Research Article

Development and Functionality of the ABC Tool (Automated Boolean Composer) to Simplify and Standardize Search Strategy Construction in Bibliographic Databases

Gabriel Rada[1,2], Camila Ávila-Oliver[1,2], Diana Biscay[1], Iván Jara[1], Daniel Nava-Mosler[1], José Tomás Ramos-Rojas[1,2], Ana María Rojas-Gómez[1], Álex Silva-Zapata[1], Juan Vásquez-Laval[1], Camilo Vergara[1]

1. Epistemonikos Foundation, Chile; 2. Universidad del Desarrollo, Chile

The quality and comprehensiveness of search strategies are critical determinants of the rigor and reliability of systematic reviews and other forms of evidence synthesis. However, constructing effective Boolean strategies is complex, time-consuming, and requires specialized expertise. This article describes the development and functionality of the Automated Boolean Composer (ABC) Tool, an automated strategy generator integrated into the advanced search interfaces of Epistemonikos's bibliographic databases. Built upon the Epistemonikos Evidence Taxonomy (EET), which is a curated hierarchical vocabulary, the tool provides more than 9,500 pre-developed Boolean search strategies covering a wide range of health-related concepts. Each strategy is based on rigorously tested Boolean logic and validated using relative recall against systematic review datasets to ensure high sensitivity and efficiency. The tool automatically inserts these pre-formulated strategies into the search environment as editable drafts and offers features such as real-time term suggestions, a taxonomy-based term browser, and automated combination line generation. By reducing workload, minimizing human error, and supporting users with varying levels of expertise, the ABC Tool has the potential to substantially enhance the efficiency, reproducibility, and methodological quality of evidence synthesis workflows.

Corresponding author: Gabriel Rada, radagabriel@epistemonikos.org

 doi.org/10.32388/6FLT1Z

Introduction

Summarizing evidence to support informed decision-making is becoming increasingly challenging. A variety of approaches have been proposed to address this, including systematic reviews, health technology assessments, clinical guidelines, and other forms of evidence synthesis. These evidence-based products play a central role in guiding policy and practice decisions [1].

A fundamental, yet sometimes underappreciated, component of any evidence-based product is the adequacy of the underlying evidence base. This refers to the collection of studies and data that support its conclusions. Getting this right starts with search strategies. A well-designed search can mean the difference between a review that offers a comprehensive picture and one that, unintentionally, leaves out key findings [2].

Poorly designed search strategies can undermine the reliability and validity of a review in several ways. If insufficiently sensitive, they may fail to capture all relevant evidence, leading to incomplete or biased conclusions. Conversely, if lacking precision, they may retrieve large volumes of irrelevant records, increasing the screening burden and reducing the overall efficiency of the review process. In addition, inadequately reported search strategies hinder transparency and reproducibility, making it difficult for others to assess, replicate, or update the review. Ensuring both methodological rigor and clear documentation in the search process is therefore essential for producing trustworthy and efficient evidence syntheses [3].

That said, crafting a solid search strategy can be surprisingly tricky. It calls for more than just knowing how to string together Boolean operators. You also need to understand the subject matter deeply enough to anticipate synonyms, related concepts, and even regional spelling differences. Once you've built the strategy, reporting it in a way that others can follow (and trust) brings its own set of demands [2].

The process can be tedious, repetitive, and highly detail-oriented. Mistakes such as misused operators or missed terms are easy to make and hard to catch. People without formal training in information retrieval may struggle, and even experienced reviewers are not immune to bias or inconsistencies [3]. To make matters worse, refining a strategy often takes several rounds of trial and error. This iterative process can drag on, consuming a disproportionate amount of time. Some estimates suggest it can take upwards of 100 hours just to finalize a search strategy for a systematic review [4].

To help ease some of these burdens, the Epistemonikos Foundation developed the ABC Tool (Automated Boolean Composer), a tool that allows users to insert and combine search terms right into the advanced search interface of a bibliographic database. It is currently available in ED-Trials, a database of randomized trials (https://trials.epistemonikos.org) [5], and will soon be available across other databases managed by the Foundation (the full list of available databases can be found at https://www.epistemonikos.org).

What follows is an overview of the development process and core functionalities of the *ABC Tool*.

Methods

The ABC Tool is built on a set of pre-developed search strategies organized within the Epistemonikos Evidence Taxonomy (EET), which is a hierarchical framework of standardized terms designed to support evidence synthesis. The EET covers a wide range of concepts, with a particular emphasis on health-related topics relevant to systematic reviews and related methodologies. Details on the structure, scope, and procedures for updating and expanding the EET are available elsewhere [6].

Development of the search strategies

Each search strategy was developed by a team of expert information specialists working in collaboration with subject-matter experts. For every term, the team compiled relevant synonyms and related concepts. This process involved non-systematic searches of key websites and documents, as well as the use of a proprietary tool that applies Word2Vec technology to the Epistemonikos Database corpus [7], [8]. Based on the terms identified, Boolean search strategies were constructed using the logical operators AND and OR, along with truncation, exact phrase matching, and appropriate nesting with parentheses.

Validation of the search strategies

The search team assessed each term to decide whether formal validation of the Boolean search strategy's performance was needed. For straightforward terms, such as the specific name of a drug, validation was generally deemed unnecessary. In contrast, for broader or more complex terms, such as entire drug classes, validation was considered essential to ensure accuracy and relevance.

The benchmark for validation was to develop search strategies with a minimum sensitivity of 95 percent, while keeping the number needed to read as low as possible. This metric refers to the number of search results that must be reviewed to find one relevant study. In other words, it reflects the efficiency of the

search: the lower the number, the less irrelevant material needs to be screened. In cases where the number needed to read exceeded 50, a lower sensitivity threshold was accepted to maintain feasibility. Each search strategy was defined as the combination of the target term and its relevant subterms. For instance, a strategy for a drug class included all terms related to the individual drugs that belong to that class.

Validation of each search strategy was carried out using the *relative recall* method. In this approach, a sample of primary studies included in published systematic reviews was used as the reference standard to assess the strategy's performance [9].

Relative recall is defined as the proportion of relevant studies retrieved by the search strategy, calculated using the following formula:

Relative Recall =

(Number of studies retrieved by the search strategy) ÷ (Total number of included studies in the reference standard)

This metric provides an estimate of the sensitivity of the search strategy in a real-world context, based on studies already identified as relevant through systematic review processes.

To generate the reference standard for each term, we used a sample of approximately 50,000 systematic reviews from the Epistemonikos Database, each with its list of included studies already uploaded [8]. For each target term, we identified all relevant reviews, and the reference standard was defined as all included studies with both a title and an abstract, drawn from at least ten systematic reviews and comprising at least 50 studies addressing the corresponding concept. When fewer than ten reviews were available, or the pool of included studies was smaller than 50, the team conducted additional searches to meet these thresholds. If these searches did not yield sufficient studies, all included studies from the eligible reviews were used as the reference standard.

The reference standard sample was divided into two subsets using random sampling. The first subset, representing 20 percent of the studies, was used to develop and refine the search strategy. Once the strategy was finalized, it was tested on the remaining 80 percent, which served as the validation set. If the strategy failed to meet the target sensitivity of 95 percent, the missed records were reviewed to determine whether they contained references to the concept of interest. When relevant studies had been overlooked, the strategy was further adjusted to capture those records and then retested on a newly generated validation set following the same procedure.

Number of available search strategies

The development of search strategies occurs alongside the ongoing expansion of the Epistemonikos Evidence Taxonomy (EET) which, while already comprehensive, is still rapidly growing [6]. Every term created within the taxonomy is reviewed by the search team. Search strategies are only omitted for broad terms where achieving adequate sensitivity is challenging or where a strategy would not be meaningful. Consequently, the number of available search strategies roughly corresponds to the number of terms in the EET. As of November 2025, this total stands at approximately 9,500.

System Development and Underlying Principles

The ABC Tool was developed as an integral part of the advanced search interface of an electronic database. It is designed to bridge the gap between complex search requirements and user accessibility. Currently, it is available in the advanced search interface of ED-Trials, a bibliographic database of randomized trials (https://trials.epistemonikos.org) [5], and it will soon be accessible through other databases managed by the Foundation. A complete list of available databases can be found at www.epistemonikos.org.

Editable drafts for customization

Recognizing that each evidence synthesis project has unique requirements, the generated strategies are offered as editable drafts. This flexibility allows users to refine, broaden, or narrow the search based on specific inclusion and exclusion criteria or according to the desired balance between sensitivity and specificity.

Enhanced reproducibility

The tool enhances reproducibility by providing a standardized and well-documented starting point for every search strategy. Even when users modify or tailor the strategy to their specific needs, the automatically generated baseline remains traceable and transparent. This feature plays a crucial role in maintaining the transparency and reliability of evidence synthesis processes.

Support for diverse skill levels

The tool is designed to support users with different levels of expertise in information retrieval. For those who lack advanced search skills, it provides guided access to well-constructed strategies, minimizing

common errors that often occur with manual Boolean formulation. For experienced users, it offers an efficient starting point that can be quickly customized and optimized.

Key features

The ABC Tool is still in its early stages, and user testing is currently ongoing. Therefore, changes and improvements can be expected in the near future. As of November 2025, several instructional materials are available. Within ED-Trials [5], users can access a brief interactive tour that introduces the core functionalities of the Advanced Search Interface, including the ABC Tool. Additionally, a comprehensive user guide is provided, featuring instructional videos and practical examples to help users make the most of the ABC tool.

The following features are currently available:

Direct entry of terms

When users select Boolean strategy from the field menu, they can type the concept of interest directly. As they type, the system offers suggestions drawn from all available pre-developed strategies. The resulting Boolean search string can then be inserted into either the Title field, to search within titles only, or the Title and Abstract field, depending on the desired search scope.

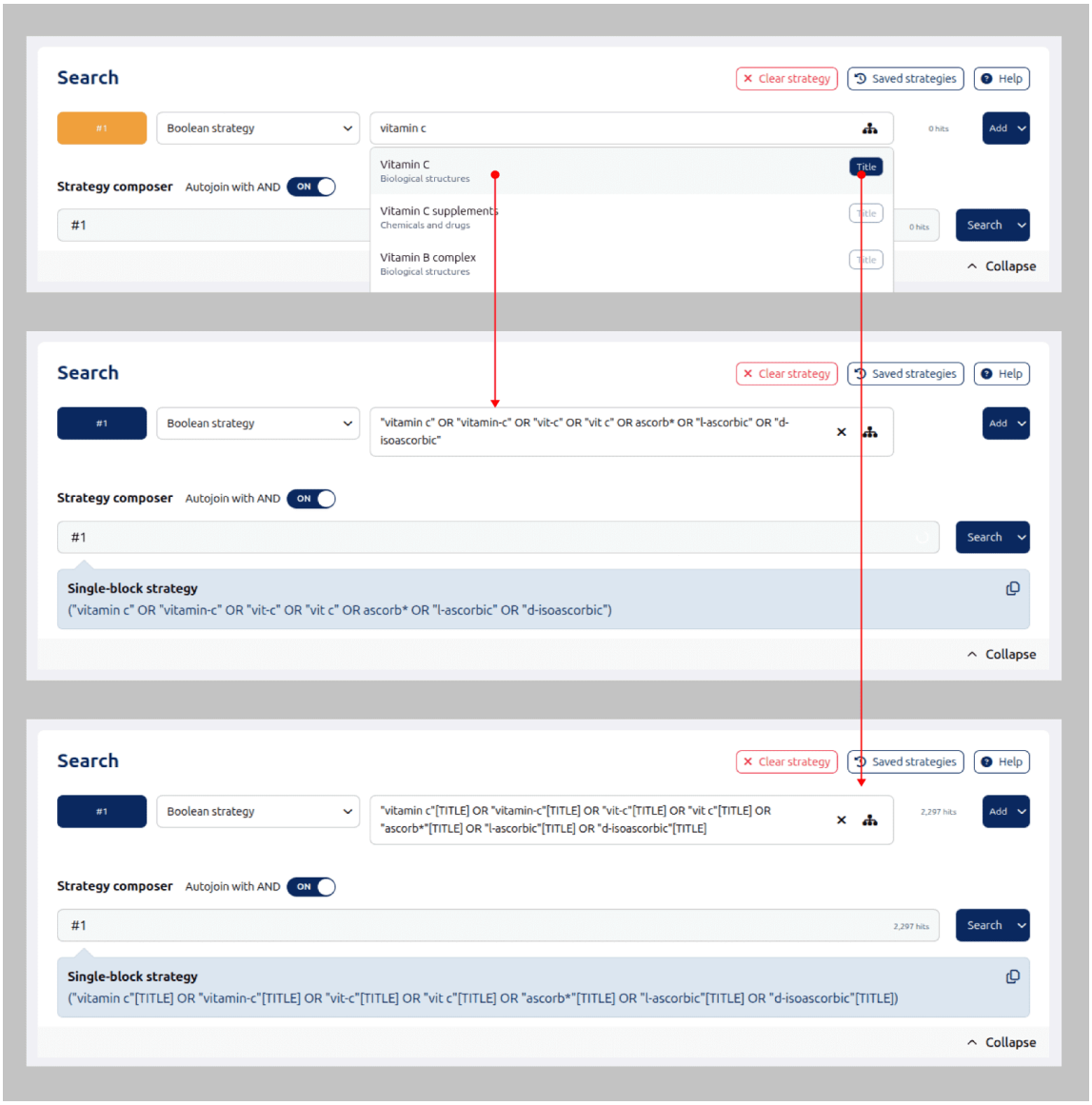

Figure 1. Direct entry of terms into the search interface.

Term browser

A taxonomy tree icon opens a navigation panel that allows users to browse all concepts according to the hierarchical structure of the Epistemonikos Evidence Taxonomy (EET) [6]. Users can select multiple terms from this interface, which are then inserted as separate lines in the advanced search interface.

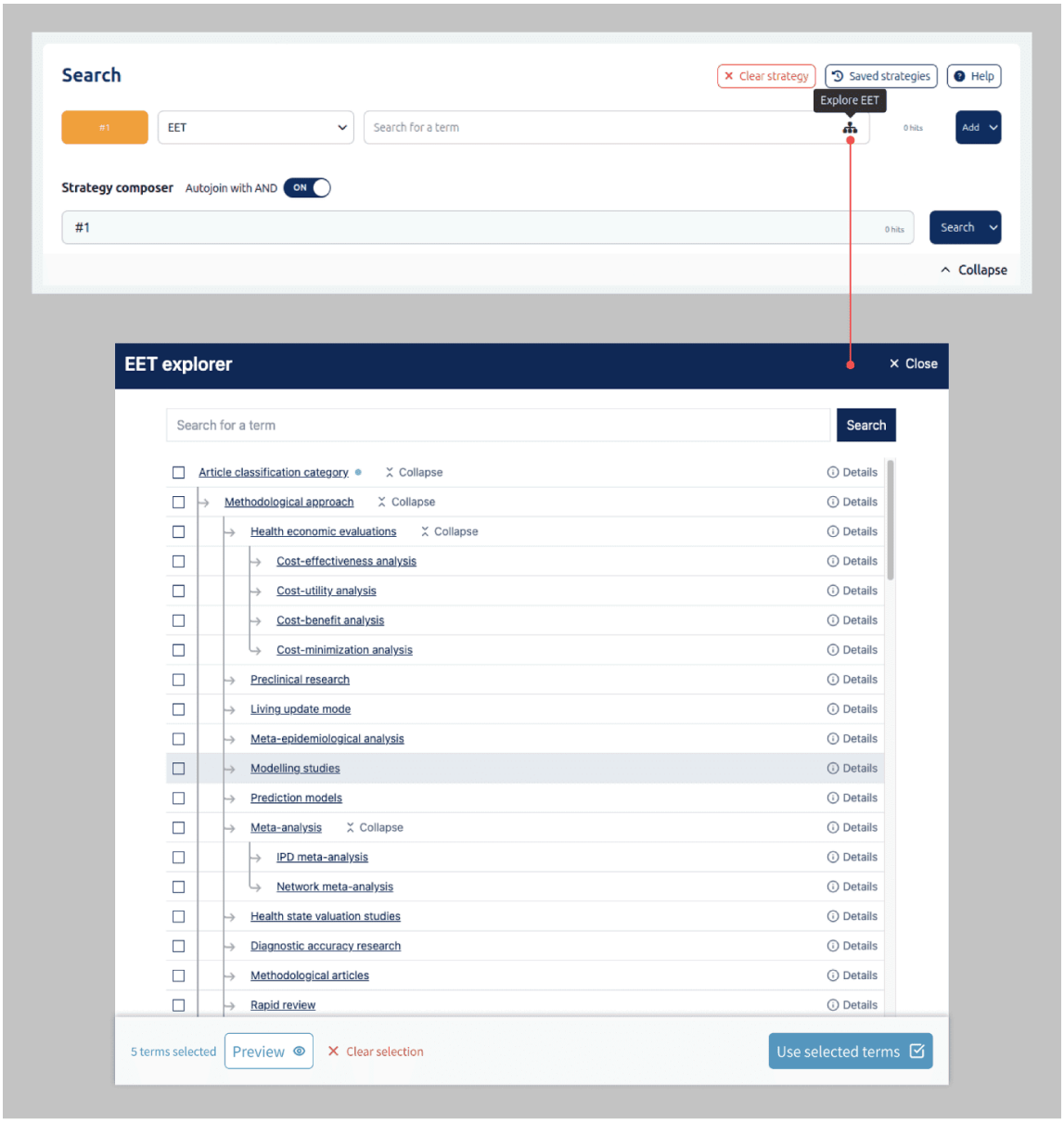

Figure 2. The EET Browser allows users to navigate the concepts, organized around the taxonomy.

Combination line

It is common practice to create a line that combines multiple lines related to the same concept. To minimize human error when manually referencing and typing line numbers (e.g., #1 OR #2 OR #3), the tool allows users to automatically insert a combination line. This line dynamically includes all preceding lines up to the previous combination line or the beginning of the strategy, streamlining the process and reducing the likelihood of mistakes (see Figure 3).

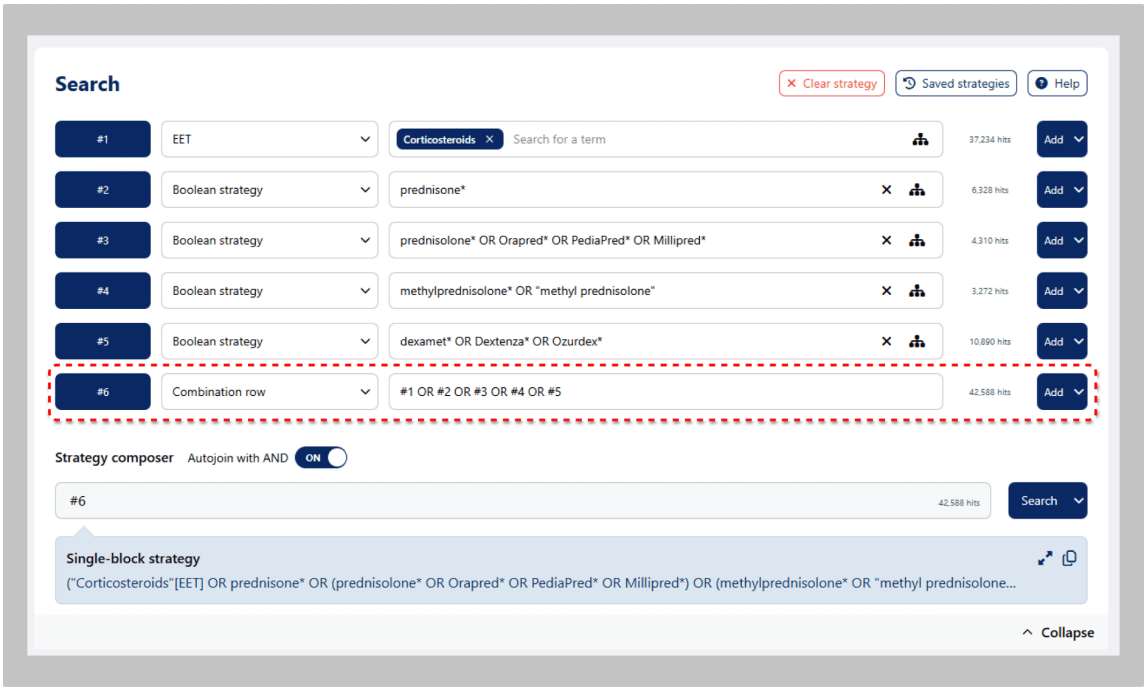

Figure 3. Semi-automated insertion of a line to combine all lines related to the same concept.

Strategy composer and autojoin feature

The final step in building a search strategy includes a line that combines the previously created lines using Boolean logic. The autojoin feature simplifies this process by automatically inserting the AND operator between individual lines (see Figure 4). When combination lines have been defined earlier in the strategy, the system prioritizes these and inserts the AND operator between them, ensuring a coherent and logically structured final query.

doi.org/10.32388/6FLT1Z

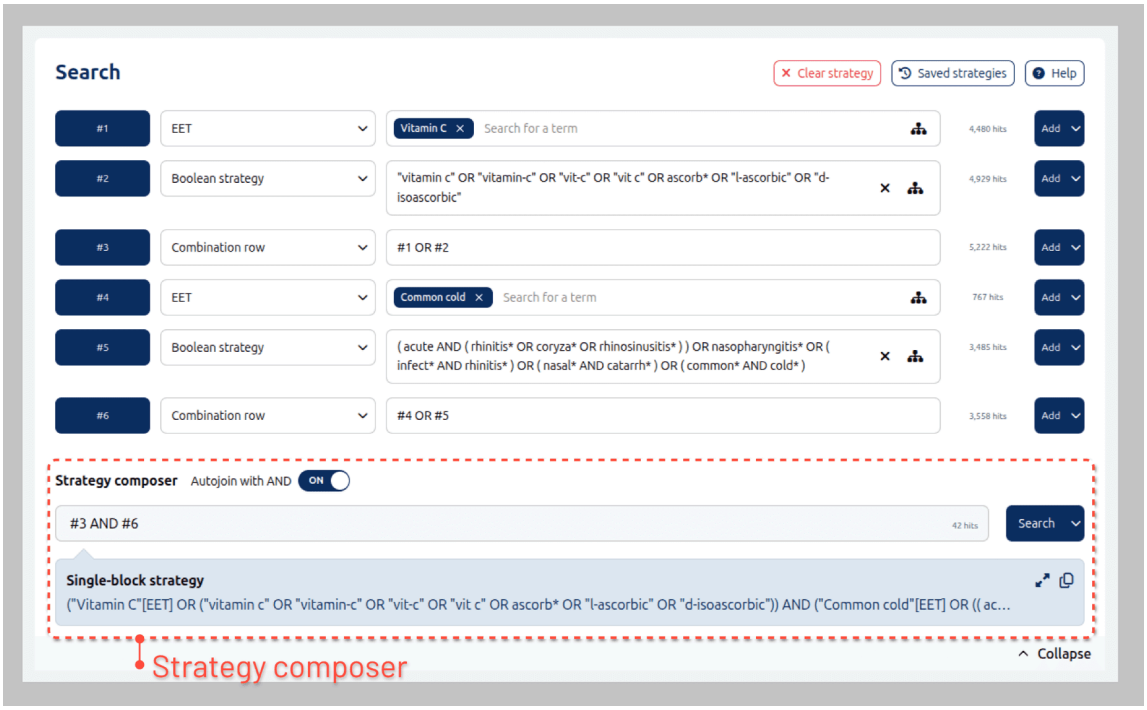

Figure 4. The strategy composer and autojoin feature facilitate the final combination of the different concepts of the search strategy.

Discussion and future directions

The ABC Tool represents a promising advancement in improving both the efficiency and standardization of search strategies for bibliographic databases. Its practical functionalities, such as direct term entry with real-time suggestions, a term browser built on the EET hierarchy, and features like combination line generation and autojoin, streamline the construction of Boolean queries while reducing the risk of human error. These tools collectively address long-standing challenges in evidence synthesis, particularly the complexity and time demands involved in developing search strategies manually. For example, the automated generation of combination lines mitigates a common source of error, where incorrect referencing of line numbers can compromise the logic of a search. By prioritizing usability and integrating automation, the ABC Tool can significantly cut down the time traditionally required to build high-quality strategies. To further support adoption and effective use, the tool includes a comprehensive user guide and an interactive tutorial, facilitating onboarding for users with varying levels of expertise.

Several ongoing initiatives aim to store and share pre-developed, high-quality search strategies for reuse. These include searchRxiv [10], the OSF Search Strategy Repository Prototype [11], and the Repository of

search strategies on the Virtual Health Library [12]. The ABC Tool, however, is not intended to replicate these efforts but to complement them. Unlike these repositories, which primarily host finalized, expert-developed strategies, the ABC Tool provides structured drafts intended to support and streamline the work of those who create or refine such strategies.

The ABC Tool offers several distinct advantages. It is highly comprehensive, with over 9,500 concept-specific strategies—substantially more than any existing initiative. Its integration with the Epistemonikos Evidence Taxonomy (EET) [6], a curated and continuously updated hierarchical vocabulary, provides a robust foundation for the consistent expansion and refinement of searchable terms. Each strategy also undergoes a systematic validation process, enhancing both reliability and coverage. This methodologically rigorous approach helps minimize the risk of overlooking relevant literature, a common issue with ad hoc or inconsistently designed searches. Additionally, the tool's full integration into an advanced search interface improves usability and reduces the potential for user error. While some overlap exists among these initiatives, future collaboration will be explored once the ABC Tool has completed its validation phase.

While the ABC Tool offers considerable benefits, several limitations and areas for further development warrant attention.

To maintain simplicity and ensure broad applicability, the current version of the tool relies on basic Boolean logic, namely the use of operators such as AND and OR, truncation, exact phrases, and parentheses. This approach facilitates compatibility across a wide range of topics and supports future translation into different database syntaxes. However, more advanced searches may require additional features, such as proximity operators, which are not yet supported. Expanding the tool's functionality to accommodate more complex search needs could enhance its utility for expert users.

The tool's integration with the Epistemonikos Evidence Taxonomy (EET) is a notable strength, but it also introduces a dependency. The tool's effectiveness is directly tied to the scope, accuracy, and regular updating of the taxonomy. Continued efforts to expand and refine the EET will be essential to maintain the quality and relevance of the generated strategies.

Although each strategy generated by the tool undergoes a rigorous internal validation process based on relative recall, external validation, whether by independent users or alternative methodological approaches, could further strengthen trust in the tool's outputs. Future research should also explore the

tool's real-world performance in systematic review workflows, including comparative evaluations against manual methods or other search automation tools.

In sum, the public release of the ABC Tool represents a meaningful advancement in the automation and standardization of literature searching. Its potential impact on the efficiency, transparency, and reproducibility of evidence synthesis is significant. Ultimately, the tool's value will depend on its ability to meet the practical needs of its users and to earn their confidence. If it succeeds in doing so, the ABC Tool will become increasingly essential as the scale and complexity of the scientific literature continue to grow.

Conclusion

The *Automated Boolean Strategy Creator* is an innovative tool designed to empower users in conducting high-quality, reproducible literature searches for evidence synthesis. By providing expertly crafted, editable Boolean strategies, it addresses critical challenges in information retrieval, making the process more efficient, accurate, and accessible, thereby strengthening the foundation of evidence-based practice.

Statements and Declarations

Funding

The salary of all the authors is totally or partially paid by the Epistemonikos Foundation. The authors received no additional financial support for the research, authorship, and/or publication of this article.

Potential Competing Interests

The Epistemonikos ABC Tool was developed and is maintained by the Epistemonikos Foundation, a non-profit organization with a strict policy to avoid financial conflicts of interest. Details about this policy and financial support can be found on the website of the Foundation (https://foundation.epistemonikos.org/). The salary of all the authors is totally or partially paid by the Epistemonikos Foundation.

All authors, as developers of the tool, have some degree of academic conflict of interest with this article.

Author Contributions

Gabriel Rada: Conceptualization, Writing - Original Draft preparation, Validation. **Camila Ávila-Oliver**: Writing - Review & Editing, Validation. **José Tomás Ramos-Rojas**: Writing - Review & Editing, Validation.

Juan Vásquez-Laval: Writing - Review & Editing, Software. **Camilo Vergara**: Writing - Review & Editing, Software. **Ana María Rojas-Gómez**: Data Curation. **Iván Jara**: Software. **Daniel Nava-Mosler**: Software. **Álex Silva-Zapata**: Software. **Diana Biscay**: Data Curation.

About the Authors

Epistemonikos (from the Greek “what is worth knowing”) is an independent, non-profit organization dedicated to providing reliable evidence for healthcare decision-making. One of its core areas of work is the development of systems that accelerate the identification, processing, and visualization of scientific evidence.

References

1. ^*Moher D, Liberati A, Tetzlaff J, Altman DG; PRISMA Group (2009). "Preferred Reporting Items for Systematic Reviews and Meta-Analyses: The PRISMA Statement." PLoS Med. **6**(7):e1000097. doi:10.1371/journal.pmed.1000097.*
2. a, b*Sampson M, McGowan J, Cogo E, Grimshaw J, Moher D, Lefebvre C (2009). "An Evidence-Based Practice Guideline for the Peer Review of Electronic Search Strategies." J Clin Epidemiol. **62**(9):944–52. doi:10.1016/j.jclinepi.2008.10.012.*
3. a, b*Cooper C, Booth A, Varley-Campbell J, Britten N, Garside R (2018). "Defining the Process to Literature Searching in Systematic Reviews: A Literature Review of Guidance and Supporting Studies." BMC Med Res Methodol. **18**(1):85. doi:10.1186/s12874-018-0545-3.*
4. ^*Saleh AA, Ratajeski MA, Bertolet M (2014). "Grey Literature Searching for Health Sciences Systematic Reviews: A Prospective Study of Time Spent and Resources Utilized." Evid Based Libr Inf Pract. **9**(3):28–50. doi:10.18438/B8DW3K.*
5. a, b, c*Rada G, Avila-Oliver C, Ramos-Rojas JT, Vasquez-Laval J, Vergara C, Cantor-Cruz F, et al (2025). "The Epistemonikos Database of Trials: A Comprehensive and Freely Accessible Bibliographic Source of Randomized Trials in Health." medRxiv. doi:10.1101/2025.11.16.25340330.*
6. a, b, c, d*Epistemonikos Foundation (2025). "Epistemonikos Evidence Taxonomy (EET)." Epistemonikos Foundation. Available at https://eet.epistemonikos.org.*
7. ^*Github repository (2025). "Github Repository." Github. Available at https://github.com/dperezrada/keywords2vec.*

8. ^a, ^b*Rada G, Pérez D, Araya-Quintanilla F, Ávila C, Bravo-Soto G, Bravo-Jeria R, et al; Epistemonikos project (2020). "Epistemonikos: A Comprehensive Database of Systematic Reviews for Health Decision-Making." BM C Med Res Methodol. **20**(1):286. doi:10.1186/s12874-020-01157-x.*
9. ^*Sampson M, Zhang L, Morrison A, Barrowman NJ, Clifford TJ, Platt RW, et al (2006). "An Alternative to the Hand Searching Gold Standard: Validating Methodological Search Filters Using Relative Recall." BMC Med Res Methodol. **6**:33. doi:10.1186/1471-2288-6-33.*
10. ^*Craven HJ, Hinrichs RJ, Stumpff JC (2024). "Using SearchRxiv for Depositing Evidence Syntheses Searches." Medical Institutional Repositories in Libraries (MIRL) Symposium. **13**. Available from: https://scholarlycommons.henryford.com/mirl/2024/program/13.*
11. ^*Ghezzi-Kopel K, Young S, Eldermire ERB, Johns EM, Kocher M (2024). "Search Strategy Repository Prototype." OSF. Available from: osf.io/wda5p.*
12. ^*Repository of search strategies on the Virtual Health Library (2025). "Repository of Search Strategies on the Virtual Health Library." Virtual Health Library. Available at https://bvsalud.org/queries/.*

Declarations

Funding: The salary of all the authors is totally or partially paid by Epistemonikos Foundation. The authors received no additional financial support for the research, authorship, and/or publication of this article.

Potential competing interests: Epistemonikos ABC Tool was developed and is maintained by Epistemonikos Foundation, a non-profit organization with a strict policy to avoid financial conflict of interests. Details about this policy and financial support can be found on the website of the Foundation (https://foundation.epistemonikos.org/). The salary of all the authors is totally or partially paid by Epistemonikos Foundation. All authors, as developers of the tool, have some degree of academic conflict of interests with this article.