

Differences in Search Strategies Across Various Scientific Databases: A Case Study on Hormonal and Cancer Therapy with Red Clover Extracts

Alexandru D. COSTIN*, Leon MUNTEAN, Andreea D. ONA, and Ioana V. BERINDEAN

Department of Genetics and Plant Breeding, University of Agricultural Sciences and Veterinary Medicine, 3-5 Mănăştur Street, 400372 Cluj-Napoca, Romania.

E-mails: alexdan.costin@gmail.com (A.D.C.); andreea.ona@usamvcluj.ro (A.D.O.); leon.muntean@usamvcluj.ro (L.M.); ioana.berindean@usamvcluj.ro (I.V.B.)

* Author to whom correspondence should be addressed;

Received: 4 June 2024 / Accepted: 30 June 2024 / Published online: 30 June 2024

Abstract

Introduction: Finding complex and relevant information in the specialty literature is a very important aspect for researchers, this representing the first step in designing and starting a new research. *Aim:* This study aimed to test four different search strategies within three databases to get a higher number of valid and reliable articles able to reflect the state of the art. *Methods:* We performed a literature search on 25 May 2024, in PubMed, Web of Science, and ScienceDirect scientific databases. We used four Boolean search strings (one sort and one long with quotation mark and one short and one long with parentheses) that combined relevant keywords related to hormonal therapy, cancer treatment, and red clover. The article types targeted were open and non-open access articles and also review and research articles. Filters as publishing year, text availability, article type and additionally subject area - for Science Direct, were applied. *Results:* The 1st short Boolean search strategy with quotation marks identified 1369 articles, and the 2nd short search strategy with parentheses identified 2734 potentially relevant articles. The 1st long Boolean search strategy with quotation marks identified 112 articles, and the 2nd long search strategy with parentheses identified 185 potentially relevant articles. The highest number of open access articles (106) and the highest number of review articles (253) was retrieved by ScienceDirect. Most of the determined articles classified as irrelevant were retrieved from ScienceDirect for all the 4-search strategy, compared to WoS and PubMed, and were excluded after filtering process, since they did not present related characteristics to the subject of interest. A significant association between article type and the 4 strategies for ScienceDirect had been identified. *Conclusions:* The short search strategy with parentheses returned a higher number of articles than the short strategy with quotation marks and the long strategies with quotes and parentheses. ScienceDirect returned the largest number of articles, but upon verification, the most irrelevant articles were also observed within this database. Regarding open access and review articles type, the short strategy with parentheses returned the highest number for PubMed and ScienceDirect, and for WoS, the highest number of such articles was obtained following the use of the short strategy with quotation marks.

Keywords: Databases; Searching strategies; Hormone therapy; Cancer; Red clover

Introduction

Finding relevant information as easy as possible in specialty literature is a very important aspect for researchers, this representing the first step in designing and starting new research. Although there are studies on red clover medicinal properties, some relevant articles are not found due to the way they are searched, thus leading to gaps in the overall picture of this species regarding its medicinal properties. A good documentation on the studied subject brings a major benefit to the research, because through documentation, researchers can find the existing gaps and manage to develop or test new hypotheses that benefit the research field. The search aim is to retrieve all

potentially relevant publications that are connected to the study objective to minimize the bias in forming conclusions [1]. Regarding red clover, even if its medicinal properties have not been extensively studied, in the last years, interest in this plant as an important source of bioactive compounds or as a basis for natural medicines and dietary supplements, was observed [2]. It exhibits antispasmodic, estrogenic and expectorant properties [3-5]. Dietary use of young red clover inflorescences can prevent cancer and reduce the menopausal symptoms severity [6]. Other studies highlight its traditional use in the treatment of bronchitis, burns, sedation [7], cardiac disorders [8], polycystic ovarian syndrome [9] and as an antidiabetic medicine [10]. When searching for information, sometimes errors can appear and can confuse or make the research difficult [11], for example, the search does not return all the relevant studies in the respective field, or it return certain studies from scientific literature that are either poorly designed or are insufficient detailed, making it impossible to test and confirm the findings [12]. To some extent, certain published research can be deceptive, influencing the use of the research results [13]. In searching strategies, search strings can be used that contains a set of keywords connected by Boolean operators (AND, OR and NOT) rather than complete sentences [14]. Also, quotation marks (that are more restrictive) and parentheses can be used in the construction of search strategies

Researchers usually combine terms for the studied problem with terms for the solving method to search in different databases. However, depending on the studied subject, the number of articles retrieved can be large and much may not be suitable for that study. Filters comprising text words or terms indexed in database, can improve the search process by increasing precision [15].

Each bibliometric database present different characteristics and various categories of filters through to assist the researchers in obtaining the articles of interest as easily as possible.

Some known databases used by researchers are PubMed, Web of Science and ScienceDirect. All the three databases have indexed journals that allow open access publication of articles.

PubMed – is a free database containing over 37 million abstracts and citations of biomedical literature [16]

Web of Science – is a subscription-based multidisciplinary scientific database presenting over 9000 journals from over 200 disciplines in fields like social sciences, arts and humanities [17].

Science Direct – is a subscription-based full-text database which is provided by the medical and scientific publishing company Elsevier, presenting over 2000 peer-reviewed journals [18].

We aimed to test four search strategies within three databases in order to get a higher number of valid and reliable articles for hormonal and cancer therapy based on red clover.

Materials and Methods

Study Design

Our team conducted a comprehensive systematic literature search on 25 May 2024 in scientific databases, including PubMed, Web of Science (WoS), and ScienceDirect. These databases were chosen because they are better known and used in the academic environment. PubMed was accessed directly since it was free, and WoS and ScienceDirect since they are subscription-based, were accessed through the Enformation platform offered by the educational institution UASVM (University of Agricultural Science and Veterinary Medicine) Cluj-Napoca. We used four Boolean search strings, one shorter and one longer, that combine relevant keywords related to hormone and cancer therapy, and red clover. These Boolean search strings were created following certain steps: first we determined a clear and focused question (Is red clover used in medicine as hormonal and cancer therapy?), then we decided which keyword should be used for the best results (finding articles from the specialized literature as efficiently as possible). The four search strategies were used the same in all three databases. To compare the search strings, criteria like completeness (if the strategy guaranteed to find a solution when there is one) and optimality (if the strategy find the best solution when there are several different solutions) were used.

For all four search strategies, we used two design type: first design presented quotation marks (which have a more restrictive effect, retrieving less articles) and second presented only parentheses to group words into complex terms.

- Shorter search string:
 - With quotation marks ("Hormonal" OR "Cancer" AND "Therapy") AND ("Red clover");

- With parentheses ((Hormonal OR Cancer) AND (Therapy)) AND (Red clover)
- Longer search string:
 - With quotation marks ("Hormonal" OR "Cancer" AND "Therapy") AND ("Red clover" OR "Trifolium pratense") AND ("Phytotherapy").
 - With parentheses ((Hormonal OR Cancer) AND (Therapy)) AND ((Red clover) OR (Trifolium pratense)) AND (Phytotherapy).

The filter applied first for all three databases was the publishing year (for studies from January 1, 2000, till May 31, 2024, representing the period with newer research) and was used to determine the total number of relevant articles. Additional filter in this step, subject area – was used just for ScienceDirect database (selecting medicine related fields since it still retrieved articles assigned to other domains). Other two filters were used to check the article general characteristics and they were text availability (open access - represent articles that can be fully accessed for free) and article type (review articles and research articles). We also checked the retrieved articles after search, if they are relevant (if they present information about the use of red clover in hormonal and cancer therapy). The presence of the established keywords in the title, abstract or content of the articles was verified, and the articles that did not contain these keywords were categorized as irrelevant articles.

Statistical Methods

The results were obtained using R4.3.1 (Beagle Scouts) statistical software [19]. To illustrate the qualitative data absolute and relative frequencies were used. We set the significance threshold to 0.05 (95% confidence level). Chi-squared test was used to test associations between text availability and article type regarding strategy string types for results retrieved from all three used databases, and for theoretical frequencies lower than 5, Fisher exact test was used.

Results

The 1st Boolean search strategy with quotation marks identified 1369 potentially relevant articles that can be used for a new study. Many of the determined articles (267) were excluded after filtering process, since they did not fit the search criteria (Figure 1).

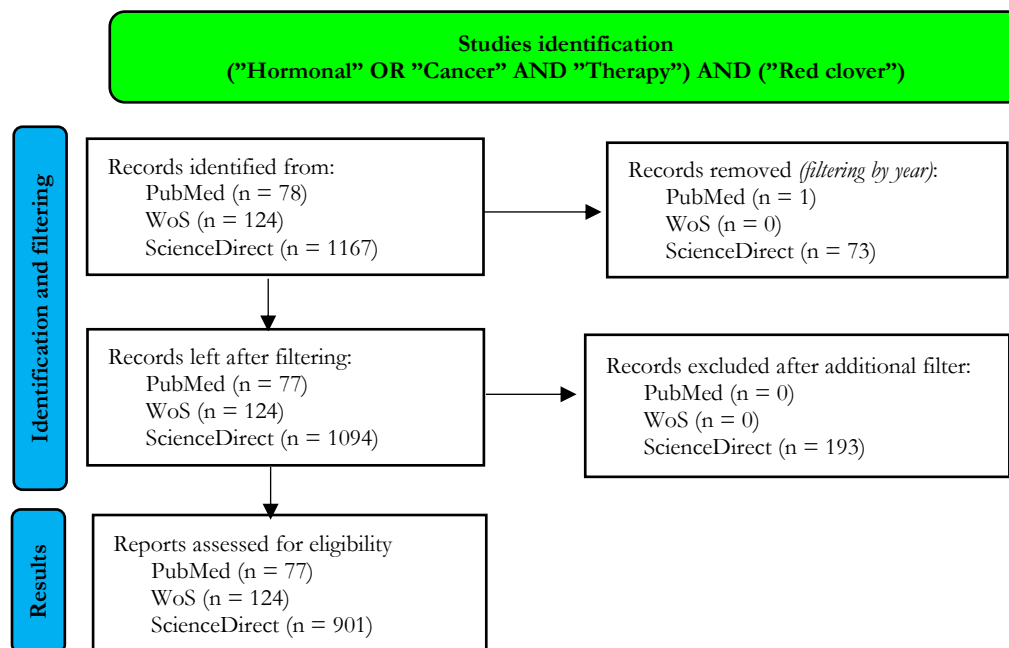


Figure 1. Results obtained with the 1st Boolean search string with quotation marks

The 1st Boolean search strategy with parentheses identified 2734 potentially relevant articles from all the three databases. Many of the determined articles (984) were excluded after filtering process, since they did not fit the search criteria (Figure 2).

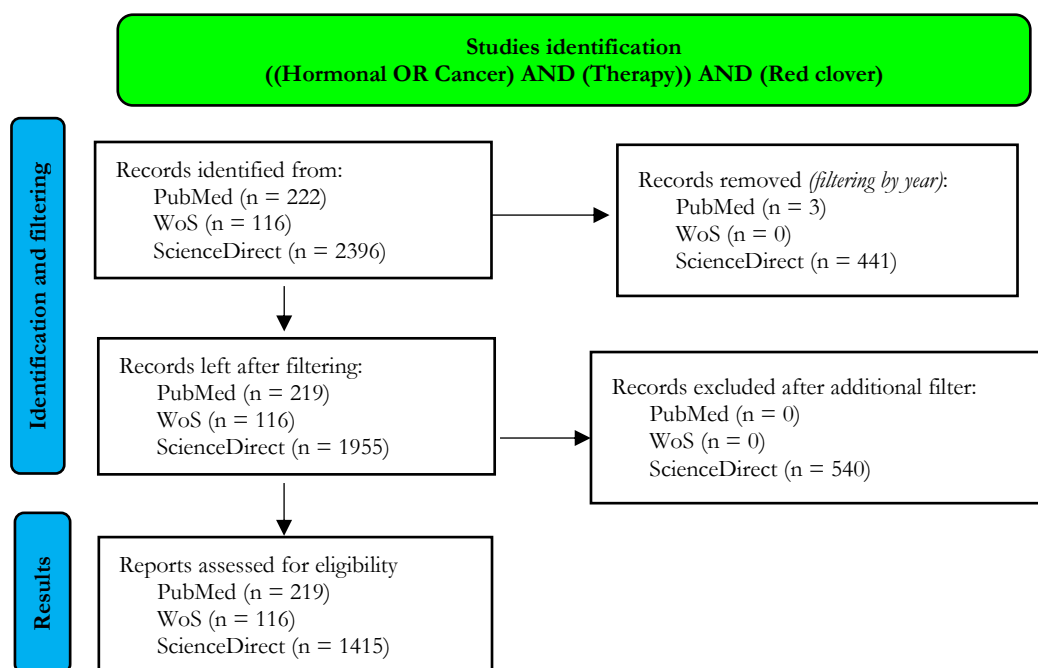


Figure 2. Results obtained with the 1st Boolean search string with parentheses

The 2nd Boolean search strategy with quotation marks identified 112 potentially relevant articles that can be used for a new study. Some of the determined articles (13) were excluded after the filtering process, since they did not fit the search criteria (Figure 3).

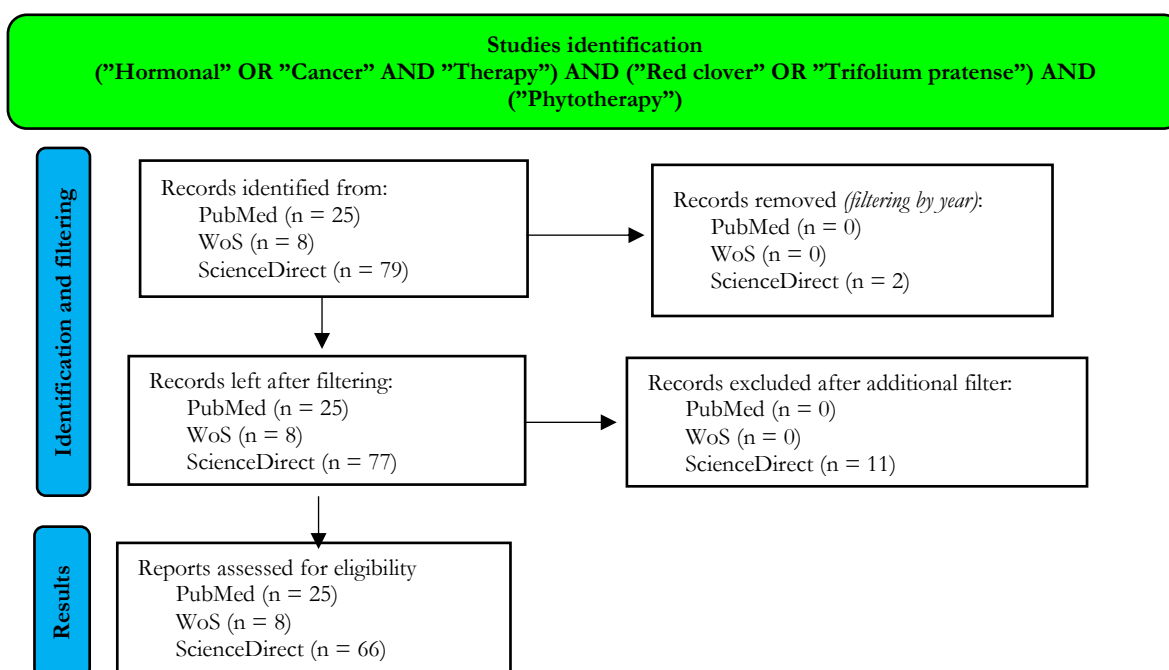


Figure 3. Results obtained with the 2nd Boolean search string using quotation marks

The 2nd Boolean search strategy with parentheses identified 185 potentially relevant articles from all the three databases. Some of the determined articles (24) were excluded after the filtering process, since they did not fit the search criteria (Figure 4).

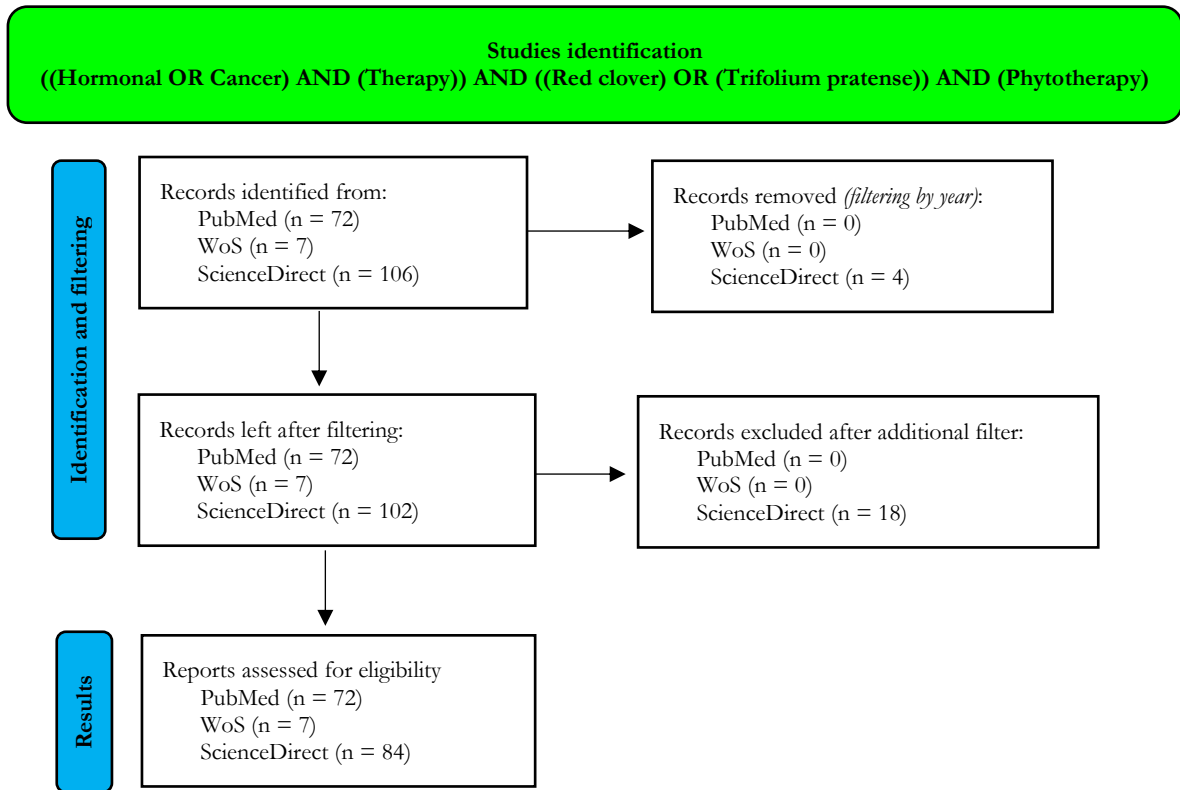


Figure 4. Results obtained with the 2nd Boolean search string using parentheses

The difference was also tested regarding the number of articles found in the searched databases that were or were not open access and in terms of the article type (review or research type) (Tables 1-3).

Table 1. Results obtained from PubMed

Characteristics		1 st strategy with quotation marks	1 st strategy with parentheses	2 nd strategy with quotation marks	2 nd strategy with parentheses	p-value
Text availability, n (%)	Open access	30 (39)	76 (34)	8 (32)	18 (25)	0.328
	Non-Open access	47 (61)	146 (66)	17 (68)	54 (75)	
Article type, n (%)	Review articles	32 (76)	77 (73)	14 (82)	33 (72)	0.866
	Research articles	10 (24)	28 (27)	3 (18)	13 (28)	

The association between text availability and article type with the two strategies does not reach the significance threshold of PubMed (Table 1) and WoS (Table 2). Also, the association between text availability and the 2 strategies for ScienceDirect does not reach the significance threshold. However, a significant association had been identified between article type and the 2 strategies for ScienceDirect (Table 3).

Table 2. Results obtained from WoS

Characteristics		1 st strategy with quotation marks	1 st strategy with parentheses	2 nd strategy with quotation marks	2 nd strategy with parentheses	p-value
Text availability, n (%)	Open access	47 (38)	42 (36)	1 (13)	0 (0)	0.111
	Non-Open access	77 (62)	74 (64)	7 (87)	7 (100)	
Article type, n (%)	Review articles	41 (34)	40 (35)	3 (38)	2 (29)	0.991
	Research articles	80 (66)	74 (65)	5 (62)	5 (71)	

Table 3. Results obtained from ScienceDirect

Characteristics		1 st strategy with quotation marks	1 st strategy with parentheses	2 nd strategy with quotation marks	2 nd strategy with Parentheses	p-value
Text availability, n (%)	Open access	106 (12)	206 (15)	8 (12)	10 (12)	0.265
	Non-Open access	795 (88)	1209 (85)	58 (88)	74 (88)	
Article type, n (%)	Review articles	253 (46)	396 (48)	20 (59)	29 (66)	0.042
	Research articles	298 (54)	429 (52)	14 (41)	15 (34)	

Between the three selected databases, ScienceDirect presented a higher number of irrelevant articles compared to the other two databases (Table 4).

Table 4. Relevance evaluation of retrieved articles

	PubMed	WoS	ScienceDirect
The 1st Boolean search strategy with quotation marks	No irrelevant articles	No irrelevant articles	Some irrelevant articles
The 1st Boolean search strategy with parentheses	Few irrelevant articles	Few irrelevant articles	More irrelevant articles
The 2nd Boolean search strategy with quotation marks	No irrelevant articles	No irrelevant articles	Some irrelevant articles
The 2nd Boolean search strategy with parentheses	Few irrelevant articles	Few irrelevant articles	More irrelevant articles

Discussion

The first used search strategy (the shorter one) with parentheses returned the larger number of results in PubMed and ScienceDirect databases, compared to the first search strategy with quotation marks and with both the second search strategies (the longer ones) with quotation marks and with parentheses that returned a lower number of potentially relevant articles (Figure 1 to Figure 4). Regarding WoS database, the first used search strategy (the shorter one) with quotation marks returned the larger number of results, compared to the other three search strategies that returned a lower number of potentially relevant articles. This highlights the fact that shorter search strategies will return a larger number of articles, but the number of articles returned will be also influenced by the databases compatibility with certain types of search strategies construction. Following the use of quotation marks

in the search strategies, the number of irrelevant articles was reduced, but it also led to the elimination of some articles that were related to the subject even though they did not exactly present the given keywords, due to the more restrictive action of these quotation marks. Of the 3 tested databases, ScienceDirect returned a higher number of articles compared to PubMed and WoS (Figure 1 to Figure 4), but also presented the highest number of irrelevant articles. Similar results regarding the number of articles retrieved by ScienceDirect compared to other databases, including PubMed, were obtained in a study in which was tested the quality of Randomized Clinical Trial based on Endometriosis Treatment Evidence [12]. Other two studies, one that reviewed the appropriate methods for teaching conflict management to medical students [20] and one that focused on accessing biomedical literature in the information landscape [21], presented a higher number of retrieved article from ScienceDirect compared to WoS and PubMed. In our search, ScienceDirect database needed an extra filter due to the presence of some irrelevant articles after first round of filters. After the filtering process from PubMed and ScienceDirect, more studies retrieved by all four used strategies were excluded due to the imposed criteria to the tested subject. An exception was WoS database which did not require the removal of any items after filtering step. Both PubMed and WoS after filtering process, returned relevant articles to the tested subject.

Regarding the text availability, PubMed, WoS and ScienceDirect presented a higher number of non-open access articles. This aspect indicates that although the three databases PubMed, WoS and ScienceDirect have indexed journals that allow open access publication of articles [22-24], they still have a higher number of non-open access journals. Even that Science Direct provided the highest number of non-open access articles, it provides the highest number of open access articles as well compared to WoS and PubMed for the first (the shorter one) search strategy (both with quotation marks and parentheses). As for article type in all four search strategies, PubMed presented a higher number of review articles than research articles and WoS presented a higher number of research articles compared to review articles. Compared to PubMed and WoS, in ScienceDirect, there was a difference of retrieved articles number between article type (research articles and review articles) and search strategy type. Shorter search strategies retrieved more research articles compared to the longer search strategies that retrieved more review articles. All these results were obtained for the studied subject, namely the use of red clover in hormonal and cancer therapy.

Upon verification for relevance of retrieved articles, the most irrelevant articles were observed within ScienceDirect database and were obtained using the short strategy with parentheses.

Study Limitations

Limits of this study were represented by the small number of databases used in the searching process (PubMed, WoS and ScienceDirect).and the published article period (were selected only articles published after 2000). Also using filters in searching process, had a contribution to the identified articles number. The criteria that were used to evaluate the relevance of retrieved articles, was decided based on what the authors considered to be appropriate for the study. ScienceDirect has the largest number of indexed journals compared to the other two databases, resulting in a larger number of articles [20,21], but this may change due to the ongoing expansion of these databases.

Future researchers testing more searching strategies in more databases, will be able to improve and generate better and more accurate results.

Conclusions

The short search strategy with parentheses returned a higher number of articles than the short strategy with quotation marks and the long strategies with quotes and parentheses. Of the three databases, ScienceDirect returned the largest number of articles, but upon verification, the most irrelevant articles were also observed within this database. Regarding open access and review articles, the short strategy with parentheses returned the highest number for PubMed and ScienceDirect, and for WoS, the highest number of such articles was obtained following the use of the short strategy with quotation marks.

List of Abbreviations: HT – Hormone therapy; WoS – Web of Science, UASVM – University of Agricultural Science and Veterinary Medicine.

Author Contributions: ADC defined the research's aim and the experiment's design. ADC, LM, ADO and IVB participated in the design of the study. ADC and ADO carried out the experiments and performed the statistical analysis. LM and IVB coordinated and helped to draft the manuscript. All authors read and approved the final manuscript.

Funding: This research received no funding.

Ethics Statement: Not applicable.

Data Availability Statement: Not applicable.

Conflict of Interest: The authors declare no conflict of interest.

References

1. Lefevre C, Manheimer E, Glanville J. Searching for studies. In: Higgins J, Green S, editors. *Cochrane handbook for systematic reviews of interventions* [Internet]. Version 5.1.0. West Sussex, England, UK: John Wiley & Sons; 2011. pp. 95–150. doi: 10.1002/9780470712184.ch6.
2. Kolodziejczyk-Czepas J. Trifolium species - the latest findings on chemical profile, ethnomedicinal use and pharmacological properties. *J Pharm Pharmacol*. 2016;68(7):845-61. doi: 10.1111/jphp.12568.
3. Leung AY, Foster S. *Encyclopedia of Common Natural Ingredients Used in Food, Drugs, and Cosmetics*; 2nd ed.; Wiley: New York, 1996; ISBN 978-0-471-50826-7.
4. Lin LZ, He XG, Lindenmaier M, Yang J, Cleary M, Qiu SX, Cordell GA. LC-ESI-MS study of the flavonoid glycoside malonates of red clover (*Trifolium pratense*). *J Agric Food Chem*. 2000;48(2):354-65. doi: 10.1021/jf991002+.
5. Booth NL, Pierson CE. Red Clover (*Trifolium Pratense*). In *Encyclopedia of Dietary Supplements*; CRC Press: New York, USA, 2005; p. 16. ISBN 978-0-429-17117-8.
6. Zielinska M, Zapotoczny P, Bialobrzewski I, Zuk-Golaszewska K, Markowski M. Engineering Properties of Red Clover (*Trifolium Pratense L.*) Seeds. *Industrial Crops and Products*. 2012;37:69-75. doi: 10.1016/j.indcrop.2011.12.002.
7. Wagay NA. Medicinal Flora and Ethno-Botanical Knowledge of Baramulla Tehsil in Jammu and Kashmir, India. *Int. J. Adv. Biotechnol. Res*. 2014;5(3):539-46.
8. Sõukand R, Kalle R. Where does the border lie: locally grown plants used for making tea for recreation and/or healing, 1970s-1990s Estonia. *J Ethnopharmacol*. 2013;150(1):162-74. doi: 10.1016/j.jep.2013.08.031.
9. Goswami PK, Khale A, Ogale S. Natural Remedies for Polycystic Ovarian Syndrome (PCOS): A Review. *Int. J. Pharma. Phytopharma. Res*. 2012; 1:396–402.
10. Bahmani M, Zargaran A, Rafieian-Kopaei M, Saki K. Ethnobotanical study of medicinal plants used in the management of diabetes mellitus in the Urmia, Northwest Iran. *Asian Pac J Trop Med*. 2014;7S1:S348-54. doi: 10.1016/S1995-7645(14)60257-1.
11. Salvador-Oliván JA, Marco-Cuenca G, Arquero-Avilés R. Errors in search strategies used in systematic reviews and their effects on information retrieval. *J Med Libr Assoc*. 2019;107(2):210-221. doi: 10.5195/jmla.2019.567.
12. Capraş RD, Urda-Cîmpean AE, Bolboacă SD. Is Scientific Medical Literature Related to Endometriosis Treatment Evidence-Based? A Systematic Review on Methodological Quality of Randomized Clinical Trials. *Medicina (Kaunas)*. 2019;55(7):372. doi: 10.3390/medicina55070372.
13. Heneghan C, Mahtani KR, Goldacre B, Godlee F, Macdonald H, Jarvies D. Evidence based medicine manifesto for better healthcare: A response to systematic bias, wastage, error and fraud in research underpinning patient care. *Evid Based Med*. 2017;22(4):120-122. doi: 10.1136/ebmed-2017-j2973rep.
14. Ugwu CN, Opah AC. Use of Boolean search strategy for accessing the databases of university of technology libraries by postgraduate students in South-East, Nigeria. *J Libr Serv and Technol*. 2023;5(2):24-35. doi: 10.47524/jlst.v5i2.25.
15. Beynon R, Leeflang MM, McDonald S, Eisinga A, Mitchell RL, Whiting P, Glanville JM. Search strategies to identify diagnostic accuracy studies in MEDLINE and EMBASE. *Cochrane Database Syst Rev*. 2013;2013(9):MR000022. doi: 10.1002/14651858.MR000022.pub3.

16. PubMed Overview. National Library of Medicine. [Internet] [accessed on 25 June 2024]. Available from: <https://pubmed.ncbi.nlm.nih.gov/about/>
17. Vieira ES, Gomes JANF. A comparison of Scopus and Web of Science for a typical university. *Scientometrics*. 2009; 81(2):587–600. doi: 10.1007/s11192-009-2178-0.
18. Tober M. PubMed, ScienceDirect, Scopus or Google Scholar – Which is the best search engine for an effective literature research in laser medicine? *Medical Laser Application*. 2011;26(3):139–144. doi: 10.1016/j.mla.2011.05.006.
19. R Core Team [Internet]. R: a language and environment for statistical computing (Vienna, Austria: R Foundation for Statistical Computing). 2023. Available at: <https://www.R-project.org/>. (Accessed 25 May 2024).
20. Mohseni F, Mohammadi A, Khajavirad N, Basiri K, Khabaz Mafinejad M. Teaching Conflict Management Skills to Medical Students: A Scoping Review. *Med J Islam Repub Iran*. 2024;38:24. doi: 10.47176/mjiri.38.24.
21. Khare R, Leaman R, Lu Z. Accessing biomedical literature in the current information landscape. *Methods Mol Biol*. 2014;1159:11-31. doi: 10.1007/978-1-4939-0709-0_2.
22. Scientific and academic research. CLARIVATE™. [Internet] [accessed on 27 June 2024]. Available from: <https://clarivate.com/products/scientific-and-academic-research/open-research/>
23. Open Access - Enabling a transition to open access. ELSEVIER. [Internet] [accessed on 27 June 2024]. Available from: <https://www.elsevier.com/open-access>.
24. 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.