

LETTER

Open Access



Exponential growth of systematic reviews assessing artificial intelligence studies in medicine: challenges and opportunities

Thilo von Groote^{1*} , Narges Ghoreishi^{2†}, Maria Björklund³, Christian Porschen¹ and Livia Puljak⁴

Abstract

The evidence-based medicine (EBM) movement is stepping up its efforts to assess medical artificial intelligence (AI) and data science studies. Since 2017, there has been a marked increase in the number of published systematic reviews that assess medical AI studies. Increasingly, data from observational studies are used in systematic reviews of medical AI studies. Assessment of risk of bias is especially important in medical AI studies to detect possible “AI bias”.

Keywords: Artificial intelligence, Evidence-based medicine, Systematic reviews

Dear Editor,

With digitalization and enhanced computing power, the scientific community is amassing data at an unprecedented rate. With this big data, clinicians and biomedical researchers collaborate with computer scientists to use artificial intelligence (AI) to detect signals from noise [1]. AI and data science are expected to contribute to significant improvements in healthcare and medicine [2]. Therefore, it appears essential to synthesize evidence from medical AI studies and assess the quality of these new data-driven interventions and tools.

Among various types of evidence, systematic reviews and meta-analyses are the standard for guideline development and guide researchers, clinicians, and policymakers alike [3]. Furthermore, increased quality of evidence-synthesis may support patients and physicians to trust the AI applications and their adoption in the healthcare sector.

The number of systematic reviews assessing studies on medical AI is growing rapidly

Since its beginning in the 1980s, the EBM movement has fostered the development of evidence syntheses, which is reflected in the rapidly growing number of systematic reviews published each year (see the black line, Fig. 1). Starting some years later, the number of medical AI studies has been proliferating similarly with a rapid pace since 2000 and a marked increase from 2017 onwards (see the gray line, Fig. 1). However, we could not find reports in the literature about the number and characteristics of systematic reviews which include medical AI studies. To assess the number of such studies and whether their number was growing in line with the medical AI studies, we performed an extensive literature search that analyzed first the number of all medical publications in PubMed and EMBASE. Then, both absolute numbers and percentages were compared in the three groups: systematic reviews overall, medical AI studies overall, and systematic reviews with a medical AI topic (Fig. 1). Percentages are shown in Fig. 1 for ease of comparison.

Compared to the overall number of systematic reviews published and to the immense growth in medical AI research, the proportion of systematic reviews specifically investigating medical AI studies has been keeping

[†]Thilo von Groote and Narges Ghoreishi contributed equally to this work.

*Correspondence: thilo.vongroote@ukmuenster.de

¹ Department of Anaesthesiology, Intensive Care and Pain Medicine, University Hospital Münster, Münster, Germany
Full list of author information is available at the end of the article



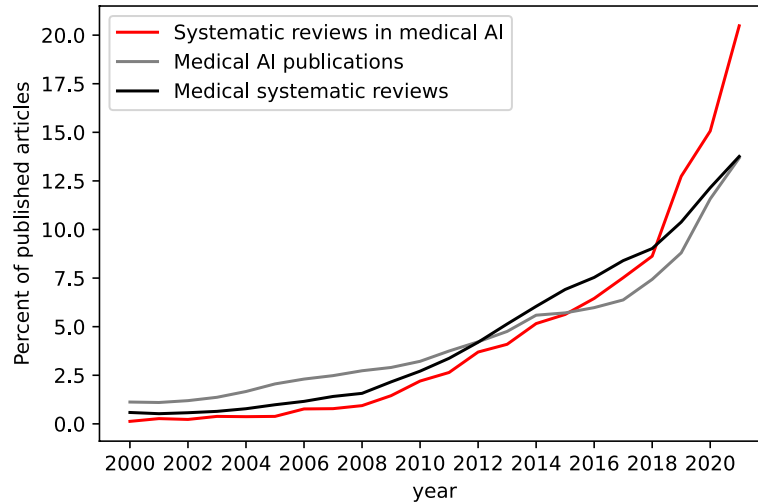


Fig. 1 Medical articles on artificial intelligence (AI), systematic reviews overall, and systematic reviews specifically investigating medical AI studies as a percentage of published articles overall; indexed per year in PubMed and EMBASE, from 2000 to 2021. Supplementary file 1 reports search strategies and software used to retrieve and analyze these records. Supplementary file 2 reports tabular results of the search

up in rate and even exceeds the pace of the former two in recent years (See the red line, Fig. 1).

The importance of systematic reviews on medical AI studies: challenges and opportunities

Systematic reviews on medical AI studies are of high importance to guide the implementation of AI tools in the healthcare sector and to provide an overview of AI-provided evidence for medical researchers, physicians, and patients. However, there are multiple challenges associated with this type of evidence synthesis. One is the

potential risk of bias [4]. Especially in machine-learning models, the training population has significant implications for an algorithms’ performance, generalizability, and its equity or discrimination, with a significant risk for the so-called AI bias [4–6]. The risk of selection bias with AI data is an important methodological consideration. The growing number of observational studies in medical AI systematic reviews (Fig. 2) calls for more attention to bias due to the non-randomized format of these studies. In addition, medical AI methods tend to require big data and, as a result, are heavily based on secondary data (data

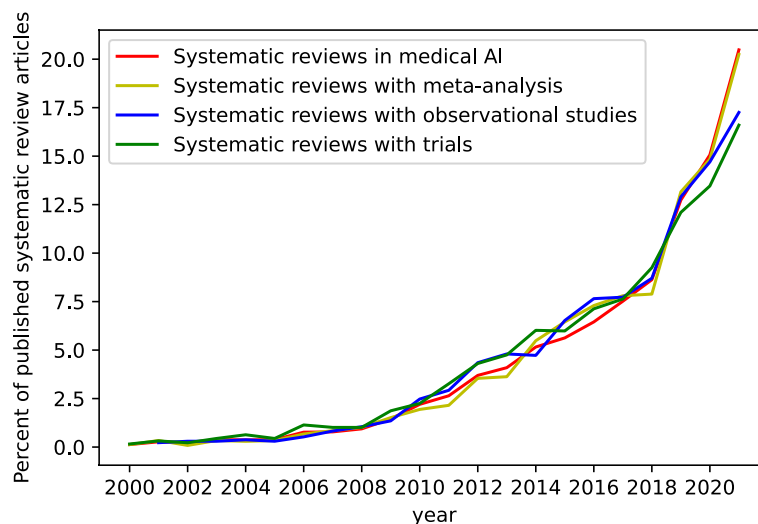


Fig. 2 Medical AI systematic reviews in terms of content: containing a meta-analysis, containing observational studies, and containing randomized controlled studies as a percentage of published articles overall; indexed per year in PubMed and EMBASE, from 2000 to 2021

that is not collected for the purpose of the research), such as electronic medical records [7].

Since its inception, the EBM movement has played a vital role in challenging bias and scrutinizing the scientific evidence from clinical trials and biomedical studies. Given the special importance of systematic reviews for the assessment of the evidence from medical AI studies, the rapid growth in the number of such reviews must be acknowledged as a positive development. In our opinion, this field holds much potential and room for further quality improvement. Therefore, a focus on investment in and development of adequate training and tools for EBM researchers to assess medical AI studies through high-quality systematic reviews would be worthwhile. Now is the time to intensify these efforts.

Abbreviations

AI: Artificial intelligence; EBM: Evidence-based medicine; EMBASE: Excerpta Medica database.

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s13643-022-01984-7>.

Additional file 1. Search strategy to identify publications on AI/ML.

Additional file 2.

Acknowledgements

We would like to acknowledge the Open Access Publication Fund of the University of Münster.

Authors' contributions

All authors have made substantial contributions to the manuscript and fulfill the criteria for authorship as defined by the ICMJE. The authors contributed as follows: conceptualization (TvG, NG); data curation (NG, MB); formal analysis (TvG, NG, CP, MB); funding acquisition (n/a); investigation (TvG, NG, CP, MB, LP); methodology (TvG, NG, CP, MB, LP); project administration (TvG, LP); resources (n/a); software (EndNote 20, Microsoft Word, Rayyan, Python); supervision (TvG, NG, LP); visualization (TvG, NG, CP, MB); writing—original draft (TvG, NG); and writing—review & editing (TvG, NG, CP, MB, LP). The author(s) read and approved the final manuscript.

Authors' information

A multidisciplinary and international team of researchers conducted this work. Dr. von Groote and Dr. Porschen are medical doctors, working in Anaesthesiology and Critical Care. Both are active clinical trial researchers. Additionally, they lead a group of junior researchers who employ data science methods and artificial intelligence to clinical data of critically ill patients, as well as conducts systematic reviews. Mrs. Ghoreishi is an epidemiologist and data scientist at the German Federal Institute for Risk Assessment (BfR). She has a background in pharmacy and epidemiology. Ms Björklund works as a research librarian and information specialist at Lund University in Sweden. Prof. Puljak is a researcher in the field of research methods, evidence synthesis, epidemiology, and pain. In addition to her faculty position at the Centre for Evidence-Based Medicine and Healthcare at the Catholic University of Croatia in Zagreb, she also holds a leadership role in Cochrane Croatia.

Funding

We acknowledge the kind support of the Open Access Publication Fund of the University of Münster (Germany) which supports Open Access Publication. The funders had no role in the study design, data collection, and analysis, decision to publish, or preparation of the manuscript. Dr. von Groote is supported

by a rotational position of the Clinical Research Unit 342 (KFO 342), funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) - ZA428/18-1. There was no additional funding received for this study.

Availability of data and materials

The search strategy is available in the supplementary files. Obtained data are available upon request.

Declarations

Ethics approval and consent to participate

The study included no primary patient data, but published studies were included. Therefore, no ethics approval or informed consent was necessary.

Consent for publication

All authors approved the final version of the manuscript. The authors have agreed to the publication of this manuscript.

Competing interests

The authors declare that they have no competing interests.

Author details

¹Department of Anaesthesiology, Intensive Care and Pain Medicine, University Hospital Münster, Münster, Germany. ²German Federal Institute for Risk Assessment, Berlin, Germany. ³Library/ICT, Faculty of Medicine, Lund University, Lund, Sweden. ⁴Center for Evidence-Based Medicine and Healthcare, Catholic University of Croatia, Zagreb, Croatia.

Received: 28 February 2022 Accepted: 15 May 2022

Published online: 28 June 2022

References

1. Topol EJ. High-performance medicine: the convergence of human and artificial intelligence. *Nat Med*. 2019;25(1):44–56.
2. Meskó B, Görög M. A short guide for medical professionals in the era of artificial intelligence. *NPJ Digit Med*. 2020;3(1):126.
3. Ioannidis JPA. Meta-research: why research on research matters. *PLoS Biol*. 2018;16(3):e2005468.
4. Challen R, Denny J, Pitt M, Gompels L, Edwards T, Tsaneva-Atanasova K. Artificial intelligence, bias and clinical safety. *BMJ Qual Saf*. 2019;28(3):231.
5. DeBrusk C. The risk of machine-learning bias (and how to prevent it). *MIT Sloan Manag Rev*. 2018.
6. Sun W, Nasraoui O, Shafto P. Evolution and impact of bias in human and machine learning algorithm interaction. *PLoS One*. 2020;15(8):e0235502.
7. Abidi SSR, Abidi SR. Intelligent health data analytics: A convergence of artificial intelligence and big data. *Healthc Manage Forum*. 2019;32(4):178–82. <https://doi.org/10.1177/0840470419846134>.

Publisher's Note

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