Bibliometric analysis using R

Tengku Hanis
PhD Student
Department of Community Medicine, USM

To do

Please install these packages to follow along in hands-on in R:

- bibliometrix
- tidyverse

Material available for download:

- https://github.com/tengku-hanis/webinar_biblio24-09-2020

What is bibliometrics?

- Measurement of the impact of research outputs using quantitative measures
- Evaluate research progress
- Frequently used in the field of library and information science

Example of papers

A Bibliometric Analysis of the Top 30 Most-cited Articles in Gestational Diabetes Mellitus Literature (1946-2019)

Pulwasha M. Iftikhar [1], Fatima Ali [2], Mohammed Faisaluddin [3], Azadeh Khayyat [4], Maribel De Gouvia De Sa [5], Tanushree Rao [6]

1. Health Sciences, St. John's University, New York, USA 2. Obstetrics and Gynecology, Shifa International Hospital, Islamabad, PAK 3. Obstetrics and Gynecology, Deccan College of Medical Sciences, Hyderabad, IND 4. Internal Medicine, Ahvaz Jundishapur University of Medical Sciences, Ahvaz, IRN 5. Obstetrics and Gynecology, Aberdeen Royal Infirmary, Aberdeen, GBR 6. Obstetrics and Gynecology, Liverpool Hospital, New South Wales, AUS

Corresponding author: Pulwasha M. Iftikhar, pulwashamaria@gmail.com
can be found in Additional Information at the end of the article

Bibliometric assessment of cancer research in Puerto Rico, 1903–2005

Ana P. Ortiz[1,2], William A. Calo[3], Carlos Suárez-Balseiro[4], Mariano Maura-Sardo[5], and Erick Suárez[2,3]

[1]Cancer Control and Population Sciences Program, University of Puerto Rico Comprehensive Cancer Center, San Juan, Puerto Rico.

[2]Biostatistics and Epidemiology Department, Graduate School of Public Health, Medical Sciences Campus, University of Puerto Rico, San Juan, Puerto Rico.

[3]Puerto Rico Cancer Center, University of Puerto Rico, San Juan, Puerto Rico.

[4]Observatory of Informatics Related Studies, Graduate School of Information Sciences and Technologies, Río Piedras Campus, University of Puerto Rico, San Juan, Puerto Rico.

Deanship of Graduate Studies and Research, Río Piedras Campus, University of Puerto Rico, San Juan, Puerto Rico.

Country Comparisons of Human Stroke Research Since 2001
A Bibliometric Study

Kjell Asplund, MD; Marie Eriksson, PhD; Olle Persson, PhD

Background and Purpose—This is the first bibliometric comparison between countries of the development of stroke research over time.
Methods—Clinical and epidemiological articles on stroke published 2001 to mid-2011 were identified in Science Citation Index Expanded. Article fractions, citation fractions, h-index, and international collaboration were calculated using the BibExcel software and adjusted for population size and gross domestic product.
Results—The United States dominated with 28.7% of the sum of article fractions and 36.2% of the sum of citation fractions. The United States, Japan, the United Kingdom, and Germany together accounted for 52.1% of articles and 61.0% of citations. When adjusted for population size or gross domestic product, several small European countries, together with Israel and Taiwan, ranked the highest. Per population, there was a negative association (*r*=0.60) between burden of stroke (disability-adjusted life-years lost) and number of articles per population. In China, South Korea, and Singapore, the annual growth of stroke articles was more than twice the worldwide average. Whereas multinational collaboration was common within Europe and North America, it was relatively uncommon between Asian countries.
Conclusions—The Big 4 in scientific literature on stroke, as to both number of articles and citations, are the United States, Japan, the United Kingdom, and Germany. Many small European countries have, in relation to their size, a high scientific production. Several countries with rapidly expanding economies have very fast growth of scientific production. Our results emphasize the need for stroke research in countries with a high population burden of stroke and highlight the role of multinational collaboration. **(Stroke. 2012;43:830-837.)**

Key Words: bibliometrics ■ h-index ■ scientific production ■ stroke

A bibliometric analysis of tuberculosis research, 2007–2016

Vaidehi Nafade[1,2⊛], Madlen Nash[1,2⊛], Sophie Huddart[1,2], Tripti Pande[2], Nebiat Gebreselassie[3], Christian Lienhardt[3], Madhukar Pai[1,2]*

1 Department of Epidemiology, Biostatistics and Occupational Health, McGill University, Montreal, Canada, 2 McGill International TB Centre, McGill University, Montreal, Canada, 3 Global TB Programme, World Health Organization, Geneva, Switzerland

⊛ These authors contributed equally to this work.
* madhukar.pai@mcgill.ca

Softwares

- **R (mainly bibliometrix package)**
- Excel
- VOSviewer
- Many others

Field tags in bibliometrix

Field Tag	Description
AU	Authors
TI	Document Title
SO	Publication Name (or Source)
JI	ISO Source Abbreviation
DT	Document Type
DE	Authors' Keywords
ID	Keywords associated by SCOPUS or ISI database
AB	Abstract
C1	Author Address
RP	Reprint Address
CR	Cited References
TC	Times Cited
PY	Year
SC	Subject Category
UT	Unique Article Identifier
DB	Bibliographic Database

More field tags on: https://www.bibliometrix.org/documents/

Field_Tags_bibliometrix.pdf

(a few) Measures in bibliometrix

- "Numerical" measures:
  - Publication per year
  - Most productive authors
  - Most productive countries
  - Most productive institutions
  - etc
- Visualization:
  - Thematic map
  - Keyword co-occurrence network
  - institutions/countries collaboration
  - etc
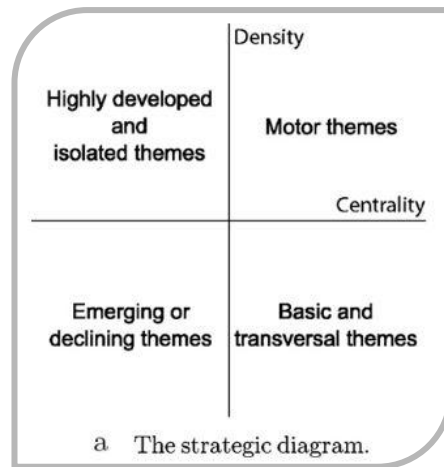
From cobo et al. (2011)
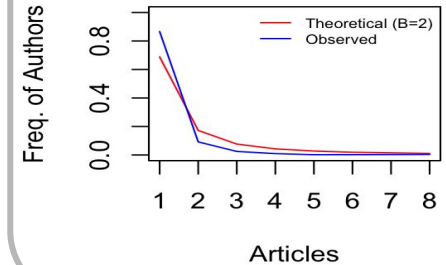
- Laws @ theories:
  - Lotka's law
    - Describes the frequency of publication by authors in any given field as an inverse square law, where the number of authors publishing a certain number of articles is a fixed ratio to the number of authors publishing a single article
    - Estimates Lotka's law coefficients for scientific productivity (Lotka A.J., 1926)
  - Bradford's law of scattering zones
    - Estimates the exponentially diminishing returns of searching for references in science journals
    - To determine core journal

Scientific Productivity

Freq. of Authors — Articles

Theoretical (B=2)
Observed

Zone 1:
- 3 sources
- 130 articles

Zone 2:
- 9 sources
- 130 articles

Zone 3:
- 27 sources
- 130 articles

Bibliometrix compatible databases

1. **Clarivate Analytics Web of Science database**
2. **SCOPUS database**
3. Dimensions database
4. PubMed/Medline database
5. Cochrane Library database of systematic reviews

Extension of bibliometrics (ie; scientometrics)

- Bibliographic coupling (same reference)
- Bibliographic co-citation (2 papers are cited in the third paper)
- Textual analysis
  - Clustering
  - TF-IDF (Term Frequency-Inverse Document Frequency)
- Topic models
- More advanced RQ:
  - Correlation between research funding and total citation
  - etc

References

- https://library.leeds.ac.uk/info/1406/researcher_support/17/measuring_research_impact
- https://cran.r-project.org/web/packages/bibliometrix/vignettes/bibliometrix-vignette.html
- https://bibliometrix.org/documents/bibliometrix_Report.html
- Lotka, A.J. (1926). The frequency distribution of scientific productivity. Journal of Washington Academy Sciences, 16, 317–323.
- Cobo, M. J., Lopez-Herrera, A. G., Herrera-Viedma, E., & Herrera, F. (2011). An approach for detecting, quantifying, and visualizing the evolution of a research field: A practical application to the fuzzy sets theory field. Journal of Informetrics, 5(1), 146-166.

QnA

Hands-on in R