Identifications of similarity metrics for cancer patients: Protocol for a Scoping Review

Iryna Manuilova, Jan Bossenz, Annemarie Weise, Dominik Boehm, Cosima Strantz, Philipp Unberath, Niklas Reimer, Patrick Metzger, Thomas Pauli, Silke D. Werle, Susann Schulze, Sonja Hiemer, Arsenij Ustjanzew, Hans A. Kestler, Hauke Busch, Benedikt Brors, Jan Christoph

Submitted to: JMIR Research Protocols
on: April 11, 2024

Table of Contents

Identifications of similarity metrics for cancer patients: Protocol for a Scoping Review

Iryna Manuilova[1, 2] MS; Jan Bossenz[1]; Annemarie Weise[1]; Dominik Boehm[3, 4] MS; Cosima Strantz[5] MS; Philipp Unberath[3, 6] Dr.rer.biol.hum.; Niklas Reimer[7, 8, 9] MS; Patrick Metzger[10, 11] Dr rer nat; Thomas Pauli[10] Dr rer nat; Silke D. Werle[12] Dr rer nat; Susann Schulze[13] Dr med; Sonja Hiemer[13] Dr med; Arsenij Ustjanzew[14] MS; Hans A. Kestler[12] Dr rer nat; Hauke Busch[7, 8] Dr rer nat; Benedikt Brors[15, 16, 17, 18] Dr rer nat; Jan Christoph[19, 5, 2] Dr. rer. biol. hum.

[1]Junior Research Group (Bio-)Medical Data Science, Faculty of Medicine, Martin-Luther-University Halle-Wittenberg Halle(Saale) DE
[2]Data Integration Centre of the University Hospital Halle(Saale) Halle(Saale) DE
[3]Medical Center for Information and Communication Technology, Universitätsklinikum Erlangen, Friedrich-Alexander-Universität Erlangen-Nürnberg Erlangen DE
[4]Bavarian Cancer Research Center (Bayerisches Zentrum für Krebsforschung) Erlangen DE
[5]Chair of Medical Informatics, Friedrich-Alexander-Universität Erlangen-Nürnberg Erlangen DE
[6]SRH Fürth University of Applied Sciences, Fürth, Germany Fürth DE
[7]Medical Systems Biology Group, Lübeck Institute of Experimental Dermatology, Universität zu Lübeck Lübeck DE
[8]University Cancer Center Schleswig-Holstein, University Hospital Schleswig-Holstein Lübeck DE
[9]Medical Data Integration Center, University Hospital Schleswig-Holstein Lübeck DE
[10]Institute of Medical Bioinformatics and Systems Medicine, Medical Center-University of Freiburg, Faculty of Medicine, University of Freiburg Freiburg DE
[11]German Cancer Research Center (DKFZ) Heidelberg, Clinical Trial Office Heidelberg DE
[12]Institute of Medical Systems Biology, Ulm University Ulm DE
[13]Krukenberg Cancer Center Halle(Saale) Halle(Saale) DE
[14]Institute of Medical Biostatistics, Epidemiology and Informatics (IMBEI), University Medical Center of the Johannes Gutenberg-University Mainz Mainz DE
[15]Division Applied Bioinformatics, German Cancer Research Center (DKFZ) Heidelberg DE
[16]German Cancer Consortium Heidelberg DE
[17]National Center for Tumor Diseases (NCT) Heidelberg DE
[18]Medical Faculty Heidelberg and Faculty of Biosciences, Heidelberg University Heidelberg DE
[19]Junior Research Group (Bio-) Medical Data Science, Faculty of Medicine, Martin Luther-University Halle-Wittenberg Halle (Saale) DE

Corresponding Author:
Iryna Manuilova MS
Junior Research Group (Bio-)Medical Data Science, Faculty of Medicine, Martin-Luther-University Halle-Wittenberg
Magdeburger Str. 8
Halle(Saale)
DE

Abstract

Background: Understanding the similarities of cancer patients is essential to advancing personalized medicine, improving patient outcomes, and developing more effective and individualized treatments. It enables researchers to discover important patterns, biomarkers, and treatment strategies that can have a significant impact on cancer research and oncology. In addition, the identification of previously successfully treated patients supports oncologists in making treatment decisions for a new patient who is clinically or molecularly similar to the previous patient.

Objective: The planned review aims to systematically summarize, map, and describe existing evidence to understand how patient similarity is defined and used in cancer research and clinical care.

Methods: To systematically identify relevant studies and to ensure reproducibility and transparency of the review process, a comprehensive literature search will be conducted in several bibliographic databases, including Web of Science, PubMed, LIVIVIVO, and MEDLINE, covering the period from 1998 to February 2024. After the initial duplicate deletion phase, a study selection phase will be applied using Rayyan, which consists of three distinct steps: Title and Abstract Screening, Disagreement Resolution, and Full-Text Screening. To ensure the integrity and quality of the selection process, each of these steps is preceded

by a pilot testing phase. This methodological process will culminate in the presentation of the final research results in a structured form according to the PRISMA-ScR flowchart. The protocol has been registered in the Journal of Medical Internet Research (JMIR).

Results: This protocol outlines the methodologies employed in conducting the scoping review. A search of the specified electronic databases and after removing duplicates resulted in 1,183 unique records. As of March 2024, the review process has moved to the full-text evaluation phase. At this stage, data extraction will be conducted using a pre-tested chart template.

Conclusions: The scoping review protocol, centered on these main concepts, aims to systematically map the available evidence on patient similarity among cancer patients. By defining the types of data sources, approaches, and methods used in the field, and aligning these with the research questions, the review will provide a foundation for future research and clinical application in personalized cancer care. This protocol will guide the literature search, data extraction, and synthesis of findings to achieve the review's objectives.

(JMIR Preprints 11/04/2024:58705)
DOI: https://doi.org/10.2196/preprints.58705

Preprint Settings

1) Would you like to publish your submitted manuscript as preprint?

✓ **Please make my preprint PDF available to anyone at any time (recommended).**
Please make my preprint PDF available only to logged-in users; I understand that my title and abstract will remain visible to all users.
Only make the preprint title and abstract visible.
No, I do not wish to publish my submitted manuscript as a preprint.

2) If accepted for publication in a JMIR journal, would you like the PDF to be visible to the public?

✓ **Yes, please make my accepted manuscript PDF available to anyone at any time (Recommended).**
Yes, but please make my accepted manuscript PDF available only to logged-in users; I understand that the title and abstract will remain v
Yes, but only make the title and abstract visible (see Important note, above). I understand that if I later pay to participate in <a href="http

Original Manuscript

Authors

Iryna Manuilova[1,2], Jan Bossenz[1], Annemarie Weise[1], Dominik Boehm[3,4], Cosima Strantz[5], Philipp Unberath[3,6], Niklas Reimer[7,8,9], Patrick Metzger[10,11], Thomas Pauli[10], Arsenij Ustjanzew[12], Silke D. Werle[13], Susann Schulze[14], Sonja Hiemer[14], Hans A. Kestler[13], Hauke Busch[7,8], Benedikt Brors[15,16,17,18], Jan Christoph[1,2,5]

[1] Junior Research Group (Bio-)Medical Data Science, Faculty of Medicine, Martin Luther-University Halle-Wittenberg, Halle (Saale), Germany
[2] Data Integration Centre of the University Hospital Halle(Saale),Halle (Saale), Germany
[3] Medical Center for Information and Communication Technology, Universitätsklinikum Erlangen, Friedrich-Alexander-Universität Erlangen-Nürnberg, Erlangen, Germany
[4] Bavarian Cancer Research Center (Bayerisches Zentrum für Krebsforschung)
[5] Chair of Medical Informatics, Friedrich-Alexander-Universität Erlangen-Nürnberg, Erlangen, Germany
[6] SRH Fürth University of Applied Sciences, Fürth, Germany
[7] Medical Systems Biology Group, Lübeck Institute of Experimental Dermatology, Universität zu Lübeck, Lübeck, Germany
[8] University Cancer Center Schleswig-Holstein, University Hospital Schleswig-Holstein, Lübeck, Germany
[9] Medical Data Integration Center, University Hospital Schleswig-Holstein, Lübeck, Germany
[10] Institute of Medical Bioinformatics and Systems Medicine, Medical Center-University of Freiburg, Faculty of Medicine, University of Freiburg, Freiburg, Germany
[11] German Cancer Research Center (DKFZ) Heidelberg, Clinical Trial Office, Heidelberg, Germany
[12] Institute of Medical Biostatistics, Epidemiology and Informatics (IMBEI), University Medical Center of the Johannes Gutenberg-University Mainz, Mainz, Germany
[13] Institute of Medical Systems Biology, Ulm University, Ulm, Germany
[14] Krukenberg Cancer Center Halle (Saale), Halle (Saale), Germany
[15] Division Applied Bioinformatics, German Cancer Research Center (DKFZ), Heidelberg, Germany
[16] German Cancer Consortium, Heidelberg, Germany
[17] National Center for Tumor Diseases (NCT), Heidelberg, Germany
[18] Medical Faculty Heidelberg and Faculty of Biosciences, Heidelberg University, Heidelberg, Germany

Corresponding Author:
Iryna Manuilova
Magdeburger Str. 8
06112 Halle (Saale), Germany
Phone: +49-3455-572651
Iryna.Manuilova@uk-halle.de

Identifications of similarity metrics for cancer patients: Protocol for a Scoping Review

Abstract

Background: Understanding the similarities of cancer patients is essential to advancing personalized medicine, improving patient outcomes, and developing more effective and individualized treatments.

It enables researchers to discover important patterns, biomarkers, and treatment strategies that can have a significant impact on cancer research and oncology. In addition, the identification of previously successfully treated patients supports oncologists in making treatment decisions for a new patient who is clinically or molecularly similar to the previous patient.

Objective: The planned review aims to systematically summarize, map, and describe existing evidence to understand how patient similarity is defined and used in cancer research and clinical care.

Methods: To systematically identify relevant studies and to ensure reproducibility and transparency of the review process, a comprehensive literature search will be conducted in several bibliographic databases, including Web of Science, PubMed, LIVIVIVO, and MEDLINE, covering the period from 1998 to February 2024. After the initial duplicate deletion phase, a study selection phase will be applied using Rayyan, which consists of three distinct steps: Title and Abstract Screening, Disagreement Resolution, and Full-Text Screening. To ensure the integrity and quality of the selection process, each of these steps is preceded by a pilot testing phase. This methodological process will culminate in the presentation of the final research results in a structured form according to the PRISMA-ScR flowchart. The protocol has been registered in the Journal of Medical Internet Research (JMIR).

Results: This protocol outlines the methodologies employed in conducting the scoping review. A search of the specified electronic databases and after removing duplicates resulted in 1,183 unique records. As of March 2024, the review process has moved to the full-text evaluation phase. At this stage, data extraction will be conducted using a pre-tested chart template.

Conclusions: The scoping review protocol, centered on these main concepts, aims to systematically map the available evidence on patient similarity among cancer patients. By defining the types of data sources, approaches, and methods used in the field, and aligning these with the research questions, the review will provide a foundation for future research and clinical application in personalized cancer care. This protocol will guide the literature search, data extraction, and synthesis of findings to achieve the review's objectives.

Keywords: Patient similarity; cancer research; patient similarity applications; precision medicine; cancer similarity metrics; scoping review protocol.

Introduction

Rapid advances in precision medicine have revolutionized cancer research, opening new opportunities to develop an unprecedented new, personalized view of each patient. The concept of precision medicine is seemingly simple: similar patients with similar characteristics share similar outcomes. By identifying important patient characteristics and traits, the search for similar patients contributes to the pursuit of precision medicine that may determine clinical outcomes through more precise targeting of treatment by genetic, biomarker, phenotypic, or psychosocial characteristics that differentiate a given patient from others with similar clinical presentations [1-4]. The ever-increasing volume and availability of health-related data is currently challenging the broad definitions of patient groups set out in the clinical practice guidelines. Defining a similarity measure that can handle the high-dimensional space of patient data is an essential step to enable stratification of patients into clinically meaningful subgroups [4-6]. The complex interaction between personalized patient treatment and the application of aggregate data underlines the fundamental understanding of modern oncology, which is based on the main principle that each patient has a deeply individual nature of their illness, and each case is special [7-9]. However, there is a parallel paradigm that demonstrates the essential role of applying existing data in improving the understanding of the individuality of cancer and optimizing the approach to personalized treatments. It suggests that a deep understanding of each patient's unique characteristics and subsequent selection

of therapeutic strategies can be greatly improved by identifying similarities between cancer patients. This approach indicates that the most effective individualized treatment strategies do not develop independently but instead result from comprehensive comparison and analysis of aggregate patient data [10, 11].

Patient similarity is a topic of significant interest and research in various areas of precision medicine, including cancer research. Some studies have explored the concept of patient similarity across different dimensions, such as genomics, clinical characteristics, treatment responses, and outcomes [1, 5, 12, 13].

Despite the extensive interest in this area, there is currently no systematic approach to clarify precisely what is understood by the concept of "patient similarity" in cancer research [6]. While individual studies may use various methodologies and metrics to assess patient similarity, there is a lack of consensus on combined approaches and definitions [4, 6, 9]. This creates an opportunity for further research to explore and define patient similarity more comprehensively.

Additionally, the definition and evaluation of common similarity metrics in cancer research that involves careful evaluation of both quantitative and qualitative factors need to be systemized. These metrics can serve as a powerful method for furthering the understanding of cancer and improving personalized patient care [6, 9]. Faced with all these research gaps, we want to conduct a scoping review.

Aim and research questions

The goal of our planned research is to collect and describe the existing knowledge that could help in defining and exploring how patient similarity is determined in cancer research and care. The scoping review addresses the main research question:

- What is understood by the concept of "patient similarity" in cancer research?

Several secondary questions have been developed to support and coordinate the analysis:

- What types of data sources are used to identify similarities between cancer patients
    - Molecular genetic data
    - Clinical data
    - Therapies or treatment
    - Histological data
- What different approaches and methods are used to identify and analyze similarities between cancer patients and which clinical relevance they have?
- Which types of cancer have been the most frequently researched when it comes to finding similarities between patients?
- What challenges and limitations have been observed in the existing literature when identifying similarities between cancer patients?

To the best of our knowledge, no scoping review has addressed the research questions proposed by this review.

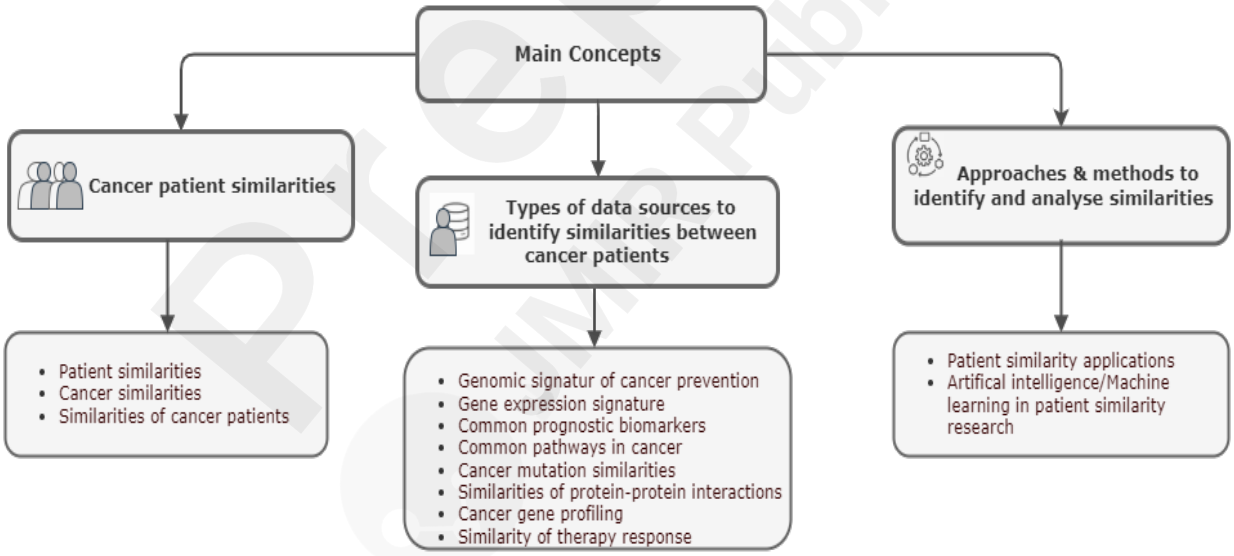
Methods

To ensure a transparent review process, our methodology will follow the “Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for Scoping Reviews (PRISMA-ScR) Checklist” and the Joanna Briggs Institute (JBI) Reviewer’s Manual on scoping reviews [14, 15]. The methodological process of conducting a scoping review will be iterative. Given this, it is expected that there may be some deviations from the originally developed a priori protocol, as a natural part of the iterative process, to refine and improve the review as it progresses. To ensure transparency in the conduct of the review, any deviations from the original protocol will be explicitly documented and reflected in the final manuscript of the review.

Main Concepts and Keywords

To guide the literature search, ensure the relevance of included studies, and improve the efficiency of the planned review process, three basic concepts and corresponding keywords were defined, which are graphically represented in Figure 1. Based on the main goals and research questions of this review, three basic concepts with corresponding keywords were selected (Figure 1), and peer review was sought from experts in the field to validate the keyword selection. These keywords guided the literature search, ensured the relevance of included studies, and improved the efficiency of the planned review process. The first concept is “Cancer patient similarities”. It focuses on understanding and summarizing the various dimensions in which cancer patients can be similar. The main keywords are “patient similarity,” “similarities of cancer patients,” and “cancer similarity metrics.” All these terms are broad and encompass any research comparing various aspects of patients, such as molecular genetic features, clinical features, treatment outcomes, survival rates, and applications used to investigate these similarities. The second concept is “Types of data sources to identify similarities between cancer patients”. The keywords in this category expand on the first category by providing detailed perspectives that define cancer similarity metrics. These are essential for identifying studies that explore similarities from different views, such as common prognostic biomarkers, protein-protein interaction pathways, gene expression analysis, and cancer gene profiling. The third concept is “Approaches & methods to identify and analyze similarities”, with the overarching keyword being “patient similarity applications.” This category aims to identify potential methods, technologies, algorithms that can be applied in cancer research regarding patient similarities.

Figure 1. Main concepts of planned Scoping Review and corresponding keywords.



Eligibility criteria: Inclusion and Exclusion

Even though our primary goal is to cover a wide range of studies to ensure broad coverage of studies on similarities to patients with cancer, we adhere to minimum exclusion criteria to maintain the quality and relevance of included studies. Pilot testing of Eligibility Criteria was performed on a sample of studies. Based on the findings inclusion and exclusion criteria were refined to ensure robustness and effectiveness in capturing relevant studies. For selecting pertinent studies for planned scoping review, we have established the following inclusion and exclusion criteria, outlined in Table 1.

Table 1. Eligibility Criteria.

Item	Inclusion criteria	Exclusion criteria
Type of studies	All types of studies providing substantial evidence or data relevant to cancer patient similarities.	Publications not addressing the aspects of similarity of cancer patients as defined in objectives. Bachelor's and master's theses, as well as unpublished manuscripts.
Population	Studies focusing on cancer patients of all ages, genders and ethnicities with different diagnoses.	Studies focusing on non-cancer conditions or animal studies.
Publications	Studies published within the last 25 years	Studies published more than 25 years ago.
Language	Studies published in English or German. (We chose English and German as the primary languages since they are the working languages of our team. This decision aims to facilitate better internal discussion and ensure the quality of results.)	Studies published in other languages

Types of evidence

To identify potentially relevant studies, and to ensure reproducibility and transparency of the planned review process, the following bibliographic databases were searched for literature coverage from 1998 through February 2024: Web of Science, PubMed, LIVIVO, MEDLINE. These databases were chosen for their comprehensive coverage of the biomedical and healthcare literature, ensuring a thorough review of studies regarding cancer patient similarities over 25 years. This approach ensures reproducibility and transparency of the review, facilitating detailed analysis of existing evidence and identification of research gaps in the field [14, 16].

Search strategy

As a result of the numerous discussions, the team developed a search strategy with three important steps: keyword search, snowball system and manual search. In the planned systematic review, keyword searching will serve as the primary method for identifying relevant studies. This approach involves the use of carefully selected keywords and keyword combinations defined using a nesting approach involving Boolean operators and field tags to provide precision (see Multimedia Appendix 1). Initially, the search will be conducted in the Web of Science database and after it, the

search queries will be refined and adapted for subsequent use in other chosen databases to identify relevant information on the research topic effectively. To reduce irrelevant findings in our research and to make it more exact [17, 18], we integrated the MeSH option (MeSH=neoplasms) with keyword searches. This was directly applied in Ovid MEDLINE and PubMed. However, we did not apply it in the Web of Science, which employs its unique indexing system [19]. To uncover literature that may have been missed after the initial keyword search, a "snowballing" method was applied [20, 21]. This involved reviewing the references of the searched articles to identify additional studies not covered in the initial database search. In addition to the keywords and snowballing searches, a manual search will be conducted. This will involve manually scanning relevant journals, conference proceedings, and other literature sources to identify studies that are not indexed in mainstream databases or published in less accessible formats. Applying this triangular search strategy to the main concepts identified in Figure 1 will provide a robust review of the existing literature and will allow the fullest possible range of studies to be integrated to identify potential similarities between cancer patients. The search options used in the individual databases are optimized to the strengths and specific functions of each platform to maximize the effectiveness and comprehensiveness of the literature selection.

Data extraction

Following the search, all identified references will be collected and uploaded to the reference management software package, EndNote 20.2.1, where duplicates will be removed. Subsequently, we will employ a selection process by our multidisciplinary team as proposed by Levac et al. [22, 23] using Rayyan, a web-based software designed to facilitate the process of conducting various types of reviews [24]. The study selection process will consist of three stages: Title-Abstract Screening, Disagreement Resolution, and finally, Full-Text Screening, outlined in Textbox 1. To ensure the quality of the overall study selection process, Pilot Testing will precede each step, and the following calibrated forms will be applied. The final results will be represented using the PRISMA-ScR flowchart [14].

Textbox 1. Stages of the study selection process.

Stage 1: Title-Abstract Screening
In this first step, we will screen titles and abstracts to quickly filter out publications that are not relevant to our research questions. This step will significantly reduce the volume of work required in the subsequent full-text review phase. To ensure objectivity, at least two reviewers in Blind Mode will screen each article. A total of 11 reviewers participated in the Title-Abstract Screening process.
Stage 2: Disagreement Resolution
Discrepancies during data extraction will be resolved based on the decision of an additional reviewer, as suggested by von Elm et al. [15]. Each reviewer will extract data independently. In case of disagreements, the independent reviewer will be consulted and make the final decision. All discrepancies and their resolutions will be documented for transparency.

Stage 3: Full-Text Screening

After the primary selection, we will conduct a full-text review of the remaining articles to further refine our selection based on specific inclusion and exclusion criteria directly related to our research questions. A total of 9 independent reviewers will be involved at this stage, as well as in the Data Extraction Process.

Management of data charting

From all publications that will be included in the research after the Full-Text Screening stage, data will be extracted. The data extraction process is manual, performed by our team using a predefined template to ensure systematic and accurate data capture. By using a predefined data extraction template, we can systematically capture all relevant information from each study, maintaining consistency and minimizing errors. Pilot testing of the template was conducted to ensure robustness and accuracy. Final draft extraction form is provided in Table 2.

Table 2. Data extraction table for the Scoping Review.

Item	Description	Keypoints
Metadata		
Title [a]	Title	
Details [a]	Author(1st), Journal, DOI	
Year of Publication [a]	YYYY	
Publication Type [a]	Type of publication	
Institute [a]	Corresponding institute	
Objective [a]	Main objective of the publication	
Methods [a]	Summary of the proposed methodological approach	
Results [a]	Short description of the results	
Conclusion [a]	Summarizing the main points and findings	
Keywords [a]	Main Keywords of the publication	
Research Findings		
Main Research Question		
What is understood by the concept of "patient similarity" in cancer research?	Key definition of "patient similarity" in the context of the publication.	Explanation of how the study defines patient similarity in the context of cancer research. This can include genetic, clinical, histological treatment-related similarities or view from methodological approaches.

		Determining the aspects of patient similarity that this publication focused on.
Secondary Questions		
1. What types of patient data are used to identify similarities?	Short description of the data (molecular genetics, clinical, histologic and treatment-related) used to define patient similarity.	Categorization of the types of patient data used to identify similarities.
2. What different approaches and methods are used to identify and to analyze similarities between cancer patients and which clinical relevance they have?	The approaches and methods employed to analyze and identify similarities (e.g. software, tools, algorithms). Information how this findings contribute to personalized medicine	Typification of the tools used to identify similarities. Clinical relevance of the methods and suitability for practical application.
3. Which types of cancer have been the most frequently researched when it comes to finding similarities between patients?	A list of cancer types that can be related as a basis for identifying similar cancer metrics	Identification of cancer types associated with the patient similarities in this study.
4. What challenges and limitations have been observed in the existing literature when identifying similarities between cancer patients?	List of potential limitations and challenges	Determination of the limits, future challenges, and unexplored areas in this field of research.

[a] mandatory field

This template is designed with several sections to capture essential information from the studies: ‘Metadata’ includes general information about the publication, and ‘Research Findings’ summarizes the main findings from each paper, specific to the research questions and objectives of the planned scoping review. The process of data charting, as in the case of the selection of sources of evidence, will start with a calibration step, which will help us prevent errors and ensure high inter-rater agreement [14].

Summarizing and presenting results

To comprehensively answer the main research question and related secondary questions, our findings will be summarized and presented using a structured approach to ensure clarity, consistency, and alignment with overarching objectives. Detailed narrative synthesis and descriptive analysis will provide the basis for summarizing and presenting the findings of the studies [14, 24, 25] included in the review, focusing on how "patient similarity" is conceptualized and operationalized within cancer

research. This process will summarize key findings, and thematic categories and establish links between approaches to 'cancer patient similarity' across studies. Additional graphical and tabular forms will be used to visualize and systematically present the collected data. For this purpose, we are planning to include flowcharts representing the study selection process, diagrams, and bar charts illustrating the intersection of different types of data sources or showing the frequency of studies of different cancer types.

Results

The review protocol, which outlines the methodology for the review, began with a database search, identifying 1183 unique papers after the removal of duplicates. As the review advanced to full-text screening by March 2024, the selection process led to 734 (62%) papers being excluded and 151 (13%) papers being earmarked for conflict resolution. Consequently, 235 (20%) papers were initially considered for inclusion, with the number rising to 258 (22%) after resolving conflicts. This indicates that approximately 25% of the selected papers significantly contribute to the analysis of the review and align with the research questions and objectives. Currently, a full-text analysis is underway using a pre-tested chart template to ensure that each selected study contributes to the comprehensive understanding the review aims to establish.

Discussion

Our planned scoping review will offer insights into how the concept of patient similarity in cancer patients has been defined and interpreted thus far. Additionally, the review aims to provide an overview of the methods used to identify similarities and differences among cancer patients. It will also specify the types of data utilized in these methods. Furthermore, it will provide an overview of the types of cancer addressed in the studies we cover. However, the scoping review of similarity measures for cancer patients may face limitations, including the possibility of missing specific study details due to its broad coverage, variability in study design, diversity of data sources, and possible publication bias. In addition, rapid advances in the field and subjectivity in study selection may affect the comprehensiveness and accuracy of the review. Despite these limitations, it is important to note that the advantages and benefits of conducting such studies far outweigh the possible disadvantages, offering valuable insights into personalized cancer treatment strategies. Firstly, it facilitates a more nuanced understanding of cancer's biological diversity, recognizing that while each case is unique, there are often underlying similarities that can guide treatment [5] Additionally, the benefits of studying patient similarities include also the potential for more effective and targeted therapies, improving prognostic models, and discovering new approaches. Finally, identifying indicators of similarity supports ongoing treatment by allowing one to act more efficiently and effectively, armed with knowledge drawn from a broader data set [11, 26]. Our review will examine the scope, range, and nature of cancer patient similarity studies. It will highlight key findings, identify research gaps, and explore new methods for assessing patient similarity. It will also suggest future directions for research. These directions may include patient-centered approaches by incorporating patient-reported outcomes and experiences into the definition of similarity measures. Thereby, we aim to ensure the relevance and applicability of findings in clinical practice. This is of special importance when the focus is on rare cancers, which are often underrepresented in studies. To understand if and how similarities can be identified and utilized in these cases, we will also consider longitudinal studies exploring patient similarities and its impact on treatment outcomes.

Acknowledgments

This research was funded by the German Federal Ministry for Education and Research through the PM4Onco project (01ZZ2322S, 01ZZ2322N, 01ZZ2322R, 01ZZ2322Q, 01772322B and 01ZZ2322O). This work was performed in fulfillment of the requirements for obtaining the degree

"Dr. rer. medic." at the Martin Luther-University Halle-Wittenberg.

Ethical Considerations

There was no requirement for ethical approval because only literature was being evaluated.

Conflicts of Interest

None declared.

Abbreviations

PRISMA-ScR: Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for Scoping Reviews

Multimedia Appendix 1

Proposed search queries for Web of Science, PubMed, LIVIVO, and MEDLINE.

References

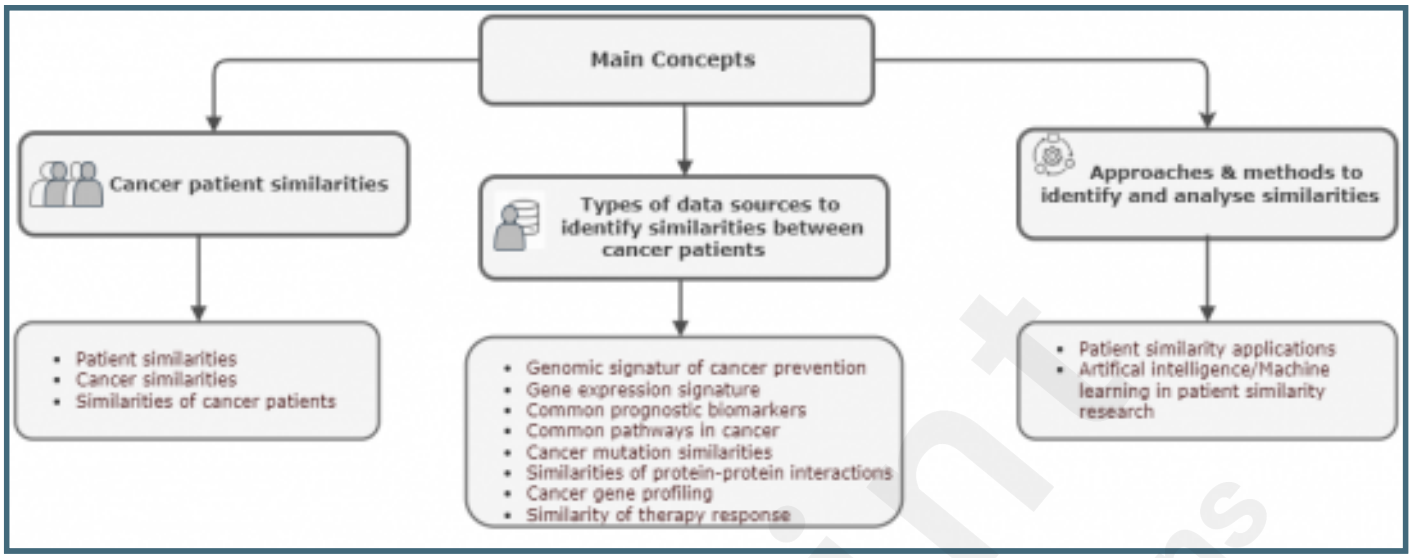
1. Dai, L., H. Zhu, and D. Liu, *Patient similarity: methods and applications.* 2020.
2. Brown, S.-A., et al., *Patient similarity and other artificial intelligence machine learning algorithms in clinical decision aid for shared decision-making in the Prevention of Cardiovascular Toxicity (PACT): a feasibility trial design.* Cardio-Oncology, 2023. **9**(1): p. 7.
3. Brown, S.A., *Patient Similarity: Emerging Concepts in Systems and Precision Medicine.* Front Physiol, 2016. **7**: p. 561.
4. Murciano-Goroff, Y.R., et al., *Precision Oncology: 2023 in Review.* Cancer Discovery, 2023. **13**(12): p. 2525-2531.
5. Parimbelli, E., et al., *Patient similarity for precision medicine: A systematic review.* J Biomed Inform, 2018. **83**: p. 87-96.
6. Li, H., et al., *A Patient Similarity Network (CHDmap) to Predict Outcomes After Congenital Heart Surgery: Development and Validation Study.* JMIR Medical Informatics, 2024. **12**: p. e49138.
7. Corti, C., et al., *Artificial Intelligence in Cancer Research and Precision Medicine: Applications, limitations and priorities to drive transformation in the delivery of equitable and unbiased care.* Cancer Treat Rev, 2023. **112**.
8. Mateo, J., et al., *Delivering precision oncology to patients with cancer.* Nature Medicine, 2022. **28**(4): p. 658-665.
9. Fang, H.S.A., et al., *Patient similarity analytics for explainable clinical risk prediction.* BMC Medical Informatics and Decision Making, 2021. **21**: p. 207.
10. Victoir, B., et al., *Targeted Therapeutic Strategies for the Treatment of Cancer.* Cancers, 2024. **16**(2): p. 461.
11. Xu, J., P. Yang, and S. Xue, *Translating cancer genomics into precision medicine with artificial intelligence: applications, challenges and future perspectives.* Hum Genet, 2019. **138**: p. 109-124.
12. Boniolo, G., P.P. Di Fiore, and S. Pece, *Patient Similarity in the Era of Precision Medicine: A Philosophical Analysis.* Erkenn, 2023. **88**: p. 2911-2932.
13. Wang, F. and J. Sun, *PSF: A Unified Patient Similarity Evaluation Framework Through Metric Learning With Weak Supervision.* IEEE J Biomed Health Inform, 2015. **19**: p. 1053-

1060.
14. Tricco, A.C., et al., *PRISMA Extension for Scoping Reviews (PRISMA-ScR): Checklist and Explanation.* Ann Intern Med, 2018. **169**(7): p. 467-473.
15. Elm, E.V., et al., *Methodische Anleitung für Scoping Reviews (JBI-Methodologie).* Z Evid Fortbild Qual Gesundhwes, 2019. **143**: p. 1-7.
16. Bramer, W.M., et al., *Optimal database combinations for literature searches in systematic reviews: a prospective exploratory study.* Systematic Reviews, 2017. **6**(1): p. 245.
17. Baumann, N., *How to use the medical subject headings (MeSH).* Int J Clin Pract, 2016. **70**(2): p. 171-174.
18. Richter, R.R. and T.M. Austin, *Using MeSH (medical subject headings) to enhance PubMed search strategies for evidence-based practice in physical therapy.* Phys Ther, 2012. **92**(1): p. 124-132.
19. Johnson, E., L. Wang, and M. Gomez, *Comparative Analysis of Search Retrieval Efficacy Across Five MEDLINE Platforms: A Longitudinal Study.* Journal of Medical Informatics Research, 2024. **31**(2): p. 156-174.
20. Doe, J. and J. Smith, *Experiences from using snowballing and database searches in systematic literature studies.* Journal of Systematic Literature Studies, 2023. **10**(2): p. 150-165.
21. Rønn, C., et al., *Circular Business Model for Digital Health Solutions: Protocol for a Scoping Review.* JMIR Res Protoc, 2023. **12**: p. e47874.
22. Levac, D., H. Colquhoun, and K.K. O'Brien, *A Systematic Approach to Conducting Review Studies: An Assessment of Content Analysis in 25 Years of IB Research.* Journal of International Business Studies, 2010. **41**(3): p. 925-940.
23. Schwenker, R., et al., *Identifying patients with psychosocial problems in general practice: a scoping review protocol.* BMJ Open, 2021. **11**: p. e051383.
24. Ouzzani, M., et al., *Rayyan—a web and mobile app for systematic reviews.* Syst Rev, 2016. **5**: p. 210.
25. Khalil, H., D. Ameen, and A. Zarnegar, *Tools to support the automation of systematic reviews: a scoping review.* J Clin Epidemiol, 2022. **144**: p. 22-42.
26. Kim, S., J.D. Herazo-Maya, and D.D. Kang, *Integrative phenotyping framework (iPF): integrative clustering of multiple omics data identifies novel lung disease subphenotypes.* BMC Genomics, 2015. **16**: p. 924.

Supplementary Files

Figures

Main concepts of planned Scoping Review and corresponding keywords.



Multimedia Appendixes

Proposed search queries for Web of Science, PubMed, LIVIVO, and MEDLINE.
URL: http://asset.jmir.pub/assets/3676d4ad48e68bb8f8eab747822fa726.docx