPC over the past decade. The whole genome is a hot topic that may prove to be a promising research area in the future. With our bibliometric analysis and THM, researchers would be able to gain a better understanding of the trends and articles worth reading in the NPC field.

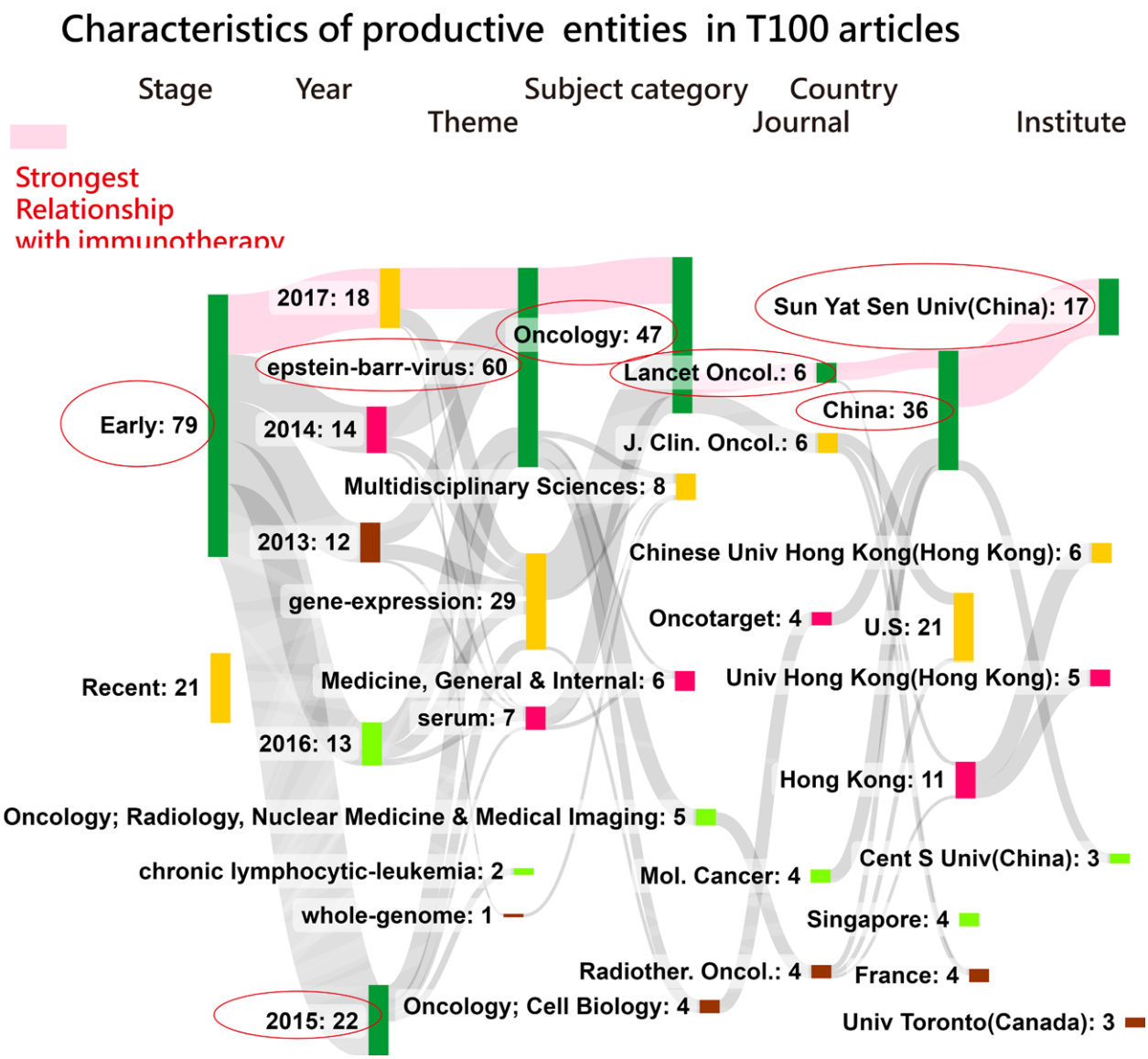

Figure 5. Using the Sankey diagram to understand the relationship between article entities by count.

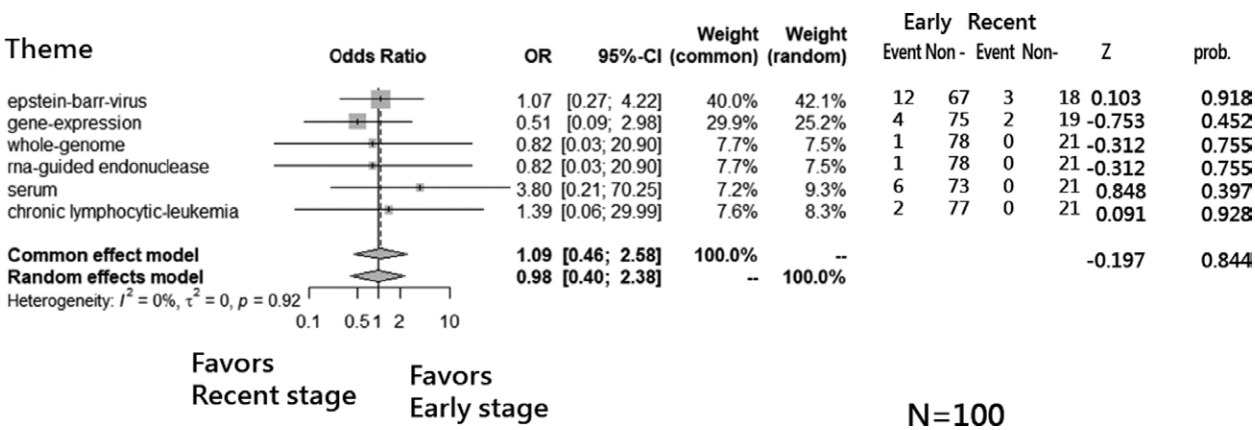

Figure 6. Comparison of difference in proportional counts of themes between the 2 early and recent stages using the forest plot (note. 2013 to 2017 and 2018 to 2022 for the early and recent stages, respectively).

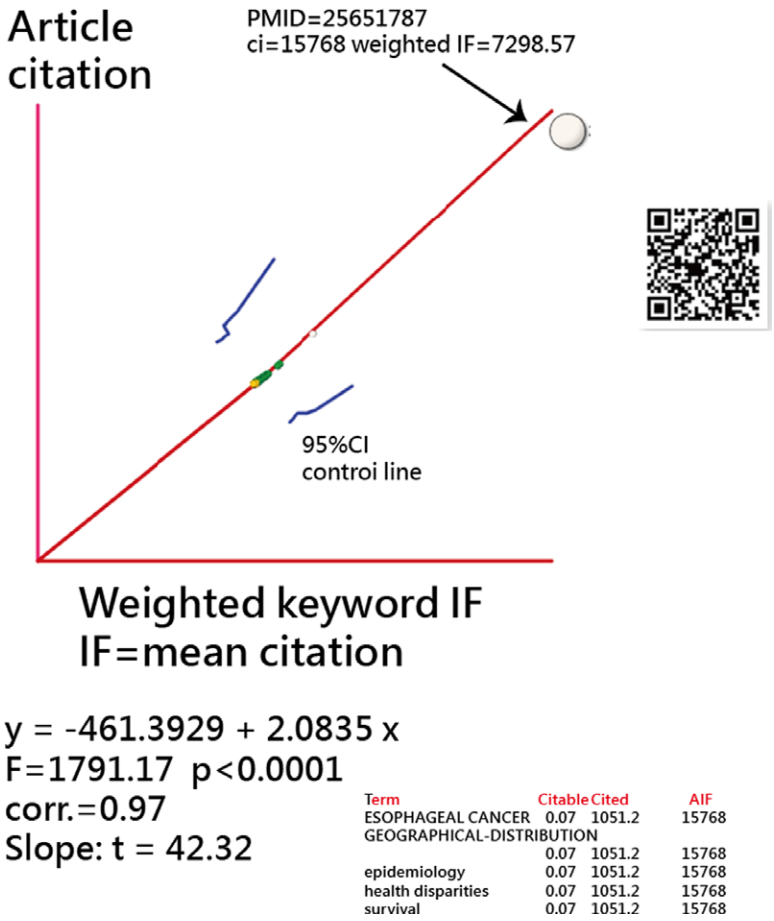

Figure 7. Using the weighted mean citations of keywords to predict article citation using the scatter plot to display.

Temporal heatmap in T20NPC

Author	Citation	Year	2013	2014	2015	2016	2017	2018	2019	2020	2021	2022	Type	BS	Growth	PMID
Torre, Lindsey A.(U.S)	15768	2015	0	0	416	2080	3083	3531	3165	1570	1160	762	4	4.34	-0.93	25651787
Fitzmaurice, Christina(U.S)	1961	2015	0	0	22	194	394	391	332	289	226	113	4	3.55	-0.98	26181261
Chen, Yu-Pei(China)	940	2019	0	0	0	0	0	0	14	188	381	356	1	3.32	0.92	31178151
Chua, Melvin L. K.(Singapore)	849	2016	0	0	1	29	76	125	191	182	161	83	4	2.91	-0.91	26321262
Chen, Benjamin J.(U.S)	585	2013	4	18	55	84	87	69	84	76	73	35	4	2.67	-0.89	23674495
Li, Fang(China)	545	2015	0	0	4	29	90	125	119	78	70	30	4	2.93	-0.97	25749389
Sun, Ying(China)	523	2016	0	0	0	3	48	77	77	107	126	85	1	2.91	0.25	27686945
Blanchard, Pierre(France)	474	2015	0	0	7	35	58	76	78	78	94	48	4	2.69	-0.50	25957714
Jiang, Xianjie(China)	471	2019	0	0	0	0	0	0	28	122	166	154	1	3.1	0.87	30646912
Lee, Anne W. M.(Hong Kong)	459	2015	0	0	1	20	48	66	78	96	83	67	4	2.89	-0.49	26351355
Sever, Richard(U.S)	440	2015	0	0	8	19	10	22	69	88	119	102	1	2.72	0.79	25833940
O'Sullivan, Brian(Canada)	429	2016	0	0	0	9	54	68	87	77	86	48	4	2.61	-0.76	26936027
Tang, Yu-Jin(China)	416	2016	0	0	0	4	54	90	106	106	31	25	4	2.78	-0.91	27703190
Morad, Samy A. F.(U.S)	399	2013	13	47	42	44	33	49	39	52	47	33	4	2.41	-0.35	23235911
Young, Lawrence S.(U.K)	391	2016	0	0	0	0	28	60	57	76	102	68	1	2.76	0.40	27687982
Jiang, Peiyong(Hong Kong)	386	2015	0	0	10	25	39	64	61	72	70	45	4	2.58	-0.52	25646427
Shield, Kevin D.(France)	375	2017	0	0	0	0	11	48	68	103	80	65	4	2.77	-0.24	28076666
Sun, Xueming(China)	372	2014	0	2	21	32	51	55	49	56	67	39	4	2.59	-0.21	24231245
Zhang, Yuan(China)	357	2019	0	0	0	0	0	0	12	100	136	109	1	2.99	0.79	31150573
Chan, K. C. Allen(Hong Kong)	350	2017	0	0	0	0	5	52	76	76	78	63	4	2.77	-0.69	28792880

: declining trend; BS: burst strength; Burst spot: red font; Burst point: at the beginning of numbers in red font; Number: citations observed over the years; Growth: correlation coefficient between citations and the corresponing series of years over the last 4 years: PMID: PubMedunique identifier: Author: 1st author of the article; : increasing trend

Figure 8. Top 20 highly cited articles shown on a temporal heatmap to select articles worth reading.

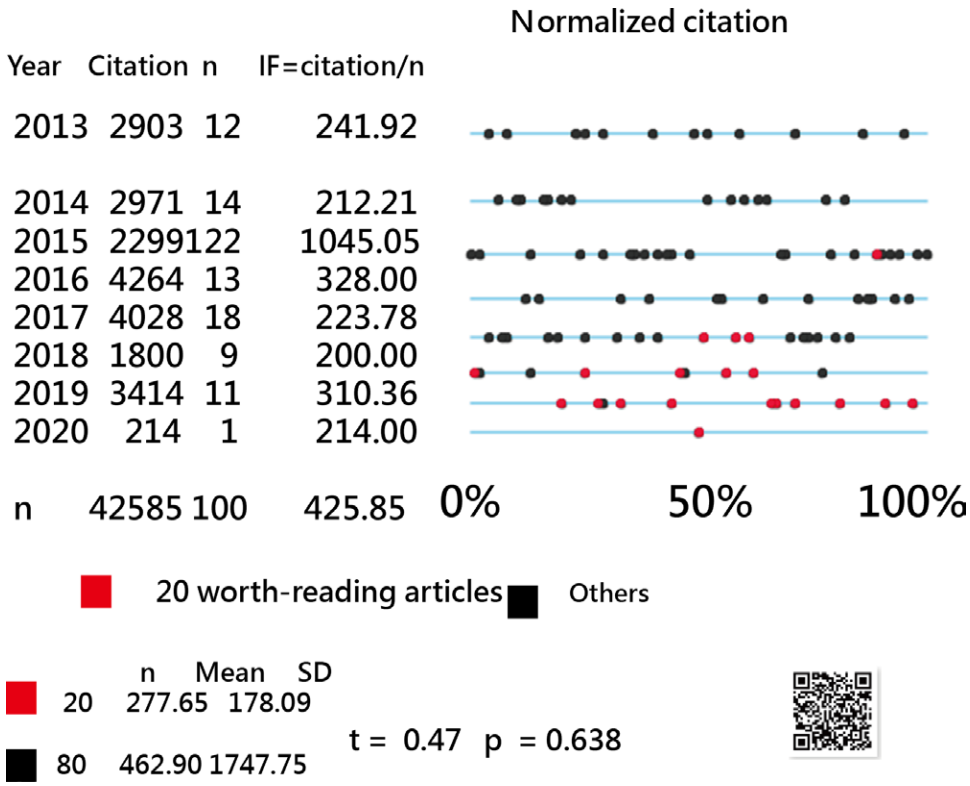

Figure 9. T100NPC on the IBP (note. top 20 articles with higher increasing citation trend in red dot and other 20 articles in bkaje dot). IBP = impact beam plot, T100NPC = 100 top-cited articles and reviews related to nasopharyngeal carcinomas.

Acknowledgments

We thank Enago (www.enago.tw) for the English language review of this manuscript.

Author contributions

Conceptualization: Chung-Chia Hung.
Data curation: Julie Chi Chow.
Formal analysis: Mei-Yu Tu.
Investigation: Cheng-Yao Lin, Willy Chou.
Methodology: Tsair-Wei Chien.

References

[1] Chen Y-P, Chan ATC, Le Q-T, et al. Nasopharyngeal carcinoma. Lancet. 2019;394:64–80.
[2] Bray F, Ferlay J, Soerjomataram I, et al. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA Cancer J Clin. 2018;68:394–424.
[3] Chen W, Zheng R, Baade PD, et al. Cancer statistics in China, 2015. CA Cancer J Clin. 2016;66:115–32.
[4] Fang L, Shi L, Wang W, et al. Identifying key genes and small molecule compounds for nasopharyngeal carcinoma by various bioinformatic analysis. Medicine (Baltim). 2021;100:e27257.
[5] Khetan P, Boffetta P, Luce D, et al. Occupations and the risk of head and neck cancer: a pooled analysis of the international head and neck cancer epidemiology (INHANCE) consortium. J Occup Environ Med.2019;61:397–404.
[6] Chua MLK, Wee JTS, Hui EP, et al. Nasopharyngeal carcinoma. Lancet. 2016;387:1012–24.
[7] Wu Q, Yuan T, Zhang Z, et al. The top 100 most impactful articles and recent trends in nasopharyngeal carcinoma from 1970 to 2018: a bibliometric analysis. J Int Med Res. 2020;48: 300060519896149.
[8] Over 1773 articles regarding NPC were published in PubMed. Available at: https://pubmed.ncbi.nlm.nih.gov/?term=Nasopharyngeal+carcinomas%5BMeSH+Major+Topic%5D. [Access date December 2, 2022].
[9] Agarwal A, Durairajanayagam D, Tatagari S, et al. Bibliometrics: tracking research impact by selecting the appropriate metrics. Asian J Androl. 2016;18:296–309.
[10] Moed HF. New developments in the use of citation analysis in research evaluation. Arch Immunol Ther Exp (Warsz). 2009;57:13–8.
[11] Tan KK, Chien TW, Kan WC, et al. Research features between Urology and Nephrology authors in articles regarding UTI related to CKD, HD, PD, and renal transplantation. Medicine (Baltim). 2022;101:e31052.
[12] Lee YL, Chien TW, Wang JC. Using Sankey diagrams to explore the trend of article citations in the field of bladder cancer: research achievements in China higher than those in the United States. Medicine (Baltim). 2022;101:e30217.
[13] Wang CY, Chien TW, Chou W, et al. Predicting the number of citations of polycystic kidney disease with 100 top-cited articles since 2010: Bibliometric analysis. Medicine (Baltim). 2022;101:e30632.
[14] Yang TY, Chien TW, Lai FJ. Citation analysis of the 100 top-cited articles on the topic of hidradenitis suppurativa since 2013 using Sankey diagrams: Bibliometric analysis. Medicine (Baltim). 2022;101:e31144.
[15] Yeh CH, Chien TW, Lin JJ, et al. Comparing the similarity and differences in MeSH terms associated with spine-specific journals using the forest plot: a bibliometric analysis. Medicine (Baltim). 2022;101:e31441.
[16] Kavanagh RG, Kelly JC, Kelly PM, et al. The 100 classic papers of pediatric orthopedic surgery: a bibliometric analysis. J Bone Joint Surg Am. 2013;95:e134.
[17] Badhiwala JH, Nassiri F, Witiw CD, et al. Highly cited works in spinal disorders: the top 100 most cited papers published in spine journals. Spine (Phila Pa 1976). 2018;43:1746–55.
[18] Heng Wong MY, Tan NYQ, Sabanayagam C. Time trends, disease patterns and gender imbalance in the top 100 most cited articles in ophthalmology. Br J Ophthalmol. 2019;103:18–25.
[19] Coelho DH, Edelmayer LW, Fenton JE. A century of citation classics in otolaryngology-head and neck surgery journals revisited. Laryngoscope. 2014;124:1358–62.
[20] Coelho DH, Edelmayer LW, Fenton JE. Citation analysis of otorhinolaryngology journals: follow-up study. J Laryngol Otol. 2015;129:489–93.
[21] Wolvius C. Predict the future, understand the past: the four types of data analysis. Available at: https://datajourney.akvo.org/blog/the-four-types-of-data-analysis. [Access date December 3,2022].

[22] Chen C. Searching for intellectual turning points: progressive knowledge domain visualization. Proc Natl Acad Sci United States America (PNAS). 2004;101(suppl. 1):5303–10.
[23] Chen C. CiteSpace II: detecting and visualizing emerging trends and transient patterns in scientific literature. J Am Soc Inf Sci Technol. 2006;57:359–77.
[24] van Eck NJ, Waltman L. Software survey: Vosviewer, a computer program for bibliometric mapping. Scientometrics. 2010;84:523–38.
[25] Callaham M, Wears RL, Weber E. Journal prestige, publication bias, and other characteristics associated with citation of published studies in peer-reviewed journals. JAMA. 2002;287:2847–50.
[26] Birkle C, Pendlebury DA, Schnell J, et al. Web of Science as a data source for research on scientific and scholarly activity. Quant Sci Stud. 2020;1:363–76.
[27] Chien TW. T100NPC listed in this study. Available at: http://www.healthup.org.tw/html100/ent202211.htm. [Access date December 2, 2022].
[28] Shao Y, Chien TW, Jang FL. The use of radar plots with the Yk-index to identify which authors contributed the most to the journal of Medicine in 2020 and 2021: a bibliometric analysis. Medicine (Baltim). 2022;101:e31033.
[29] Hsieh WT, Chien TW, Chou W. The 100 most cited articles have fewer citations than other bibliometric articles: a pairwise comparison using a temporal bubble graph. Medicine (Baltim). 2022;101:e32101.
[30] Yang DH, Chien TW, Yeh YT, et al. Using the absolute advantage coefficient (AAC) to measure the strength of damage hit by COVID-19 in India on a growth-share matrix. Eur J Med Res. 2021;26:61.
[31] Wang LY, Chien TW, Chou W. Using the IPcase Index with inflection points and the corresponding case numbers to identify the impact hit by COVID-19 in China: an observation study. Int J Environ Res Public Health. 2021;18:1994.
[32] Kuo SC, Chien TW, Chou W. Questions to the article: demonstrating the ascendancy of COVID-19 research using acronyms. Scientometrics. 2021;126:8761–4.
[33] Yeh JT, Shulruf B, Lee HC, et al. Faculty appointment and promotion in Taiwan's medical schools, a systematic analysis. BMC Med Educ. 2022;22:356.
[34] Belikov AV, Belikov VV. A citation-based, author- and age-normalized, logarithmic index for evaluation of individual researchers independently of publication counts. F1000Research. 2015;4:884.
[35] Ho YS. Bibliometric analysis of adsorption technology in environmental science. J Environ Prot Sci. 2007;1:1–11.
[36] Ho YS, Satoh H, Lin SY. Japanese lung cancer research trends and performance in science citation index. Intern Med. 2010;49:2219–28.
[37] Yie KY, Chien TW, Yeh YT, et al. Using social network analysis to identify spatiotemporal spread patterns of COVID-19 around the World: online dashboard development. Int J Environ Res Public Health. 2021;18:2461.
[38] Kan WC, Chou W, Chien TW, et al. The most-cited authors who published papers in JMIR mHealth and uHealth using the authorship-weighted scheme: bibliometric analysis. JMIR Mhealth Uhealth. 2020;8:e11567.
[39] Finnegan A, Sao SS, Huchko MJ. Using a chord diagram to visualize dynamics in contraceptive use: bringing data into practice. Glob Health Sci Pract. 2019;7:598–605.
[40] Huang YP, Pao JL, Chien TW, et al. Thematic analysis of articles on artificial intelligence with spine trauma, vertebral metastasis, and osteoporosis using chord diagrams: a systematic review and meta-analysis. Medicine (Baltim). 2022;101:e32369.
[41] Li MJ, Chien TW, Liao KW, et al. Using the Sankey diagram to visualize article features on the topics of whole-exome sequencing (WES) and whole-genome sequencing (WGS) since 2012: bibliometric analysis. Medicine (Baltim). 2022;101:e30682.
[42] Lee YL, Chien TW, Wang JC. Using Sankey diagrams to explore the trend of article citations in the field of bladder cancer: research achievements in China higher than those in the United States. Medicine (Baltim). 2022;101:e30217.
[43] Chou PH, Lin JCJ, Chien TW. Text mining and forest plots were used to identify similarities and differences between two spine-related journals based on medical subject headings (MeSH terms) and author-specified keywords in 100 top-cited articles. Scientometrics. 2023;128:1–17.
[44] Yeh CH, Chien TW, Lin JJ, et al. Comparing the similarity and differences in MeSH terms associated with spine-specific journals using the forest plot: a bibliometric analysis. Medicine (Baltim). 2022;101:e31441.
[45] Wu JW, Yan YH, Chien TW, et al. Trend and prediction of citations on the topic of neuromuscular junctions in 100 top-cited articles since 2001 using a temporal bar graph: a bibliometric analysis. Medicine (Baltim). 2022;101:e30674.
[46] Wang CY, Chien TW, Chou W, et al. Predicting the number of citations of polycystic kidney disease with 100 top-cited articles since 2010: Bibliometric analysis. Medicine (Baltim). 2022;101:e30632.
[47] Chen CH, Chien TW, Yu-Chieh Ho S, et al. Predicting article citations using data from 100 top-cited publications in the field of Psoriasis vulgaris and biological agents (PVBA) since 1991: a bibliometric analysis. Medicine (Baltim). 2022;101:e29396.
[48] Cho H, Nishimura K, Watanabe K, et al. Effective pseudolabeling based on heatmap for unsupervised domain adaptation in cell detection. Med Image Anal. 2022;79:102436.
[49] Chow JC, Chien TW, Chou W. Suggestions to the article: demonstrating the ascendancy of COVID-19 research using acronyms. Scientometrics. 2022;127:2897–9.
[50] Ho SY, Chien TW, Shao Y, et al. Visualizing the features of inflection point shown on a temporal bar graph using the data of COVID-19 pandemic. Medicine (Baltim). 2022;101:e28749.
[51] Chien TW. How to determine the inflection point on a given ogive curve. Available at: http://www.healthup.org.tw/studioclassroom.asp. [Access date November 24, 2022].
[52] Yang DH, Chien TW, Yeh YT, et al. Using the absolute advantage coefficient (AAC) to measure the strength of damage hit by COVID-19 in India on a growth-share matrix. Eur J Med Res. 2021;26:61.
[53] Author impact beam plots in Web of Science Author Records. Available at: https://www.youtube.com/watch?v=dcXgx5wxUp4. [Access date December 3, 2022].
[54] Chen YP, Chan ATC, Le QT, et al. Nasopharyngeal carcinoma. Lancet. 2019;394:64–80.
[55] Sun Y, Li WF, Chen NY, et al. Induction chemotherapy plus concurrent chemoradiotherapy versus concurrent chemoradiotherapy alone in locoregionally advanced nasopharyngeal carcinoma: a phase 3, multicenter, randomized controlled trial. Lancet Oncol. 2016;17:1509–20.
[56] Jiang X, Wang J, Deng X, et al. Role of the tumor microenvironment in PD-L1/PD-1-mediated tumor immune escape. Mol Cancer. 2019;18:10.
[57] Torre LA, Bray F, Siegel RL, et al. Global cancer statistics, 2012. CA Cancer J Clin. 2015;65:87–108.
[58] Block JH, Fisch C. Eight tips and questions for your bibliographic study in business and management research. Manag Rev Q. 2020;70:307–12.
[59] Hsu SY, Chien TW, Yeh YT, et al. Citation trends in ophthalmology articles and keywords in mainland China, Hong Kong, and Taiwan since 2013 using temporal bar graphs (TBGs): bibliometric analysis. Medicine (Baltim). 2022;101:e32392.
[60] Wu L, Wu H, Ou T, et al. Mapping theme trends and recognizing hot spots in viral pneumonia: a bibliometric analysis of global research. Am J Transl Res. 2022;14:2972–87.
[61] Forman D, Bray F, Brewster DH, et al. Cancer Incidence in Five Continents. Vol X. Available at: https://ci5.iarc.fr/Default.aspx. [Access date December 9, 2013].
[62] Yu MC, Yuan JM. Nasopharyngeal cancer. In: D Schottenfeld, J Fraumeni, eds. Cancer Epidemiology and Prevention. 3rd ed. New York: Oxford University Press; 2006: 620–626.
[63] Chang ET, Adami HO. The enigmatic epidemiology of nasopharyngeal carcinoma. Cancer Epidemiol Biomarkers Prev. 2006;15:1765–77.
[64] Klein G. Nasopharyngeal carcinoma (NPC) is an enigmatic tumor. Semin Cancer Biol. 2002;12:415–8.
[65] Buell P. The effect of migration on the risk of nasopharyngeal cancer among Chinese. Cancer Res. 1974;34:1189–91.
[66] Henle W, Henle G. The cause of infectious mononucleosis. A review. In: PM Briggs, G De-The, LM Payne, eds. Oncogenesis and Herpes Viruses. Lyon, France: IARC Press; 1972: 269–274.
[67] Chen YP, Chan ATC, Le QT, et al. Nasopharyngeal carcinoma. Lancet. 2019;394:64–80.
[68] Sun Y, Li WF, Chen NY, et al. Induction chemotherapy plus concurrent chemoradiotherapy versus concurrent chemoradiotherapy alone in locoregionally advanced nasopharyngeal carcinoma: a phase 3, multicenter, randomized controlled trial. Lancet Oncol. 2016;17:1509–20.
[69] Jiang X, Wang J, Deng X, et al. Role of the tumor microenvironment in PD-L1/PD-1-mediated tumor immune escape. Mol Cancer. 2019;18:10.
[70] Shen L, Xiong B, Li W, et al. Visualizing collaboration characteristics and topic burst on international mobile health research: bibliometric analysis. JMIR Mhealth Uhealth. 2018;6:e135.
[71] Wang S, Wu K, Zhang Z, et al. Mapping theme trends and recognizing research hot spots in the use of ultrasound in orthopedics: a bibliometric analysis of global research. Am J Transl Res. 2021;13:9892–911.

[72] Ho SY, Chien TW, Shao Y, et al. Visualizing the features of inflection point shown on a temporal bar graph using the data of COVID-19 pandemic. Medicine (Baltim). 2022;101:e28749.
[73] Linacre JM. Percentages with continuous Rasch Models. Rasch Meas Trans. 2001;14:771–4.
[74] Shao Y, Nadkarni S, Niu N, et al. A note on the Newton–Raphson iteration method in the Rasch Model. Rasch Meas Trans. 2022;35:1851–6.
[75] Chien TW. How to determine the inflection point on a given ogive curve. Available at: http://www.healthup.org.tw/studioclassroom.asp. [Access date November 24, 2022].
[76] Over 410 articles with the title of 100 top-cited in PubMed since1999. Available at: https://pubmed.ncbi.nlm.nih.gov/?term=100%5Btitle%5D%20and%20cited%5Btitle%5D&sort=date&timeline=expanded. [Access date December 2, 2022].
[77] Lu Y, Lin HH, et al. Classification and citation analysis of the 100 top-cited articles on adult spinal deformity since 2011: a bibliometric analysis. J Chin Med Assoc. 2022;85:401–8.
[78] Vavryčuk V. Fair ranking of researchers and research teams. PLoS One. 2018;13:e0195509.
[79] Sekercioglu CH. Quantifying coauthor contributions. Science. 2008;322:371.
[80] Chien TW, Wang HY, Lai FJ. Applying an author-weighted scheme to identify the most influential countries in research achievements on skin cancer: observational study. JMIR Dermatol. 2019;2:e11015.
[81] Kan WC, Chou W, Chien TW, et al. The most-cited authors who published papers in JMIR mHealth and uHealth using the authorship-weighted scheme: bibliometric analysis. JMIR Mhealth Uhealth. 2020;8:e11567.