How to publish a bibliometric analysis: Some tips and hands-on using R

Tengku Muhd Hanis Mokhtar
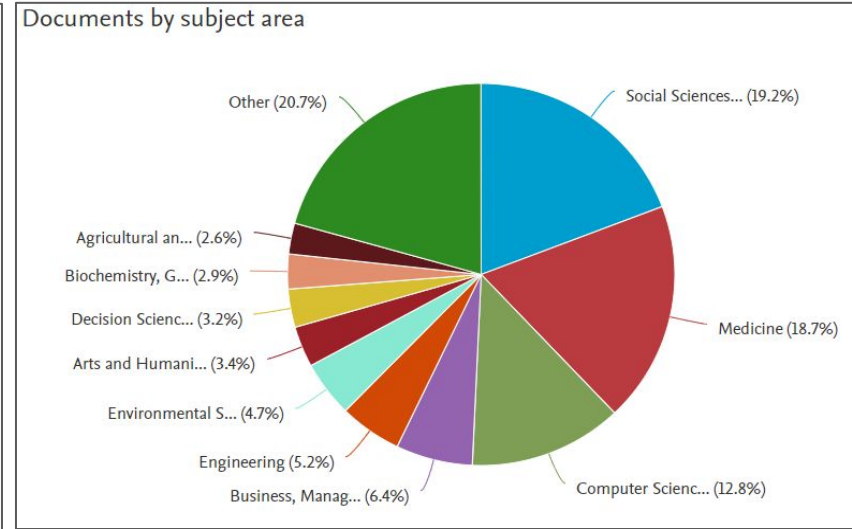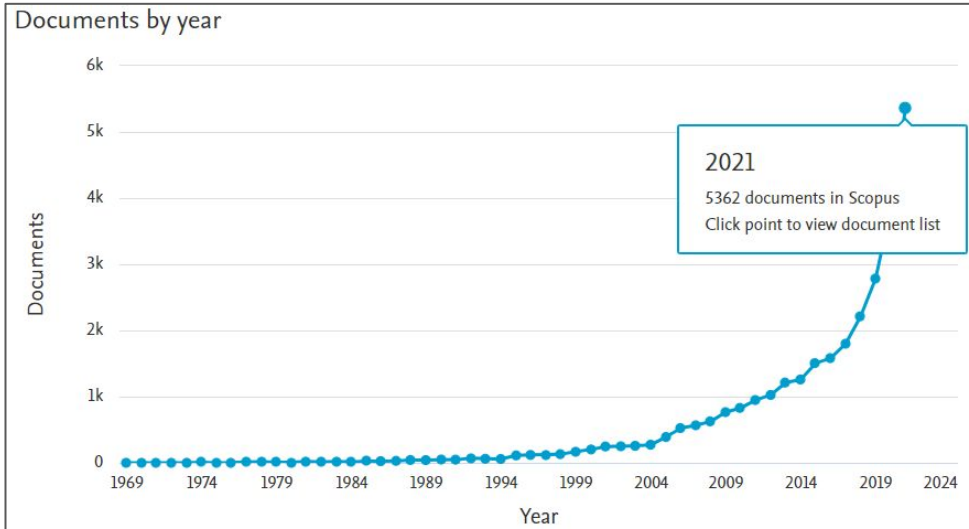Jan 18, 2022

Download material:

https://github.com/tengku-hanis/bibliometrics-Jan18-2022

Bibliometric analysis

● Scopus: 29, 684 bibliometric publication (17-Jan-2022)

Documents by year

2021
5362 documents in Scopus
Click point to view document list

Documents by subject area

Other (20.7%)
Social Sciences... (19.2%)
Agricultural an... (2.6%)
Biochemistry, G... (2.9%)
Decision Scienc... (3.2%)
Arts and Humani... (3.4%)
Environmental S... (4.7%)
Engineering (5.2%)
Business, Manag... (6.4%)
Medicine (18.7%)
Computer Scienc... (12.8%)

What is a bibliometric analysis?

- Analysis of "bibliographic information"
- Quantitatively assess a large research output (> 200 papers)
- Evaluate research progress
- Frequently used in the field of library and information science
- Not a new things (as early as 1896)
- Bibliographic information is depend on which database:
  - Web of science
  - Scopus
  - PubMed
  - etc

Data for bibliometrics

Net survival differences of breast cancer between stages at diagnosis and age groups in the east coast region of West Malaysia: a retrospective cohort study

Tengku Muhammad Hanis [1], Najib Majdi Yaacob,[1] Suhaily Mohd Hairon,[2] Sarimah Abdullah[1]

[1]Unit of Biostatistics and Research Methodology, School of Medical Sciences, Universiti Sains Malaysia - Kampus Kesihatan, Kubang Kerian, Malaysia
[2]Department of Community Medicine, School of Medical Sciences, Universiti Sains Malaysia - Kampus Kesihatan, Kubang Kerian, Malaysia

Correspondence to
Dr Najib Majdi Yaacob;

ABSTRACT
Objective Estimation of the net survival of breast cancer helps in assessing breast cancer burden at a population level. Thus, this study aims to estimate the net survival of breast cancer at different cancer staging and age at diagnosis in the east east region of West Malaysia.
Setting Kelantan, Malaysia.
Participants All breast cancer cases diagnosis in 2007 and 2011 identified from Kelantan Cancer Registry.
Design This retrospective cohort study used a relative survival approach to estimate the net survival of patients with breast cancer. Thus, two data were needed; breast cancer data from Kelantan Cancer Registry and general population mortality data for Kelantan population.
Primary and secondary outcome measures Net survival according to stage and age group at diagnosis at 1, 3 and 5 years following diagnosis.
Results The highest net survival was observed among stage I and II breast cancer cases, while the lowest net survival was observed among stage IV breast cancer cases. In term of age at diagnosis, breast cancer cases aged 65 and older had the best net survival compared with the other age groups.
Conclusion The age at diagnosis had a minimal impact on the net survival compared with the stage at diagnosis. The finding of this study is applicable to other populations with similar breast cancer profile.

Strengths and limitations of this study
► This population-based study used state cancer registry data to describe the net survival of breast cancer in the east coast region of West Malaysia.
► The use of relative survival approach to estimate the net survival is considered as a standard practice for a population-based study.
► This study expanded an abridged life table of the general population mortality data into a complete life table to estimate the net survival due to unavailability of a complete life table.
► The use of different method to expand the mortality data leading to a lack of standardisation in this research area.
► This study lacked socioeconomic, sociodemographic and clinical information which may further explain our finding.

INTRODUCTION
Breast cancer is the most common cancer and the leading cause of cancer-related death among women globally.[1] In the Asia region, the incidence of breast cancer had increased at least moderately in the Eastern and South-eastern region over the last two decades.[2] Several studies have suggested that the increased incidences of breast cancer in Asia are due to factors such as economic growth and adoption of a more westernised lifestyle including no breastfeeding, reduced parity and increased animal fat consumption.[3][4] However, the risk of developing breast cancer in Malaysia is still lower compared with the

8, respectively.[5] In term of breast cancer survival, a 5-year relative survival between 2005 and 2009 in Malaysia was 67.8%, while other Asian countries such South Korea (82.7%), Indonesia (77.7%) and Thailand (71.3%) had a better survival.[6] Breast staging and age are important prognostic factors of breast cancer. Early-stage patients with breast cancer are expected to have better survival compared with those diagnosed at a later stage.[7] Several studies considered age as an important prognostic factor of breast cancer,[8–10] although several Malaysian studies did find it otherwise.[11–13] There is a need to provide more information on breast cancer survival, generally in the Asia Pacific region and especially in Malaysia as it will help in planning effective public health management and control of the disease.

Net survival (NS) is a hypothetical measure in which the only possible cause of death is the disease of interest.[14][15] NS aims to reflect

---

the Kelantan general population mortality data from the DOSM was only available in an abridged life table format. Thus, the mortality data were expanded by the researcher in this study. Other studies may use a different method to expand the mortality data leading to a lack of standardisation in this research area. Admittedly, the expanded life table remained as an approximation to an actual complete life table. However, the use of different life tables in the relative survival analysis had been shown to have a minor change in the overall estimate of the NS rate.[19] Currently, the approach used in our study may serve as an alternative for populations without a complete life table. In facts, previous cancer studies had used a similar approach to estimate an NS statistic.[58–60] Also, this study was restricted by the limited information due to the used of secondary data to further explore the finding of this study. Information such as lymph node involvement, degree of metastasis and hormone receptor status was not available in this study.

CONCLUSIONS
This study presents the NS difference of patients with breast cancer according to cancer staging and age at diagnosis in the east-coast region of West Malaysia. The best survival was observed among patients with breast cancer aged 65 and older and those diagnosed at stages I and II. The age at diagnosis appeared to had a minimal impact on the NS compared with the stage at diagnosis. The NS is multifactorial in nature, thus detailed sociodemographic and clinical information on the breast cancer profile in this region is needed to extend this finding to other neighbouring populations.

Department and Medical Research and Ethics Committee, Ministry of Health Malaysia.

ORCID iD
Tengku Muhammad Hanis http://orcid.org/0000-0003-1466-7765

REFERENCES
1. Ginsburg O, Bray F, Coleman MP, et al. The global burden of women's cancers: a grand challenge in global health. Lancet 2017;389:847–60.
2. Youlden DR, Cramb SM, Yip CH, et al. Incidence and mortality of female breast cancer in the Asia-Pacific region. Cancer Biol Med 2014;11:101–15.
3. Yip CH, Bhoo Pathy N, Teo SH. A review of breast cancer research in Malaysia. Med J Malaysia 2014;69:8–22.
4. Fan L, Goss PE, Strasser-Weippl K. Current status and future projections of breast cancer in Asia. Breast Care 2015;10:372–8.
5. Kirubakaran R, Chee Jia T, Mahamad Aris N. Awareness of breast cancer among surgical patients in a tertiary hospital in Malaysia. Asian Pac J Cancer Prev 2017;18:115–20.
6. Allemani C, Weir HK, Carreira H, et al. Global surveillance of cancer survival 1995-2009: analysis of individual data for 25,676,887 patients from 279 population-based registries in 67 countries (CONCORD-2). Lancet 2015;385:977–1010.
7. Iqbal J, Ginsburg O, Rochon PA, et al. Differences in breast cancer stage at diagnosis and cancer-specific survival by race and ethnicity in the United States. JAMA 2015;313:165–73.
8. Wong FY, Tham WY, Nei WL, et al. Age exerts a continuous effect in the outcomes of Asian breast cancer patients treated with breast-conserving therapy. Cancer Commun 2018;38:39–11.
9. Bharat A, Aft RL, Gao F, et al. Patient and tumor characteristics associated with increased mortality in young women (< or =40 years) with breast cancer. J Surg Oncol 2009;100:248–51.
10. Partridge AH, Hughes ME, Warner ET, et al. Subtype-dependent relationship between young age at diagnosis and breast cancer survival. J Clin Oncol 2016;34:3308–14.
11. Ibrahim NI, Dahlui M, Aina EN, et al. Who are the breast cancer survivors in Malaysia? Asian Pac J Cancer Prev 2012;13:2213–8.
12. Nordin N, Yaacob NM, Abdullah NH, et al. Survival time and prognostic factors for breast cancer amongWomen in north-east Peninsular Malaysia. Asian Pac J Cancer Prev 2018;19:497–502.
13. Taib NA, Akmal M, Mohamed I. Improvement in survival of breast cancer patients - trends over two time periods in a single institution in an Asia Pacific country, Malaysia. Asian Pac J Cancer Prev 2011;12:345–9.
14. Mariotto AB, Noone A-M, Howlader N, et al. Cancer survival: an overview of measures, uses, and interpretations. J Natl Cancer Inst Monogr 2014;2014:145–86.
15. Lambert PC, Dickman PW, Rutherford MJ. Comparison of different approaches to estimating age standardized net survival. BMC Med Res Methodol 2015;15:1–13.
16. UKIACR. Standard operating procedure: guidelines on population based cancer survival analysis, 2016. Available: http://ukiacr.org/publication/guidelines-population-based-cancer-survival-analysis
17. Sarfati D, Blakely T, Pearce N. Measuring cancer survival in populations: relative survival vs cancer-specific survival. Int J Epidemiol 2010;39:598–610.
18. Eloranta S, Adolfsson J, Lambert PC, et al. How can we make cancer survival statistics more useful for patients and clinicians: an illustration using localized prostate cancer in Sweden. Cancer Causes Control 2013;24:505–15.
19. Schaffar R, Rachet B, Belot A, et al. Estimation of net survival for cancer patients: relative survival setting more robust to some assumption violations than cause-specific setting, a sensitivity analysis on empirical data. Eur J Cancer 2017;72:78–83.
20. National Cancer Institute. Malaysian National Cancer Registry Report 2007 - 2011. Putrajaya, 2015. http://nci.moh.gov.my

Type of analysis

- Performance analysis - account for contribution in research area
  - Publication-related metrics (proxy of productivity):
    - Total publication
    - Single-authored publications
    - Co-authored publications
    - Number of active years of publication
    - etc
  - Citation-related metrics (proxy of influence)
    - Total citations
    - Average citations
    - Collaboration index/collaboration coefficient
    - etc

Type of analysis (cont.)

- Science mapping - assess relationship in research area (author, article, etc)
  - Citation analysis - most influential publications
  - Co-citation analysis
    - Relationship among cited publication
    - foundation/basic theme (look in the past)
  - Bibliographic coupling
    - Relationship among publications with same references
    - Latest development
  - Co-word analysis
    - Certain words (keywords, tittle, etc) shared by publications
    - May indicate future trend or support co-citation analysis and bibliographic coupling

Type of analysis (cont.)

- ○ Co-authorship analysis
  - ■ Interaction among researchers
  - ■ Also can be applied to institutions, countries
- Theory-related metrics
  - ○ Lotka's law
  - ○ Bradford's law
- Other measures
  - ○ Thematic map
  - ○ Evolutional thematic map
  - ○ etc

Type of analysis (cont.)
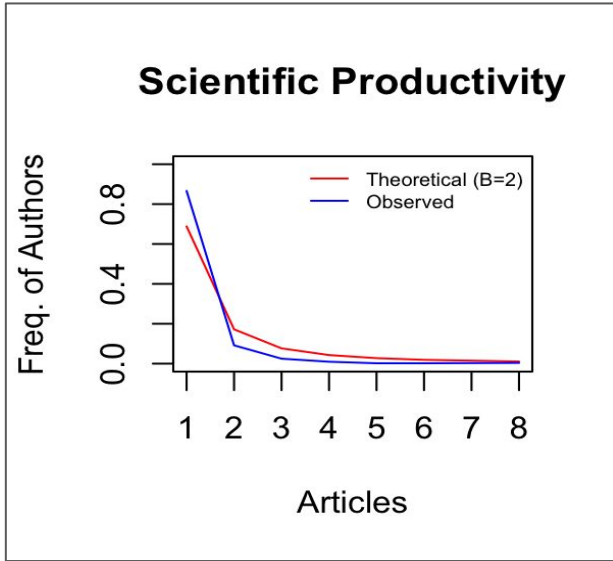
Lotka's law

Bradford's law

Thematic map

Bibliometric analysis

Main techniques

Performance analysis

Publication-related metrics
- Total publications (TP)
- Number of contributing authors (NCA)
- Sole-authored publications (SA)
- Co-authored publications (CA)
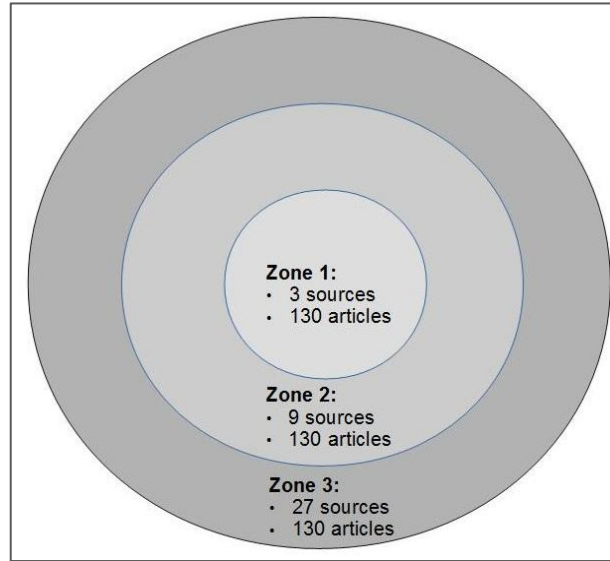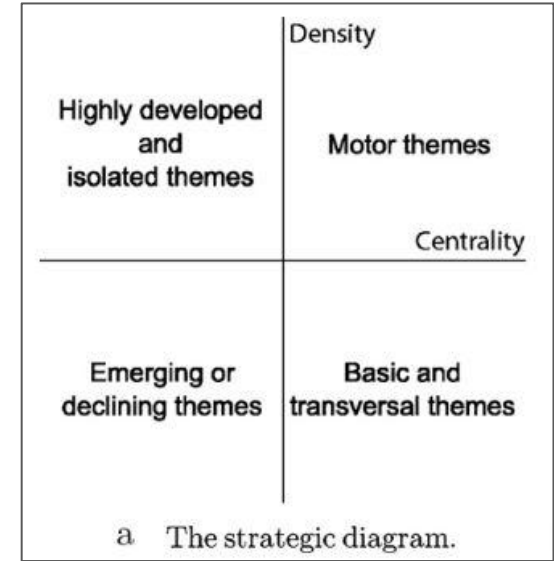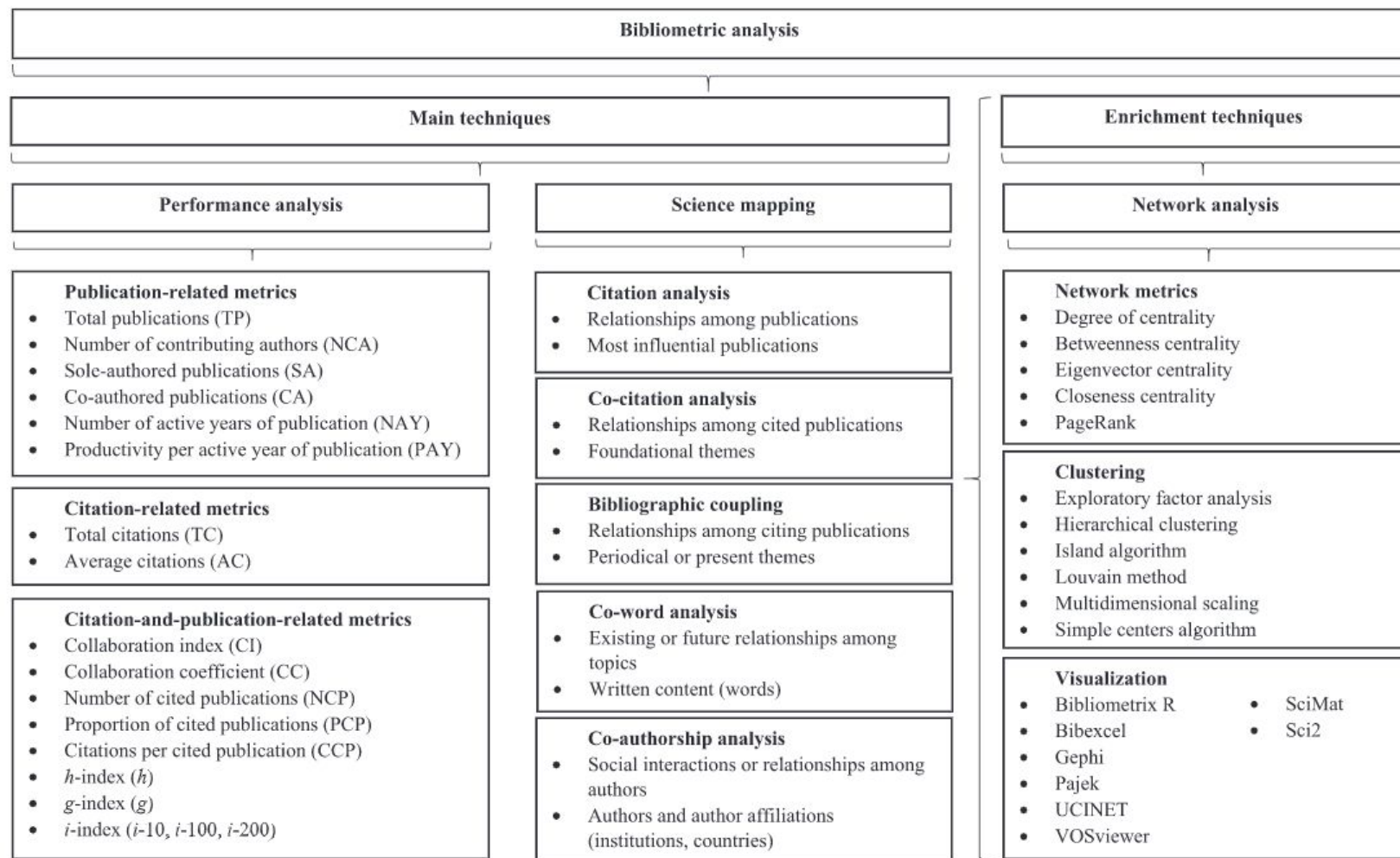- Number of active years of publication (NAY)
- Productivity per active year of publication (PAY)

Citation-related metrics
- Total citations (TC)
- Average citations (AC)

Citation-and-publication-related metrics
- Collaboration index (CI)
- Collaboration coefficient (CC)
- Number of cited publications (NCP)
- Proportion of cited publications (PCP)
- Citations per cited publication (CCP)
- h-index (h)
- g-index (g)
- i-index (i-10, i-100, i-200)

Science mapping

Citation analysis
- Relationships among publications
- Most influential publications

Co-citation analysis
- Relationships among cited publications
- Foundational themes

Bibliographic coupling
- Relationships among citing publications
- Periodical or present themes

Co-word analysis
- Existing or future relationships among topics
- Written content (words)

Co-authorship analysis
- Social interactions or relationships among authors
- Authors and author affiliations (institutions, countries)

Enrichment techniques

Network analysis

Network metrics
- Degree of centrality
- Betweenness centrality
- Eigenvector centrality
- Closeness centrality
- PageRank

Clustering
- Exploratory factor analysis
- Hierarchical clustering
- Island algorithm
- Louvain method
- Multidimensional scaling
- Simple centers algorithm

Visualization
- Bibliometrix R
- Bibexcel
- Gephi
- Pajek
- UCINET
- VOSviewer
- SciMat
- Sci2

N. Donthu et al., 2021

Comparison of major review methods.

Review type	Goal	When to use	When not to use	Scope	Dataset	Analysis
Bibliometric analysis	• Summarizes large quantities of bibliometric data to present the state of the intellectual structure and emerging trends of a research topic or field.	• When the scope of review is broad. • When the dataset is too large for manual review.	• When the scope of review is specific. • When the dataset is small and manageable enough that its content can be manually reviewed.	• Broad	• Large	• Quantitative (evaluation and interpretation) • Qualitative (interpretation only)
Meta-analysis	• Summarizes the empirical evidence of relationship between variables while uncovering relationships not studied in existing studies.	• When the focus of review is to summarize results rather than to engage with content, which may be broad or specific. • When studies in the field are homogenous. • When the number of homogeneous studies available is sufficiently high. • When the number of homogeneous studies remaining after removing low quality studies is sufficiently high.	• When studies in the field are heterogeneous. • When the number of homogenous studies is relatively low. • When the number of high-quality homogeneous studies is relatively low.	• Broad • Specific	• Large • Small but adequate	• Quantitative (evaluation and interpretation)
Systematic literature review	• Summarizes and synthesizes the findings of existing literature on a research topic or field.	• When the scope of review is specific. • When the dataset is small and manageable enough that its content can be manually reviewed.	• When the scope of review is broad. • When the dataset is too large for manual review.	• Specific	• Small	• Qualitative (evaluation and interpretation)

More or less similar analysis

- Scientometrics:
  - More general and most often overlapped with bibliometrics
- Informetrics
  - More general than bibliometrics and scientometrics

More advanced (not really a bibliometrics)

- Text analysis
  - Word /document clustering - kmeans, hierarchical clustering, etc
  - Term frequency - inverse document frequency (TF-IDF)
  - Topic models - Latent dirichlet allocation (LDA), latent semantic analysis (LSA)
  - Wordcloud
  - etc
- Network analysis

A few ideas on bibliometrics

- General bibliometrics
    - Bibliometric assessment of breast cancer research
- Specific bibliometrics - limited to topics (need to review), number of papers, regions
    - Breast cancer and machine learning
    - Top 100 papers in machine learning
- Bibliometric comparison
    - Between countries, regions, etc

A few ideas on bibliometrics (cont.)

Research

BMJ Open — Bibliometric analysis of the top-cited gastroenterology and hepatology articles

Samy A Azer,[1] Sarah Azer[2]

ABSTRACT
Objective: To identify the top-cited articles in gastroenterology and hepatology, and analyse their characteristics.
Methods: Two searches were conducted in the Science Citation Index Expanded database; a search of 69 journals under the category 'Gastroenterology and Hepatology' (list A) and a keyword search of all journals (list B). The search results were analysed and the inter-rater coefficient of agreement was measured using Cohen κ.
Results: The number of citations varied from 1049 to 2959 in list A and from 1929 to 5500 in list B. In both lists, the majority of articles were research papers. No significant correlations were found between the number of citations and the number of years since publication (R²=0.00992, p=0.473 and R²=0.00202,

Strengths and limitations of this study
► Two searches were conducted in the Science Citation Index Expanded database.
► The search was based on journals with high impact factor and only those in the English language.
► Analysis explored a range of parameters in the assessment.

measurement to assess the work of researchers and impact of research,[3] and to rank researchers on the basis of differences in citation indices.[3][4] Recently, Nicholson and Ioannidis[5] explored whether there is a link between highly cited research and

---

BMJ Global Health — A bibliometric analysis of COVID-19 research in Africa

Fatuma Hassan Guleid,[1] Robinson Oyando,[2] Evelyn Kabia,[2] Audrey Mumbi,[2] Samuel Akech,[3] Edwine Barasa [2,4]

ABSTRACT
Background The COVID-19 pandemic has led to an unprecedented global research effort to build a body of knowledge that can inform mitigation strategies. We carried out a bibliometric analysis to describe the COVID-19 research output in Africa in terms of setting, study design, research themes and author affiliation.
Methods We searched for articles published between 1 December 2019 and 3 January 2021 from various databases including PubMed, African Journals Online, medRxiv, Collabovid, the WHO global research database and Google. All article types and study design were included.

Key questions
What is already known?
► Africa's contribution to global health research is low (1.3%) considering the high burden of infectious disease on the continent.
What are the new findings?
► Africa is contributing to the generation of COVID-19 knowledge by publishing primary and secondary research articles and editorial and commentary-type articles.
► African authors have made significant contrib

RESEARCH ARTICLE

University-Industry Collaboration in China and the USA: A Bibliometric Comparison

Ping Zhou[1]*, Robert Tijssen[2]©, Loet Leydesdorff[3]©

1 Department of Information Resources Management, School of Public Affairs, Zhejiang University, Hangzhou, Zhejiang, China, 2 Center for Science and Technology Studies (CWTS), Leiden University, Leiden, The Netherlands, 3 Amsterdam School of Communications Research (ASCoR), University of Amsterdam, Amsterdam, The Netherlands

© These authors contributed equally to this work.
* pingzhou@zju.edu.cn

Abstract

BMC Women's Health

RESEARCH ARTICLE — Open Access

Contribution of Arab countries to breast cancer research: comparison with non-Arab Middle Eastern countries

Waleed M Sweileh[1]*, Sa'ed H Zyoud[2], Samah W Al-Jabi[2] and Ansam F Sawalha[1]

Abstract
Background: Breast cancer is one of the most common types of cancers affecting women worldwide. The main objective of this study was to assess and compare research activity in breast cancer in Arab countries with non-Arab Middle Eastern countries.
Methods: Publications about "breast cancer" as a research topic were retrieved using the ISI Web of Science database. Analysis was confined to original research and review articles. Research productivity was assessed by assessing number of publications and time trend of these publications, names of journals, citation analysis, top 10 active institutions as well as country contribution to breast cancer research. The quantity and quality of publications from Arab countries in addition to 3 other Middle East countries (Turkey, Iran and Israel) were assessed and compared using the h-index tool.
Results: A total of 1658 original research and review articles about "breast cancer" were published from Arab countries

---

antioxidants — MDPI

Review

A Bibliometric Review of Publications on Oxidative Stress and Chemobrain: 1990–2019

Taylor McElroy [1,2,3] and Antiño R. Allen [1,2,3],*©

1 Division of Radiation Health, University of Arkansas for Medical Sciences, Little Rock, AR 72205, USA; TMMCELROY@uams.edu
2 Department of Pharmaceutical Sciences, University of Arkansas for Medical Sciences, Little Rock, AR 72205, USA
3 Department of Neurobiology & Developmental Sciences, University of Arkansas for Medical Sciences, Little Rock, AR 72205, USA
* Correspondence: aallen@uams.edu; Tel.: +1-501-686-7553

check for updates

Abstract: Oxidative stress is considered one of the possible mechanisms behind chemobrain or the cognitive dysfunction persistent after chemotherapy treatment. Breast cancer patients have reported chemobrain symptoms since the 1990s. In this present bibliometric review, we employed

---

Research

BMJ Open — Global research trends in spinal ultrasound: a systematic bibliometric analysis

Xiao Zhai,[1] Jin Cui,[1,2] Jie Shao,[1] Qijin Wang,[3] Xiao Chen,[1] Xianzhao Wei,[1] Xiaoyi Zhou,[1] Ziqiang Chen,[1] Yushu Bai,[1] Ming Li[1]

ABSTRACT
Background In recent years, there has been increased interest in the use of ultrasound technology in the evaluation of spinal and paraspinal regions.
Objective This study aimed to investigate trends in spinal ultrasound research from 1994 to 2015 and compare the contributions of such research from different countries and authors.
Study design Bibliometric analysis.
Setting Publications related to spinal ultrasound from 1994 to 2015 were retrieved from the Web of Science

Strengths and limitations of this study
► This bibliometric study included data on publication number, citation frequency, relative research interest, mean number of citations per paper, and H-index.
► The fitting mathematical model was used to analyse the time trend of the publications by using GraphPad Prism 5.
► VOSviewer, a Java programme, was used to visualise and analyse hotspots, masterpiece clusters, and

Steps for bibliometric analysis

1. Define scope, objectives and selection criteria
   - Find "similar" papers as reference
   - Avoid country limitation
   - The story should be somewhat "interesting" - can provide insight, and not only descriptive
2. Specify techniques (based on objectives)
3. Comes up with search terms - always check with the databases whether the terms valid or not
4. Data searching on databases:
   - **SCOPUS, WOS**, Digital Science Dimensions, PubMed, Lens or Cochrane
5. Review downloaded data/abstracts if needed (brief review)
6. Run bibliometric analysis and report finding

Suggested readings

- **Donthu, N., Kumar, S., Mukherjee, D., Pandey, N., & Lim, W. M. (2021). How to conduct a bibliometric analysis: An overview and guidelines. Journal of Business Research, 133, 285–296. https://doi.org/10.1016/j.jbusres.2021.04.070**
- **Aria, M., & Cuccurullo, C. (2017). bibliometrix : An R-tool for comprehensive science mapping analysis. Journal of Informetrics, 11(4), 959–975. https://doi.org/10.1016/j.joi.2017.08.007**
- Hood, W.W., Wilson, C.S. The Literature of Bibliometrics, Scientometrics, and Informetrics. Scientometrics 52, 291 (2001). https://doi.org/10.1023/A:1017919924342

Brief bibliometric example in [Scopus](#)

- Search terms:
  - Male breast cancer
  - Male breast neoplasm
- Make sure we get the right papers
- Save searched results

Hands-on in R