Editorial

The KEYWORDS Framework: Standardizing Keyword Selection for Improved Big Data Analytics in Biomedical Literature

*I*n the fast-paced world of modern science, where the explosion of Big Data has revolutionized research, the selection of appropriate keywords for scientific manuscripts continues to be one critical yet often-overlooked detail. While keywords have traditionally served as indexing and search engine tools, their importance now extends far beyond these simple functions.[1,2] Keywords are becoming the building blocks of Big Data analyses, such as bibliometric analyses, in the biomedical field, yet the approach to choosing keywords remains remarkably inconsistent and heavily based on the authors' judgment[3] as researchers are rarely provided with a clear guidance on selecting the most impactful terms. The editorial aims to draw attention to the importance of a systematic, standardized approach in keyword selection and propose a framework to guide researchers in choosing the most effective terms.

The Evolving Role of Keywords in Research

As the field of bibliometrics continues to advance, researchers increasingly rely on keywords to map the intellectual landscape of scientific fields. Machine learning algorithms use keyword frequencies and associations to identify research trends, predict future research directions, and even assist in hypothesis generation.[1,4,5] In addition, well-chosen keywords can bridge disciplines, revealing unexpected connections.[4] Moreover, policymakers rely on Big Data analysis to make informed decisions, with keywords playing an imperative role in directing and filtering data toward relevant issues. Choosing appropriate keywords ensures policy decisions are grounded in accurate and relevant insights.

Why a Framework?

The reason why we need a framework for something as simple as keyword selection is the need for a consistent and uniform selection strategy. When keywords are chosen without a clear strategy, they become unreliable data points, making it difficult to conduct accurate and meaningful analyses across large datasets. This, in turn, limits the potential for large-scale analyses to yield meaningful insights.[5,6] A structured framework ensures that keywords consistently capture the core aspects of a study. The approach not only creates a more interconnected and easily navigable scientific literature

landscape but also facilitates comprehensive Big Data analyses, such as bibliometric studies, by enhancing the comparability of research and reducing missing data, ultimately leading to effective evidence synthesis across multiple studies.[1,2]

Introducing KEYWORDS Framework: An Answer to the Call for Standardization

As we move deeper into the era of Big Data, the research community needs to recognize that there is a need for standardization in the keyword selection process. Inspired by established frameworks for structuring research questions [PICO], systematic reviews [PRISMA], and qualitative research [SPIDER],[7-9] the KEYWORDS framework offers a structured approach to keyword selection:

K—Key concepts (Research Domain)
E—Exposure or Intervention
Y—Yield (Expected Outcome)
W—Who (Subject/sample/problem/phenomenon of interest)
O—Objective or Hypothesis
R—Research Design
D—Data analysis tools
S—Setting (Conducting site and setting)

The process of developing the KEYWORDS framework involves the following steps:

Identification of critical elements

Since there is no existing framework specifically for keyword selection, the following frameworks were selected as foundational frameworks for generating the KEYWORDS framework as they include critical elements that capture the core aspects of biomedical studies- Population, Intervention, Comparison, Outcome (PICO), Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA), and Sample, Phenomenon of Interest, Design, Evaluation, Research type (SPIDER). Critical elements from these three frameworks were identified and listed.[6-8]

Inspection and selection

Each element was carefully examined for its relevance to keyword selection in research studies. Appropriate elements that potentially capture the core aspects of biomedical research were selected.

Deduplication

Overlapping or duplicate elements across the guidelines were identified and removed.

Structuring and acronym creation

The selected elements were organized into a logical sequence that follows the typical flow of a research study and arranged to form the memorable acronym KEYWORDS, with each letter representing a crucial aspect of the study for keyword selection. By systematically covering key elements of a study, the KEYWORDS framework helps ensure systematic, consistent, uniform, and relevant keyword selection.

PRACTICAL APPLICATION OF THE KEYWORDS FRAMEWORK

At least eight relevant keywords, one from each category, are recommended for use. The selection of keywords should be based on the type of the study. for example, in original research, the sample refers to research subjects and the intervention is treatment provided, whereas in bibliometrics, the sample consists of research publications and the intervention is analogous to the data analysis techniques used to evaluate the publications. The examples of applying the KEYWORD framework for different study types are presented in Table 1. In addition, while selecting

Table 1: Applying the KEYWORDS framework for keyword selection for different study types		
Study type	Study description	Keyword suggestions according to the KEYWORDS framework
Experimental Study	**Title:** Effect of Probiotic Supplementation on Gut Microbiota Composition in Patients with IBS: An RCT **Description:** A study investigated the impact of probiotic supplementation on gut microbiota composition in IBS patients through an RCT conducted in a clinical setting. The intervention involved daily probiotic supplementation for 12 weeks. Data were analyzed using SPSS software.	**K**ey Concepts: Gut microbiota **E**xposure/ Intervention: Probiotics **Y**ield: Microbiota and Symptom Relief **W**ho: Irritable Bowel Syndrome **O**bjective: probiotics efficacy **R**esearch Design: Randomized Controlled Trial, Quantitative **D**ata Analysis Tools: SPSS **S**etting: Clinical Setting
Observational Study	**Title:** Experiences of Living with Chronic Pain: A Qualitative Study of Patient Narratives **Description:** A qualitative study exploring the experiences of chronic pain patients through semi-structured interviews conducted in a community setting focusing on the impact of chronic pain on their lives and coping strategies. Thematic analysis was done using NVivo software.	**K**ey Concepts: Chronic Pain **E**xposure: Daily Challenges **Y**ield: Coping Strategies, Quality of Life **W**ho: Chronic Pain Patients **O**bjective: Patient Experience **R**esearch Design: Qualitative Research, Observational Study, Thematic Analysis **D**ata Analysis Tools: NVivo **S**etting: Community setting
Review (Systematic Review)	**Title:** Systematic Review of Antimicrobial Resistance in Dental Biofilms **Description:** Systematic review synthesizing research on antimicrobial resistance in dental biofilms focused on studies published in PubMed and Scopus databases. Common resistance patterns and research gaps were identified. Meta-analysis was performed using RevMan.	**K**ey Concepts: Antimicrobial Resistance **E**xposure/ Intervention: Antimicrobial Agent **Y**ield: Resistance Patterns **W**ho: Dental Biofilms **O**bjective: Research Gaps, Drug Resistance **R**esearch Design: Systematic Review, Meta-Analysis **D**ata Analysis Tools: RevMan **S**etting: PubMed and Scopus
Bibliometric Analysis	**Title:** Trends and Impact of Clinical Trials on Oral Biofilm in Dental Medicine: A Bibliometric Analysis **Description:** This bibliometric study analyzes research trends and citation impacts of clinical trials on the oral biofilm from 2000 to 2023. Data were retrieved from citation databases such as Web of Science and Scopus. VOSviewer software was used for mapping research networks, while citation metrics (e.g., H-index and citation counts) were analyzed to assess the research impact.	**K**ey Concepts: Oral Biofilm, Dental Medicine **E**xposure/ intervention: Network Analysis, Citation Analysis **Y**ield: Citation Impact, Research Trends **W**ho: Clinical trials **O**bjective: H-index, Research Networks **R**esearch Design: Bibliometrics **D**ata Analysis Tool: VOSviewer **S**etting: Global, Web of Science, and Scopus

All the studies mentioned in the table are hypothetical examples.

RCT = randomized controlled trial, IBS = irritable bowel syndrome

keywords, it is important to balance the specificity and generality to ensure visibility and relevance. Also, using the standardized terminology, such as MeSH, enhances consistency for data analysis. Moreover, terms that bridge disciplinary boundaries can promote collaboration and innovation across fields.[10]

Strengths and Limitations of the Framework

The framework—KEYWORDS demonstrates the potential for high content validity due to its comprehensive coverage of crucial biomedical research elements and flexibility in accommodating various study designs. It ensures the inclusion of often-overlooked aspects of biomedical research and provides a more complete representation of the study in its keywords, which improves the integrity, utility, and comparability of data. Moreover, it is specifically designed for keyword selection, aiming to fill information gaps during Big Data analysis. While the framework is best suited for well-designed experimental studies, observational studies, reviews, and bibliometric analysis in the biomedical field, it is inappropriate for theoretical, opinion-based, descriptive, methodological, historical, or philosophical articles. Although based on established practices,[7-9] this framework requires testing and validation in real-world applications to confirm its effectiveness for its intended purpose. While developed for biomedical research, the framework may be adaptable to other fields with similar research structures but should be evaluated for appropriateness before application.

A Call to Action

Given the multifaceted importance of keywords, authors should approach their selection with the same rigor they apply to their research methodology. The KEYWORDS framework offers a simple yet comprehensive tool for achieving this goal. I urge researchers, editors, and publishers to adopt this approach, recognizing that something as seemingly simple as keyword selection can have a profound impact on the accuracy of Big Data analysis. By adopting the KEYWORDS framework, we can create more analyzable scientific literature, supporting the next wave of scientific discoveries driven by data analytics. Let us give keywords careful consideration they deserve as these are not merely metadata: they are rather a gateway to discoveries in our data-driven world.

Namrata Dagli

Center for Global Health Research, Saveetha Medical College, Saveetha Institute of Medical and Technical Sciences, Chennai, India

Address for correspondence: Dr. Namrata Dagli, Center for Global Health Research, Saveetha Medical College, Saveetha Institute of Medical and Technical Sciences, Chennai, India.
E-mail: dr.namrata.dagli@gmail.com

References

1. Pottier P, Lagisz M, Burke S, Drobniak SM, Downing PA, Macartney EL, *et al*. Keywords to success: A practical guide to maximise the visibility and impact of academic papers. bioRxiv 2023;2023:10.
2. Zhang Y, Chen F, Suk J, Yue Z. WordPPR: A researcher-driven computational keyword selection method for text data retrieval from digital media. Commun Methods Meas 2023:1-17. 10.1080/19312458.2023.2278177
3. Lu W, Liu Z, Huang Y, Bu Y, Li X, Cheng Q. How do authors select keywords? A preliminary study of author keyword selection behavior. Journal of Informetrics 2020;14:101066.
4. Abdullah KH, Roslan MF, Ishak NS, Ilias M, Dani R. Unearthing hidden research opportunities through bibliometric analysis: A review. Asian J Res Educ Soc Sci. 2023;5:251-62.
5. Shu X, Ye Y. Knowledge discovery: Methods from data mining and machine learning. Soc Sci Res 2023;110:102817.
6. Yasser CM. An analysis of problems in metadata records. J Libr Metadata. 2011;11:51-62. https://doi.org/10.1080/1938638 9.2011.570654.
7. Moher D, Liberati A, Tetzlaff J, Altman DG, Prisma G. Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement. Int J Surg 2010;8:336-41.
8. Schardt C, Adams MB, Owens T, Keitz S, Fontelo P. Utilization of the PICO framework to improve searching PubMed for clinical questions. BMC Med Inform Decis Mak 2007;7:1-6.
9. Cooke A, Smith D, Booth A. Beyond PICO: The SPIDER tool for qualitative evidence synthesis. Qual Health Res 2012;22:1435-43.
10. Bailey KD. Towards unifying science: Applying concepts across disciplinary boundaries. Syst Res Behav Sci 2001;18:41-62.