Informação em Pauta

Fortaleza, CE v. 9 2024 ISSN 2525-3468

ARTIGO

DATABASES AND OTHER SOURCES OF INFORMATION IN SYSTEMATIC REVIEWS: practical recommendations

BASES DE DADOS E DEMAIS FONTES DE INFORMAÇÃO EM REVISÕES SISTEMÁTICAS: recomendações práticas

Wánderson Cássio Oliveira Araújo[1]

Fernanda Machado Lopes[2]

Natália Martins Dias[3]

Andréa Barbará da Silva Bousfield[4]

Mariana Luiza Becker da Silva[5]

Andrea Valéria Steil[6]

[1] Doctoral student in Information Science at the Federal University of Santa Catarina (UFSC). Librarian at the Federal University of Ceará in the Health Sciences Library.

E-mail: wcassio@ufc.br

[2] PhD in Psychology. Professor of the Postgraduate Program in Psychology at the Federal University of Santa Catarina (UFSC).

E-mail: femlopes23@gmail.com

[3] PhD in Developmental Disorders. Professor of the Postgraduate Program in Psychology at the Federal University of Santa Catarina (UFSC).

E-mail: natalia.m.dias@ufsc.br

[4] PhD in Psychology. Professor of the Postgraduate Program in Psychology at the Federal University of Santa Catarina (UFSC).

E-mail: andreabs@gmail.com

[5] Doctoral student, master's and undergraduate in Psychology in the Graduate Program in Psychology at the Federal University of Santa Catarina (UFSC).

E-mail: marianaluiza_b@hotmail.com

[6] PhD in Production Engineering. Professor of the Postgraduate Program in Psychology at the Federal University of Santa Catarina (UFSC).

E-mail: andrea.steil@ufsc.br

Conflito de interesses: Os autores declaram que não há conflito de interesses.

Financiamento: Não há.

Declaração de Disponibilidade dos dados: Todos os dados relevantes estão disponíveis neste artigo.

Recebido em: 31 jan 2024.
Aceito em: 31 jun. 2024.
Publicado em: 04 nov. 2024.

Como citar este artigo:

ARAÚJO, W. C. O.; LOPES, F. M.; DIAS; N. M.; BOUSFIELD, A. B. S.; SILVA, M. L. B.; STEIL, A. V. Databases and other sources of information in systematic reviews: practical recommendations. **Informação em Pauta**, Fortaleza, v. 9, p.1-17, 2024.

ABSTRACT

Bibliographic databases are the main source of information for identifying and retrieving primary studies to be synthesized in systematic reviews (SR). After analyzing the information sources used in 305 SR in psychology, this article explains how the information sources have been reported and presents practical solutions

to the gaps and inconsistencies identified. The article classifies academic bibliographic information sources, defines bibliographic databases, and demonstrates how to name information sources in SR correctly. It also discusses the number of databases in SR, the construction of search strategies, the use of controlled vocabulary, and the databases burst. Finally, the article discusses gray literature, the use of complementary sources, the regionality of information sources, and the librarian's role in SR. It is hoped that the issues discussed, and the practical recommendations of this article contribute to improving SR performed by researchers in all areas of knowledge.

Keywords: systematic review; information sources; database; information retrieval.

RESUMO

Bases de dados bibliográficas são a principal fonte de informação para a identificação e a recuperação de estudos primários a serem sintetizados em revisões sistemáticas (RS). A partir da análise das fontes de informação utilizadas em 305 RS em psicologia, o objetivo deste artigo é explicitar como as fontes de informação têm sido reportadas em RS em psicologia e apresentar soluções práticas para as lacunas e incoerências identificadas. O artigo apresenta uma classificação das fontes de informação acadêmicas bibliográficas, define base de dados bibliográfica e demonstra como nomear corretamente as fontes de informação em RS. Também discorre sobre a quantidade de bases em RS, sobre a construção de estratégias de busca, sobre o uso de vocabulário controlado e sobre a explosão em bases de dados. Por fim, o artigo discute a questão da literatura cinza, o uso de fontes complementares, a questão da regionalidade das fontes de informação e o papel do bibliotecário nas RS. Espera-se que as questões discutidas e as recomendações práticas desse artigo contribuam para o aperfeiçoamento das RS realizadas por pesquisadores em todas as áreas do conhecimento.

Palavras-chave: revisão sistemática; fontes de informação; bases de dados; recuperação da informação.

1 INTRODUCTION

Systematic reviews (SR) began to be published in large numbers in the health area in 1980 (Hansen; Trifkovic, 2013). In the last two decades, SRs have gained space, volume, and importance in different areas of knowledge. PubMed, one of the most relevant sources of information in Health Sciences, had 2,274 SRs indexed until the year 2000. Twenty years later, this number jumped to 163,237 indexed SRs. In 2022 alone, 38,503 SRs were indexed in this source of information, presenting the highest indexation of SR publications in the historical series. Such figures denote the constant growth of this type of research.

An SR uses explicit and systematic procedures to collect, combine, and synthesize the results of primary studies that answer a formulated question clearly (Higgins et al., 2021). It is configured in a logical and linear process in which each stage is necessarily connected to the others (Purssell; McCrae, 2020). Its objective is to generate an objective, empirically based answer to a research question by identifying, synthesizing, and evaluating all available evidence (Patole, 2021). This type of review can be applied to different types of research questions (Newman; Gough, 2020) and different fields of knowledge (Patole, 2021). Systematic reviews are allocated at the top of the evidence pyramid (Murad et al., 2016). In other words, when an SR is appropriately conducted, the answers found are highly reliable scientific evidence and are usually practically applied in their area of knowledge.

SRs have played an important role in the scientific development of different areas of knowledge by synthesizing evidence from studies already carried out, pointing out research gaps, helping to organize theoretical principles, connecting theoretical approaches with practical results, and subsidizing clinical decisions, among others. The use of SR has also grown due to its association with evidence-based practice in the

health area since one of its pillars is the search for the best available research evidence, integrated into the clinical experience and adapted to the characteristics of the client/patient (Melnik; Souza; Carvalho, 2014).

However, the quantitative profusion of SRs is not proportional to the expected methodological quality (Park et al., 2022; Ramasamy, 2022; Steil et al., 2022). Researchers in the areas of health, education, engineering, computing, and psychology have indicated that the quality of a systematic review is associated with how it is conducted (Borrego; Foster; Froyd, 2015; Grainger et al., 2020; Higgins; Green, 2011; Siddaway et al., 2019). There are several updated directives and manuals with guidelines for conducting systematic reviews, such as Cochrane Handbook for Systematic Reviews of Interventions (Higgins et al., 2022), the JBI Manual for Evidence Synthesis (Aromataris; Munn, 2020), the Systematic Reviews: CRD Guidance for Undertaking Reviews in Health Care (Centre for Reviews and Dissemination, 2009), and the Campbell Systematic Reviews: Policies and Guidelines (The Campbell Collaboration, 2019), among others. The well-known PRISMA statement (Preferred Reporting Items of Systematic Reviews and Meta-Analyses, versions 2009 and 2020) provides detailed guidance and examples on reporting an SR to facilitate disclosing the SR transparently and completely (Page et al., 2021).

These directives, books, guides, and statements allow researchers to conduct themselves within the international standards expected for quality SR. However, even with extensive detailing of each process, there are still many reviews self-titled as systematic that do not follow part of the proposed directives and do not adequately report all expected items (Ramasamy, 2022; Siddaway et al., 2019; Steil et al., 2022). This failure to follow international directives for SR tends to generate biased, non-reproducible, and not very transparent results, consequently synthesizing knowledge with low scientific and practical validity.

For example, a recent scope review identified many discrepancies between what is reported by SRs and what is recommended by PRISMA from a sample of 305 SRs in psychology. One of the items not fully complied with is the selection and reporting of information sources used for identifying and retrieving primary studies to be synthesized in the SR (Steil et al., 2022). The lack of more specific details for identifying information sources is potentially one of the reasons for the lack of standardization in the choice and way of reporting sources in SRs. Based on this review's findings, this article aims to explain how the sources of information have been reported in SRs in psychology and to present practical solutions to the gaps and inconsistencies identified.

The practical solutions indicated in this article were elaborated from methodological guidelines of the reference institutions in the subject. We considered the different skill levels of researchers in using digital resources related to information sources so that the article is useful to as many researchers as possible. Although the data analyzed are from the area of Psychology, we emphasize that the recommendations presented here apply to all areas of knowledge since there is no differentiation or segmentation of recommendations on the use of information sources for performing SR according to the area of knowledge.

2 BIBLIOGRAPHIC DATABASES, SEARCH SERVICES, AND OTHER SOURCES OF INFORMATION

The development of an SR is entirely dependent on the use of information sources. These sources allow recovering the documents necessary to synthesize evidence of research findings. Bibliographic databases are the preferred source for the review search process, but other non-academic sources of information can also be included.

One of the strengths of bibliographic databases is the existence of search systems

that allow multiple ways of retrieving documents. Every database has a search engine, but not every search engine is a database. This statement is intended to make an operational distinction from the framework of the types of information sources for SRs.

Bibliographic databases can be defined as a collection of electronic or digitized documents organized from metadata, selected through clear criteria, formalized in an indexing policy, and retrieved in multiple ways using a search system. The collection of this type of information source is organized to make available articles from scientific journals, conference annals, patents, books and book chapters, and legal publications, among other academic and scientific documents (Guinchat; Menou, 1994). This definition excludes institutional repositories, digital and virtual libraries, search engines, institutional websites, and databases, which are sources of information but not necessarily bibliographic databases.

Chart 1 classifies, in a simplified manner, how academic bibliographic information sources can be divided considering the format in which the content is made available (if referential or full text), the type of access to the information source (if restricted or free), and the type of access to documents in the collection of the information source (if paid, open or hybrid) as criteria. The same source of information can be classified into more than one category.

Chart 1 - Classification of bibliographic academic information sources

CATEGORIES		DESCRIPTION	EXAMPLES
CONTENT FORMAT	Reference	They provide access to the record or metadata of the document, such as title, abstract, authors, journal, year, and keywords, among other items that are indexed in an information source. They usually have a link that directs to an external source, such as a full-text database, the journal that holds the document, or an institutional repository.	ERIC Embase PubMed PsycINFO Scopus Web of Science
	Full Text	Provides full access to the document for reading and/or download. The document can be available in the original format from scientific journals, such as PDF or EPUD, or in HTML format if the content is available in the web browser.	Annual Reviews Emerald JSTOR Science Direct SciELO
ACCESS TO INFORMATION SOURCE	Access Restricted	Restrict their access to users who subscribe to the service. Access via contracts from universities, libraries, or hospitals that subscribe to the service for local use is also possible. Remote access through broader contracts, such as in Brazil, with the Capes Journal Portal, which provides access to paid databases for individuals linked to specific education and research bodies, is possible. Although access to the database or information source is paid, its content can make open access and paid access documents available.	Academic Search Premier CINAHL Embase Engineering Village Scopus Web of Science
	Access Free	They allow free access to the search system and/or its collection without the user or institutions needing subscriptions or financial investment. Although access to the database is free of charge, its content is not always open access. In some cases, the recovered documents may be of paid access, requiring financial investment to acquire the document or using an intermediary platform, such as the Capes Journal Portal. Access to the database does not necessarily entitle you to the entire collection.	Cochrane Dimensions Lilacs PubMed DOAJ ERIC SciELO Repositórios institucionais

ACCESS TO THE COLLECTION	**Paid Access**	The documents in the collection belong to a scientific publisher and require financial investment for acquisition and use. Although the collection is mostly paid access, initiatives are underway to make part of it available in open access.	ACM Digital Library Annual Reviews Emerald JSTOR Science Direct
	Open access	The documents available in the collection are under open access licenses, such as *Creative Commons* or similar, without requiring financial investment to acquire the document. The same document can be found in different sources, as the license allows sharing the document.	DOAJ SciELO Lilacs Repositórios institucionais
	Hybrid Access	The collection consists of both paid and open-access documents. It is a model adopted by several free-access databases and has gained strength in restricted-access bibliographic databases due to the open-access movement.	Cochrane Embase PubMed Scopus Web of Science

Source: Prepared by the authors.

The sources of bibliographic information used in SRs are primarily academic. However, depending on the objectives and structuring of the review, non-academic sources, such as government websites, statistical data repositories, institutional report collections, documents from organizational sources, and even consultation with specialists, may be used.

All manuals cited in the introduction of this work have items dedicated to recommendations for using information sources in SRs. However, due to the complexity and constant updating of search engines and functionalities, such recommendations are generic and, in some cases, may even be limiting for more creative searches. Even in a scenario of multiple options of information sources and different ways of performing the search, these two items must be clearly and consistently reported. For this purpose, using PRISMA 2020 (Page et al. 2021) can be an important mechanism for a comprehensive and replicable description of the use of information sources.

3 METHODOLOGICAL PROCEDURES

This article's data on the use of information sources in SR comes from a recent scope review on how SR is reported in Psychology (Steil et al., 2022). This review evaluated how much the self-declared systematic reviews in Psychology follow the recommendations of the PRISMA checklist.

The search was performed using the databases PsycINFO and MEDLINE (via PubMed). PsycINFO was chosen because it is specific to Psychology and MEDLINE due to the wide range of articles in the Health area. A librarian assisted in developing the search strategy for each database, which included the terms psychology and systematic review OR meta-analysis together with free terms as central subject headings for reviews published between January 1st, 2019, and November 9th, 2020 (search date) without language restriction. All subject headings were exploded, and 2,487 SRs were recovered.

Four independent authors examined the studies in a double-blind process, reading the titles and abstracts using the Rayyan® software (Ouzzani et al., 2016). A sample calculation was performed on the OpenEpi platform (available at: https://www.openepi.com/SampleSize/SSPropor.htm), considering the significant number of articles that met the eligibility criteria in the screening phase (N=1472), a 95% confidence interval, and a 5% margin of error. This procedure resulted in a sample of 305 SRs to be included in the

review. A random number generator was used to select the specific articles from the 1,472 initially screened, which went on to full-text analysis. Four authors also independently read the full texts in a double-blind process. Disagreements were resolved in a meeting until a consensus was reached. Data was extracted from the 305 SRs using an Excel® spreadsheet based on PRISMA items. Descriptive statistics were calculated for all data found.

Excel® spreadsheets were created with detailed descriptions of each SR's items according to the PRISMA items. Specifically for this work, the authors analyzed the data regarding the use of information sources.

4 RESULTS AND DISCUSSIONS

From the original data search and extraction by Steil et al. (2022), the present study focuses on the raw information extracted concerning the sources of information, as reported by the SR authors. The results presented and discussed are organized into categories.

4.1 Bibliographic database nomenclatures

Our data analysis suggests that it can be challenging to identify where the research was performed depending on the source of information used. Take PubMed as an example, commonly referred to as a database. Although this idea is common to some researchers, the National Library of Medicine, owner of PubMed, defines it as a search service encompassing three subsets of data: MEDLINE, PubMed Central, and NCBI BookShelf. In other words, when indicating PubMed as a database, the researcher incurs a practical, conceptual, and naming error. The database is MEDLINE since it is responsible for the indexing policy, an inherent characteristic of a bibliographic database, via the Literature Selection Technical Review Committee (National Library of Medicine, 2020). However, MEDLINE does not have its own search system, requiring the use of intermediate search services such as PubMed, Livivo, Embase, Cochrane, Ovid, EBSCO Host, and Biblioteca Virtual em Saúde (BVS), among others. Search services are reduced to an interface that provides a mechanism for retrieving information from a selected, organized, and indexed collection. Unlike a bibliographic database, search services are generally not part of and do not influence the editorial process necessary for forming the collection, available in the sources of information they make available to the user.

The data collected in the scope review (Steil et al., 2022) demonstrate the use of non-standardized nomenclatures for the same source of information. We use the case of the MEDLINE database to illustrate this point. The scope review indicated that this database was named MEDLINE without indicating the search service (n=82), MEDLINE via PubMed (n=18), MEDLINE via Ovid (n=01), PubMed without indicating the MEDLINE data subset (n=70), and MEDLINE and PubMed simultaneously (n=19) (Steil et al., 2022). Thus, due to the nomenclature used, it is unclear which search service researchers use to access the MEDLINE collection in many SRs in psychology.

Although we have used MEDLINE as an example, other bibliographic databases are subject to this standardization problem since multiple search services can access them. For example, we can cite the Educational Resources Information Center (ERIC), the Food Science and Technology Abstracts, GEOBASE, and PsycArticles.

The correct indication of the database is an important factor in analyzing the search process, in addition to better understanding and replicability. This point is important because the same dataset, accessed through different search services, tends to retrieve various documents in quantity and quality (Katchamart et al., 2011), partly due to the thematic

representation. Therefore, the correct indication of the database nomenclature helps the researcher to understand and replicate the search process according to the bibliographic database and/or search service used.

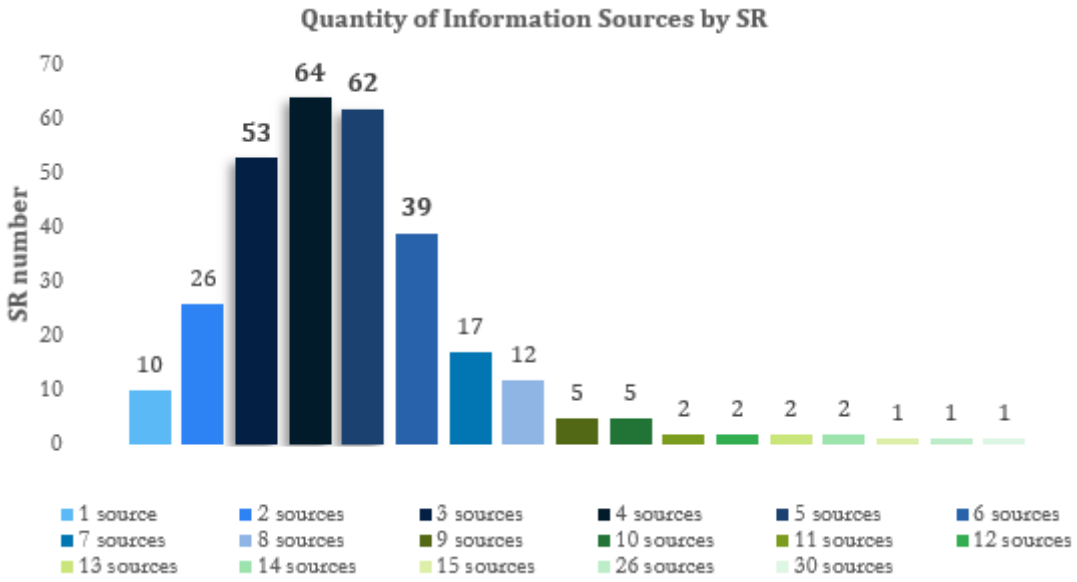
We suggest indicating the intermediary search service used to access a database, when appropriate, to replicate the search reliably. One way to describe the database and intermediary search service is by indicating both MEDLINE via PubMed if the restriction for that data subset occurs. It is important to make this indication in search services that allow selecting MEDLINE but also search other databases. One example is EBSCO. In this case, instead of informing that the search was done in EBSCO, a single search service for several databases (the EBSCO Host), it is interesting to indicate MEDLINE via EBSCO, when applicable. This orientation is valid for researchers who use MEDLINE through any other search service and for other examples of existing bibliographic databases in various areas of knowledge.

4.2 Database quantitative

Although there is no consensus on how many bibliographic databases are necessary to execute a systematic review (Bramer et al., 2017), there is an understanding that a more significant number of relevant information sources can reduce possible risks of bias in the data collection process (Higgins et al., 2021). The data of the review by Steil et al. (2022) drew attention to the significant quantitative difference of bibliographic databases used: the numbers ranged from 1 to 30 databases in a single SR. We do not intend to determine a specific number of databases to be used in an SR. However, the data collected showed a predominance of use of 3 to 6 databases in the SR of the psychology area.

Graph 1 shows the quantitative database used in the sampled studies.

Graph 1 - Quantitative of bibliographic databases used in SR in the area of Psychology



Source: Prepared by the authors based on Steil et al. (2022).

The SR predominantly used three (n=53), four (n=64), five (n=62), or six (n=39) sources of information to perform the document search (Steil et al., 2022). We consider that using only three databases may not cover different sets of important research data to answer the SR question. We suggest choosing at least two types

of database, when possible: a) multidisciplinary and b) specific to the research area of knowledge, with the addition of a regional database. Let's take the area of education as an example. The authors of an SR study could choose Scopus and Web of Science as bibliographic databases of the multidisciplinary type and the Educational Resources Information Center (ERIC), Education Research Complete via EBSCO, Education Database via ProQuest, and Educ@ as specific bibliographic databases in the area of education, the latter being a regional type, totaling six bibliographic databases.

We understand that this suggestion might not be fully feasible depending on the SR question, the type of researcher access, the content of bibliographic databases or information sources, and the researcher's search skills. In these cases, we suggest seeking the help of an information professional, preferably a librarian.

4.3 Search strategies

Search strategy, also known as string or search key, is the representation of the research question, objectives, or hypotheses of a study in a language understandable by the database search system. The researcher must translate their question into a set of authorized terms known as subject headings. These headings can be combined with Boolean operators, wildcard characters, and metadata identification tags, among other mechanisms, to retrieve the documents available in the database collection that contain the terms defined in the search strategy (Macfarlane; Russell-Rose; Shokraneh, 2022). Subject headings are the terms that make up a controlled vocabulary of a domain of knowledge, such as a thesaurus, and are previously defined by a multidisciplinary technical team. They aim to enable thematic representation of a document in a collection, that is, to describe the subjects that the document addresses and to assist the user in retrieving that document when searching by subject (Lazarinis, 2014).

The scope review data (Steil et al., 2022) show that 95.73% (n=292) of SR in Psychology reported the use of a search strategy to retrieve documents. Of this total, 73.97% (n=216) presented the search strategy in its format of use in the database, either in the body of the text or in supplementary material. The remaining 26.03% (n=76) presented only the terms used, without the complete structure of the search strategy, making its replication impossible. Considering the need for replicability in the SR, this would be possible only in about 70.81% of the total reviews analyzed.

It is important to understand that a well-designed search strategy tends to reduce search time, retrieve documents of greater utility, and allow locating implicit ideas of the search question. The better the search strategy, the better the use of the bibliographic database or information source resources. This importance is stated in the study by Salvador-Oliván, Marco-Cuenca, and Arquero-Avilés (2019), where the authors verified that 92.7% of SR indexed in MEDLINE via PubMed have some flaw in their search strategies.

Although there is a recommendation to present the complete strategies used in all bases, such as in PRISMA 2020 (Page et al., 2021) and MECIR (Higgins et al., 2021), some journals only request the presentation of a strategy (usually the one used at MEDLINE, in the case of SR in Health Sciences). We recommend that authors present all search strategies in the formats used in each database. This availability tends to improve replicating information retrieval in database search services since a) it helps in the construction of more effective search strategies by less experienced researchers, b) it has the power to standardize specific searches based on the use of already validated strategies, and c) it optimizes the researcher's time in elaborating the search strategy and effective use of search systems of bibliographic databases and other sources of information.

We draw your attention to the possibility of replicating the search strategy only when it is carried out in the database in which the original search strategy was applied. This is because different bibliographic databases use different search engines and different vocabularies. To illustrate the different vocabularies, let us take as an example an SR that wants to recover studies with women. Considering the use of thesaurus subject headings specific to each database, we have a) in PubMed, the subject heading for the term woman is "women", b) in Embase, it is "female", and c) in PsycInfo, it is "human females".

We would have 533 documents when simulating an Embase search using the PsycINFO subject heading (human females). However, if we use the correct subject heading for this base (female), taken from EMTREE, this number rises to 11,079,902 documents. In this sense, for a better use of the databases, regarding the use of search strategies, we suggest:

- Selecting the field or tag corresponding to the search by subject heading, according to the standard vocabulary of the database, if any, to retrieve the documents indexed by that term. For example, (a) in PubMed, you should use the term plus the [MeSH Terms] tag (or its abbreviation [MH]) as follows: women [MeSH Terms]; (b) in databases that have an individualized search field, such as Embase, Cinahl, or PsyInfo, select the field individually and include the subject heading. This process must be done individually for each database, respecting the standard vocabulary.
- When searching in other fields, such as title, abstract, and keywords, making different combinations of terms in natural and controlled language (standard database vocabulary) is interesting. However, this combination must respect semantic, relational, and logical meanings to maintain the consistency of the search specificity. Combining different terms and vocabularies helps to sensitize the search when considering different fields of search.
- In databases and information sources with no standard vocabularies for indexing, such as Web of Science, Scopus, Science Direct, Emerald Insight, and Scielo, you can choose to use the search strategy with the combined terms of different vocabularies in conjunction with terms in natural language.

Peer Review of Electronic Search Strategies (PRESS) can be considered at this stage. The PRESS is a practical guide for structuring search strategies with their differential of undergoing a peer review that must be conducted blindly and in pairs by librarians and/or other information specialists. It is divided into six interrelated sections that cover the best practice aspects of information retrieval in databases: search question translation; boolean and proximity operators; subject headings; word search; spelling, syntax, and number of lines; and boundaries and filters (Mcgowan et al., 2016).

4.4 Controlled vocabularies

A point to be highlighted when choosing and using databases is the existence of a controlled vocabulary linked to your search system. Only bibliographic information sources with controlled vocabularies for organizing their collection, regardless of the type of vocabulary, can expressly be considered a bibliographic database. This is one of the criteria for distinguishing between a bibliographic database and other types of information sources.

Controlled vocabularies are standardized terminological structures through strict organization criteria that make it possible to level the communication process between different human actors or between humans and machines. These structures are composed of terms or phrases that consistently allow one to describe a domain of knowledge terminologically. Its function is to enable the thematic representation of

documents through indexing, using the terms present in the vocabulary, and, consequently, facilitate the information retrieval in information sources since the search will partially consider the terms already predefined in the vocabulary, thus avoiding the use of terms that do not necessarily represent the subject to be retrieved. Examples of controlled vocabularies are subject heading lists, authority lists, thesaurus, and ontologies (Chatterjee, 2017).

The terms that make up the thesaurus (or controlled vocabulary indicated by the source of information) must be used for information retrieval to occur efficiently in bibliographic databases. Broadly speaking, these terms are named as authorized terms. Specifically, the authorized terms that make up a thesaurus are called subject headings or descriptors, the latter a nomenclature used in Brazil.

Although our suggestions are based on the use of subject headings for information retrieval, it is necessary to clarify that not all information sources used in SRs necessarily have a controlled vocabulary, preferably a thesaurus, linked to their search process. As previously exemplified, Scopus, Web of Science, Science Direct, Emerald Insight, and Scielo are information sources that do not have a standard thesaurus linked to the search process.

The importance of a linked thesaurus is due to the standardization of communication between the information retrieval system and the researcher, facilitating that the search strategy thematically corresponds to the previously indexed subject (Chatterjee, 2017; Lazarinis, 2014). In a simplified way, vocabularies function as a standard language between humans and machines. Thus, controlled vocabularies facilitate communication between these two entities, improving the search system's understanding of what the researcher wants to find.

In this regard, our suggestion is that preference should be given to information sources with controlled vocabularies whenever possible. However, we understand that this suggestion is more feasible for bibliographic databases in the health area. In the multidisciplinary field, we can exemplify a set of databases available in the EBSCO Host as Academic Search Premier, SocINDEX, and Library, Information Science & Technology as bibliographic databases with their own controlled vocabularies.

We suggest using thesauri and lists of free terms, also called natural language terms, to improve the retrieval process in databases without controlled vocabularies (and even in those that have them but require expanding the search terms in text). The free terms are synonyms or variants that can assist in the terminological standardization and expansion of search terms, although they are not authorized terms. Chart 2 presents some possibilities of controlled vocabularies that can help choose search terms.

Chart 2 - Controlled Vocabularies

AREA OF KNOWLEDGE	CONTROLLED VOCABULARY
Arts and Architecture	Thesaurus of Art & Architecture
Information Science	Library and Documentation Thesaurus LISA Thesaurus LISTA Thesaurus Tesauro Brasileiro de Ciência da Informação Tesauro de Ciencias de la Documentación
Health Sciences	APA Thesaurus CINAHL Thesaurus DECS (Descritores em Ciências da Saúde) Dentistry & Oral Subjects EMTREE (Embase Tree) MeSH (Medical Subject Headings) SPORTDiscus Dictionary of Synonyms
Applied Social Sciences	Business Source Premier Thesaurus EconLit Thesaurus SocINDEX Subject Terms

Communication	Communication Source Thesaurus
Law	Tesauro do Supremo Tribunal Federal
Education	Australian Thesaurus of Education Descriptors BRASED (Thesaurus Brasileiro da Educação) ERIC Thesaurus European Education Thesaurus
Engineering and Mathematics	Agrovoc Engineering Village Thesaurus GeoBase Thesaurus IEEE Thesaurus MathSciNet Thesaurus
Multidisciplinary	Academic Search Premier Subject Terms Catálogo de Autoridades da Biblioteca Nacional EuroVoc (EU multilingual and multidisciplinary thesaurus) Library of Congress Subject Headings ProQuest Thesaurus Science Direct Topics (list of multidisciplinary terms) UNESCO Thesaurus UNBIS (United Nations Bibliographic Information System)

Source: Prepared by the authors.

These vocabularies, even if they are not directly linked to databases, allow retrieving review documents to follow a pattern in the different sources of information used. This will expand and improve the quality of the search terms in the text. Although our suggestion has practical value, it is necessary to consider that vocabularies from other areas of knowledge can and should be used in an interdisciplinary way. The combination with natural language is essential for retrieving the most sensitive information.

4.5 Database explosion

Although our data cannot support a more accurate analysis of the database explosion, it is an important factor in choosing databases. When this technique is used, it is necessary to explicitly report it in the SR.

The search with Explosion can be defined as the method of retrieving information in databases that expands the search by the primary subject heading and extends it to all its subordinate terms. For example, if you search for the term Psychology in PubMed using the tag for MeSH Terms, by standard, the search system will consider the search term plus all of its subordinate terms in the base thesaurus. In the case of the term Psychology, the subordinate terms in the search via PubMed will be Cognitive Science; Cognitive Neuroscience; Economics, Behavioral; Environmental Psychology; Ethnopsychology; Forensic Psychology; Psychology, Adolescent; Psychology, Child; Psychology, Clinical; Psychology, Comparative; Psychology, Developmental; Psychosocial Functioning; Psychology, Educational; Psychology, Experimental; Psychology, Industrial; Presenteeism; Time Management; Psychology, Medical; Psycho-Oncology; Psychology, Positive; Psychology, Social; and Psychology, Sports.

In this example, retrieving a single search term (Psychology) was automatically expanded to twenty-two other subordinate terms that are not necessarily those the researcher had as their primary search objective. This extended search process will only have functionality for indexing metadata fields directly related to the controlled vocabulary of the bibliographic database, which is not valid for other search fields, such as title, abstract, or author keywords.

Regarding the explosion, we suggest that when choosing the databases for the SR, the researchers verify the search behavior concerning the explosion of terms. The researcher must analyze if the explosion occurs in the standard search,

as is the case with PubMed, or if it is necessary to add a callsign for the explosion to occur, as in the Cumulative Index to Nursing and Allied Health Literature (CINAHL). This verification is important because, depending on the search system, it will be necessary to a) add characters, such as the + symbol in CINAHL; b) select the type of explosion, as in Embase; c) or edit tags to prevent the explosion in search systems that explode by default, as in PubMed.

Although it is useful in some cases, when greater sensitivity is needed in the search, the explosion can be a big issue when the primary term has several subordinate terms with no direct relationship with the main theme of the search. In some cases, the explosion makes the search impossible, given that the results are quantitatively expressive, complicating the document selection process. Opt for explosion search when the set of terms subordinate to the subject heading is entirely relevant to your search. When part of the subordinate terms are irrelevant, we suggest using the primary subject heading in conjunction with the actually pertinent subordinate subject headings.

It is important to note that it is only possible to use this functionality in search systems and bibliographic databases that have a thesaurus integrated into their retrieval mechanism, such as Medline via PubMed, Medline via EBSCO, Embase, CINAHL, Cochrane, PsyInfo, and OVID. Information sources such as Web of Science and Scopus do not make the explosion possible since they do not have an integrated thesaurus. These two databases choose an automated indexing system, called KeyWords Plus, and the interoperability of different vocabularies, respectively. They are culled from the original sources in which the documents are made available.

4.6 Grey literature

Although the conceptualization of gray literature may lack standardization between different areas of knowledge, it can be understood as the set of documents, printed or digital, produced by different actors, such as government, business, industrial, and academic entities, that are not under the control of commercial publishers or that have not been submitted to peer review, but that contain useful information that complements the analysis of a study and/or serve to mitigate possible problems of scientific bias (Higgins et al., 2021; Paez, 2017). Many documents can be classified as gray literature, including theses and dissertations, research reports, committee instructions and guides, government reports, articles from conferences and ongoing research, book chapters, unpublished data, personal correspondence, and data of a political nature, among others (Hopewell et al., 2007; Paez, 2017).

Most of the SR in psychology analyzed (65.57%, n=200) did not include any gray literature sources in their search. Only 34.43% (n=105) of the studies used gray literature (Steil et al., 2022). Although gray literature can often be indicated as lower quality literature or considered challenging to recover (Woods; Phillips; Dudash, 2020), this perception is not shared by manuals for systematic reviews. A clear example is the use of gray literature as a source of information by PRISMA 2020.

Also regarding the quality of the studies, it is noteworthy that it is not appropriate to limit searches in information sources for journals according to rankings such as Qualis CAPES, Impact Factor, H-Index, and Cite Score, among others. This procedure limits finding documents that may contain important data to synthesize evidence. In addition, an SR synthesizes all available evidence that answers their review question. The risk of bias analysis, or methodological quality assessment, of the included studies will present information on the potential biases of the individual studies included or on their methodological quality. In turn, synthesizing the accumulated evidence will show the certainty of the synthesized evidence in the SR.

An additional recommendation concerning gray literature is that its form of retrieval differs from databases, as it is usually available in repositories, search engines, and databases specific to this purpose. While most databases use thesauri as the standard controlled vocabulary for the thematic representation of documents, the representation process in the gray literature is not standardized between the different platforms. A controlled vocabulary is generally not used.

We suggest using alternative search strategies with reduced format, which represent the main theme of the research as a solution for a more effective retrieval. Some of these sources limit the number of terms that can be searched, making it impossible to use the same strategy as in a bibliographic database. In some cases, using more than one strategy is necessary to extract better search results.

Google Scholar is a source of information often used to retrieve gray literature, presenting one of these limitations. The Google Scholar search box is limited to 256 characters, with spaces. The system ignores search terms that exceed this limit and may present non-standard search results or false positives (Boeker; Vach; Motschall, 2013). Of 105 studies indicating the use of gray literature, 39 (37.14%) SR used Google Scholar as one of the sources of gray literature (Steil et al., 2022). Taking this information as a basis, our recommendation is to verify, when possible, the search instructions of the information source so that it can be used according to the standard search parameters to possibly minimize its limitations.

For the reader's knowledge, in addition to Google Scholar, considered a search engine that allows finding gray literature (in addition to other materials that are not gray literature since they are indexed in databases, for example), there are specific sources for searching for gray literature, such as Open Grey, Grey Guide, and Networked Digital Library of Theses and Dissertations (NDLTD), Open Access Theses and Dissertations (OATD), and Biblioteca Digital de Teses e Dissertações (BDTD), among many others.

4.7 Use of complementary sources

Approximately 67.54% (n=206) of the papers analyzed in the scope review (Steil et al., 2022) used some type of complementary source, such as a list of references, contact with experts, or previous review studies. Of this total, 35.92% (n=74) indicate the use of reference lists as a complementary source (Steil et al., 2022), respecting one of the guidelines for the use of complementary sources in SR (Higgins et al., 2021). Although reference lists were the preferred method of access to complementary sources in the SR analyzed, other options have been little explored. One possibility is using the tool similar articles or similar records (depending on the database), used only by one study of the scope review. When selecting a specific article, this tool presents a set of suggestions from other articles with thematic relationships with the subject indexed in the originally selected article; that is, the platform offers suggestions from similar documents.

It is possible that certain documents may not be recovered even when using well-designed search strategies. This may be due to the non-inclusion of a term in the search strategy, the possible restriction of a Boolean operator, or a thematic representation that does not necessarily correspond to the document's content. We ran a test with the article Systematic Review and Meta-Analysis of Self-Serving Attribution Biases in the Competitive Context of Organized Sport, using the tool similar articles in PubMed and Embase. We obtained 99 and 249,470 similar results, respectively. Given this scenario, which is quantitatively impossible to analyze thoroughly, we suggest that the researcher organize the documents indicated by relevance and define a standard number for reading as a complementary source.

We draw attention to the cautious use of the tool similar articles since it can (a)

indicate documents that are already present in the set retrieved in the primary search; (b) suggest articles with no logical relation to the search question of the review since the algorithm considered the presence of specific subject headings or general terms and did not necessarily respect the logic of Boolean operators; and (c) indicate documents other than those defined in the search protocol of the search.

4.8 Regionality of information sources

The results of the scope review (Steil et al., 2022) indicate a reduced (or limited) use of regional databases. An additional evaluation identified that, of 305 studies, only 7.86% (n=24) used at least one regional information source.

There is a clear preference for more consolidated databases in countries of greater scientific relevance. However, regional initiatives can and should be considered relevant sources of information for data collection in SR, such as Literatura Latino-Americana e do Caribe em Ciências da Saúde (Lilacs), Scientific Electronic Library (SciELO), China National Knowledge Infrastructure, and Africa-Wide Information. A regional information source indexes only content from a specific country or geographic region in one or more languages.

The SR published by Brazilian and Chinese researchers presented in the scope review by Steil et al. (2022) is an example. In the first case, out of seven studies, 71.42% (n=5) used at least one source of information at the regional level. In the second case of 19 studies, only 31.57% (n=6) used some source of regional information. We do not have enough data to draw a solid conclusion. Still, in our analysis, the greater use of regional databases in studies with Brazilian researchers may be a direct result of years of investment in human resources, technological infrastructure, marketing, and scientific dissemination by the Centro Latino-Americano e do Caribe de Informação em Ciências da Saúde - Bireme and the SciELO initiative, among other regional initiatives. Specifically in the Chinese case, we can assume that the need to publish in foreign journals is a variable of influence since three publications were made in regional journals in the Brazilian case, all using regional information sources. Although Chinese researchers used regional information sources, none of the studies in the sample were published in a regional journal.

We suggest that researchers include at least one regional database in the data collection process whenever possible. This suggestion exists in other guidelines, such as MERCIR (Higgins et al., 2021), but little is present in other international guidelines.

By performing the search exclusively in international databases, the researcher may unintentionally exclude relevant documents, which, for some reason, are not indexed in these sources of information (Betrán et al., 2005). In practice, there is no reason to exclude regional databases, given that their professionalization in terms of infrastructure, indexing criteria, and usability and qualification of human resources have increasingly become equivalent to international bibliographic databases.

4.9 Role of the librarian

Regardless of the research skill level of the SR team, a librarian's participation is recommended to assist with the entire research data management process. This professional can be present in all phases of the research to enable better use of the tools and functionalities of the bibliographic databases, as recommended by several review guidelines (Aromataris; Munn, 2020; Higgins et al., 2022). Although it is a recommendation, only 9.83% (n=30) of the SRs in Psychology had this professional on their team (Steil et al., 2022).

Studies with a librarian were more likely to present the search strategy (80%) than those without (69.81%).

The average number of bibliographic databases used was similar (4.6 with the participation of a librarian and 4.9 without a librarian).

Although our results cannot affirm whether the participation of this professional plays a role in the informational quality of SR, our suggestion is to have one on the research team whenever possible since "the contribution of a research librarian or information scientist can be invaluable in designing and refining research" (Aromataris; Munn, 2020, without pagination, our translation).

5 CONCLUSION

This article presented solutions to problems identified in published systematic reviews regarding using and reporting bibliographic information sources. From our analysis, we suggest the following solutions:

- Correctly indicate in which source of information the data collection was made, reporting the name of the bibliographic database, source of information, or search service used to search the data set;
- Define between three and six databases for data collection and, when possible, combine multidisciplinary bibliographic databases with disciplinary databases;
- Give preference to bibliographic databases that use a controlled vocabulary as a standard of thematic representation to obtain a more consistent search process and a more effective retrieval;
- When choosing databases with integrated controlled vocabulary, verify the explosion availability of the search terms and inform whether or not the explosion was used in the search report in the SR methodological procedures section;
- When the gray literature is used as a source of information, make adaptations in search strategies considering that the search system of these tools is different from that found in databases;
- When using complementary sources through databases, it is suggested to use the search tool "similar documents", considering it to be a viable and fast option;
- Use at least one regional database since relevant studies may be published in regional sources but, for some reason, not be retrieved in international bibliographic databases;
- Whenever possible, consult a librarian specialized in information retrieval to assist in the search protocol and retrieval process, following information quality standards to take advantage of all bibliographic database functionalities.

Finally, we understand that the methodological complexity of developing an SR imposes challenges on the researcher. In this sense, we hope that this article's conceptual clarifications and practical recommendations can improve SR conducted by researchers in all areas of knowledge.

REFERÊNCIAS

AROMATARIS, Edoardo; MUNN, Zachary (ed.). **JBI Manual for Evidence Synthesis**. Adelaide: JBI, 2020. DOI: https://doi.org/10.46658/JBIMES-20-01.

BRAMER, Wichor M. et al. Optimal database combinations for literature searches in systematic reviews: a prospective exploratory study. **Systematic Reviews**, v. 6, n. 1, p. 1-12, Dec. 2017. DOI: http://dx.doi.org/10.1186/s13643-017-0644-y.

BETRÁN, Ana P. et al. Effectiveness of different databases in identifying studies for systematic reviews: experience from the WHO systematic review of maternal morbidity and mortality. **BMC Medical Research Methodology**, v. 5, n. 1, p. 1-5, 28 Jan. 2005. DOI:

REFERÊNCIAS

http://dx.doi.org/10.1186/1471-2288-5-6.

BOEKER, Martin; VACH, Werner; MOTSCHALL, Edith. Google Scholar as replacement for systematic literature searches: good relative recall and precision are not enough. **BMC Medical Research Methodology,** v. 13, n. 1, p. 1-12, 26 out. 2013. DOI: http://dx.doi.org/10.1186/1471-2288-13-131.

BORREGO, Maura; FOSTER, Margaret J.; FROYD, Jeffrey E. What is the State of the-Art of Systematic Review in Engineering Education? **Journal Of Engineering Education**, [v. 104, n. 2, p. 212-242, abr. 2015. DOI: http://dx.doi.org/10.1002/jee.20069.

CENTRE FOR REVIEWS AND DISSEMINATION. **Systematic Reviews**: CRD's Guidance for Undertaking Reviews in Health Care. York, England: University of York, 2009.

CHATTERJEE, Amitabha. **Elements of Information Organization and Dissemination**. Cambrige: Chandos Publishing, 2017.

GRAINGER, Rebecca et al. Issues in reporting of systematic review methods in health app-focused reviews: a scoping review. **Health Informatics Journal**, v. 26, n. 4, p. 2930-2945, Sep. 2020. DOI: http://dx.doi.org/10.1177/1460458220952917.

GUINCHAT, Claire; MENOU, Michel. **Introdução geral as ciências e técnicas da informação e documentação**. Tradução Míriam Vieira da Cunha. 2. ed. Brasília, DF: IBICT, 1994

HANSEN, Henrik; TRIFKOVIC, Neda. **Systematic Reviews**: Questions, Methods and Usage. Copenhagen, Denmark: Ministry Of Foreign Affairs Of Denmark, 2013.

HIGGINS, Julian Pt; GREEN, Sally (ed.). Cochrane **Handbook for Systematic Reviews of Interventions version 5.1.0:** updated March 2011. London, UK: The Cochrane Colaboration, 2011.

HIGGINS, Julian P.T.; Lasserson T; CHANDLER, Jackie; Tovey D; THOMAS, James; Flemyng E, Churchill R. **Methodological Expectations of Cochrane Intervention** Reviews. London, UK: Cochrane, 2021.

HIGGINS, Julian P.T.; THOMAS, James; CHANDLER, Jackie; Cumpston M, Li T, Page MJ, Welch VA (ed.). **Cochrane Handbook for Systematic Reviews of Interventions versão 6.3** (Last updated: 4 August, 2022). Cochrane, 2022. Disponível em: https://training.cochrane.org/handbook. Aceso em: 30 jan. 2023.

HOPEWELL, Sally et al. Grey literature in meta-analyses of randomized trials of health care interventions. **Cochrane Database of Systematic Reviews**, p. 1-15, Apr. 2007. DOI: http://dx.doi.org/10.1002/14651858.mr000010.pub3.

KATCHAMART, Wanruchada et al. PubMed had a higher sensitivity than Ovid-MEDLINE in the search for systematic reviews. **Journal of Clinical Epidemiology,** v. 64, n. 7, p. 805-807, July 2011. DOI: http://dx.doi.org/10.1016/j.jclinepi.2010.06.004.

LAZARINIS, Fotis. **Cataloguing and Classification**: an introduction to AACR2, RDA, DDC, LCC, LCSH and MARC 21 standards. Cambrige: Chandos Publishing, 2014.

MACFARLANE, Andrew; RUSSELL-ROSE, Tony; SHOKRANEH, Farhad. Search strategy formulation for systematic reviews: issues, challenges and opportunities. **Intelligent Systems With Applications**, v. 15, p. 1-10, Sep. 2022. DOI: https://doi.org/10.1016/j.iswa.2022.200091.

MCGOWAN, Jessie et al. PRESS Peer Review of Electronic Search Strategies: 2015 guideline statement. **Journal of Clinical Epidemiology**, v. 75, p. 40-46, July. 2016. DOI: http://dx.doi.org/10.1016/j.jclinepi.2016.01.021.

MELNIK, Tamara; SOUZA, Wanderson Fernandes de; CARVALHO, Marcele Regine de. A importância da prática da psicologia baseada em evidências: aspectos conceituais, níveis de evidência, mitos e resistências. **Revista Costarricense de Psicología**, San José, Costa Rica, v. 33, n. 2, p. 79-92, jul./dez. 2014.

MURAD, M Hassan et al. New evidence pyramid. **Evidence Based Medicine**, v. 21, n. 4, p. 125-127, June 2016. DOI: http://dx.doi.org/10.1136/ebmed-2016-110401.

NATIONAL LIBRARY OF MEDICINE. MEDLINE, **PubMed, and PMC (PubMed Central)**: how are they different? 2020. Disponível em: https://www.nlm.nih.gov/bsd/difference.html. Acesso em: 30 jan. 2023.

NEWMAN, Mark; GOUGH, David. Systematic Reviews in Educational Research: methodology, perspectives and application. In: ZAWACKI-RICHTER, Olaf et al (ed.). **Systematic Reviews in Educational Research**: methodology, perspectives and application. Cham, Switzerland: Springer, 2020. p. 3-22.

OUZZANI, Mourad; HAMMADY, Hossam; FEDOROWICZ, Zbys; ELMAGARMID, Ahmed. Rayyan—a web and mobile app for systematic reviews. **Systematic Reviews,** v. 5, n. 1, p. 1-10, Dec. 2016. DOI: http://dx.doi.org/10.1186/s13643-016-0384-4.

PAEZ, Arsenio. Gray literature: an important resource in systematic reviews. **Journal of Evidence-Based Medicine**, v. 10, n. 3, p. 233-240, Aug. 2017. DOI: http://dx.doi.org/10.1111/jebm.12266.

PATOLE, Sanjay. Systematic Reviews, Meta-Analysis, and Evidence-Based Medicine. In: PATOLE, Sanjay (ed.). **Principles and Practice of Systematic Reviews and Meta-Analysis.** Cham, Switzerland: Springer, 2021. p. 1-10.

PARK, Ho Young et al. Quality Reporting of Systematic Review and Meta-Analysis According to PRISMA 2020 Guidelines: results from recently published papers in the Korean Journal of Radiology. **Korean Journal of Radiology,** v. 23, n. 3, p. 355-369, 2022. DOI: http://dx.doi.org/10.3348/kjr.2021.0808.

PAGE, Matthew J. et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. BMJ, v. 372, p. 1-9, Mar. 2021. **BMJ**. DOI: http://dx.doi.org/10.1136/bmj.n71.

PURSSELL, Edward; MCCRAE, Niall. **How to Perform a Systematic Literature Review**: a guide for healthcare researchers, practitioners and students. Cham, Switzerland: Springer, 2020.

RAMASAMY, Akilesh. PRISMA 2020: key changes and implementation aspects. **Journal of Oral and Maxillofacial Surgery**, v. 80, n. 5, p. 795-797, May 2022. DOI: http://dx.doi.org/10.1016/j.joms.2021.12.018.

SALVADOR-OLIVÁN, José Antonio; MARCO-CUENCA, Gonzalo; ARQUERO-AVILÉS, Rosario. Errors in search strategies used in systematic reviews and their effects on information retrieval. **Journal of The Medical Library Association**, v. 107, n. 2, p. 210-221, Apr. 2019. DOI: http://dx.doi.org/10.5195/jmla.2019.567.

SIDDAWAY, Andy P.; WOOD, Alex M.; HEDGES, Larry V. How to do a Systematic Review: A Best Practice Guide for Conducting and Reporting Narrative Reviews, Meta-Analyses, and Meta-Syntheses. **Annual Review of Psychology**, v. 70, n. 1, p.747-770, Jan. 2019. DOI: http://dx.doi.org/10.1146/annurev-psych-010418-102803.

STEIL, Andrea Valéria et al. Reporting characteristics of systematic reviews in Psychology: a scoping review. **Journal Of Health Psychology**, v. 27, n. 13, p. 2964-2981, Jan. 2022.

THE CAMPBELL COLLABORATION. **Campbell Systematic Reviews**: Policies and Guidelines. Oslo, Norway: The Campbell Collaboration, 2019. DOI: http://dx.doi.org/10.4073/cpg.2016.1.

WOODS, Stephen; PHILLIPS, Kathleen; DUDASH, Andrew. Grey literature citations in top nursing journals: a bibliometric study. **Journal of the Medical Library Association,** v. 108, n. 2, p. 262-269, Apr. 2020. DOI: http://dx.doi.org/10.5195/jmla.2020.760.