ORBIT ASSOCIATION

Journal of Contemporary Decision Science

Journal homepage: www.cds-journal.org/
ISSN: 3104-3380

Comprehensive Insights into Bibliometric Analysis: Tools, Techniques, Applications, Challenges, and Emerging Solutions for Research Excellence

Gülay Demir[1,*]

[1] High School of Health Vocational Services, Sivas Cumhuriyet University, Sivas, Türkiye

ARTICLE INFO

Article history:
Received 17 March 2026
Received in revised form 21 April 2026
Accepted 23 June 2026
Available online 27 June 2026

Keywords:
Bibliometric analysis; Research evaluation; Science mapping; Research excellence; Reporting guidelines.

ABSTRACT

Due to the rapid growth in the volume of scientific literature, there is an increasing need for systematic methods to evaluate scientific output, identify research trends, and support evidence-based decision-making. Bibliometric analysis has become one of the most widely used tools for evaluating scientific productivity, research impact, collaboration networks, and knowledge structures. However, despite its growing popularity, the field of bibliometrics still lacks consistency in terms of the tools, methodologies, reporting practices, and technologies used in bibliometric analysis. The aim of this study is to provide a comprehensive overview of the current state of bibliometrics and propose a practical guide to conducting rigorous and comprehensive bibliometric analyses. Through a systematic literature review and synthesis, the study examines four main aspects of bibliometric analysis: tools, methodologies, application areas, and reporting quality. Major bibliometric software platforms, namely VOSviewer, Bibliometrix/Biblioshiny, CiteSpace, and Gephi, are compared in terms of their functionality, strengths, and limitations. In addition, the study identifies the major stages of bibliometric research, including search strategy development, data collection, data preprocessing, science mapping, scientific performance analysis, visualization, and reporting. The importance of transparency and reproducibility in bibliometric research is highlighted through the presentation of the GLOBAL reporting guideline and the PRISMA-ScR framework. The results show that no single tool can perform all types of bibliometric analyses effectively, while combining multiple tools provides a more comprehensive understanding of the research landscape. Furthermore, the study identifies methodological rigor—particularly the selection of an appropriate database, the development of a robust search strategy, and thorough data cleaning—as a key determinant of high-quality bibliometric research. The study demonstrates how bibliometrics has evolved from a descriptive evaluation instrument into a strategic research assessment tool that supports policy development and promotes research excellence. Finally, emerging developments in artificial intelligence, open science, altmetrics, and open citations are expected to have a significant impact on the future of bibliometric research. The proposed framework can serve as a practical guide for researchers, editors, research managers, and policymakers conducting bibliometric analyses.

* *Corresponding author.*
E-mail address: gulaydemir@cumhuriyet.edu.tr

https://doi.org/10.67334/cds31202635

Journal of Contemporary Decision Science
Volume 00, Issue 00 (2026) 1-26

1. Introduction

1.1. The Information Explosion and the Need for Research Excellence

In the modern era, information is created, distributed, and consumed in volumes that have never been seen before. With about 4.10% annual growth in scientific publications and a doubling time of 17.3 years, the researchers, institutions, and policymakers face the challenge of maneuvering around vast volumes of information [1, 2]. In this case, the problem does not pertain to information access alone but is connected to defining, measuring, and improving research excellence.

The concept of excellence research has become a multi-dimensional phenomenon in the scholarly environment of the current day. When judging whether a research piece or scientist is excellent, it is necessary to take into consideration multiple criteria, including not only scientific output but also impact, innovativeness, collaboration capacity, and social value [3]. At the same time, it should be admitted that multi-dimensionality of the concept of excellence being measurable, comparable, and trackable requires certain methodological approaches [4].

It appears that bibliometric analysis can be viewed as one of the most effective ways to quantitatively evaluate scientific productivity and impact. Bibliometric analysis allows assessing research performance, mapping scientific cooperation, and identifying research trends by analyzing publication and citation data [5, 6]. Stated differently, bibliometric analysis serves as a compass helping to achieve research excellence.

1.2. Growth in Bibliometric Analysis and Reporting of Quality

There has been considerable growth in the use of bibliometric analysis within the academic community recently. According to the Web of Science database search, the use of the term "bibliometric analysis" is growing exponentially, especially after 2021. From 2022 to 2025, there were 4,064 and 8,617 publications respectively, Figure 1. This shows that the use of bibliometric analysis is not just limited to methodology but has become a field of research.

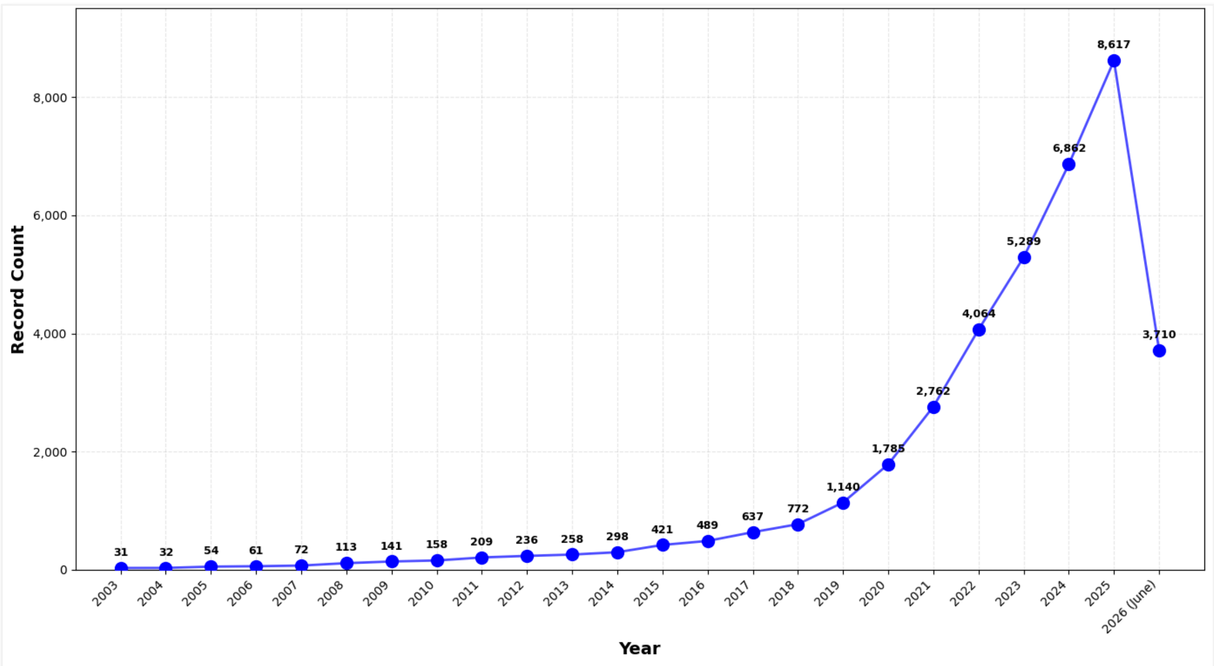

Fig. 1. Publication Records by Year (2003-2026/June)

However, the driving factors behind the rise seen in Figure 1 can be many. First, there has been an increasing use of bibliometric indicators in the framework of research assessment. Various decision-making procedures in different settings, whether related to promotion, funding, ranking, or science policies on a national scale, are increasingly based on bibliometric information [7]. Second, there has been a rising development of convenient bibliometric software programs. Programs like VOSviewer, Biblioshiny, and CiteSpace allow researchers with minimal programming skills to perform comprehensive bibliometric analysis [8]. Third, the open science agenda and data availability play a role in this regard. Such databases as Web of Science, Scopus, and Google Scholar, along with open-access systems such as Dimensions, OpenCitations, and CrossRef, offer valuable bibliometric data [9].

To better understand the thematic trends in the bibliometric analysis literature, the trending topics of studies searched with the keyword "bibliometric analysis" in the Web of Science database were examined. The TreeMap visualization presented in Figure 2 was created using Biblioshiny (the web interface of the Bibliometrix R package).

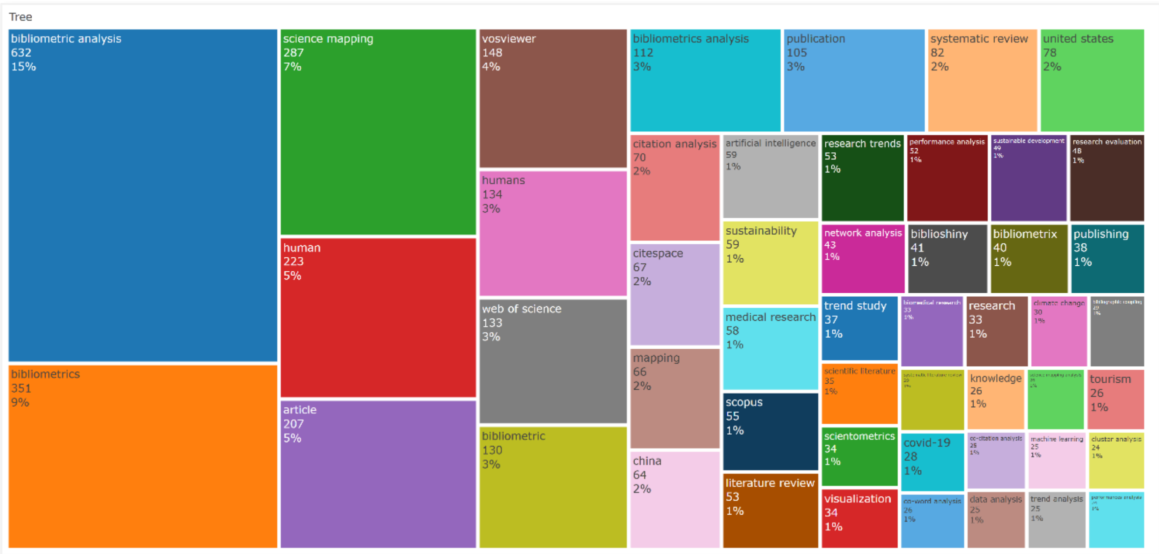

Fig. 2. Treemap visualization of studies searched with the keyword “Bibliometric Analysis”

According to Figure 2, the most dominant terms are "bibliometric analysis" (632, 15%), "bibliometrics" (351, 8%), and "science mapping" (287, 7%). This indicates that bibliometric analysis studies are largely focused on method development and science mapping. A noteworthy finding is the significant presence of the term’s "humans" (134, 3%) and "human" (223, 5%). These terms point to an increase in bibliometric analyses, particularly in human-centered disciplines such as health sciences, psychology, and social sciences in recent years. The frequent appearance of software names such as "VOSviewer" (148, 4%) and "CiteSpace" (41, 1%) also confirms the widespread use of these tools in bibliometric analyses. Furthermore, the prominence of terms such as "science mapping" (7%), "bibliometrics" (8%), "citation analysis" (2%), and "research evaluation" (1%) in relation to the term "bibliometric analysis" reveals the multifaceted nature of bibliometric analysis and its connections with different subfields. To analyze the conceptual relationships in the research area more deeply, a network map was created based on the frequency of co-occurrence of keywords.

The VOSviewer network visualization presented in Figure 3 reveals the relationships and thematic clustering among keywords in the bibliometric analysis literature.

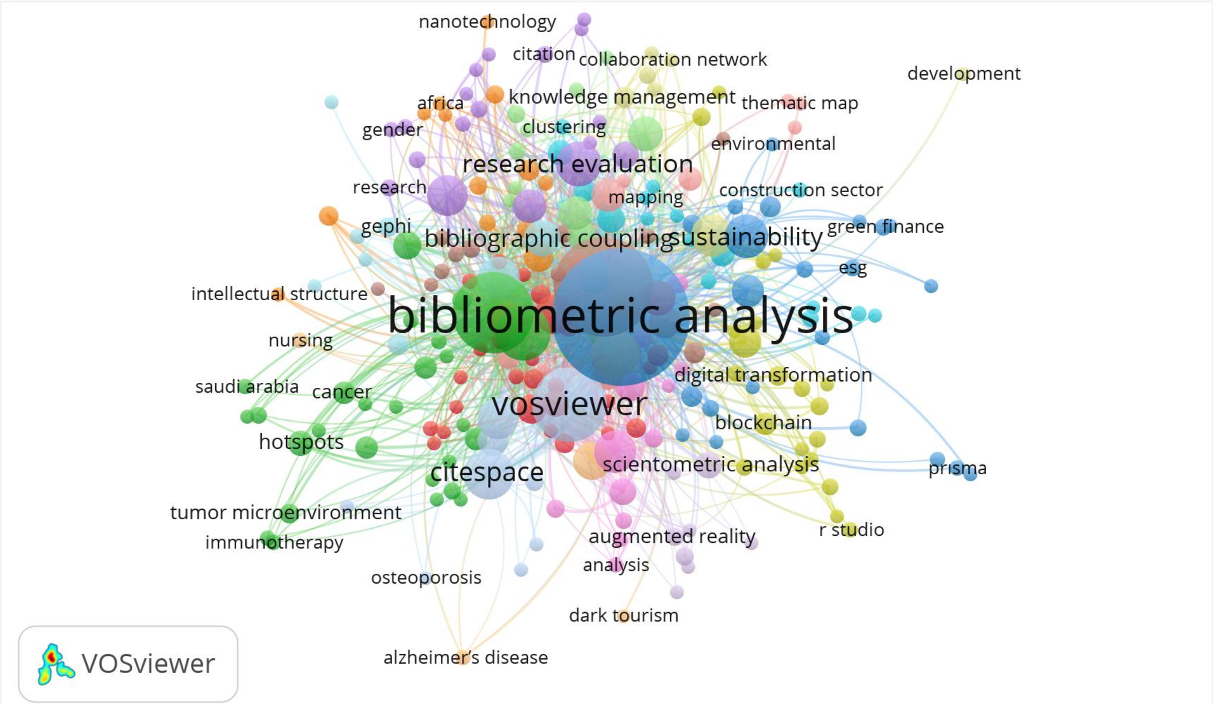

Fig. 3. Keyword Association Network Map of Studies Searched with the Keyword "Bibliometric Analysis"

In Figure 3, keywords sharing similar themes are grouped under the same-colored clusters. A total of five main thematic clusters were identified. The first and largest is the Methodological Terms cluster, represented in blue; this cluster contains fundamental methodological concepts of the field such as "bibliometric analysis," "bibliometrics," and "science mapping." The second cluster, represented in red, is the Human-Centered Terms cluster, emphasizing the human-centered nature of bibliometric applications, particularly in health and social sciences, with terms like "humans" and "human." The third cluster, represented in green, is the Analysis Tools cluster; software such as "VOSviewer," "CiteSpace," and "Bibliometrix" are included in this cluster and constitute the technological infrastructure of bibliometric analysis. The fourth cluster, represented in purple, is the Evaluation-Oriented Terms cluster, referring to applications in the context of research evaluation and science policy with terms such as "citation analysis," "research evaluation," and "science policy." The fifth and final cluster is the Current Research Trends cluster, represented in orange; it encompasses current and future research topics such as "sustainable development," "sustainability," and "trend study." The five-cluster structure proves the transformation of bibliometric analysis from a purely methodological tool into a multifaceted research area with an interdisciplinary approach to various subject matters. However, there is one drawback related to the rapid development of the research area: the quality of reporting of the bibliometric studies varies. Ng *et al.,* [10] carried out a detailed scoping review and discovered that there is no evidence-based guidance on how to report the results of bibliometric analysis. In this study, which involved the screening of 48,750 papers, only 23 papers met the inclusion criteria and 17 (74%) of them provided opinion-based suggestions, whereas only one study used a systematic search and the Delphi technique based on the EQUATOR guidance. [10, 11].

1.3. The Role of Bibliometric Analysis on the Road to Research Excellence

The idea of research excellence has turned out to be a crucial controversial issue in the current context of academia. However, the problem is that no common understanding of the concept of research excellence, its evaluation and incentives is available [12]. The contribution of bibliometric analysis to moving toward research excellence is achieved with the help of the following mechanisms:

An opportunity for Objective Evaluation: Unlike subjective peer review evaluation, bibliometrics provide an objective approach for evaluating researchers' work. Citation counts, h-index and impact factor of journals are some of the widely used metrics for measuring the impact of research results [13].

Comparison of Performance: With the help of bibliometric analysis, researchers' performance can be compared between different researchers, institutes and countries; thus, strengths and weaknesses will be identified, and strategic decisions will be made [14].

Detection of Research Trends: Bibliometric analysis is a very efficient tool for identifying research trends, emerging issues and possible future directions in every specific field [15].

Mapping of Collaboration Network: As modern science tends to become collaborative, there appears a need to visualize and analyze collaboration of researchers via co-authorship networks [16].

Identifying Research Gaps: Bibliometric analysis maps the gaps and understudy issues in every field [17].

1.4. Scope of the Article, Research Questions and Method

The current research offers a scientific overview of the bibliometric analysis ecosystem, which covers tools, methodology, application areas, and reporting quality issues. The research addresses the following questions:

RQ1: Which tools are used in bibliometric analysis? What are the advantages and disadvantages of these tools? Which tool is the best one to use for kinds of analysis?

RQ2: What methodological approaches are the basis of bibliometric analysis? How to prove validity and reliability of these methodologies? What are their advantages and disadvantages?

RQ3: In which academic fields is bibliometric analysis applied? How is it connected to research evaluation? What roles does it play in shaping the directions towards research excellence?

RQ4: In what way can bibliometric research be reported better? What does the GLOBAL guide offer? What practical suggestions can be offered to researchers and evaluators?

RQ5: How will the future of bibliometric analysis look like in the era of artificial intelligence, big data, and open science? What will change in research evaluation?

In the framework of the present study, peer-reviewed articles in English published between 2000 and 2025 and dedicated to the aspects of tools, methodology, application areas, and reporting of bibliometric analysis were analyzed. Keywords "bibliometric analysis," "bibliometric tools," "science mapping," "research evaluation," "research excellence," and "reporting guideline" were used to search for the relevant articles in the Web of Science, Scopus, Google Scholar, and Dimensions databases. The results obtained within the framework of the GLOBAL project were specifically used together with Ng *et al.,* [10].

1.5. Bibliometric Analysis Ecosystem: A Four-Dimensional Model

Bibliometric analysis can be described as an intricate system of interrelated processes involving four basic elements. This study proposes a four-dimensional model, as shown in Figure 4, by adding the reporting quality dimension to the traditional three-dimensional model (tools, methodologies, applications).

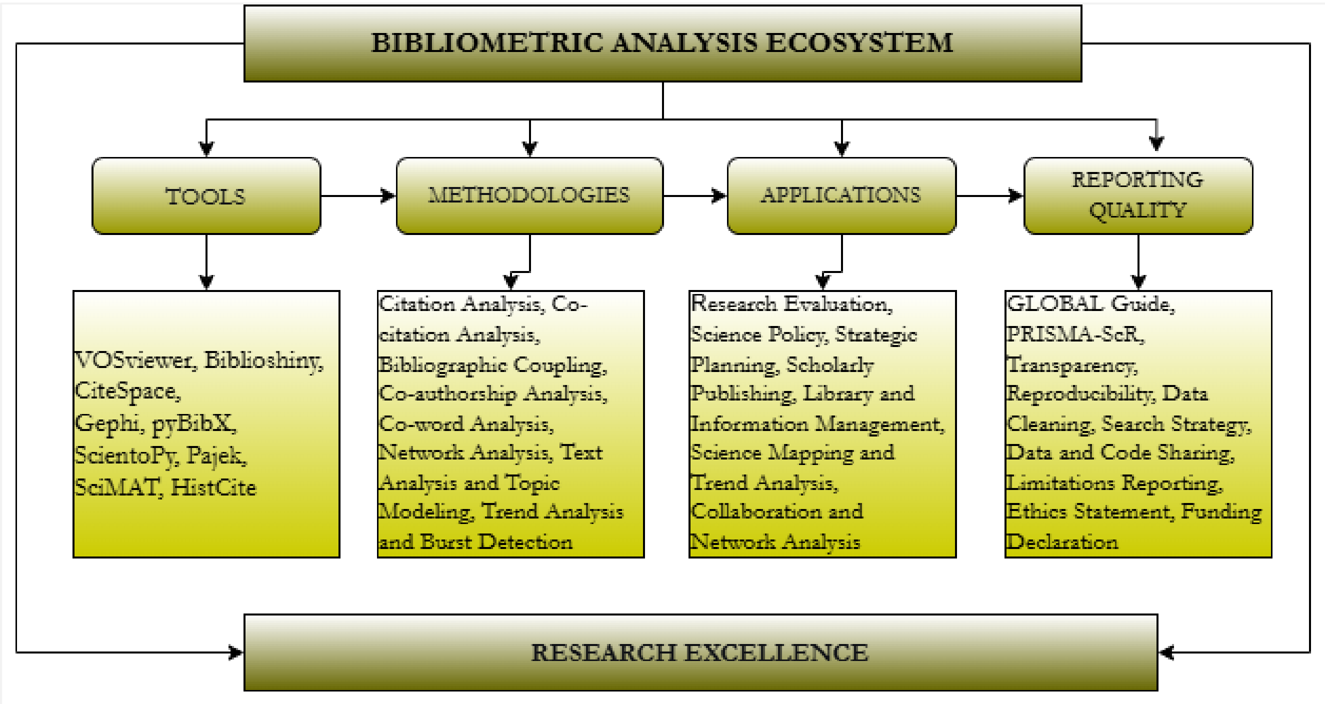

Fig. 4. Bibliometric Analysis Ecosystem - 4 Dimensions (Created by the author)

The Reporting Quality dimension in Figure 4 was added based on the findings of the GLOBAL project. The 32 items identified by Ng *et al.,* [10] are critical for transparent, complete and reproducible reporting of bibliometric studies.

Over time, initiatives such as the Leiden Manifesto, Hicks *et al.,* [18], DORA [19] and GLOBAL [10] have provided important guidance in reporting bibliometric studies. GLOBAL aims to bring about a revolutionary standard in the reporting of bibliometric studies.

2. Bibliometric Analysis Process: A Step-by-Step Guide

Bibliometric analysis is a multi-stage process that requires systematic and rigorous implementation. This section details the steps to follow from planning to reporting a bibliometric study, the challenges that may be encountered at each stage, and proposed solutions to these challenges. The process basically consists of four main stages: (1) Planning and Design, (2) Data Collection, (3) Data Analysis, and (4) Reporting. This four-stage model is visualized in Figure 5.

Stage 1: Planning and Design

The success of a bibliometric analysis largely depends on the decisions made during the planning phase. Errors at this stage can negatively impact the entire process and jeopardize the validity of the results. Every bibliometric study should begin with a clear and focused research question. The research question directly influences the scope of the study, the data sources to be used, the analysis methods, and the reporting strategy. The quality of the bibliometric analysis depends on the scope, quality, and suitability of the data source used. The main data sources and their characteristics are compared in Table 1.

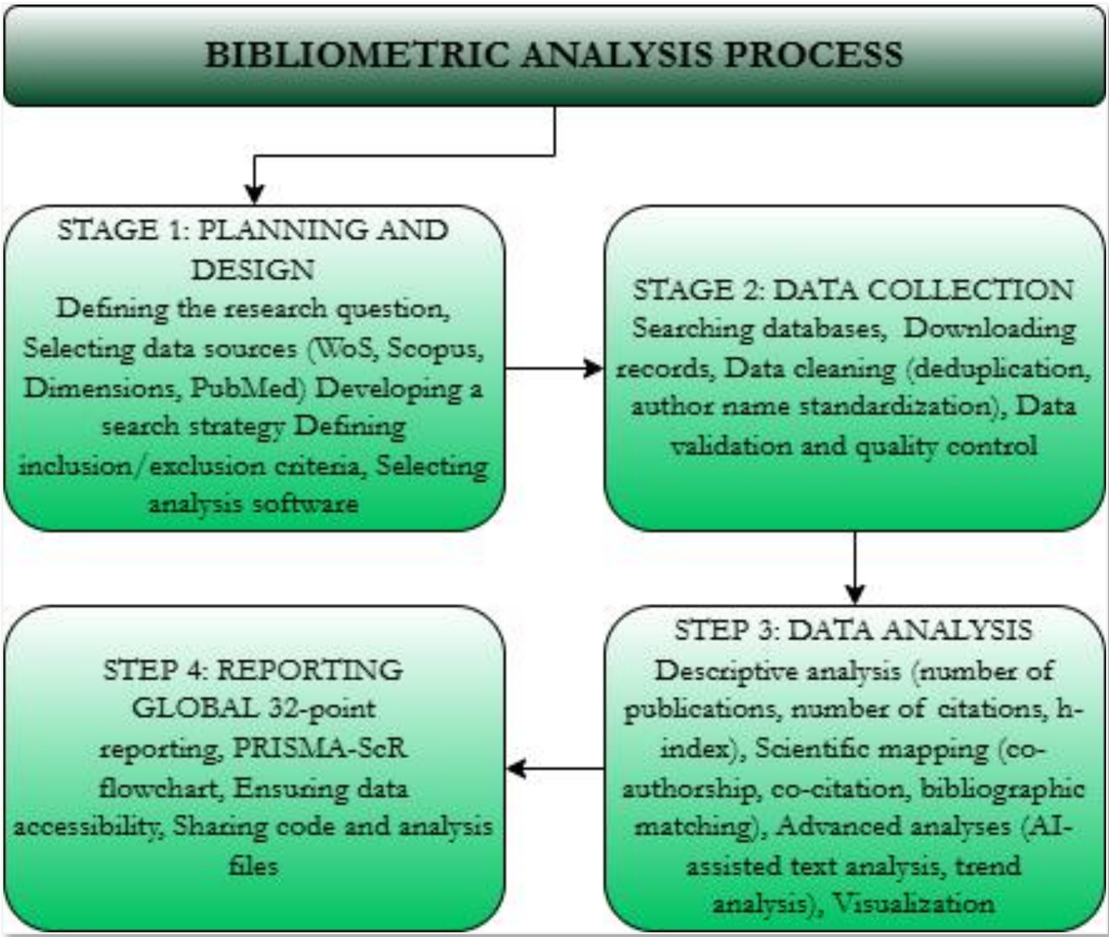

Fig. 5. Bibliometric Analysis Process - Four-Step Model (Created by the author)

Table 1
Comparison of Major Bibliometric Data Sources

Features	Web of Science (WoS)	Scopus	Google Scholar	Dimensions	PubMed
Scope	~34,000 journals	~25,000 journals	Extensive (journals, books, theses, web)	~130 million documents	~30 million biomedical articles
Historical Depth	From 1900 to present	1970 to present	Variable	Extensive	1966 to present
Language	Primarily English	Multilingual	Multilingual	Multilingual	Primarily in English
Citation Data	Available (rich)	Available (rich)	Available (limited)	Available	Available (via PubMed Central)
Export Formats	Plain text, RIS, BibTeX	RIS, CSV, BibTeX	RIS, BibTeX	CSV, JSON, RIS	XML, MEDLINE
Access	Subscription	Subscription	Free	Partially free	Free
Strengths	High quality, standardized	Comprehensive, user-friendly	Comprehensive, free	Comprehensive, open access	Biomedical focused
Weaknesses	Narrow scope, cost	Not fully comprehensive	Data quality issues	Limited citation data	Biomedical only

Sources: Adapted from AlRyalat *et al.,* [20] and Gusenbauer [21].

Search strategy is one of the most critical steps in bibliometric analysis and is among the most frequently emphasized items in the GLOBAL guideline [10]. To develop an effective search strategy, the following steps should be followed:

Keyword Identification: Define the key concepts in the research area, identify synonyms, related terms and abbreviations, Utilize MeSH (Medical Subject Headings) or other controlled dictionaries, Review keywords used in pioneering studies. Use of Boolean Operators: AND: Combines different concepts (example: "bibliometric*" AND "AI"), OR: Combines synonyms (example: "AI" OR "artificial intelligence"), NOT: Excludes unwanted terms (example: "AI" NOT "robotics")

Field Restrictions: Searches can be made in the title, abstract, keywords, or full text fields. Restrictions can also be applied to fields such as author name, institution, and journal.

Time Limitation: A specific time range can be determined according to the purpose of the study. Generally, the last 5-10 years are preferred, but wider ranges can be used for a historical perspective.

Document Type Restriction: Generally, articles, reviews, and conference proceedings are included. Editorial papers, letters, and book reviews may be excluded.

PRESS Check: The search strategy should be peer-reviewed using the PRESS (Peer Review of Electronic Search Strategies) tool developed by McGowan *et al.,* [22]. PRESS evaluates the following components: Translation and accuracy, Correct use of Boolean operators, Appropriateness of field constraints, Use of controlled dictionaries, Repeatability of the search strategy.

Inclusion/exclusion criteria are the rules that determine which studies are included in the analysis. An example set of inclusion/exclusion criteria is presented in Table 2.

Table 2
Inclusion/Exclusion Criteria

Criteria	Inclusion	exclusion
Publication Type	Article, review, conference paper	Editorial, letter, book review, abstract
Language	English or Turkish	Other languages
Time Range	2015-2025	Before 2015 or after 2025
Subject	AI and health related	Other topics
Data Access	Full text accessible	No full text access
Citation Status	Received at least 1 citation	Never cited (if necessary)

These criteria in Table 2 should be shaped according to the research question and reported transparently.

Stage 2: Data Collection

The data collection stage is the process of collecting, cleaning, and preparing data from selected data sources for analysis. Search results are downloaded in the format supported by the selected bibliometric software. Common formats include; Plain text (WoS), RIS (Scopus, Google Scholar), CSV, BibTeX. Maximum download limits (e.g., 500 for WoS, 2000 for Scopus) should be considered.

Data cleaning is one of the most neglected but critical stages of bibliometric analysis. OpenRefine, R (tidyverse), Python (pandas), and Excel/Google Sheets are data cleaning tools.

Stage 3: Data Analysis

Data analysis is the stage where bibliometric data is transformed into meaningful information. This phase includes descriptive analysis, scientific mapping, and advanced analysis. Descriptive analysis reveals the basic features of the data set: Publication analysis, citation analysis, author analysis, institution analysis, journal analysis are performed. Scientific mapping aims to visualize the intellectual structure of the research field. Co-authorship Analysis, Co-citation Analysis, Bibliographic Coupling, Keyword Co-occurrence Analysis and Text Analysis are used in scientific mapping. Advanced Analyzes can be grouped as Trend Analysis, Performance Analysis, AI Supported Analyzes. Visualization enables bibliometric findings to be presented effectively and understandably: It can be categorized as Network Visualizations, Thematic Maps, Time Series Visualizations, Geographic Visualizations.

Stage 4: Reporting

Reporting is the presentation of bibliometric analysis in a transparent, complete and reproducible manner. The GLOBAL 32 items provide guidance at this stage [10]. PRISMA-ScR [23] is a reporting guide developed for scoping reviews. Since bibliometric studies involve a systematic data collection process, the transparency of the study can be increased using the PRISMA-ScR flowchart. The GLOBAL 32 items should be used as a checklist in reporting bibliometric studies [10]. The reproducibility of bibliometric analyses is critical for scientific reliability: All code used in the analysis (R, Python) should be shared on an open access platform. Platforms such as GitHub, OSF, Zenodo can be used. The study protocol can be registered beforehand (OSF Registries, EQUATOR).

2.1. Practical Recommendations and Common Mistakes

These mistakes can be grouped into categories such as errors in data source selection, search strategy, data cleaning, analysis, and reporting.

Common Errors in Data Source Selection: Different databases have varying scopes and quality. Multiple data sources should be used whenever possible. While Google Scholar has broad scope, it contains data quality issues. Exaggerating the scope undermines the credibility of the study. Citation numbers change over time. The data collection date must be reported.

Common Errors in Search Strategy: A narrow keyword set that does not include synonyms and related terms can exclude important studies. Controlled dictionaries like MeSH and Emtree provide more comprehensive and standardized search results. Failure to peer-review the search strategy leads to errors being overlooked.

Common Errors in Data Cleaning: Author-based analyses yield erroneous results when the same author appears in different ways. Different spellings of the same institution affect institution-based rankings. Repetitions from different data sources inflate analysis results. If data cleaning decisions are not documented, the study cannot be replicated.

Common Errors in Analysis and Reporting: Bibliometric indicators show correlation, not causality. Citation count alone is not an indicator of quality. Default settings are not suitable for every study. Parameters should be tailored to the research question. Color, size, and layout should not be misleading. Every bibliometric study has limitations, which should be reported transparently.

3. Bibliometric Analysis Tools

Bibliometric analysis tools play a critical role in the collection, processing, analysis, and visualization of scientific data. In recent years, the number of both traditional and AI-powered tools has increased significantly. This section presents the available tools in categories, provides guidance on their basic features, and determines which tool is more suitable for which type of analysis. The choice of analysis software depends on the scope of the study, the type of analysis, and the researcher's technical proficiency. The classification of bibliometric analysis tools into four main categories is given in Table 3.

Table 3
Bibliometric analysis tools

Tool	License	Key Features	The analyses performed
VOSviewer	Free (Academic)	• Co-authorship, co-citation, bibliographic matching, keyword association • Network, density, distribution views • Clustering and tagging • WoS, Scopus, PubMed, RIS, CSV	• Bibliographic network visualization • Co-authorship networks • Keyword maps • Quick visual presentations

Table 3
Continued

Tool	License	Key Features	The analyses performed
Bibliometrix	Open Source (GPL)	• WoS, Scopus, Dimensions, PubMed, Lens, Cochrane • Publication, citation, collaboration analysis • Three-field charts, thematic maps • Time series and trend • Automated report • Data cleaning	• Comprehensive analyses • Reproducible research • Time series and trends • Customized visualization • R users
Biblioshiny	Open Source (GPL)	• Bibliometrix web interface • No coding knowledge required • Drag-and-drop data upload • Interactive visualization • Automated reporting	• Non-coding users • Fast analysis • Basic bibliometric analyses
CiteSpace	Free (Academic)	• Time-axis visualization • Burst detection • Co-citation and intellectual mapping • Dual network mapping • Cluster labeling (LLR, TF-IDF, MI) • Network metrics	• Research trends • Scientific revolutions • Time series analysis • Identifying burst terms
Gephi	Open Source (GPL)	• Advanced network visualization (ForceAtlas2, Yifan Hu) • Dynamic filtering • Network statistics • Plugin support • Millions of nodes	• Complex collaboration networks • Large-scale network analyses • Publication-quality visualizations • Interactive presentations
pyBibX	Open Source	• Embedding vectors • Topic modeling (BERTopic, LDA, NMF) • Text summarization (PEGASUS, BERT, GPT) • NLP (Sentence-BERT, ChatGPT) • Traditional analyses • Visualization	• Text analysis and NLP • Automatic summarization • Semantic network analysis • AI-powered science mapping
ScientoPy	Open Source	• Publication and citation trends • Keyword trends • Author, institution, and country analyses • Line and bar charts, word clouds • Data preprocessing • CSV, WoS, Scopus	• Trend analysis • Time series studies • Performance indicators • Data exploration and preliminary analysis
Pajek	Free (Academic)	• Processing millions of nodes • Advanced network metrics • Network segmentation • Block modeling • Network reduction	• Very large network analysis • Complex network metrics • Network theory research
SciMAT	Free (Academic)	• Thematic evolutionary analysis • Strategic diagrams • Performance analysis • Longitudinal analysis • Multiple data sources	• Thematic evolutionary analysis • Strategic management • Field structural analysis • Comparison of time periods
HistCite	Free	• Citation history • Chronological network visualization • Most influential articles • WoS data	• Historical development analysis • Identification of key studies • Chronology of scientific discoveries

Table 3
Continued

Tool	License	Key Features	The analyses performed
Publish or Perish	Free (Academic)	• Google Scholar data retrieval • h, g, e, i10 indexing • Personal/organizational analysis • Citation reports	• Individual performance • Corporate evaluation • Fast metric calculation
Dimensions Analytics	Commercial / Free (Partial)	• 130M+ documents • Open access data • Integrated analytics and visualization • Citation and altmetric data • Corporate analytics • Funds and patents	• Corporate evaluation • Comprehensive data analysis • Fund and patent analysis

Sources: Adapted from Moral Muñoz *et al.,* [24]; Aria and Cuccurullo [25].

In addition to the information in Table 3, bibliometric analysis tools can be summarized into 4 groups: Established/Traditional Tools; VOSviewer, Bibliometrix, CiteSpace, Gephi, SciMAT. AI-Powered/Next-Generation Tools; pyBibX, ScientoPy. Special Purpose/Niche Tools; Pajek, HistCite, Publish or Perish, and Platform/Web-Based Tools; Biblioshiny, Dimensions Analytics.

3.1. Future Tool Development Trends

The future of bibliometric analysis tools will be shaped by the following trends:

Artificial Intelligence and Deep Learning: As in the case of pyBibX, AI integration will increase, and automated summarization, topic modeling, and semantic analysis will become standard.

Real-Time Analysis: Tools will offer real-time analysis and monitoring capabilities with instant data streams.

Multiple Data Source Integration: Data integration from multiple sources such as WoS, Scopus, Dimensions, PubMed, and OpenAlex will be facilitated.

Altimetric Integration: In addition to traditional citation data, altimetric data such as social media, news, and blogs will be included in the analyses.

Open Science and Reproducibility: Tools will offer built-in features to ensure the reproducibility of analyses (code generation, version control).

Cloud-Based Operation: Cloud-based, collaborative platforms that do not require local installation will become widespread.

4. Reporting Bibliometric Analyses: GLOBAL and PRISMA-Scr

The scientific validity and reliability of bibliometric analyses depend not only on the methods used, but also on how these methods are reported. Reporting quality directly affects reproducibility, critical evaluation, and contribution to scientific knowledge [26, 27].

Reporting quality enables scientific studies to fulfill three basic functions:

Transparency: The reader can fully understand what was done, how it was done, and why it was done.

Reproducibility: Other researchers can repeat the same study and verify the results.

Critical Evaluation: Readers and reviewers can evaluate the strengths and weaknesses of the study.

Two basic guidelines used in reporting bibliometric analyses have been examined in detail. Transparency, reproducibility, and critical evaluation are the foundations of scientific validity. The study by Ng *et al.,* [10] reveals serious deficiencies in the reporting quality of bibliometric analyses.

i. GLOBAL Guidelines: This 32-item guideline is the first comprehensive guideline developed for reporting bibliometric studies. The items cover all sections, from title and methods to findings and discussion.
ii. PRISMA-ScR Guidelines: Developed for scoping reviews, this guideline can also be adapted to bibliometric studies. The PRISMA-ScR flowchart is ideal for visualizing the data collection process.

GLOBAL and PRISMA-ScR are complementary guidelines. When used together, the reporting quality of bibliometric studies is significantly improved. Data and code sharing, transparent methodology reporting, visualization principles, and reporting of limitations are key strategies that enhance reporting quality. The scientific contribution of bibliometric analyses depends not only on their findings but also on how those findings are reported. The GLOBAL and PRISMA-ScR guidelines aim to increase the reliability, reproducibility, and impact of bibliometric studies by systematizing this reporting process.

5. Bibliometric Indicators and Metrics

Bibliometric indicators and metrics are tools that enable the quantitative measurement of scientific productivity, impact, collaboration, and research quality. These indicators are widely used to evaluate the performance of researchers, institutions, countries, and journals, to identify research trends, and to shape science policies [15, 28]. This section describes the process of evaluating the most used bibliometric indicators, explaining how they are calculated and analyzing their strong and weak points.

5.1. Classification of Indicators

Bibliometric indicators can be classified into five main categories:

Publication Indicators: Productivity and publication performance. Journal Impact Factor (JIF): The average number of citations received by journal articles published in the last two years. SCImago Journal Rank (SJR): An indicator of journal impact based on Elsevier's Scopus data, weighted according to the quality of the sources of citations. Source Normalized Impact per Paper (SNIP): A field-normalized impact indicator that considers the interdisciplinary differences of journals. CiteScore: The average number of citations over the last 4 years, based on Scopus data [29].

Citation Indicators: Scientific impact and visibility. Citation indicators are used to measure the impact, visibility, and importance of scientific studies. Total Citations (TC): The total number of citations received by an article, author, or institution. Average Citations per Paper (ACPP): The total number of citations is divided by the number of publications. Average Citations per Year (ACY): The total number of citations divided by the number of years published. Field-Weighted Citation Impact (FWCI): The citation performance of a publication compared to other publications in the same field, year, and document type. Percentile Index: The percentile of a publication's citation performance compared to other publications in the reference set [30].

Author Indicators: Individual performance and productivity. Author indicators are used to evaluate the performance of individual researchers. h-index: The highest h value where an author has at least h articles cited. g-index: The highest g value where the total number of citations for the author's most cited articles is greater than or equal to g^2. m-index: The h-index divided by the author's number of active years. e-index: The number of excess citations above the h-index for articles included in the h-index. i10-index: The number of articles that have been cited at least 10 times. Number of Publications Per Author: The total number of publications of an author. Average Number of Authors Per Article: Calculated as Total number of authors / Total number of publications. Collaboration Index (CI): The average number of authors per article. Co-authorship Index: The change in collaboration trends in a particular field over time. International Collaboration Rate: Calculated as Number of internationally co-authored articles / Total number of articles [31].

Journal Indicators: These are the effectiveness and quality of a journal. Journal indicators are used to evaluate the quality, impact, and prestige of academic journals. Journal Impact Factor (JIF): Calculated as (Number of citations received by articles in the previous 2 years in the current year) / (Number of articles in the previous 2 years). 5-Year Impact Factor: The average number of citations for articles in the previous 5 years. Eigenfactor: An indicator that measures the centrality of journals in citation networks. Article Influence Score: A measure of the average impact of articles in the journal. CiteScore: The average citation over the last 4 years, based on Scopus data. SJR (SCImago Journal Rank): A weighted citation count based on source quality, based on Scopus data. SNIP (Source Normalized Impact per Paper): An impact indicator that normalizes disciplinary differences. Google Scholar Metrics: h5-index and h5-median based on Google Scholar data [30].

Network Indicators: Collaboration, centrality, and network structures. Network indicators are used to measure the position, importance, and relationships of actors (authors, institutions, countries, keywords) in bibliometric networks. Degree Centrality: The number of direct connections to a node. Betweenness Centrality: How often a node is located on the shortest paths between other pairs of nodes. Closeness Centrality: The inverse of the average distance of a node to all other nodes. Eigenvector Centrality: Calculation of a node's centrality based on the centrality of the nodes it is connected to. PageRank: An algorithm developed by Google that measures the importance of nodes in the network. Clustering Coefficient: The ratio of a node's neighbors to each other. Network Density: The ratio of connections to potential connections in the network. Modularity: The degree to which the network is clustered. Average Path Length: The average shortest distance between all pairs of nodes in the network. Small-World Properties: A measure of whether the network has small-world properties (high clustering, low path length) [32].

Bibliometric indicators are powerful tools for measuring research performance, impact, and collaboration. However, no single indicator is sufficient on its own. Indicators should be used in combination, interpreted in context, and their limitations should be clearly reported. Initiatives such as the Leiden Manifesto and DORA provide guidance for the responsible use of metrics [7, 33].

6. Data Sources and Databases

This section systematically examines and compares the most frequently used data sources in bibliometric analyses, providing guidance for selecting data sources according to researchers' needs. The quality and reliability of bibliometric analyses largely depend on the scope, quality, and suitability of the data source used.

6.1. Classification of Data Sources

Bibliometric data sources can be classified into four main categories: commercial/subscription-based databases, open-access/free databases, special-purpose databases, and citation and metadata platforms:

Commercial/Subscription-Based Databases: WoS, Scopus. A comprehensive understanding of the strengths and weaknesses of both databases is critical for researchers to make informed data source selections. In this context, Table 4 presents a comparative analysis of WoS and Scopus in terms of various criteria.

Table 4
Comparison of WoS and Scopus

Criteria	Web of Science	Scopus
Institution	Clarivate Analytics	Elsevier
Year of Establishment	1900	1970
Number of Active Journals	~34,000	~25,000
Book Coverage	~120,000	~250,000
Conference Papers	Available (CPCI)	Available
Historical Depth	Excellent (1900-)	Good (1970-)
Broadness of Coverage	Good	Excellent
Quality Control	Very Reliable	Tough
Update Rate	Weekly	Daily
Download Capacity	500 records	20,000 records
Author Profiles	ORCID, ResearcherID	ORCID, Scopus Author ID
Institution Profiles	Available	Available (rich)
Citation Data	Excellent	Excellent
Data Standardization	Excellent	Good
Language Diversity	English-focused	Multilingual
Humanities	Good	Medium
Cost	High	High
Analytical Tools	InCites, JCR	SciVal, Analytic

Sources: AlRyalat *et al.,* [20]; Gusenbauer [21]; Clarivate, [34]; Elsevier, [35].

Table 4 aims to guide researchers in selecting the appropriate database for the purpose and requirements of their studies.

Open Access/Free Databases: Google Scholar, Dimensions (partially), PubMed, OpenAlex. In addition to commercial databases, open access and free databases are playing an increasingly important role in the evaluation of academic research. These databases have become indispensable tools, especially for researchers with limited resources, studies adopting open science principles, and extensive literature reviews. While open access databases generally have broader coverage than their commercial counterparts, they do have some limitations in terms of data quality, standardization, and reliability. Each open access data source has its own strengths and weaknesses. Therefore, researchers need to choose the data source or combination of sources that suit the purpose, scope, and technical requirements of their studies. Table 5 presents a comparative analysis of these four open access data sources in terms of various criteria.

Table 5
Comparison of Open Access Data Sources

Criteria	Google Scholar	Dimensions	PubMed	OpenAlex
Scope	Very Broad	Large	Biomedical	Very Large
Number of Documents	~200 million	~130 million	~30 million	~250 million
Citation Data	Limited, Not Reliable	Good	Limited	Good
Data Quality	Low	Medium	High	Medium - High
Standardization	None	Medium	High (MeSH)	Medium
API Access	None	Partial	Yes (NCBI E-utilities)	Yes (REST API)
Free Access	Yes (Fully)	Partial (Basic Version)	Yes (Fully)	Yes (Fully)
Language Support	Multilingual	Multilingual	English	Multilingual
Update Rate	Real-time	Daily	Daily	Weekly
Open Source	No	No	Partial	Yes
Community Support	Large	Medium	Large	Developing

Table 5
Continued

Criteria	Google Scholar	Dimensions	PubMed	OpenAlex
Optimal Use Case	Exploration and Preliminary Analysis	Comprehensive Analytics	Biomedical Research	Big Data Analytics
Data Export Formats	RIS, BibTeX, EndNote	CSV, JSON, RIS, BibTeX	MEDLINE (XML), RIS, Plain Text	JSON, CSV (API)
Institution/Author Profiles	Google Scholar Profile	Dimensions Profile	PubMed (ORCID)	OpenAlex Profile
Research Funding Data	No	Yes	No	Yes
Patent Data	Yes (limited)	Yes	No	Yes
Policy Document	No	Yes	Limited	Yes
Historical Depth	Broad (1990+)	Medium (1970+)	High (1966+)	Medium (1990+)
Web Interface	Simple	Advanced	Advanced	Developing

Sources: Gusenbauer [21]; Digital Science [36]; NCBI [37]; Priem *et al.,* [38].

Table 5 compares the four most commonly used open access data sources in bibliometric analyzes (Google Scholar, Dimensions, PubMed, and OpenAlex) in terms of scope, number of documents, citation data, data quality, standardization, API access, free access status, language support, update rate, open source status, community support, best-suited areas, and other technical features.

Special Purpose Databases: IEEE Xplore (engineering), PubMed (biomedical), arXiv (physics, mathematics). In addition to commercial and open-access general databases, special-purpose databases focusing on specific disciplines or specific publication types also play an important role in bibliometric analyses. While these databases allow for in-depth examination of research in a specific field, they provide content at a level of expertise that general databases cannot provide. Each special purpose database has its own unique strengths and weaknesses. Table 6 presents these three special-purpose databases comparatively in terms of various criteria and guides the selection of appropriate databases for different research purposes.

Table 6
Comparative Analysis of Special Purpose Databases

Criteria	IEEE Xplore	PubMed	arXiv
Developer/Institution	IEEE (Institute of Electrical and Electronics Engineers)	NCBI (National Center for Biotechnology Information) / NLM	Cornell University
Area of Expertise	Electrical-Electronics, Computer, Engineering	Biomedical, Health Sciences, Life Sciences	Physics, Mathematics, Computer Science, Biology, Statistics, Finance
Scope	~6 million Documents	~30 million Articles	~2.5 million Preprints
Historical Depth	From 1988 to Present	1966 to Present	1991 to Present
Content Types	Content Types • Journal articles • Conference proceedings • Technical standards • Books • Educational materials	• Journal articles • Reviews • Clinical trials • Editorials (selected) • PubMed Central (open access)	• Preprints • Physics (hep-th, cond-mat, etc.) • Mathematics • Computer Science (cs.XX) • Biology (q-bio) • Statistics (stat) • Finance (q-fin)
Peer Review/Quality Control	Yes (Strict)	Yes (Strict - MEDLINE)	No (Preprint)

Table 6
Continued

Criteria	IEEE Xplore	PubMed	arXiv
Citation Data	Rich	Limited (with PMC)	Limited
Access	Subscription (Corporate)	Free (Open Access)	Free (Open Access)
API Access	Available (IEEE API)	Yes (NCBI E-utilities)	Yes (arXiv API)
Language Support	English-focused	English predominantly	Multilingual (content-based)
Checked Glossary	IEEE Standards, MeSH	MeSH (Medical Subject Headings)	None (Free tagging)
Update Rate	Weekly	Daily	Everyday
Community	Engineers, Academics	Biomedical researchers, Clinicians	Physicists, Mathematicians, Computer Scientists
Data Formats	IEEE Xplore format, RIS, BibTeX	MEDLINE (XML, RIS, Plain Text, CSV	arXiv XML, TeX, PDF

Sources: IEEE [39]; NCBI [37]; arXiv [40].

Table 6 presents a comparative analysis of three purpose-built databases (IEEE Xplore, PubMed, and arXiv) used in bibliometric analysis, categorized by developer institution, area of expertise, scope, historical depth, content types, peer review status, citation data, accessibility, API accessibility, language support, glossary, update rate, community, and data formats.

Citation and Metadata Platforms: CrossRef, OpenCitations. Among the data sources that form the basis of bibliometric analysis, citation and metadata platforms are critical, in addition to direct databases. These platforms perform essential functions such as the permanent identification of academic publications, metadata sharing, and the provision of open citation data. Especially with the strengthening of the open science movement, these platforms play a significant role in increasing the transparency, reproducibility, and reliability of bibliometric analyses. Both platforms serve different purposes in bibliometric analysis. While CrossRef is an indispensable resource for DOI-based identification and metadata collection, OpenCitations is an ideal platform for analyses based on open citation data. Table 7 presents a comparative analysis of these two citation and metadata platforms in terms of various criteria, guiding the selection of the appropriate platform for different research purposes.

Table 7
Comparative Analysis of Citation and Metadata Platforms

Criteria	CrossRef	OpenCitations
Developer/Institution	Publishers International Linking Association (PILA)	Open access academic initiative
Main Function	DOI (Digital Object Identifier) system management	Providing open access citation data
Data Type	Metadata, DOI	Citation data, metadata
Data Volume	~130 million DOI (2025)	~1.5 billion Citations (COCI, 2025)
Start Year	2000	2017 (with I4OC)
Scope	• Journal articles • Books • Conference proceedings • Theses • Technical reports • Data sets	• COCI (Crossref Open Citation Index) • Citation data • Metadata • Citation networks • Indexed citations
License	Commercial/Open (mixed)	CC0 (Creative Commons Zero)
API Access	Yes (REST API, free)	Yes (REST API, free)
Open Access	Partial (API free, data subscription)	Completely free

Table 7
Continued

Criteria	CrossRef	OpenCitations
Citation Data	Limited (Documents with DOI)	Rich (~1.5 billion citations with COCI)
Metadata	Rich (author, title, abstract, publisher, etc.)	Limited (basic metadata)
Standardization	High	High
International Scope	Broad (all publishers)	Developing (selected publishers)
Support for the Open Citation Movement	Partial (with COCI)	Full (I4OC supporter)
Community Support	Large (publishers, academics)	Growing (open science advocates)
Difficulty of Use	Low (API well documented)	Medium (API)
Documentation	Excellent	Good
Sample Usage	DOI generation, metadata validation	Citation network analysis, open citation evaluation
Web Interface	Yes (CrossRef.org)	Yes (OpenCitations.net)

Sources: CrossRef [41]; OpenCitations [42]; Heibi *et al.,* [43].

Table 7 presents a comparative analysis of two primary citation and metadata platforms (CrossRef and OpenCitations) used in bibliometric analysis, considering factors such as developer organization, main function, data type, data volume, initiation year, scope, license, API access, open access status, citation data, metadata, standardization, international scope, open citation movement support, community support, difficulty of use, documentation, web interface, and data export formats.

7. Application Areas

Bibliometric analyses are increasingly used in various fields of the academic world. From research evaluation to policy making, from strategic planning to academic publishing, bibliometric methods support data-driven decision-making processes across a wide range of areas. In this section, a comprehensive analysis is presented of the major areas of application of bibliometrics analyses, describing the objectives, methods, and achievements of bibliometric studies in these applications. There are six main areas in which bibliometric applications can be classified [44]:

Research Evaluation: Evaluation on the levels of individual researchers, research organizations, and countries, used in the context of job promotions, positions, grants, and university rankings.

Science Policy and Strategy Making: Used to determine research priorities, to allocate funds, to create new research centers, and to plan staff.

Publishing: Used in journal evaluation, ranking, and improvement with measures like Journal Impact Factor (JIF), CiteScore, SCImago Journal Rank (SJR), and Source Normalized Impact per Paper (SNIP).

Libraries: Used in selection of materials for library collections and decisions about subscribing to journals.

Science Mapping and Forecasting: Identifies the intellectual, social, and conceptual structure of research fields and helps to forecast future developments and trends.

Collaboration and Social Network Analysis: Studies collaboration networks and information flows between individuals, research organizations, and countries.

Maintaining an ethical, transparent, and responsible attitude in the application areas of bibliometric analyses is crucial to prevent any possible adverse consequences caused by misuse of these tools. In this respect, Leiden Manifesto [18] (Hicks *et al.,* 2015), DORA [19] (San Francisco Research Evaluation Agreement, 2012), and the GLOBAL guide (Ng *et al.,* [10]) may serve as useful guidelines.

Leiden Manifesto specifies ten fundamental principles of responsible metrics use in research evaluation, which include: (1) combine quantitative metrics with qualitative evaluation; (2) link metrics to the mission of research; (3) normalize metrics taking into account subject areas; (4) ensure transparency of data collection and analysis process; (5) recognize explicitly all limitations of metrics; (6) avoid reductionism; (7) pay attention to any adverse effects of metrics on science; (8) evaluate risks of metric misuse; (9) use metrics not just for evaluation purposes, but also for development; and (10) use metrics in a responsible way [18].

DORA [19] recommends more specific principles to fight against misuse of journal impact factor in research evaluation: do not use the journal impact factor as the only indicator of quality of an article or researcher; take into account both qualitative evaluation and other metrics such as h-index and citations; consider publications' outputs and metrics, innovations, reproducibility, and alternative outputs; use the journal impact factor for assessment of quality of the journal, not for the quality of individual articles [19].

In turn, the GLOBAL guideline (Ng *et al*., [10]) promotes transparency, completeness, and reproducibility of bibliometric analyses through implementation of ethical principles. Specifically, Clause 8, dealing with data sources reporting; Clause 11, concerning data cleaning; Clauses 12–14, addressing the methodology and parameters of analysis reporting; and Clause 27, which refers to reporting limitations, can help achieve transparency and reliability of bibliometric studies.

8. Challenges, Risks, and Future Trends

Bibliometric analyses offer great value when it comes to evaluation, guidance, and comprehension of scientific research. However, there are several obstacles and dangers associated with the usage of bibliometric analyses. This section systematically examines the key challenges, ethical risks, and future trends faced by bibliometric analyses. Understanding these challenges will contribute to the more responsible, transparent, and effective use of bibliometric analyses.

8.1. Challenges Related to Data Sources

The reliability of bibliometric analyses largely depends on the scope and quality of the data sources used. In this context, the main challenges are:

Scope and Coverage Biases: Commercial databases (WoS, Scopus) tend to cover English-language publications, science and medicine fields, and North American and Western European institutions more than others. This leads to biases based on language, discipline, geography, publication type, and time. Publications in the humanities, social sciences, and developing countries are particularly negatively affected by these biases.

Data Quality Problems: Deficiencies in the standardization of author name, institution name and journal name, errors in citation data and missing metadata threaten the reliability of the analysis results. The appearance of the same author in different ways, the recording of the same institution with different spellings and inconsistencies in citation counts cause serious problems, especially in author and institution-based analyses.

Recommended Solutions: To overcome these challenges, it is necessary to use multiple data sources (WoS, Scopus, OpenAlex) together, to meticulously implement data cleaning processes and to transparently report the scope and limitations of data sources in accordance with the GLOBAL guideline [10].

8.2. Methodological Challenges

The methodological obstacles arise during the process of designing, conducting, and presenting bibliometric analyses.

Parameters of the Analysis: The selection of such parameters as clustering algorithms (Louvain, Walktrap, Girvan-Newman), thresholds (the number of publications required, the number of citations required), and normalization techniques (Salton's cosine, Jaccard, Association strength) greatly influences the outcome of the analysis. Reproducibility Problems: Limited data accessibility, lack of sharing of analysis codes, and insufficient transparency in reporting the methodology hinder the reproducibility of bibliometric studies. This negatively affects the verification and reliability of scientific results. Proposed Solutions: Reporting the rationale for parameter selections (GLOBAL Clause 14), conducting robustness tests with different parameters, ensuring data and code sharing (GLOBAL Clauses 19, 30), and reporting the methodology in full detail (GLOBAL Clauses 8-20) are necessary to overcome these challenges.

8.3. Risks of Metric Manipulation

Bibliometric metrics can be artificially inflated using various manipulation methods listed in Table 8, threatening the reliability of evaluation processes.

Table 8
Main Types of Manipulation

Types of Manipulation	Explanation	Examples
Citation Gangs	Citation between specific researchers	Researchers A, B, and C forcing each other to cite each other
Self-Citing	Over-citing of one's own work by an author	Inflating the h-index and total citation count
Mandatory Citation	Editors or reviewers forcing citations of their own work	Requesting the inclusion of specific citations in the peer review report
Fake Publications	Publishing in fake journals	Exploiting security vulnerabilities in open access systems
Citation Trading	Buying or exchanging citations for commercial purposes	Intermediaries charging a specific fee per citation

The solution suggestions for the manipulations in Table 8 are effective strategies to reduce the risks of manipulation, reporting self-citation rates, evaluating citation quality, using multiple metrics together, and encouraging transparency through data sharing.

8.4. Misuse of Metrics

Using bibliometric metrics outside their intended context or overinterpreting them can lead to serious misjudgments, as given in Table 9.

Table 9
Main Types of Misuse

Wrong Type of Use	Explanation	Example
Fallacy of Causation	Citation count = Research quality	Interpreting correlation as causality
One-Dimensional Evaluation	Over-focusing on a single metric	Evaluating based solely on h-index or JIF
Out-of-Context Evaluation	Ignoring disciplinary and career differences	Comparing researchers in different fields using the same metrics

Solution proposals for Table 9 (Leiden Manifesto and DORA): The Leiden Manifesto [18] offers ten fundamental principles for the responsible use of metrics in research evaluation. The DORA [19] statement highlights that it is wrong to use the journal impact factor as the only measure of the quality of the researcher or the article. According to these principles, it is important to use several indicators simultaneously, to choose normalized ones and to analyze the metrics in the context.

8.5. Future Trends

With the development of artificial intelligence and machine learning technologies, it becomes possible to substantially extend the scope of bibliometric studies [45, 46]. Data cleaning, advanced text analysis (BERT, GPT), automated reports and summaries, real-time analysis and prediction of citations are some of the applications of such technologies that enhance the quality and depth of bibliometric studies. At the same time, there are certain dangers associated with the application of these technologies such as algorithmic bias, non-transparency, data privacy issues and over-automation.

The open science movement encourages bibliometric analyses to be more transparent, repeatable, and accessible [47]. Open databases such as OpenAlex (~250 million documents) and Dimensions, open citation data such as OpenCitations (~1.5 billion citations), and open reporting standards such as the GLOBAL guide contribute to the democratization of bibliometric analyses. Initiatives such as CoARA (Coalition for Advancing Research Assessment) and the Barcelona Declaration promote the use of transparent and responsible metrics in research assessment. Sub metrics are indicators that complement traditional citation-based metrics, measuring interactions across non-traditional sources such as social media, news, blogs, and policy documents [15]. They offer advantages such as rapid feedback, measurement of social and practical impact beyond traditional academic influence and capturing the interactions of different user groups. However, weaknesses such as reliability, standardization, manipulation, and scope should also be considered. It is recommended that sub metrics be used in conjunction with traditional metrics and that their limitations be reported transparently.

9. Discussion

This research analyzed the ecosystem of bibliometric analysis in a holistic manner along four major dimensions (tool, method, application, and reporting). According to the results, bibliometric analysis is becoming increasingly significant in academia, but the tools must be applied in an ethical, responsible and transparent way.

There is great diversity of bibliometric analysis tools, which allows choosing different options, but it poses a problem of determining which tool is more appropriate for what task. Nevertheless, restrictions on data cleaning and thorough analysis imply the use of Bibliometrix and pyBibX.

The rise of AI-powered tools heralds a significant transformation in the future of bibliometric analyses. As highlighted by Thayyib *et al.,* [45], these tools offer deeper insights with capabilities such as text analysis, topic modeling, and automated summarization. The integration of Bibliometrix and Biblioshiny with the R ecosystem offers a significant advantage for reproducible research and provides an approach consistent with the recommendations of the GLOBAL guideline Ng *et al.,* [10].

Seven fundamental methodologies used in bibliometric analyses are designed to answer different research questions. While traditional methods such as citation analysis and co-citation analysis are effective in revealing the intellectual structure of research areas, bibliographic matching and trend analysis are more advantageous in identifying current research trends. Next-generation methods such as text analysis and topic modeling offer researchers' deeper insights through the integration of artificial intelligence and natural language processing technologies. Transparent reporting of parameter selections, robustness testing, and data and code sharing are required to ensure methodological robustness [10].

Bibliometric analyses are used in six main application areas. Research evaluation is the most common area and provides indispensable tools for performance measurement at individual, institutional, and national levels [44]. In the field of science policy and strategic planning, it supports critical decisions such as determining research priorities and allocating funds. However, with the widespread use of bibliometric analyses, risks such as metric manipulation and misuse of metrics are also increasing. Initiatives such as the Leiden Manifesto [18], DORA [19] and GLOBAL [10] provide important guidance in managing these risks.

Reporting quality is critical for the reliability and reproducibility of bibliometric studies. The study by Ng *et al.,* [10] revealed significant shortcomings in the reporting quality of bibliometric analyses. Thirty-two components of the GLOBAL guidelines are available to provide an overall checklist for improving transparency, comprehensiveness, and reproducibility in the process of bibliometric study reporting. The most highlighted components are method reporting, providing a complete search strategy, and specifying data sources, whereas the most overlooked component is data cleaning methodology.

There are three main trends that will affect the development of the future bibliometric analyses. Artificial intelligence will make analyses more efficient through applications like automatic data cleaning, text mining, and real-time analysis. Open science movement will contribute to the democratization of bibliometric analyses by providing open data resources like OpenAlex and OpenCitations. Altmetrics, on the other hand, will make bibliometric analyses more holistic and provide measurement of interactions from alternative sources like social media and policy papers. The results obtained in this study are in line with previous studies. Donthu *et al.,* Results of [44], Ng *et al.,* [10] and Moral Muñoz *et al.,* [24] confirm our results. In summary, bibliometric analyses can be considered as useful tools in the process of achieving research excellence. Nevertheless, following guidelines like GLOBAL, Leiden Manifesto, and DORA in order to use these tools ethically and responsibly is the joint responsibility of researchers, organizations, and policymakers.

10. Conclusion

This study adds value to the field of bibliometrics through providing an integrative approach towards understanding the current situation of the bibliometric field. Differing from other review studies, this study provides an integrative view of bibliometrics analysis considering four interlinked dimensions of analysis – analytical tools used, methods applied, fields of applications, and reporting quality. Through this approach, the study presents a complete framework covering the entire process of bibliometrics starting from data collection to analysis and reporting and even research evaluation. Additionally, the study builds on existing information through incorporation of recent issues in the field such as the GLOBAL reporting guideline, open science movement, artificial intelligence, semantics analysis, and altmetrics. Through the consideration of these emerging trends within the integrated framework, the study can offer not only a theoretical understanding but also practical advice on how to conduct high-quality bibliometric studies.

It is now possible to answer the research questions that guided the study based on the detailed analysis undertaken. In respect to the first research question about the tools utilized in bibliometric analysis and their pros and cons, this study indicated that many tools are available, such as VOSviewer, Bibliometrix/Biblioshiny, CiteSpace, Gephi, pyBibX, ScientoPy, Pajek, SciMAT, HistCite, Publish or Perish, and Dimensions Analytics. VOSviewer scores highly in the aspect of visualization due to its ease of use and visual excellence; Bibliometrix is characterized by comprehensive and reproducible analysis within an R ecosystem, while pyBibX has state-of-the-art AI-based text analysis techniques. In addition, CiteSpace specializes in the analysis of trends and bursts, whereas Gephi is useful in network visualization. The choice of tools will depend on various factors like type of analysis, volume of data, skill of the researcher, and need for reproducibility.

With regards to the second research question on methodological approaches and their validity, seven basic methodological approaches used for conducting bibliometric studies have been discovered. Citation analysis is used to gauge the influence and knowledge circulation in the sciences. Co-citation analysis helps to see the intellectual structure and knowledge base of research areas using a retrospective and static approach. Bibliographic coupling analyzes research front and the structure of research using a contemporaneous and dynamic approach. Co-authorship analysis uses social network theory to understand collaborative processes and research networks. Co-word analysis identifies the thematic structure and conceptual relations of research areas. Artificial intelligence and natural language processing techniques are used for conducting text analysis and topic modeling. Trend analysis and burst detection identify temporal dynamics and turning points and make future projections. Each methodological approach has its own strength and weakness, and researchers need to choose the most suitable approach according to their research question while making explicit the choice made by them. Validity and reliability of the methodology can be ensured through explicit reporting of parameter choices, robustness tests with alternative parameter choices and sharing of data and code.

Regarding the third research question concerning application areas and contributions to research evaluation, the paper revealed that there are six major application areas where bibliometric analyses are utilized. Research evaluation is the most common application area. It helps in individual level evaluation (promotion, appointments, funding applications) as well as in institutional level evaluation (ranking, performance evaluation) and national level evaluation (science policies, competitiveness analysis). The application areas of science policy and strategic planning include the utilization of bibliometric analyses in defining research priorities, allocating funds, creating new research facilities, and making personnel plans. The application areas of scholarly publishing include journal evaluation, ranking, and development using bibliometric analysis techniques. Library and information management make use of bibliometric analyses in developing the collection, making journal subscriptions, and providing advice to researchers. Science mapping and trend analysis shows intellectual, social, and conceptual structure of research fields, detects research trends, and provides projections for the future. Collaboration and network analysis involves studying research collaboration, knowledge flows, interdisciplinary, and international collaboration.

About the fourth research question on the issue of reporting quality and solutions towards achieving transparent, comprehensive, and reproducible reporting, the study reveals that the GLOBAL guideline created by Ng *et al.,* [10] offers a systematic 32-item checklist for the purpose of reporting bibliometric analyses. A scoping review performed by Ng *et al.,* has shown the significant problems associated with the reporting quality of bibliometric studies as only 23 records among 48,750 screened met the inclusion criteria with 74% of these being characterized by providing recommendations based on opinions rather than evidence. The most discussed items are those concerned with transparent and reproducible reporting of methods, presentation of the search strategy in its completeness, and data sources. The least frequently addressed item is the issue of data cleaning methods that were included in the list due to expert opinion. Some recommendations for researchers would be to follow the GLOBAL 32 items checklist, use PRISMA-ScR flow diagrams to depict data collecting process, make public data and code on publicly available websites like OSF, Zenodo, and GitHub, as well as report limitations transparently in accordance with GLOBAL Item 27. The scientific value of the bibliometric analysis lies not only in its results but in the way these results are reported as well.

Finally, on the fifth research question related to the evolution of bibliometrics in the era of artificial intelligence, big data, and open science, the study revealed that there are three key trends that would influence the future of bibliometrics. Artificial intelligence and machine learning would increase the scope and depth of bibliometric studies by enabling data cleaning, analysis of texts using big data and language models like BERT and GPT, automated report generation and summarization, real-time analysis, citation prediction models, and semantic science mapping.

The open science and open data movement is promoting more transparent, reproducible, and accessible bibliometric analyses through open databases such as OpenAlex (~250 million documents) and Dimensions, open citation data such as OpenCitations (~1.5 billion citations), open reporting standards such as GLOBAL, and initiatives such as the Barcelona Declaration and CoARA. Altmetrics and social impact measurement are complementing traditional citation-based metrics by capturing interactions from social media, news, blogs, and policy documents. These changes will change research assessment processes, but there are risks associated with algorithm bias, transparency issues, privacy of data, and overreliance on automation that should be mitigated via the human-in-the-loop method of using artificial intelligence outputs.

Based on the results of the current study, some recommendations can be put forward for different stakeholders. First, researchers are to choose appropriate tools based on analysis types, data sets, and their technical competence. VOSviewer can be used for visualization, Bibliometrix for analysis in general, and pyBibX for text analysis. Furthermore, researchers are advised to use different tools for different aspects of bibliometric analysis to obtain more complete results. It is highly important for researchers to pay considerable attention to data cleaning, using techniques of deduplication, name standardization of authors, and institutional affiliation standardization, and report these activities according to the GLOBAL guideline. Methodological transparency should be achieved through the reporting of parameters used for the analyses, e.g., thresholds, algorithms, and normalizations, providing justification for each of them, and performing robustness tests with different parameters. Researchers must follow the GLOBAL guideline and use it as a checklist to conduct and report their studies transparently, comprehensively, and reproducibly, and visualize the data collection process using the PRISMA-ScR flow chart. Sharing data and code on open repositories such as OSF, Zenodo, or GitHub contributes to reproducibility, and the software versions used should be indicated. Limitations should be reported openly in accordance with the GLOBAL Item 27, and causation should not be claimed. Finally, researchers are to adhere to ethical standards of the Leiden Manifesto and DORA by abstaining from metric manipulation practices such as citation cartels, coercive citations, and self-citations.

It would be wise to use a multidimensional evaluation system that does not rely on one indicator, including h-index and JIF. Such an evaluation process will need to include qualitative evaluation and will involve the application of field-normalized metrics (such as Field-Weighted Citation Impact (FWCI) and Source Normalized Impact per Paper (SNIP)). In terms of implementing the open science policy, institutions need to make sure that their publications are open access, share the data and code used, and follow the guidelines presented in the GLOBAL policy. Researcher training needs to include learning about bibliometric and data sciences within the graduate curriculum, as well as providing researchers with training in data management and open science. Finally, institutions should diversify their sources of data by utilizing databases, such as OpenAlex and Dimensions, alongside the WoS and Scopus databases.

It is important for policy makers to implement reform of the research evaluation system via implementation of the Leiden manifesto and DORA guidelines within national policies, which include moving away from the assessment of research via Journal Impact Factor and including standards of research quality and impact on society. Policy makers must ensure funding for infrastructure of open science by promoting open access databases like OpenAlex and Dimensions, and open access publishing. Additionally, research funds should be allocated strategically by identifying priority areas through bibliometric analysis, investing in emerging fields, and creating mechanisms that foster international collaboration.

This study has several limitations that should be acknowledged. Regarding scope limitations, the study examined the bibliometric analysis ecosystem across the dimensions of tools, methodologies, applications, and reporting quality, while dimensions such as historical development, in-depth ethical issues, and the social psychology of citation behavior were outside the scope. In terms of data limitations, the study is based on secondary literature through systematic literature review and synthesis, with no primary data collected. The sources used are predominantly English-language, and significant studies in other languages may have been excluded. The data collection period reflects the current state as of 2025. Methodologically, the systematic literature review was conducted using thematic synthesis and does not include meta-analysis or quantitative synthesis. The search strategy was limited to selected keywords, and in-depth case studies are not included. Regarding interpretation limitations, the interpretations are limited by the authors' expertise and perspective, and researchers from different disciplines may offer different interpretations. Predictions regarding future trends are based on current trajectories and involve uncertainty. Finally, regarding currency limitations, bibliometric analysis tools and data sources are continuously updated, and future developments may necessitate updates to some of the findings.

A bibliometric analysis can serve as an important tool for navigation towards research excellence. However, it must be properly understood, placed into proper context and critically evaluated. The purpose of this article is to provide an overview for those involved in bibliometric analysis as well as to contribute to research excellence.

Acknowledgement
This research was not funded by any grant.

Conflicts of Interest
The authors declare no conflicts of interest.

References
[1] Bornmann, L., Haunschild, R., & Mutz, R. (2021). Growth rates of modern science: a latent piecewise growth curve approach to model publication numbers from established and new literature databases. Humanities and Social Sciences Communications, 8(1), 224. https://doi.org/10.1057/s41599-021-00903-w
[2] Bascur, J. P., Verberne, S., van Eck, N. J., & Waltman, L. (2025). Which topics are best represented by science maps? An analysis of clustering effectiveness for citation and text similarity networks. Scientometrics, 130(2), 1181-1199. https://doi.org/10.1007/s11192-024-05218-6
[3] Thelwall, M. (2026). Large language models and responsible research evaluation: an extension of the Leiden Manifesto. Scientometrics, 1-23. https://doi.org/10.1007/s11192-026-05552-x
[4] Cicero, T. (2026). Interpreting Bibliometric Indicators as the "Blood Tests" of Research Systems. Publications, 14(1), 9. https://doi.org/10.3390/publications14010009
[5] Nowakowska, M. (2025). A comprehensive approach to preprocessing data for bibliometric analysis. Scientometrics, 130(9), 5191-5225. https://doi.org/10.1007/s11192-025-05415-x
[6] Leoni, M. L. G., Mercieri, M., Viswanath, O., Cascella, M., Rekatsina, M., Pasqualucci, A., ... & Varrassi, G. (2025). Neuropathic pain: a comprehensive bibliometric analysis of research trends, contributions, and future directions. Current Pain and Headache Reports, 29(1), 73. https://doi.org/10.1007/s11916-025-01384-1
[7] Torres-Salinas, D., Orduña-Malea, E., Delgado-Vázquez, Á., Gorraiz, J., & Arroyo-Machado, W. (2024). Foundations of narrative bibliometrics. Journal of informetrics, 18(3), 101546. https://doi.org/10.1016/j.joi.2024.101546

[8] Kumar, R. (2025). Bibliometric analysis: comprehensive insights into tools, techniques, applications, and solutions for research excellence. Spectrum of Engineering and Management Sciences, 3(1), 45-62. https://doi.org/10.31181/sems31202535k
[9] Culbert, J. H., Hobert, A., Jahn, N., Haupka, N., Schmidt, M., Donner, P., & Mayr, P. (2025). Reference coverage analysis of OpenAlex compared to Web of Science and Scopus. Scientometrics, 130(4), 2475-2492. https://doi.org/10.1007/s11192-025-05293-3
[10] Ng, J. Y., Liu, H., Masood, M., Syed, N., Stephen, D., Ayala, A. P., ... & Moher, D. (2025). Guidance for the reporting of bibliometric analyses: A scoping review. Quantitative Science Studies, 6, 988-1001. https://doi.org/10.1162/QSS.a.12
[11] Montazeri, A., Mohammadi, S., Hesari, P. M., Ghaemi, M., Riazi, H., & Sheikh-Mobarakeh, Z. (2023). Preliminary guideline for reporting bibliometric reviews of biomedical literature (BIBLIO): A minimum requirement. Systematic Reviews, 12, 239. https://doi.org/10.1186/s13643-023-02410-2
[12] Sahoo, S. K., Fischer, S., Sahu, S., Choudhury, B. B., Dhal, P. R., & Dhar, I. (2026). Mapping Tools, Techniques, and Applications for Research Excellence: A Bibliometric Study. Spectrum of Engineering and Management Sciences, 1-26. https://doi.org/10.31181/sems41202672s
[13] Ninkov, A., Frank, J. R., & Maggio, L. A. (2022). Bibliometrics: methods for studying academic publishing. Perspectives on medical education, 11(3), 173-176. https://doi.org/10.1007/s40037-021-00695-4
[14] Al-Jamimi, H. A., BinMakhashen, G. M., & Bornmann, L. (2022). Use of bibliometrics for research evaluation in emerging markets economies: a review and discussion of bibliometric indicators. Scientometrics, 127(10), 5879-5930. https://doi.org/10.1007/s11192-022-04490-8
[15] Hassan, W., & Duarte, A. E. (2024). Bibliometric analysis: a few suggestions. Current problems in cardiology, 49(8), 102640. https://doi.org/10.1016/j.cpcardiol.2024.102640
[16] Oyewola, D. O., & Dada, E. G. (2022). Exploring machine learning: a scientometrics approach using bibliometrix and VOSviewer. SN Applied Sciences, 4(5), 143. https://doi.org/10.1007/s42452-022-05027-7
[17] Klarin, A. (2024). How to conduct a bibliometric content analysis: Guidelines and contributions of content co-occurrence or co-word literature reviews. International Journal of Consumer Studies, 48(2), e13031. https://doi.org/10.1111/ijcs.13031
[18] Hicks, D., Wouters, P., Waltman, L., de Rijcke, S., & Rafols, I. (2015). Bibliometrics: The Leiden Manifesto for researching metrics. Nature, 520(7548), 429-431.
[19] DORA. (2012). San Francisco Declaration on Research Assessment. https://sfdora.org/
[20] AlRyalat, S. A. S., Malkawi, L. W., & Momani, S. M. (2019). Comparing bibliometric analysis using PubMed, Scopus, and Web of Science databases. JoVE (Journal of Visualized Experiments), (152), e58494. https://doi.org/10.3791/58494
[21] Gusenbauer, M. (2019). Google Scholar to overshadow them all? Comparing the sizes of 12 academic search engines and bibliographic databases. Scientometrics, 118(1), 177-214. https://doi.org/10.1007/s11192-018-2958-5
[22] McGowan, J., Sampson, M., Salzwedel, D. M., Cogo, E., Foerster, V., & Lefebvre, C. (2016). PRESS peer review of electronic search strategies: 2015 guideline statement. Journal of clinical epidemiology, 75, 40-46. https://doi.org/10.1016/j.jclinepi.2016.01.021
[23] Tricco, A. C., Lillie, E., Zarin, W., O'Brien, K. K., Colquhoun, H., Levac, D., ... & Straus, S. E. (2018). PRISMA extension for scoping reviews (PRISMA-ScR): checklist and explanation. Annals of internal medicine, 169(7), 467-473. https://doi.org/10.7326/M18-0850
[24] Moral Muñoz, J. A., Herrera-Viedma, E., Santisteban Espejo, A. L., & Cobo Martín, M. J. (2020). Software tools for conducting bibliometric analysis in science: An up-to-date review. https://doi.org/10.3145/epi.2020.ene.03
[25] 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
[26] Lim, W. M., & Kumar, S. (2024). Guidelines for interpreting the results of bibliometric analysis: A sensemaking approach. Global Business and Organizational Excellence, 43(2), 17-26. https://doi.org/10.1002/joe.22229
[27] Stefanis, C., Stavropoulou, E., Stavropoulos, A., Gyriki, D., Nikolaidis, C. G., Vassos, V., ... & Bezirtzoglou, E. (2025). Innovative methodological pillars for bibliometric studies: AI screening, data normalization and dual-tool analysis. Discover Applied Sciences, 7(9), 973. https://doi.org/10.1007/s42452-025-07653-3
[28] García-Villar, C., & García-Santos, J. M. (2021). Bibliometric indicators to evaluate scientific activity. Radiología (English Edition), 63(3), 228-235. https://doi.org/10.1016/j.rxeng.2021.01.002
[29] Mahapatra, N., & Sahoo, J. (2023). Metrics employed in the evaluation of research productivity: A systematic literature review. Journal of Librarianship and Information Science, 55(3), 868-885. https://doi.org/10.1177/09610006221104798
[30] Roldan-Valadez, E., Salazar-Ruiz, S. Y., Ibarra-Contreras, R., & Rios, C. (2019). Current concepts on bibliometrics: a brief review about impact factors, Eigenfactor score, CiteScore, SCImago Journal Rank, Source-Normalised Impact

per Paper, H-index, and alternative metrics. Irish Journal of Medical Science (1971-), 188(3), 939-951. https://doi.org/10.1007/s11845-018-1936-5

[31] Ahmad, I., Shahid, A., Khan, S. U., Hussain, T., Akbar, W., Attar, R. W., & Alhazmi, A. H. (2026). Comparative analysis of h-index variants using an extensive dataset. Scientometrics, 1-22. https://doi.org/10.1007/s11192-025-05520-x

[32] Dias, A., Ruthes, S., Lima, L., Campra, E., Silva, M., Bragança de Sousa, M., & Porto, G. (2020). Network centrality analysis in management and accounting sciences. RAUSP Management Journal, 55(2), 207-226. https://doi.org/10.1108/RAUSP-02-2019-0021

[33] Kanchan, T., & Krishan, K. (2019). The Leiden manifesto and research assessment. Science and Engineering Ethics, 25(2), 643-644. https://doi.org/10.1007/s11948-017-0012-2

[34] Clarivate. (2025). Web of Science Core Collection. https://clarivate.com/webofsciencegroup/solutions/web-of-science-core-collection/

[35] Elsevier. (2025). Scopus. https://www.elsevier.com/solutions/scopus

[36] Digital Science. (2025). Dimensions. https://www.dimensions.ai/

[37] NCBI. (2025). PubMed. https://pubmed.ncbi.nlm.nih.gov/

[38] Priem, J., Piwowar, H., & Orr, R. (2022). OpenAlex: A fully open index of scholarly works, authors, venues, institutions, and concepts. arXiv preprint arXiv:2205.01833. https://doi.org/10.48550/arXiv.2205.01833

[39] IEEE. (2025). IEEE Xplore Digital Library. https://ieeexplore.ieee.org/

[40] arXiv. (2025). arXiv.org e-Print Archive. https://arxiv.org/

[41] CrossRef. (2025). Crossref.org. https://www.crossref.org/

[42] OpenCitations. (2025). OpenCitations.net. https://opencitations.net/

[43] Heibi, I., Peroni, S., & Shotton, D. (2019). Software review: COCI, the OpenCitations Index of Crossref open DOI-to-DOI citations. Scientometrics, 121(2), 1213-1228. https://doi.org/10.1007/s11192-019-03217-6

[44] 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

[45] Thayyib, P. V., Mamilla, R., Khan, M., Fatima, H., Asim, M., Anwar, I., ... & Khan, M. A. (2023). State-of-the-art of artificial intelligence and big data analytics reviews in five different domains: a bibliometric summary. Sustainability, 15(5), 4026. https://doi.org/10.3390/su15054026

[46] pyBibX. (2025). https://github.com/valdecy/pybibx

[47] Barcelona Declaration on Open Research Information. (2023). https://barcelona-declaration.org/