

DEBATE

Open Access



Bubble effect: including internet search engines in systematic reviews introduces selection bias and impedes scientific reproducibility

Marko Ćurković^{1*}  and Andro Košec²

Abstract

Background: Using internet search engines (such as Google search) in systematic literature reviews is increasingly becoming a ubiquitous part of search methodology. In order to integrate the vast quantity of available knowledge, literature mostly focuses on systematic reviews, considered to be principal sources of scientific evidence at all practical levels. Any possible individual methodological flaws present in these systematic reviews have the potential to become systemic.

Main text: This particular bias, that could be referred to as (re)search bubble effect, is introduced because of inherent, personalized nature of internet search engines that tailors results according to derived user preferences based on unreproducible criteria. In other words, internet search engines adjust their user's beliefs and attitudes, leading to the creation of a personalized (re)search bubble, including entries that have not been subjected to rigorous peer review process. The internet search engine algorithms are in a state of constant flux, producing differing results at any given moment, even if the query remains identical. There are many more subtle ways of introducing unwanted variations and synonyms of search queries that are used autonomously, detached from user insight and intent. Even the most well-known and respected systematic literature reviews do not seem immune to the negative implications of the search bubble effect, affecting reproducibility.

Conclusion: Although immensely useful and justified by the need for encompassing the entirety of knowledge, the practice of including internet search engines in systematic literature reviews is fundamentally irreconcilable with recent emphasis on scientific reproducibility and rigor, having a profound impact on the discussion of the limits of scientific epistemology. Scientific research that is not reproducible, may still be called science, but represents one that should be avoided. Our recommendation is to use internet search engines as an additional literature source, primarily in order to validate initial search strategies centered on bibliographic databases.

Keywords: Literature searches, Reporting, Web searching, Research ethics, Scientific rigor

Background

Transparency, reproducibility, and rigor were recently reemphasized as fundamental characteristics of scientific epistemology [1]. Historically, the scientific biomedical community canonized major ethical issues through the use of safeguards, such as informed consent, external independent review process, and the peer review process.

However, issues regarding transparency and reproducibility could be seen as ethical issues per se, as they provoke questions regarding ends and means of science itself [2, 3]. In order to integrate the vast quantity of available knowledge, literature mostly focuses on systematic reviews, considered to be principal sources of scientific evidence at all practical levels [4]. Any possible individual methodological flaws present in these systematic reviews have the potential to become systemic [5].

* Correspondence: markocurak@gmail.com

¹University Psychiatric Hospital Vrapče, Bolnička cesta 32, Zagreb, Croatia
Full list of author information is available at the end of the article



Main text

Many guidelines exist on how to methodologically plan, implement and report a search of biomedical scientific literature. Current standards of reporting on search methodology have been found lacking and certain standards have already been proposed [6, 7]. Almost none of them have addressed the bubble effect issue. That is a tendency to be selectively exposed to personalized information in a way that influences individual beliefs and attitudes [8]. Systematic literature reviews start with a bibliographic database search accessing a large number of peer-reviewed scientific studies. It has become increasingly frequent to include internet search engines as access points as well [6, 7, 9]. Internet search engines can be useful in reviewing literature not found in common bibliographic databases. The need for this kind of supplemental literature is mostly justified with producing more comprehensive and applicable outcomes of scientific research [7, 9]. Although internet search engines seem immensely useful, they produce multiple sources of bias, and are ultimately irreproducible, no matter how seriously one takes transparency and rigor into account. The end user, whoever that might be, is principally concerned with the usefulness of data, and often does not pay attention to the underlying search methodology [10]. Alternative trajectories of acquiring scientific knowledge are tempting, but add to the risk of bypassing one of the primary safeguards – including entries that have not been subjected to rigorous peer review process. Moreover, internet search engines tailor search results according to derived user preferences based on unreproducible criteria. Pariser introduced the term “filter bubble” in 2011, as a personal selection bias inherent to internet search engines [8]. Some internet search engines support the use of advanced search features such as Boolean logic, but are not adequate equivalents of bibliographic databases. In contrast, internet search engines use personalized algorithms in order to evaluate and stratify results according to the trustworthiness of websites, and relevance to the search query in comparison to the end user’s search history, among other things [8, 11]. The most common example of personalization is the redirection to a country specific search engine default version. There are many more subtle ways of adjusting for the user’s beliefs and attitudes, leading to the creation of a personalized search bubble. In addition, variations and synonyms of search queries are used autonomously, detached from user insight and intent [11]. The internet search engine algorithms are in a state of constant flux, producing differing results at any given moment, even if the query remains identical. Moreover, internet content is inherently unstable. Even the most well-known and respected systematic literature reviews do not seem immune to the some negative

implications of the search bubble effect [6, 12–15]. Analyses have shown that internet searches implemented within Cochrane systematic reviews were not reported in sufficient detail to be transparent and reproducible [14]. Some of the most common issues found were inconclusive reporting on search queries and search limits, or a descriptive account of search methodology [14].

When using internet search engines for scientific purposes, some recommendations have been already stated, such as logging out of personal accounts, automatic or manual clearing of web search history, turning off web history options, using anonymous browsing options or using advanced search options [12]. Using speech marks or Verbatim options may reduce automatic reinterpreting of search queries as well as using meta-search engines that operate on different underlying settings (such as DuckDuckGo, Search Encrypt or StartPage). Simplification of retrieval and storage of scientific data from internet search engines has also been advised [16]. Even with regularly implementing these, systematic literature reviews using internet search engines remain vulnerable to the issue of reproducibility. Despite the fact that even systematic reviews that are based solely on bibliographic databases may not be entirely reproducible [17], using internet search engines has an profound, additional negative influence primarily on the processes of data searching, retrieval, storage and reporting of systematic literature reviews [6, 7, 9, 14, 15, 18]. Nonetheless, internet search engines are a beneficial method of reaching specific, predefined sources of data (such as sites of relevant agencies). Even used for those, obvious purposes, issues of data retrieval, storage and reporting of search methodology may persist.

The only solution at present, having in mind that internet search engines may be impossible to avoid as they provide valuable data that cannot be reached by other means, should be respect of the principle of transparency. Authors should disclose all relevant details of their search queries as well as their immediate context. Such an approach, in order to become relevant, demands a commitment from the entire scientific community. For now, using internet search engines and its associated lack of guidance in making internet searching reproducible fails to identify results without introducing bias. This practice may bring more harm than benefit.

Our recommendation is to use internet search engines as an additional literature source, primarily in order to validate and review initial search strategies centered on bibliographic databases. Internet search engines used with caution may be useful in the preparation phase of systematic reviews, to refine and create more robust bibliographic database search strategies. This may seem contradictory to general recommendations for conducting systematic review, which advise to search multiple

databases and use additional search strategies [19, 20], but there is no such a thing as reproducible research when it is conducted by primarily unscientific means. In other words, scientific research that is not reproducible, may still be called science, but represents one that should be avoided. When facing the issue of reproducibility, one should at least make best effort to prevent, predict and finally to control possible harms. These commitments cannot be met when using internet search engines in systematic reviews as they are out of subjects reach.

Conclusion

If we choose to encompass the entirety of knowledge, there is a good chance the (re)search bubble effect will lead us to results that have already been chosen for us. These (re)search bubble characteristics need to be addressed, as they are in stark contrast with transparency, reproducibility and rigor – the prerequisites of scientific thinking.

Acknowledgements

Not applicable.

Funding

This article did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Availability of data and materials

Not applicable.

Authors' contributions

MC provided initial idea for the manuscript. MC and AK were both included in planning, development, revision and approval of the final version of the manuscript.

Ethics approval and consent to participate

No ethical approval was needed for this manuscript.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Author details

¹University Psychiatric Hospital Vrapče, Bolnička cesta 32, Zagreb, Croatia.

²Department of Otorhinolaryngology and Head and Neck Surgery, University Hospital Center Sestre milosrdnice, Vinogradska cesta 29, Zagreb, Croatia.

Received: 3 March 2018 Accepted: 30 October 2018

Published online: 13 November 2018

References

- Munafò MR, Nosek BA, Bishop DVM, et al. A manifesto for reproducible science. *Nat Hum Behav.* 2017;1(1):0021. <https://doi.org/10.1038/s41562-016-0021>.
- Emanuel E, et al, editors. *The Oxford textbook of clinical research ethics.* New York: Oxford University; 2008.

- Bernabe R, Van Thiel G, Van Delden J. What do international ethics guidelines say in terms of the scope of medical research ethics? *BMC Med Ethics.* 2016;17:23. <https://doi.org/10.1186/s12910-016-0106-4>.
- Shamseer L, Moher D, Clarke M, et al. PRISMA-P (Preferred Reporting Items for Systematic review and Meta-Analysis Protocols) 2015 checklist : recommended items to address in a systematic review protocol. *BMJ.* 2015; 349(1):7647. <https://doi.org/10.1136/bmj.g7647>.
- Møller MH, Ioannidis JPA, Darmon M, Møller MH, JPA I, Darmon M. Are systematic reviews and meta-analyses still useful research? We are not sure. *Intensive Care Med.* 2018;44(4):518–20. <https://doi.org/10.1007/s00134-017-5039-y>.
- Adams J, Hillier-Brown FC, Moore HJ, et al. Searching and synthesising 'grey literature' and 'grey information' in public health: critical reflections on three case studies. *Syst Rev.* 2016;5:164. <https://doi.org/10.1186/s13643-016-0337-y>.
- Stansfield C, Dickson K, Bangpan M. Exploring issues in the conduct of website searching and other online sources for systematic reviews: how can we be systematic? *Syst Rev.* 2016;5:191. <https://doi.org/10.1186/s13643-016-0371-9>.
- Pariser E. *The filter bubble: what the internet is hiding from you.* London: Viking; 2011.
- Mahood Q, Van Eerd D, Irvin E. Searching for grey literature for systematic reviews: challenges and benefits. *Res Synth Methods.* 2014;5:221–34. <https://doi.org/10.1002/jrsm.1106>.
- Stevens ML, Moseley A, Elkins MR, Lin CC, Maher CG. What searches do users run on PEDro? An analysis of 893,971 search commands over a 6-month period. *Methods Inf Med.* 2016;55(4):333–9. <https://doi.org/10.3414/ME15-01-0143>.
- Blakeman K. Finding research information on the web: how to make the most of Google and other free search tools. *Sci Prog.* 2013;96:61–84.
- Ćurković M. Need for Controlling of the Filter Bubble Effect. *Sci Eng Ethics.* 2017. <https://doi.org/10.1007/s11948-017-0005-1>. Epub ahead of print.
- Ćurković M. The implications of using internet search Engines in Structured Scientific Reviews. *Sci Eng Ethics.* 2018. <https://doi.org/10.1007/s11948-017-0013-1>. Epub ahead of print.
- Briscoe S. A review of the reporting of web searching to identify studies for Cochrane systematic reviews. *Res Synth Methods.* 2017. <https://doi.org/10.1002/jrsm.1275>. Epub ahead of print.
- Briscoe S. Web searching for systematic reviews: a case study of reporting standards in the UK health technology assessment programme. *BMC Res Notes.* 2015;8:153. <https://doi.org/10.1186/s13104-015-1079-y>.
- Haddaway N, Collins A, Coughlin D, Kirk S. A rapid method to increase transparency and efficiency in web-based searches. *Environmental Evidence.* 2017;6(1):1–14. <https://doi.org/10.1186/s13750-016-0079-2>.
- Yoshii A, Plaut DA, KA KMG, Anderson MJ, Wellik KE. Analysis of the reporting of search strategies in Cochrane systematic reviews. *J Med Libr Assoc.* 2009;97(1):21–9. <https://doi.org/10.3163/1536-5050.97.1.004>.
- Cooper C, Booth A, Britten N, Garside R. A comparison of results of empirical studies of supplementary search techniques and recommendations in review methodology handbooks: a methodological review. *Syst Review.* 2017;6(1):234. <https://doi.org/10.1186/s13643-017-0625-1>.
- Bramer WM, Rethlefsen ML, Kleijnen J, Franco OH. Optimal database combinations for literature searches in systematic reviews: a prospective exploratory study. *Syst Rev.* 2017;6(1):245. <https://doi.org/10.1186/s13643-017-0644-y>.
- Cooper C, Booth A, Varley-Campbell J, Britten N, Garside R. Defining the process to literature searching in systematic reviews: a literature review of guidance and supporting studies. *BMC Med Res Methodol.* 2018;18(1):85. <https://doi.org/10.1186/s12874-018-0545-3>.

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more biomedcentral.com/submissions

