



A Bibliometric Analysis of the Minimum Data Set in Healthcare Research

Mahdie ShojaeiBaghini^{1*}

¹Information Management, Medical Informatics Research Center, Institute for Futures Studies in Health, Kerman University of Medical Sciences, Kerman, Iran

Abstract

Introduction: The minimum data set is the first important step in developing a health information system. Given that in recent years we have been faced with growth and accumulation in the minimum data set research, this study aims to perform a systematic library analysis of the minimum dataset and find out the field of research and trends.

Methods: The present study, examined the scientific publications in the minimum data set research in the PubMed on May 16, 2022, using scientific mapping and bibliometric approach. Searches were not limited by publication type, date, or language. The data were exported to Microsoft Excel 2016 and Mendeley Desktop version 1.19.8. After screening based on the PRISMA checklist, were performed bibliometric analysis and scientific mapping, using the RStudio package and the VOSviewer software tool.

Results: About 35% of the publications in the field have been published since 2017 and mainly in the Journal of the American Geriatrics Society and Journal of the American Medical Directors Association. The top and most popular authors were MOR V., Fries B.E., and J.N. Morris that had strong cooperation relationships. The United States also had the most scientific production, and most articles were published nationally. The analysis of the author's keywords also showed that the top five keywords were Humans, Aged, Female, Male, and Standards.

Conclusion: Although global studies on minimum dataset have a long history, they are still growing. Recently, it has shown promising applications in information systems and clinical research.

Keywords: MDS, Minimum Data Set, Bibliometric Analysis, Mapping, Network Analysis

Article History:

Received: 23 November 2023
Accepted: 25 December 2023

Please cite this paper

as:ShojaeiBaghini M. A
Bibliometric Analysis of the
Minimum Data Set in Healthcare
Research. Health Man & Info
Sci. 2024; 11(1): 1-8. doi:
10.30476/jhmi.2024.102269.1216.

*Correspondence to:

Mahdie ShojaeiBaghini,
Information Management, Medical
Informatics Research Center,
Institute for Futures Studies in
Health, Kerman University of
Medical Sciences, Kerman, Iran
Email: mahdiehsh@gmail.com

Introduction

The efficacy of medical decision-making is highly contingent upon the accuracy and timeliness of health information (1). In healthcare, the stakes of such decisions are invariably high, as the consequences of misinformed or inadequate patient health information can range from medical errors to suboptimal decision-making, ultimately leading to increased healthcare costs. The critical challenge lies in ensuring physicians have access to necessary and sufficient information about each patient. This balancing act between providing sufficient, yet effective, information is non-trivial, given the vast and varied nature of patient data (2). In response to this challenge, the concept of a Minimum Data Set (MDS) emerges as a pivotal tool (3). An MDS is designed to collect essential information that forms the backbone

of health information systems, facilitating disease information management (4, 5). The primary purpose of an MDS is to standardize data collection, ensuring data comparability and consistency across different settings. The comprehensiveness, accuracy, data comparison, quality of data, care quality, data audit and benchmarking care plans, and life quality can all be enhanced with an MDS (6-8). By ensuring the availability of high-quality data, an MDS enables healthcare providers to make more informed decisions, thereby improving patient outcomes. It also supports critical functions such as data auditing, care planning, and benchmarking against care standards. Additionally, the data collected through an MDS can be instrumental in enhancing the overall quality of life for patients by providing insights into effective care strategies and disease control measures (9, 10).

The burgeoning interest in MDS research is evidenced by the increasing publication of articles in this field over the years. These publications include original studies aiming at the development and application of MDS, as well as review studies that synthesize and evaluate existing research findings. Such scholarly efforts underscore the critical role of MDS in advancing healthcare research and practice (10-14).

While review studies synthesize and critique the existing studies (15), bibliometric analyses offer a statistical examination of research documents and authorship, providing a systematic and robust approach to analyzing extensive scientific data (16). In academia, bibliometric analysis has become a fundamental method for identifying research hotspots, global trends, and the impact of contributions from scholars, journals, and countries/regions (17-19). This approach not only aids in tracking the evolution and patterns of successful publications, but also explores the development of emerging fields and fosters multidisciplinary collaborations (19). Bibliometric analyses have proven invaluable for identifying new trends, assessing journal and article performance, examining collaboration patterns, and understanding the intellectual structure of specific research areas, particularly in health and biomedicine (18, 20). As a result, well-done bibliometric studies can assist academics in gaining a comprehensive picture of the field, identifying knowledge gaps, developing novel research ideas, and positioning their intended contributions to the field (19).

This study aimed to delineate the utilization, scope, trends, and emerging hotspots of MDS in healthcare research. Utilizing the PubMed citation database, known for its comprehensive coverage of biomedical and health sciences literature (21, 22), we adopted a quantitative bibliometric approach complemented by network visualization techniques. This study represents the first functional review and analysis of the citation network within MDS literature, with potential implications for guiding future research directions in this field. Accordingly, this paper sought to address the following research questions:

What are the most frequently cited articles, leading journals, prominent authors, foremost countries, and top universities in MDS research?

What are the predominant patterns of co-

citation, co-authorship, simultaneity, and lexical overlap in MDS research?

Methods

This study employed a bibliometric and scientific mapping approach to rigorously analyze the scientific publications in the Minimum Data Set research within the PubMed database as of May 16, 2022. This methodology was chosen for its capacity to evaluate the scientific production quantitatively and qualitatively in this domain, providing insights into research trends, key themes, and the overall impact of MDS studies on healthcare information management.

Data Collection

The search strategy was meticulously designed to encompass a broad spectrum of publications related to MDS, utilizing a combination of MeSH terms and keyword searches to ensure comprehensive coverage. The search query was formulated as follows: ((((((“common data element”[MeSH Terms]) OR (“Common Data Element”[Title/Abstract])) OR (“Data Element, Common”[Title/Abstract])) OR (“Data Elements, Common”[Title/Abstract])) OR (“Element, Common Data”[Title/Abstract])) OR (“Elements, Common Data”[Title/Abstract])) OR (“minimum dataset”[Title/Abstract])) OR (“minimum data set”[Title/Abstract])).

No limitation was imposed on the publication type, date, or language to maximize the inclusivity of the search results. The resulting data were saved in .csv format for subsequent analysis using Microsoft® Excel 2016 and organized using Mendeley Desktop version 1.19.8 for reference management.

Selection Criteria

From the initial search yield of 2718 records, the study focused on publications from 1979 to 2021. The decision to exclude articles published in 2022 was made to ensure the completeness of the dataset, given that the publication year was not yet concluded. Following the PRISMA checklist for systematic reviews, duplicates were identified and removed, resulting in 2707 records. For qualitative synthesis, only documents classified as articles were selected as they were deemed most likely to have undergone peer review before publication (23). Thus, conference articles, book chapters, letters to the editor, and editorials were

removed, and 2623 records remained. Further refinement was carried out by excluding 56 records with incomplete citation information or those lacking English titles, abstracts, or keywords, resulting in a final count of 2567 eligible studies. Two reviewers independently evaluated the study eligibility. The decision was made on whether the article fulfilled the eligibility based on the consensus between the two reviewers.

Statistical Analysis

Due to the limitations of PubMed regarding analyses based on addresses or citations, additional tools were employed for performance analysis and citation network analysis. Specifically, the RStudio package and VOSviewer software for Windows, version 1.6.18, were utilized. VOSviewer facilitated the visualization and navigation of network-based maps, enabling the identification of key themes and trends within MDS research. RStudio provided comprehensive tools for the statistical analysis of the dataset, allowing for a detailed examination of publication patterns, author contributions, and thematic clusters.

Results

The results of the study of the publication trend of scientific productions in the field of MDS showed that the first scientific document in this field was indexed in the PubMed citation database in 1979 and has gradually increased. Of course, it has increased significantly in recent years so that about 35% of MDS articles have been published since 2017. A more detailed study of the PubMed citation database showed that scientific documents of MDS were published in 886 journals. Of them, the Journal of the American Geriatrics Society and the Journal of the American Medical Directors Association published the largest number of articles (more than 15 percent of documents) in this field. Journal

of the American Geriatrics Society published by Wiley had an IF=7.538 and the Journal of the American Medical Directors Association published by Elsevier had an IF=7.802.

Figure 1(a) demonstrates the top 20 authors with the most academic publications. Lapane KL. with a total of 87 articles and Bernabei R. with 55 articles were the two top authors. Also, Figure 1(b) shows the amount of scientific production of each author over time. According to Mor V., Fries B.E. and J.N. Morris were the most experienced authors in the field of MDS.

More detailed analysis showed that based on the nationality of the corresponding author, 97 countries had participated in publishing research related to MDS. Figure 2(a) shows that the United States of America has the most scientific productions in the field of MDS by a large margin compared to other countries. Then, Australian, Canadian, and Iranian researchers had the most scientific outputs in MDS. Figure 2(b) shows the number of Single Country Publications (SCP) and Multiple Country Publications (MCP). Accordingly, Australia and Canada had the highest MCP ratio with an MCP ratio of 0.18.

Figure 3(a) illustrates the most preferred title words used in the titles of articles. Among the top 20 keywords, “nursing home” is repeated in the title of 648 articles, “minimum data” in 428, and “home resident” in 403 article titles. The trend of title keywords over time, based on the computing of word weight, showed that “nursing home”, which has been used in the title of MDS articles since 2004, had the highest weight in 2011. After that, “minimum” and “set” have been the most weighted since 2001 in the titles of articles used in 2007 (Figure 3(b)).

Co-occurrence analysis of words is used to measure the degree of content relationship between two words, even hidden words, and they highlight the main topics and concepts of the research field.

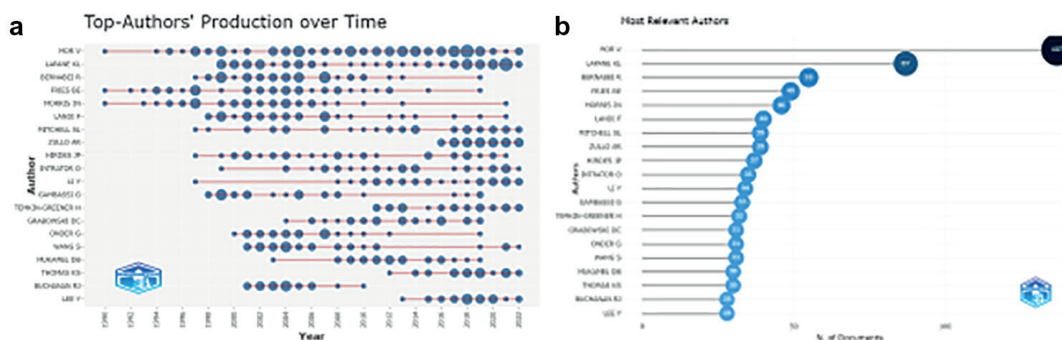


Figure 1: (a) Number of scientific productions in the field of MDS by author (b) Scientific productions of authors by year

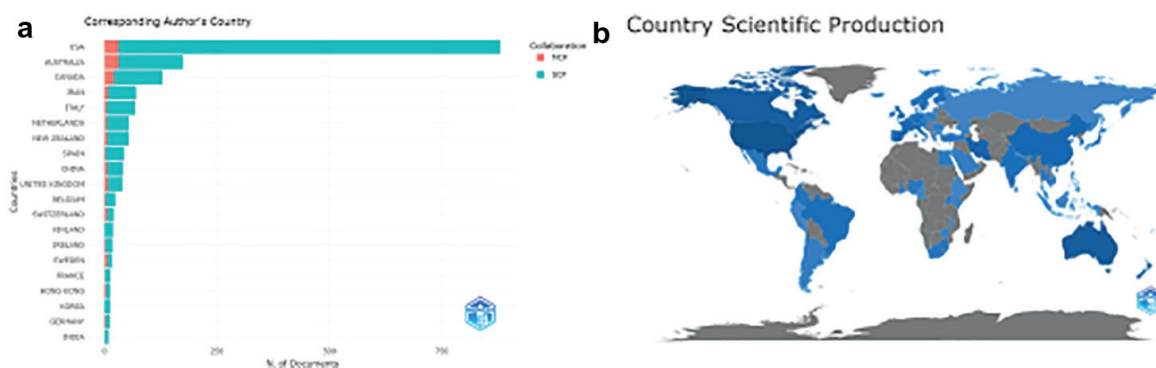


Figure 2: (a) Geographical distribution of countries in scientific production of MDS in the PubMed database (b) The most productive countries based on single-country publications (SCP), and multi-country publications (MCP)

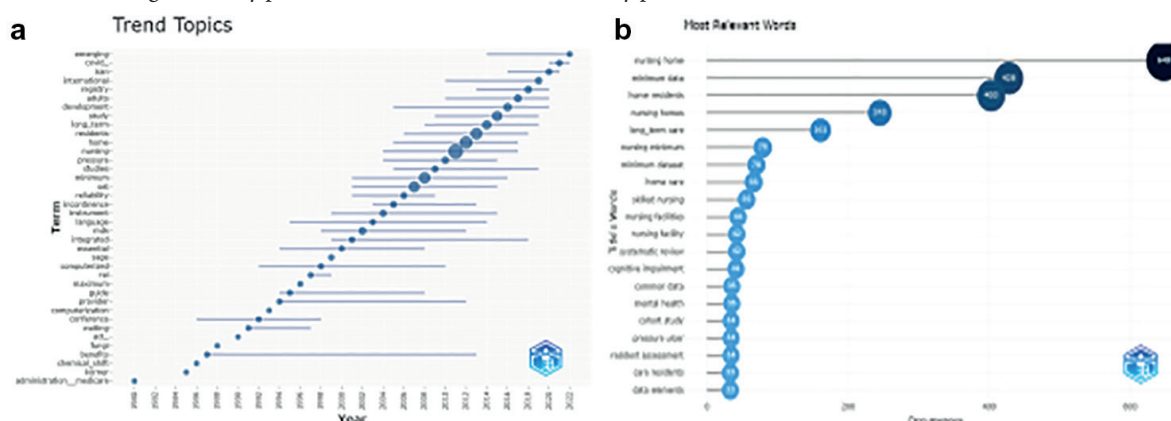


Figure 3: (a) Most common keywords in the titles of MDS articles (b) Trend of keywords in the titles of MDS articles, over time, based on word weight calculation

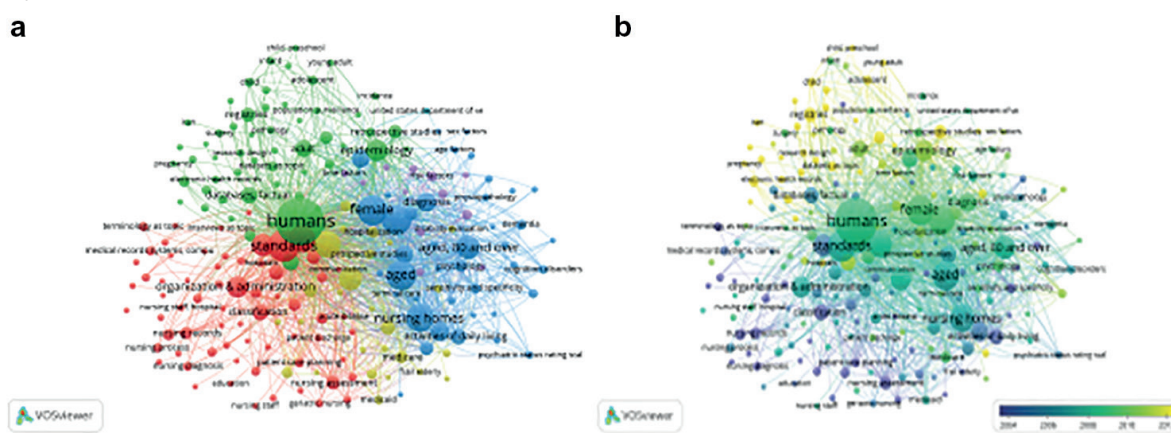


Figure 4: (a) Keywords author's network visualization (b) Overlay visualization by time (year).

Therefore, to better understand the concepts of the MDS field, co-occurrence analysis of words was also used. The co-occurrence map of the author's keywords was done using VOSviewer software. For this purpose, the number of repetitions of the author's keywords was considered 10 times. Of the total 1439 author keywords, only 201 were repeated at least 10 times in the articles. The results were presented in the form of 6 main clusters. Based on Total Link Strength (TLS=Total Link Strength), the author's top five keywords were: human (TLS=6799), age (TLS=3668), female (TLS=3633),

male (TLS=3567), and standard (TLS=3090) (Figure 4(a)). Figure 4(b) shows the co-occurrence analysis of the words based on time and over the years. In network and cluster analysis, the color indicates a specific research cluster, and the circle dimension indicates how often the keywords are repeated in the document. The thickness of the line connecting the circles demonstrates the intensity of the correlation between the keywords (24).

In addition, the co-authorship of the researchers was analyzed. Co-authorship means the participation of two or more authors in the

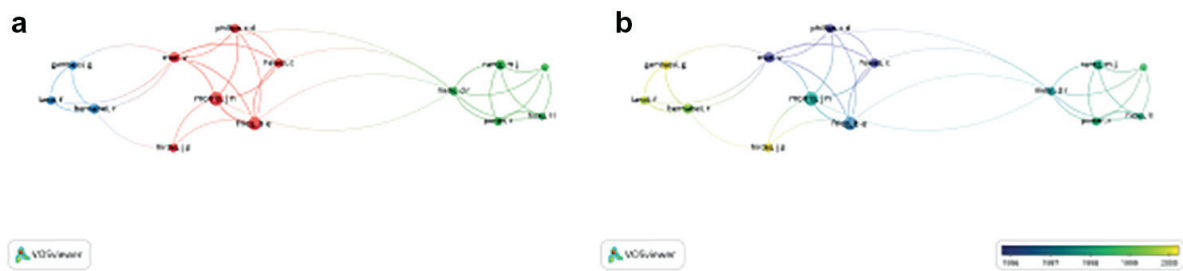


Figure 5: (a) Researchers' co-authorship network visualization (b) Overlay visualization by time (year)

production of a scientific publication. As shown in Figure 4, the author-based co-authorship map presents 14 items in three main identified clusters. The size of the circles indicates the number of documents, and the thickness of the lines indicates the amount of collaboration between the authors (25).

The maximum number of authors per article was 25, and the minimum number of articles per author was 5. In total, out of 3817 authors, only 48 reached this threshold. According to TLS J.N. Morris was (TLS=35). Then, Fries B.E. and Mor V. each with (TLS=33), Haves C. (TLS=30), and Mehr D.R (TLS=23) were in the next positions. J.N. Morris, Fries B.E., Mor V. are among the oldest authors in this field and have strong collaborative relationships. Of course, most authors had a relationship with Mehr D.R. (Figure 5).

Discussion

This bibliometric analysis delineates the evolving landscape of MDS research, revealing an upward trajectory in scientific outputs since 1986, as indexed by PubMed. This increase underscores the burgeoning interest in leveraging MDS to enhance clinical research, reflecting its growing importance in the medical and research communities. The analysis of 2623 articles highlighted not only the quantitative growth of MDS-related studies but also the qualitative advancements in this field, signifying the critical role of MDS in healthcare research and information management (8, 26).

Network analysis and scientific mapping emerged as pivotal for assessing the breadth and depth of MDS research. Network analysis and scientific maps can be considered one of the most effective methods for reviewing and analyzing literature (27). These methods, characterized by their complexity and the necessity for sophisticated mapping tools, facilitate a

comprehensive review of the literature (28). One of the scientific maps is co-authorship network analysis, which leads to the production of shared public documents. Scientific collaboration is a complex phenomenon that improves the sharing and production of scientific documents (29). Co-authorship is one of the most documented forms of scientific collaboration, which can be useful for planning future scientific collaborations (30). Co-authorship networks are a powerful tool for evaluating the collaborative process and identifying leading authors and organizations (31). Co-authorship network analysis has proven instrumental in understanding the dynamics of scientific collaboration, revealing that such partnerships are integral to the generation and dissemination of knowledge within the MDS domain.

Co-authorship network analysis showed that Morris, John N. had the strongest co-authorship relationships with the H index of 63. Researchers prefer to communicate with researchers with a high h-index (32) because they can share ideas, resources, and information by creating communication networks, producing and presenting new knowledge, and ultimately creating innovation, reducing costs, and increasing research productivity (31). This trend underscores the value of collaboration in fostering innovation, reducing research costs, and enhancing productivity. Furthermore, our findings suggest that authors, especially from developing countries, often engage in international collaborations to bolster the credibility and visibility of their work. These partnerships not only facilitate publication in prestigious journals but also pave the way for educational and professional development opportunities, thereby contributing to a more interconnected and resourceful global research community.

In bibliometric studies, keyword analysis is an important topic that can summarize the content

of articles. Furthermore, keyword analysis can be used to identify research trends as well as to determine interdisciplinary research areas (33, 34). By counting the statistical frequency of keywords, hot spots of research areas can be identified and the research area for specialized terms can be developed and expanded (35-37). The keyword analysis of the studies offers insights into the thematic evolution and research hotspots within the MDS field. By examining the frequency and emergence of specific terms over time, we observed a dynamic expansion of research themes, reflecting the responsiveness of the field to emerging challenges and opportunities. The use of the Medical Subject Headings (MeSH) thesaurus further ensures the accuracy and relevance of these keywords, aligning the identified trends with broader medical and research terminologies. MeSH includes subject headings appearing in MEDLINE/PubMed, the NLM Catalog, and other NLM databases (38).

Limitations and Future Directions

Despite its novel contributions, this study is not without limitations. The exclusive reliance on PubMed as a data source may have resulted in the omission of relevant studies indexed in other databases, potentially introducing a selection bias. Moreover, the predominance of English-language publications in PubMed could marginalize non-English research outputs, thereby imposing a linguistic bias. Additionally, by focusing solely on journal articles, this analysis might have overlooked significant contributions from other formats, such as conference papers and books.

To address these limitations and enrich the understanding of MDS research, future studies should consider a more inclusive approach by incorporating multiple databases and accommodating diverse publication types. Such endeavors would not only validate the findings of this study but also offer a more holistic view of the MDS research landscape.

Conclusion

In conclusion, this bibliometric analysis underscores the significance of MDS research in advancing healthcare information management and clinical research. Through collaborative networks and the strategic use of keywords, the MDS research community continues to evolve, responding to the changing needs of

healthcare and research. Despite its limitations, this study lays a foundation for future research, encouraging a broader examination of MDS studies to further enhance the impact of the field on global healthcare outcomes.

Ethical Consideration

Given the bibliometric nature of this study, which involved the analysis of publicly available data without direct human participation, formal ethical approval was not required. However, all analytical processes were conducted according to best practices for research integrity and reproducibility.

Financial Support or sponsorship

None

Conflict of Interest

There are no conflicts of interest

References

1. Yamin M, Alharthi S. Measuring impact of healthcare information systems in administration and operational management. *International Journal of Information Technology*. 2020;12(3):767-74.
2. Smolij K, Dun K. Patient health information management: searching for the right model. *Perspectives in health information management/AHIMA, American Health Information Management Association*. 2006;3.
3. Hajizadeh E, Shahmoradi L, Mahmoodi M, Rakhshan A, Nazari R, Barzgari S. Identification of the minimum data set to design a mobile-based application on overweight and obesity management for children and adolescents. *J Diabetes Metab Disord*. 2021;20(1):1011-20. doi: 10.1007/s40200-021-00807-1.
4. Hajesmaeel-Gohari S, Bahaadinbeigy K, Tajoddini S, S RNK. Minimum data set development for a drug poisoning registry system. *Digit Health*. 2019;5:2055207619897155. doi: 10.1177/2055207619897155.
5. Ahmadi M, Madani T, Alipour J. Development a national minimum data set (MDS) of the information management system for disability in Iran. *Disabil Health J*. 2019;12(4):641-8. doi: 10.1016/j.dhjo.2019.05.008.

6. Davey CJ, Slade SV, Shickle D. A proposed minimum data set for international primary care optometry: a modified Delphi study. *Ophthalmic Physiol Opt.* 2017;37(4):428-39. doi: 10.1111/opo.12372.
7. Mehraeen E, Safdari R, SeyedAlinaghi S, Mohammadzadeh N, Mohraz M. Common elements and features of a mobile-based self-management system for people living with HIV. *Electron Physician.* 2018;10(4):6655-62. doi: 10.19082/6655.
8. Lowe JR, Wallace SJ, Sam S, Young A. Minimum data and core outcomes for subacute rehabilitation: A scoping review. *Clin Rehabil.* 2022;36(3):388-406. doi: 10.1177/02692155211060468.
9. Sheykhotayefeh M, Safdari R, Ghazisaeedi M, Khademi SH, Seyed Farajolah SS, Maserat E, et al. Development of a Minimum Data Set (MDS) for C-Section Anesthesia Information Management System (AIMS). *Anesth Pain Med.* 2017;7(2):e44132. doi: 10.5812/aapm.44132.
10. Bernardi FA, Yamada DB, de Oliveira BM, Lima VC, Félix TM, Alves D. The minimum dataset for rare diseases in Brazil: a systematic review protocol. *Procedia Computer Science.* 2022;196:439-44.
11. Sabahi A, Asadi F, Shadnia S, Rabiei R, Hosseini A. Minimum Data Set for a Poisoning Registry: A Systematic Review. *Iran J Pharm Res.* 2021;20(2):473-85. doi: 10.22037/ijpr.2020.113869.14538.
12. Zakerabasali S, Kadivar M, Safdari R, Niakan Kalhori SR, Mokhtaran M, Karbasi Z, et al. Development and validation of the Neonatal Abstinence Syndrome Minimum Data Set (NAS-MDS): a systematic review, focus group discussion, and Delphi technique. *J Matern Fetal Neonatal Med.* 2022;35(4):617-24. doi: 10.1080/14767058.2020.1730319.
13. Halligan S, Altman DG, Taylor SA, Mallett S, Deeks JJ, Bartram CI, et al. CT colonography in the detection of colorectal polyps and cancer: systematic review, meta-analysis, and proposed minimum data set for study level reporting. *Radiology.* 2005;237(3):893-904. doi: 10.1148/radiol.2373050176.
14. Hutchinson AM, Milke DL, Maisey S, Johnson C, Squires JE, Teare G, et al. The Resident Assessment Instrument-Minimum Data Set 2.0 quality indicators: a systematic review. *BMC Health Serv Res.* 2010;10:166. doi: 10.1186/1472-6963-10-166.
15. Randolph J. A guide to writing the dissertation literature review. *Practical assessment, research, and evaluation.* 2019;14(1):13.
16. Zhang JZ, Srivastava PR, Sharma D, Eachempati P. Big data analytics and machine learning: A retrospective overview and bibliometric analysis. *Expert Systems with Applications.* 2021;184:115561.
17. Guler AT, Waaijer CJ, Palmblad M. Scientific workflows for bibliometrics. *Scientometrics.* 2016;107:385-98. doi: 10.1007/s11192-016-1885-6.
18. Peng C, He M, Cutrona SL, Kiefe CI, Liu F, Wang Z. Theme Trends and Knowledge Structure on Mobile Health Apps: Bibliometric Analysis. *JMIR Mhealth Uhealth.* 2020;8(7):e18212. doi: 10.2196/18212.
19. Donthu N, Kumar S, Mukherjee D, Pandey N, Lim WM. How to conduct a bibliometric analysis: An overview and guidelines. *Journal of business research.* 2021;133:285-96.
20. Guo Y, Hao Z, Zhao S, Gong J, Yang F. Artificial Intelligence in Health Care: Bibliometric Analysis. *J Med Internet Res.* 2020;22(7):e18228. doi: 10.2196/18228.
21. Falagas ME, Pitsouni EI, Malietzis GA, Pappas G. Comparison of PubMed, Scopus, Web of Science, and Google Scholar: strengths and weaknesses. *FASEB J.* 2008;22(2):338-42. doi: 10.1096/fj.07-9492LSF.
22. National Library of Medicine. PubMed Overview [Internet]. Internet. [cited 2023 Jan 23]. Available from: <https://pubmed.ncbi.nlm.nih.gov/about/>
23. Milian EZ, Spinola MdM, de Carvalho MM. Fintechs: A literature review and research agenda. *Electronic commerce research and applications.* 2019;34:100833.
24. Wang K, Herr I. Machine-Learning-Based Bibliometric Analysis of Pancreatic Cancer Research Over the Past 25 Years. *Front Oncol.* 2022;12:832385. doi: 10.3389/fonc.2022.832385.
25. National Library of Medicine [Internet]. Welcome to Medical Subject Headings. [cited 2022 Aug 13]. Available from: https://www.nlm.nih.gov/mesh/meshhome.html?_gl=1*423gxf*_ga*Mzk4OTIwMTQuMTU5NTE0MDUyMA..*_g_P1FP9PL4*MTY2MDM3NDU5O

- S4xLjEuMTY2MDM3NTEyNy4w
26. Stahl M, Abdel-Wahab O, Wei A, Savona M, Xu M, Xie Z. An agenda to advance research in MDS: a TOP 10 Priority List from the first international workshop in MDS (iwMDS). *Blood Adv.* 2022.
 27. Cecconi FR, Moretti N, Dejacco M, editors. *A Bibliometric Analysis on Costs Estimation of Building Retrofit*. IOP Conference Series: Earth and Environmental Science; 2019. p. 121-36.
 28. Aria M, Cuccurullo C. bibliometrix: An R-tool for comprehensive science mapping analysis. *Journal of Informetrics.* 2017;11(4):959-75.
 29. De Stefano D, Giordano G, Vitale MP. Issues in the analysis of co-authorship networks. *Quality & Quantity.* 2011;45:1091-107.
 30. Higaki A, Uetani T, Ikeda S, Yamaguchi O. Co-authorship network analysis in cardiovascular research utilizing machine learning (2009-2019). *Int J Med Inform.* 2020;143:104274. doi: 10.1016/j.ijmedinf.2020.104274.
 31. Fonseca Bde P, Sampaio RB, Fonseca MV, Zicker F. Co-authorship network analysis in health research: method and potential use. *Health Res Policy Syst.* 2016;14(1):34. doi: 10.1186/s12961-016-0104-5.
 32. Jalali ZS, Introne J, Soundarajan S. Social stratification in networks: insights from co-authorship networks. *J R Soc Interface.* 2023;20(198):20220555. doi: 10.1098/rsif.2022.0555.
 33. Wang M, Chai L. Three new bibliometric indicators/approaches derived from keyword analysis. *Scientometrics.* 2018;116:721-50.
 34. Tripathi M, Kumar S, Sonker S, Babbar P. Occurrence of author keywords and keywords plus in social sciences and humanities research: A preliminary study. *COLLNET Journal of Scientometrics and Information Management.* 2018;12(2):215-32.
 35. Grubbs JC, Glass RI, Kilmarx PH. Coauthor Country Affiliations in International Collaborative Research Funded by the US National Institutes of Health, 2009 to 2017. *JAMA Netw Open.* 2019;2(11):e1915989. doi: 10.1001/jamanetworkopen.2019.15989.
 36. Schodl K, Klein F, Winckler C. Mapping sustainability in pig farming research using keyword network analysis. *Livestock Science.* 2017;196:28-35.
 37. Huai C, Chai L. A bibliometric analysis on the performance and underlying dynamic patterns of water security research. *Scientometrics.* 2016;108:1531-51.
 38. Neveol A, Dogan RI, Lu Z. Author keywords in biomedical journal articles. *AMIA Annu Symp Proc.* 2010;2010:537-41.