SYSTEMATIC REVIEW

Scoping review of knowledge graph applications in biomedical and healthcare sciences

[version 1; peer review: 2 approved with reservations]

Sanjay Budhdeo[1-4], Joe Zhang[5,6], Yusuf Abdulle[7], Paul M Agapow[8], Douglas GJ McKechnie[9], Matt Archer[10], Viraj Shah[11], Eugenia Forte[12], Ayush Noori[13,14], Marinka Zitnik[3,14-16], Hutan Ashrafian[17,18], Nikhil Sharma[1,19]

[1]Department of Clinical and Movement Neuroscience, University College London, London, WC1N 3BG, UK
[2]National Hospital for Neurology and Neurosurgery, London, WC1N 3BG, UK
[3]Broad Institute of MIT and Harvard, Cambridge, Massachusetts, 02142, USA
[4]The Alan Turing Institute, London, NW1 2DB, UK
[5]Institute for Global Health Innovation, Imperial College London, London, SW7 2AZ, UK
[6]Department of Critical Care, Guy's and St Thomas' NHS Foundation Trust, London, SE1 7EH, UK
[7]Institute for Health Informatics, University College London, London, NW1 2DA, UK
[8]GSK, Stevenage, SG1 2NFX, UK
[9]Department of Primary Care and Population Health, UCL Medical School (Royal Free Campus, University College London Research, London, NW3 2PF, UK
[10]Oxford University Hospitals NHS Foundation Trust, Oxford, England, OX3 9DU, UK
[11]Imperial College Business School, London, SW7 2AZ, UK
[12]GKT School of Medical Education, King's College London Faculty of Life Sciences & Medicine, London, England, SE1 1UL, UK
[13]Harvard College, Cambridge, Massachusetts, 02138, USA
[14]Department of Biomedical Informatics, Harvard University, Boston, Massachusetts, 02115, USA
[15]Harvard Data Science Initiative, Harvard University, Cambridge, Massachusetts, 02134, USA
[16]Kempner Institute for the Study of Natural and Artificial Intelligence, Harvard University, Cambridge, Massachusetts, 02134, USA
[17]Imperial College London Department of Surgery and Cancer, London, SW7 2AZ, UK
[18]Institute of Global Health Innovation, Imperial College London, London, SW7 2AZ, UK
[19]BioCorteX Ltd, London, W1W 5PF, UK

v1 **First published:** 12 Feb 2025, **10**:66
https://doi.org/10.12688/wellcomeopenres.23599.1
Latest published: 12 Feb 2025, **10**:66
https://doi.org/10.12688/wellcomeopenres.23599.1

Abstract

Introduction

There is increasing use of knowledge graphs within medicine and healthcare, but a comprehensive survey of their applications in biomedical and healthcare sciences is lacking. Our primary aim is to systematically describe knowledge graph use cases, data characteristics, and research attributes in the academic literature. Our secondary objective is to assess the extent of real-world validation of findings from knowledge graph analysis.

Open Peer Review

Approval Status ? ?

	1	2
version 1 12 Feb 2025	? view	? view

1. **Yuxing Lu**, Peking University, Beijing, China

2. **Bo-Wei Zhao**, Southwest University,, Chongqing, China

Any reports and responses or comments on the article can be found at the end of the article.

Methods

We conducted this review in accordance with the PRISMA extension for Scoping Reviews to characterize biomedical and healthcare uses of knowledge graphs. Using keyword-based searches, relevant publications and preprints were identified from MEDLINE, EMBASE, medRxiv, arXiv, and bioRxiv databases. A final set of 255 articles were included in the analysis.

Results

Although medical science insights and drug repurposing are the most common uses, there is a broad range of knowledge graph use cases. General graphs are more common than graphs specific to disease areas. Knowledge graphs are heterogenous in size with median node numbers 46 983 (IQR 6 415-460 948) and median edge numbers 906 737 (IQR 66 272-9 894 909). DrugBank is the most frequently used data source, cited in 46 manuscripts. Analysing node and edge classes within the graphs suggests delineation into two broad groups: biomedical and clinical. Querying is the most common analytic technique in the literature; however, more advanced machine learning techniques are often used.

Discussion

The variation in use case and disease area focus identifies areas of opportunity for knowledge graphs. There is diversity of graph construction and validation methods. Translation of knowledge graphs into clinical practice remains a challenge. Critically assessing the success of deploying insights derived from graphs will help determine the best practice in this area.

Plain language summary

Knowledge graphs (KGs) are advanced data maps that are useful in healthcare because they are able to encode information about relationships between data. In our study, we explore KGs in the context of healthcare and biomedical sciences. We analyzed 255 articles, discovering that the two most popular uses were for identifying new scientific insights and for identifying existing drugs which may be used to treat additional diseases. We also find KGs have been applied to explore some conditions, such as COVID-19, but broaden usage to other conditions. We identify was to improve the use of KGs. These include further study to understand the best way to create KGs, use of more sophisticated machine learning techniques for analysis, and validating insights through further experiments or clinical trials.

Keywords
knowledge graph, drug repurposing, machine learning, graph statistics

Corresponding author: Sanjay Budhdeo (s.budhdeo@ucl.ac.uk)

Author roles: Budhdeo S: Conceptualization, Data Curation, Formal Analysis, Funding Acquisition, Investigation, Methodology, Project Administration, Visualization, Writing – Original Draft Preparation, Writing – Review & Editing; **Zhang J**: Conceptualization, Data Curation, Funding Acquisition, Investigation, Methodology, Project Administration, Visualization, Writing – Review & Editing; **Abdulle Y**: Data Curation, Investigation, Writing – Review & Editing; **Agapow PM**: Methodology, Writing – Review & Editing; **McKechnie DG**: Data Curation, Investigation, Writing – Review & Editing; **Archer M**: Data Curation, Investigation, Writing – Review & Editing; **Shah V**: Data Curation, Investigation, Writing – Review & Editing; **Forte E**: Data Curation, Investigation, Writing – Review & Editing; **Noori A**: Methodology, Writing – Review & Editing; **Zitnik M**: Methodology, Writing – Review & Editing; **Ashrafian H**: Supervision, Writing – Review & Editing; **Sharma N**: Supervision, Writing – Review & Editing

Competing interests: SB owns equity in Owkin Inc. JZ was employed as a senior informatician by Arcturis Data HA is the Chief Scientific Officer of Preemptive Medicine and Health Security initiative at Flagship Pioneering UK. NS is the Chief Executive of BioCorteX Inc. AN and MZ acknowledge support from Amazon Faculty Research, Google Research Scholar Program, AstraZeneca Research, Roche Alliance with Distinguished Scientists, Pfizer Research, Sanofi iDEA-iTECH Award, Chan Zuckerberg Initiative

Grant information: This work was supported by Wellcome [102186; the Charlotte and Yule Bogue Research Fellowship in honour of Sir Charles Lovatt Evans and A.J. Clark from University College London, and from The Alan Turing Institute's Enrichment Scheme to SB] [203928; to JZ]. JZ acknowledges support from the National Institute for Health Research Biomedical Research Centre based at Imperial College NHS Trust and Imperial College London. DGJM was supported by NIHR as an In-Practice Fellow (NIHR301988). AN gratefully acknowledges the support of the SPARK Fellowship from the Center for Public Service and Engaged Scholarship at Harvard College. AN and MZ gratefully acknowledge the support of NIH R01-HD108794, US DoD FA8702-15-D-0001, awards from Harvard Data Science Initiative, Amazon Faculty Research, Google Research Scholar Program, AstraZeneca Research, Roche Alliance with Distinguished Scientists, Pfizer Research, Sanofi iDEA-iTECH Award, Chan Zuckerberg Initiative, and Kempner Institute for the Study of Natural and Artificial Intelligence at Harvard University. A.N. was supported, in part, by the Summer Institute in Biomedical Informatics at Harvard Medical School.
The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

How to cite this article: Budhdeo S, Zhang J, Abdulle Y *et al.* **Scoping review of knowledge graph applications in biomedical and healthcare sciences [version 1; peer review: 2 approved with reservations]** Wellcome Open Research 2025, **10**:66 https://doi.org/10.12688/wellcomeopenres.23599.1

First published: 12 Feb 2025, **10**:66 https://doi.org/10.12688/wellcomeopenres.23599.1

Introduction

Context and importance of knowledge graphs

Data representation plays a vital role in advancing medicine: efficient organisation of increasingly large datasets enables analysis to robustly generate novel insights[1,2]. An increasingly important representation method is the knowledge graph (KG). KGs consist of the entities, relationships, and facts in a given domain, captured as a graph of nodes (representing entities or concepts) and edges (indicating the relationships or associations between them), often enriched with attributes, classifications, and semantic meanings[3].

Data representation through KGs has a historical foundation dating back multiple decades. Key accelerants to adoption were the development of Resource Description Framework (RDF) and Web Ontology Language (OWL) standards for the semantic web and Google's implementation of knowledge graphs in its search algorithms[3,4]. KGs offered distinct advantages over traditional relational databases, including greater schema flexibility and the capacity to capture nuanced capture of edge characteristics and relationships, making them valuable for knowledge retrieval analytical purposes such as recommender systems[3]. Early uses of KGs within healthcare included theoretical conceptualizations of clinical reasoning[5] and in research literature representation[6]. More recent applications in biomedicine and healthcare have included identifying drug repurposing candidates and generating novel biomedical hypotheses by established and early-stage pharmaceutical companies such as AstraZeneca[7] and Benevolent AI[8].

Motivation for review

Given the increasing use of KGs in biomedical and healthcare sciences, this review presents a comprehensive survey of research literature to detail the use of KGs within this domain. Many other reviews discuss the history of KGs and the methodologies used in their construction and analysis. There are also reviews that categorize uses for knowledge graphs[3]. This prior work includes systematic overviews of specific methods or analysis archetypes for KGs, for example, KG reasoning[9], KG completion[10], and relational machine learning[11]. There have also been reviews on data quality and methodology in KG construction across uses, including KG completeness[12] and generative KG construction[13].

There are commentaries and reviews for specific use cases within biomedicine, such as drug repurposing and adverse drug reactions[14], and focussed reviews, for example, on bioinformatic graphs and their analyses without a systematic survey of use cases[15]. In addition, some reviews have surveyed use of KGs in specific disease areas[16]. There are also commentary articles that outline potential broader use cases for biomedical KGs[17].

Despite this work, no systematic survey of use cases for knowledge graphs within the biomedical and healthcare sciences has been conducted. Our primary aim is to address this gap by providing a landscaping of disease areas in which knowledge graphs are used, and how they are used. We survey characteristics of these manuscripts, including author affiliations and funding. We also review graph characteristics: including node class and edge class numbers, node and edge numbers, node class domains and whether graphs have been made openly available. We review techniques used in KG analysis.

Our secondary research question examines the extent to which findings or insights from knowledge graphs been tested and validated in the real-world. There has been no systematic survey KG validation in the biomedical literature. KG validation includes the process of ensuring the accuracy of the data within a KG, as well as verifying insights derived from the KG. Validation is often divided into two categories: internal and external; however, these terms do not have standardized definitions. Consequently, in this article, we use the terms 'inside graph validation' and 'outside graph validation'. Inside graph validation assesses robustness of the graph based on the graph data itself. This includes hold-one-out studies or cross-validation, calculation and assessment of performance scores, or analysis with multiple algorithms on the same graph and comparing results. Outside graph validation refers to testing of insights in a different dataset, for example through in vitro testing or a clinical trial.

Aims of the review

The primary aim of this review is to establish a systematic description of the use cases, data characteristics, and research characteristics in current KG implementations in the academic literature. The secondary aim is to determine the extent to which findings or insights from KGs have been tested and validated in the real world.

Methods

Protocol

We were guided by the PRISMA extension for Scoping Reviews checklist[18]. A protocol was written *a priori* and posted on the website OSF on 5 November 2021 (https://osf.io/etg6f). We limited our search to only academic literature (i.e., publications and preprints), and did not survey other sources of insights such as patent databases. This topic is suitable for a scoping review as it represents an exploratory mapping of the literature, where it was not possible to delineate how best to categorise manuscripts *a priori*.

Literature search

Relevant studies were be identified from an academic search on databases for biomedical literature. We used MEDLINE and EMBASE (via Ovid), as well as medRxiv, arXiv (section: quantitative biology) and bioRxiv, employing free-text searches for keywords. Our search strategy was as follows: "knowledge graph" OR "graph neural network" OR "graph convolutional network" This was an adaptation of the search methodology outlined in a similar scoping review by Chatterjee *et al*[16]. The search strategy was agreed upon by the authors (SB, JZ, HA, NS).

Inclusion and exclusion criteria

Publications and preprints available until 21 November 2021 written in English were included. Manuscripts were required to (i) describe the creation of a KG or the use a KG to generate

insights; (ii) use multimodal graphs (i.e. graphs with more than one discrete node class); and (iii) directly mention medical insights and/or health outcomes in humans.

We excluded (i) neuroimaging graph theory and graph topology manuscripts, unless KGs were explicitly mentioned within the manuscript; (ii) papers analysing animal models; (iii) articles about traffic road safety where health outcomes were not explicitly mentioned; (iv) manuscripts using graph neural network or graph theory which did not make use of a KG; (v) reviews and commentary articles, opinion pieces, editorials, and any other articles that did not report original research; (vi) single modality graphs with only one node class (as defined by the authors); and (vii) papers not in English language. Review contents were used to guide additional literature discovery.

Data collection

Duplicate manuscripts were removed using Endnote and Rayyan.ai. Two reviewers (JZ, SB) carried out a title and abstract screening. Papers passing the abstract screen underwent a full-text screening by two reviewers (SB and one of JZ, YA, VS, DM, EF, and MA). Data extraction was carried out by two reviewers (SB and one of JZ, YA, VS, DM, EF, and MA).

For manuscripts passing full-text screen, we collected the following data:

(i) Demographics: (1) domain (a preliminary categorisation was refined iteratively by SB and JZ, with final categorization consisted of: medical science insights, drug repurposing, literature representation, drug interactions and toxicity, diagnostics, drug discovery, electronic health record (EHR) data representation, public health, non-EHR patient data representation, risk prediction, and drug related uses outside of discovery, repurposing, interactions and toxicity) (2) affiliation of authors (institution and country) (3) disease-area (e.g. refined iteratively categorized by SB and JZ, final categorisation included: general/non-disease specific, infectious diseases (Covid-19, cancer, neurology, mental health, diabetes/endocrine, rare disease, respiratory disease)

(ii) Graph characteristics and data sources: (1) public availability of the KG (2) node and edge types (e.g. genes, drugs, proteins, etc.) (3) named data sources (e.g. The Cancer Genome Atlas, MEDLINE scraping, private EHR data, etc.) (4) size of the graph (the number of nodes, edges, node classes and edge classes, if included in the manuscript)

(iii) Analysis: (1) analysis methodology (2) Validation: whether there was inside graph or outside graph (downstream) validation of insights (3) Future plans/next steps in the analysis

Regarding validation: inside graph validation refers to efforts to validate findings by dividing data into training, testing and validation datasets, by holding out or adding in datasets to check robustness of conclusions, using cross-validation techniques when running models, or using multiple analysis methods and comparison of performance scores. Outside graph validation refers to additional analyses to test findings generated from the original graph, which may include testing hypotheses using in vitro studies, animal experiments, human observational data, or clinical trials.

Title, abstract and full text screens

Discrepancies in the title an abstract screen were reviewed by JZ and SB after unblinding, to reach a consensus decision. For discrepancies in the full text screen, a third reviewer (JZ or YA) reviewed the manuscript.

Categorisation in data collection

Preliminary categorisations were defined *a priori* and refined during data collection. Categorization of use case, disease area, analysis type and future plans were iterative and based on periodic review and agreement between SB and JZ.

Categorisations of use case

The following definitions were used as a guide for final use case categorisations:

- medical science insights: Predictions or hypotheses regarding scientific knowledge. Examples might include protein-protein interactions, genes implicated in diseases, or a clustering of symptoms
- drug repurposing: identification of novel indications for existing drugs
- literature representation: mapping and representation of research literature. This category includes bibliometrics. This excludes papers which are better categorised under diagnostics, drug repurposing, or other drug related categories
- drug interactions and toxicity: prediction of drug side effects or adverse events
- diagnostics: use of knowledge graphs to predict patient diagnosis. This may be based on electronic health records or other data sources, such as patient data entry on a virtual platform
- drug discovery: identification of novel drugs which may have future uses as treatments
- EHR representation: Mapping and representation of clinical and/or operational data held in electronic health records. EHR representation specifically for diagnostics is included in the diagnostics category
- public health: Representation of concepts for the purpose of or regarding population health management. This includes social determinants of health, environmental factors impacting health, and infectious disease outbreaks
- non-EHR patient data: Mapping and representation of patient data where the source is not from electronic health records, and the purpose is not related to other categories, including drug related categories (drug

repurposing, drug discovery, drug interactions and toxicity, drug related - other) or diagnostics category. This includes patient data entry in cellular/mobile or web portals and use of claims data.

- risk prediction: Mapping and representations of data in order to predict or otherwise better appreciate risk of acquiring a condition, or of disease prognosis.
- drug related - other: representation of drug data for a purpose other than drug repurposing, drug discovery or drug interactions and toxicity. This includes representation of drug-target data without downstream use predictions regarding repurposing, interactions, etc.

Categorisation of analysis methodology

The following definitions were used as a guide for determining whether which category of knowledge graph analysis was used in each manuscript. Where possible only one category was chosen – for example, if a novel analysis methodology was compared to pre-existing methods to demonstrate superiority, only the novel method was categorised.

Querying: The process of extracting data from a graph using a specific query language or API. The graph might be queried to extract a subgraph. Examples may include finding all the nodes that meet certain criteria or retrieving specific information about a node.

Graph statistics: Analysing properties of a graph, including measures including but not limited to centrality, degree distribution, and clustering coefficient.

Graph convolutional networks (GCN): A type of neural network that can be used to learn from graphs by aggregating information from a node's neighbors through message passing.

Graph attentional networks: A neural network that can be used to learn from graphs by selectively attending to important nodes or edges using attention mechanisms. This allows graph attentional networks to focus on the most relevant information for a given task.

Deep learning (non-GCN, non-GAN): Deep learning is machine learning using neural networks with multiple layers. This includes methods such as graph autoencoders, Graph Recurrent Neural Networks, Graph Generative Adversarial Networks, but excludes GCNs or GANs which have a separate model label.

Embedding methods without use of GCN, GAN or deep learning architectures:

Node embedding: The process of representing a node as a vector in a low-dimensional space (i.e. simplifying complex data about the node while preserving important information), which captures the node's characteristics and context within the graph. This can be done using a variety of methods, e.g. node2vec, DeepWalk, LINE, SDNE

Graph embedding: The process of representing a graph as a fixed-length vector (i.e. string of numbers) or matrix (i.e. 2D table of numbers) that captures its structural properties and relationships between nodes. e.g. RotatE, TransE

Supervised classification methods (non-embedding): A type of machine learning that can be used to classify graphs by using labeled ådata to train a model. This includes methods such as support vector machines (SVMs), decision trees, random forests, logistic regression, K-Nearest Neighbours, or naive Bayes. Supervised learning uses labelled data, with desired output types already known.

Unsupervised graph clustering: machine learning to group or cluster nodes or graphs based on their structural similarities, without the data having labels. This includes methods such as spectral clustering, hierarchical clustering, and k-means clustering.

Analysis

Analysis was carried out by SB and JZ. Graph demographics were analyzed in Google Sheets. Harmonisation of dataset terms took place in a semi-automated fashion with Python (https://www.python.org/), Pandas (https://pandas.pydata.org/), fuzzywuzzy[19] and manual review. Harmonisation of node terms for to create gold standard concepts for grouping took place in a semi-automated fashion using Python (https://www.python.org/), BERT PubMed embedding (https://huggingface.co/microsoft/BiomedNLP-BiomedBERT-base-uncased-abstract-fulltext) and k-nearest neighbors classification with manual review of groupings. Meta-graph visualisation of co-occurring node terms was carried out with Gephi (https://gephi.org/).

Article filtering

Articles were filtered through abstract and full-text screen, as shown by the flow chart in Figure 1.

Results

This section first presents demographic features of the surveyed manuscripts, including a description of manuscript numbers over time, use case, and disease area categorisation, author affiliations and funding sources. The graph characteristics session provides descriptive statistics between node and edge numbers, and node and edge class numbers in graphs used by the articles. A meta-graph of node classes is used to provide further information on node class characteristics from included graphs. A summary of data sources used in graph construction is presented. Finally, graph analysis techniques, validation methods and stated plans for future work are summarised.

Manuscript demographics

Figure 2 demonstrates the trend in the number of manuscripts that were preprinted/published each year. There is an increasing

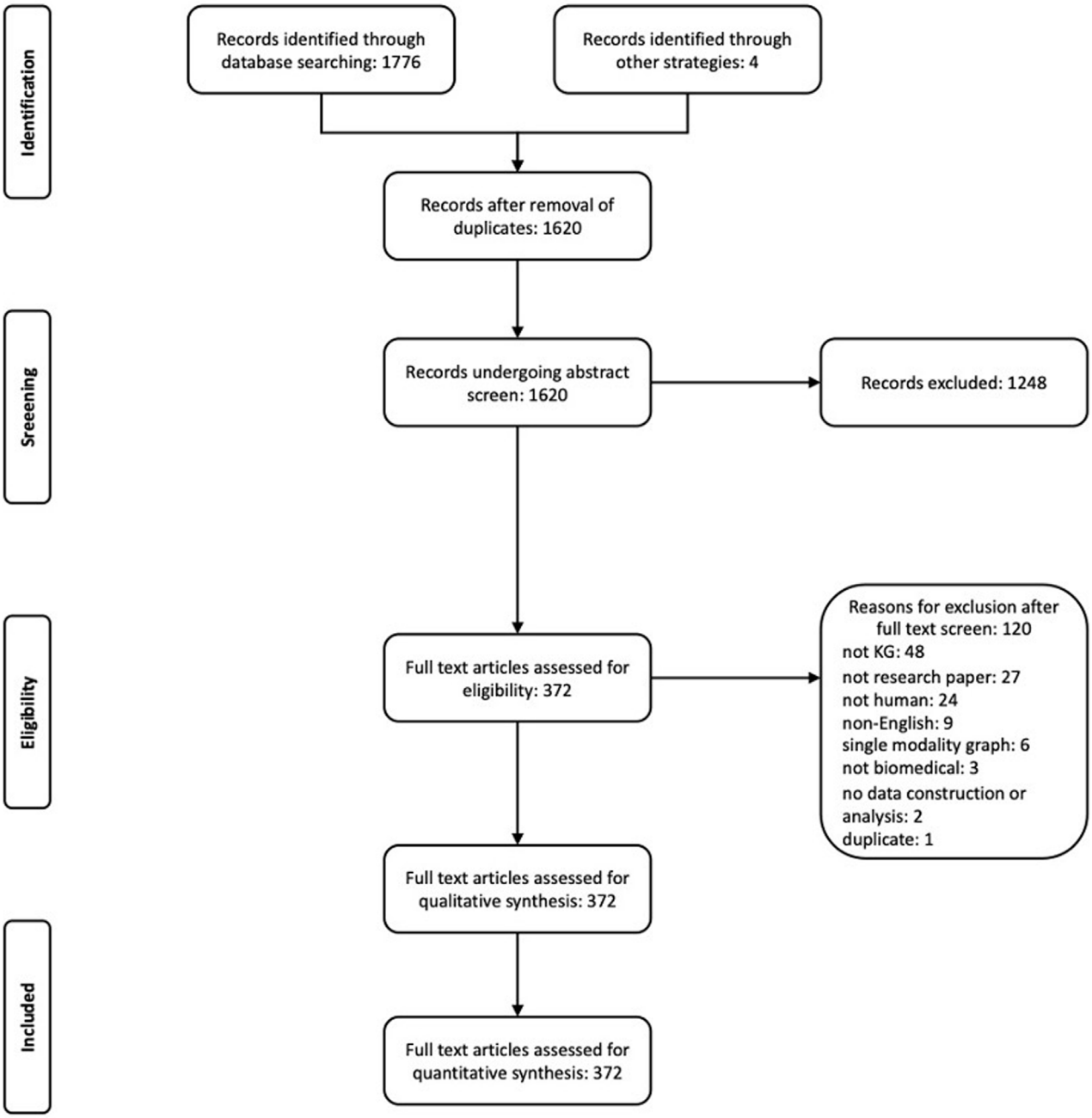

Figure 1. Flow chart of paper filtering based on PRISMA guidelines.

number of publications in this area; furthermore, the trend has accelerated in recent years.

Figure 3 represents the categories of biomedical and healthcare knowledge graphs identified, grouped by use case. The two most common use cases are medical science insights unrelated to drugs or diagnostics (74 studies) and drug repurposing (57 studies), which are both twice as frequently seen as the third most common use case of literature representation (25). Papers categorised in the use case 'medical sciences insights' develop predictions or hypotheses regarding scientific knowledge. Examples might include protein-protein interactions, genes implicated in diseases, or a clustering of symptoms. Papers categorised as 'literature representation' include mapping and representation of research literature. This category includes bibliometrics and excludes papers which would be better categorised under other use cases such as diagnostics, drug repurposing, or other drug-related use cases. There is a broad

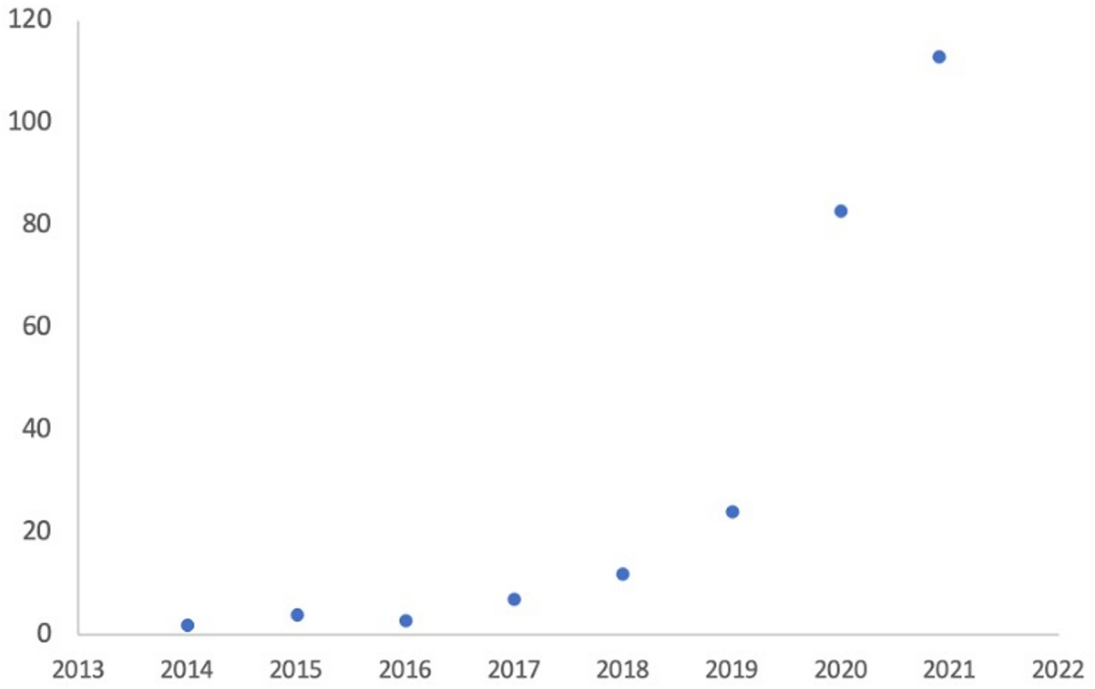

Figure 2. Number of preprints and publications that passed the full text screen, categorized by year. There is an increase in publication number over time.

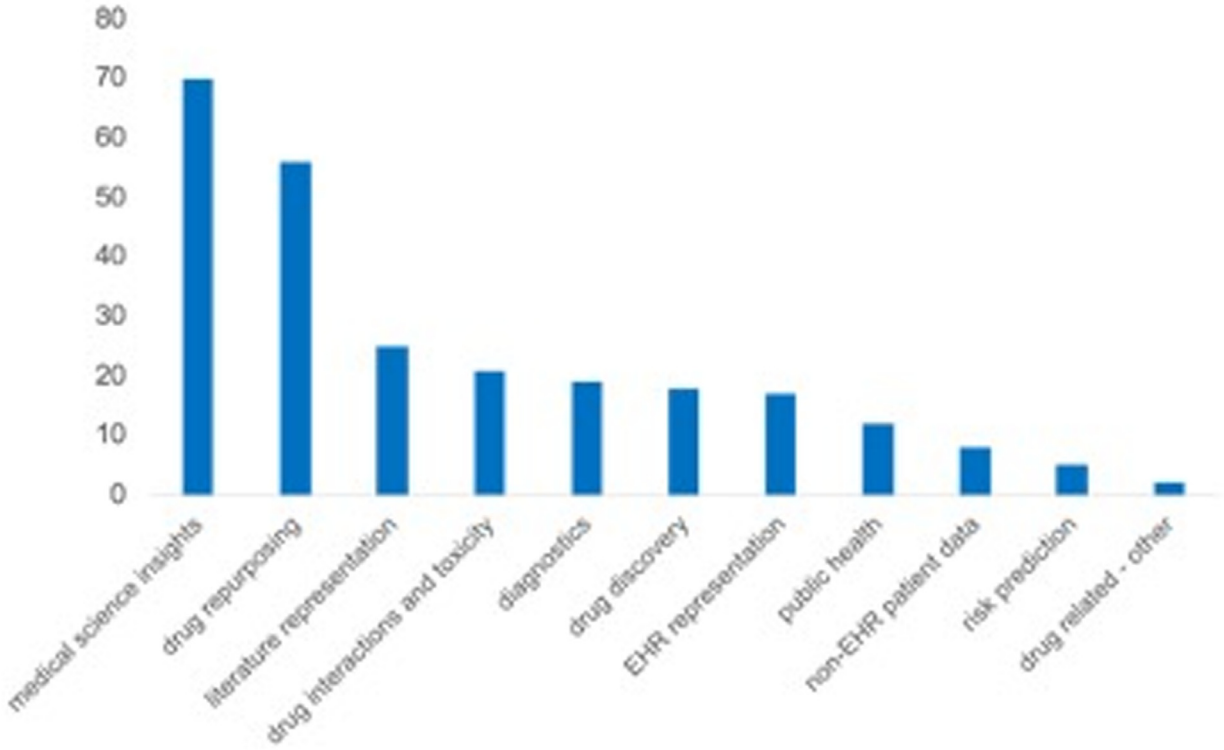

Figure 3. Use case categorisation of biomedical and healthcare science knowledge graphs. The two most common categories are non-drug, non-diagnostics medical science insights (74 studies) and drug repurposing (57 studies). There are a broad range of use cases encompasses biomedical, patient representation and population health management.

array of use cases which include biomedical (medical science insights, drug repurposing, literature representations, drug interactions and toxicity, drug discovery, and drug related-other), patient related (EHR representation, non-EHR patient data) and population health management (diagnostics, public health, risk prediction). Descriptions of the use case categories are summarized in the Methods section.

Figure 4 demonstrates the therapeutic area categorisation of identified manuscripts. General KGs are the most common category (115 manuscripts) and exceed KGs representing any individual therapeutic area. The most common therapeutic area was infectious disease (47 manuscripts); however, this was driven mainly by COVID-19 (41 manuscripts); KG for other infectious diseases (6 manuscripts) were less common. Outside of infectious diseases, manuscripts using KGs for oncology (26 manuscripts) and neurology (13 manuscripts) were also common. Graphs were often constructed to represent a single disease rather than a whole disease area.

Table 1 demonstrates an aggregated count of country of primary affiliation for first and last authors of manuscripts. The United States and China are the top two contributors, with 162 and 155 manuscripts, respectively, followed by other European and North American countries. After USA and China, the next highest author count is from the UK (33 authors).

Table 2 demonstrates the breakdown of funding declarations within the manuscripts passing review screening. Each manuscript can have more than one archetype of funding source. Most manuscripts in this area receive funding for research from government or government funded bodies. More manuscripts receive commercial funding than non-governmental non-profit funding in this area. Most papers don't have commercial affiliations. Table 3 outlines the breakdown of companies funding KG research. There is not one single dominant company in this area. Multiple different company types have funded research in this area, including big technology companies, biopharmaceutical companies, consultancies and start-ups, with no single sector dominant.

Graph characteristics

38.9% of manuscripts featured graphs that were open sourced. Our definition of open source is that the graph was downloadable or otherwise freely and openly available.

Table 4 provides summary statistics for unique node and edge classes within each graph. More manuscripts reported node classes than edge classes. In order to standardize the treatment of data across manuscripts reporting edge classes in a heterogenous manner, all edges between two node classes were counted as a single edge class. The skew and kurtosis of node and edge class counts suggest non-normally distributed data for both, so the median interquartile range (IQR) and range are presented. There is a wide range of node class numbers (2–41), but the median and IQR suggest this is due to outlying values. The mean class number is 6.0 and median is 4. The mean edge class number is 8.0, and the median is 4. The IQR for edge class numbers is 3 to 7. The range spans from a minimum of 1 to a maximum of 210, influenced by outlying values.

Table 5 summarises node and edge numbers reported by manuscripts. Fewer manuscripts report node and edge numbers compared to node class and edge class numbers. For node and edge numbers, the skewness and kurtosis of the datasets together suggest non-normal distribution, so median IQR and range are presented. The node number dataset includes 123 reports, and the edge numbers include 115 reports. The mean

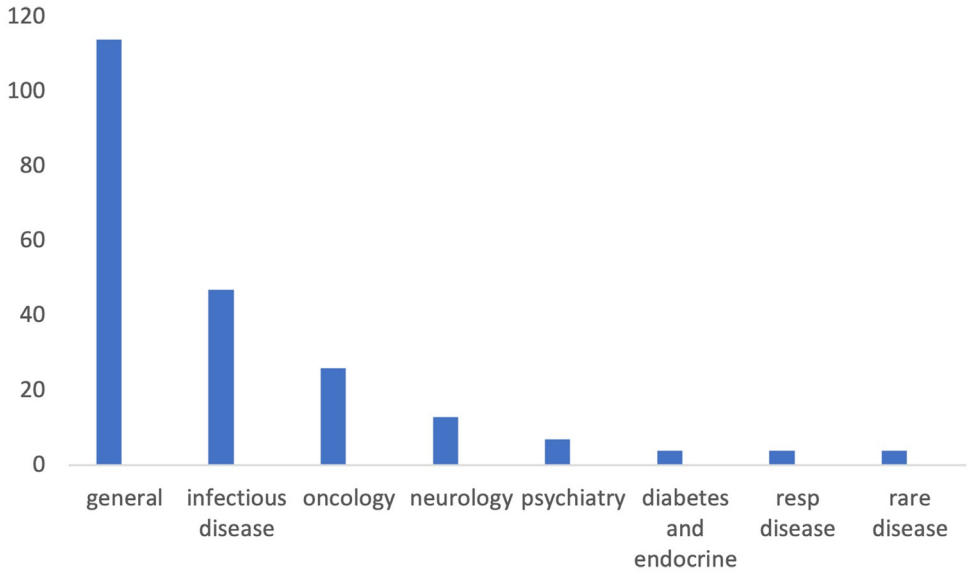

Figure 4. Disease area categorisation of manuscripts using biomedical and healthcare knowledge graphs. The most common category was general (115), followed by Covid-19 (41), oncology (26) and neurology (13). Resp disease stands for respiratory disease.

Table 1. Country breakdowns consisting of a sum of counts of first and last author country affiliation. USA and China are the countries which most frequently have contributing authors.

Country	Count
USA	162
China	155
UK	33
Germany	24
Canada	20
Netherlands	15
Italy	12
Rest of Europe	33

Table 2. Breakdown of manuscripts by funding mechanism for manuscripts in the area of knowledge graphs in biomedicine and healthcare. Results show that most manuscripts receive funding from government or government-funded bodies.

Funding mechanism	count
not stated	34
government funding	166
NGO funding	29
commercial funding	40

Table 3. Frequency of commercial affiliations. All companies reported as funding more than one manuscript are reported. This table provides an overview of companies funding knowledge graph research, demonstrating the diverse range of contributors, including big technology firms, biopharmaceutical companies, consultancies, and start-ups, with no single sector exerting dominance.

Company	Count
Amazon (including AWS)	5
AstraZeneca	5
IBM	5
Google	4
Benevolent AI	3
CoVar Applied Technologies	3
Elsevier	3
Enveda Biosciences	3
Bayer	2
Biogen	2
Causality Biomodels	2
Data2Discovery	2
Euretos	2
IQVIA	2
nference	2
QIAGEN	2
Yidu Cloud	2

Table 4. Node class and edge class numbers from knowledge graphs used by identified manuscripts. This shows a wide range of values in both datasets. IQR is interquartile range.

Value	Node class numbers	Edge class numbers
Number of values	241	207
Skewness	3.3	9.3
kurtosis	15.0	109.5
Mean	6.0	8.0
Median	4	4
Lower IQR	3	3
Upper IQR	7	7
Lower bound	2	1
Upper bound	41	210

Table 5. Node numbers and edge numbers from knowledge graphs used by identified manuscripts. Both datasets display non-normal distribution with wide range driven by upper bound outlying values.

Value	Node numbers	Edge numbers
Number of values	123	115
Skewness	9.6	10.6
kurtosis	99.5	112.8
mean	3 016 830	152 556 781
median	46 983	906 737
lower IQR	6 415	66 272
upper IQR	460 948	9 894 909
lower bound	212	238
upper bound	180 200 000	14 000 000 000

number of nodes across in KGs used by manuscripts is approximately 3 016 830, while the median is 46 983. The node number IQR is 454 533 but the range is higher at 180 199 788, suggesting outlier values. The mean number of edges across the KGs is 152 556 781, higher than the median which is 906 737, demonstrating skewed distribution. The edge number upper IQR is 9 828 637 but the range is higher at 13 999 999 762, due to outlier values. The ratio of median edge number to median node number was approximately 19, which gives some indication of the connectivity within graphs.

Figure 5 summarizes information about common node classes and their relations into a meta-graph. The nodes are aggregated node class categories extracted from the manuscripts, and edges represent the node class-node class relationships that are extracted from the manuscript. The methods for generation node class groupings and graph visualisation are further detailed in the Methods section. They demonstrate that disease-gene and disease-drug edges are most commonly seen. On visual inspection, the graph separates into two groups centred upon disease concepts: biomedical nodes (here used to mean concepts more related to molecular biology, 'wet lab' science or bioinformatics) and clinical nodes (here used to mean concepts predominantly patient related or more commonly encountered in the patient facing context).

Data sources

Table 6 provides a list of all datasets where five or more manuscripts used the source. DrugBank emerges as a frequently used data source, with utilization in 46 manuscripts, underscoring its significance in drug-related KG research. PubMed and UniProt also exhibit high usage, appearing in 26 KGs each, reflecting the importance of literature and protein data.

Graph analysis

Table 7 summarizes counts of analysis technique archetypes used in KG analysis in the identified manuscripts. There are many different methodologies being used in KG analysis, of varying sophistication. We observed a diverse range of approaches in KG research. The most prevalent method is graph querying,

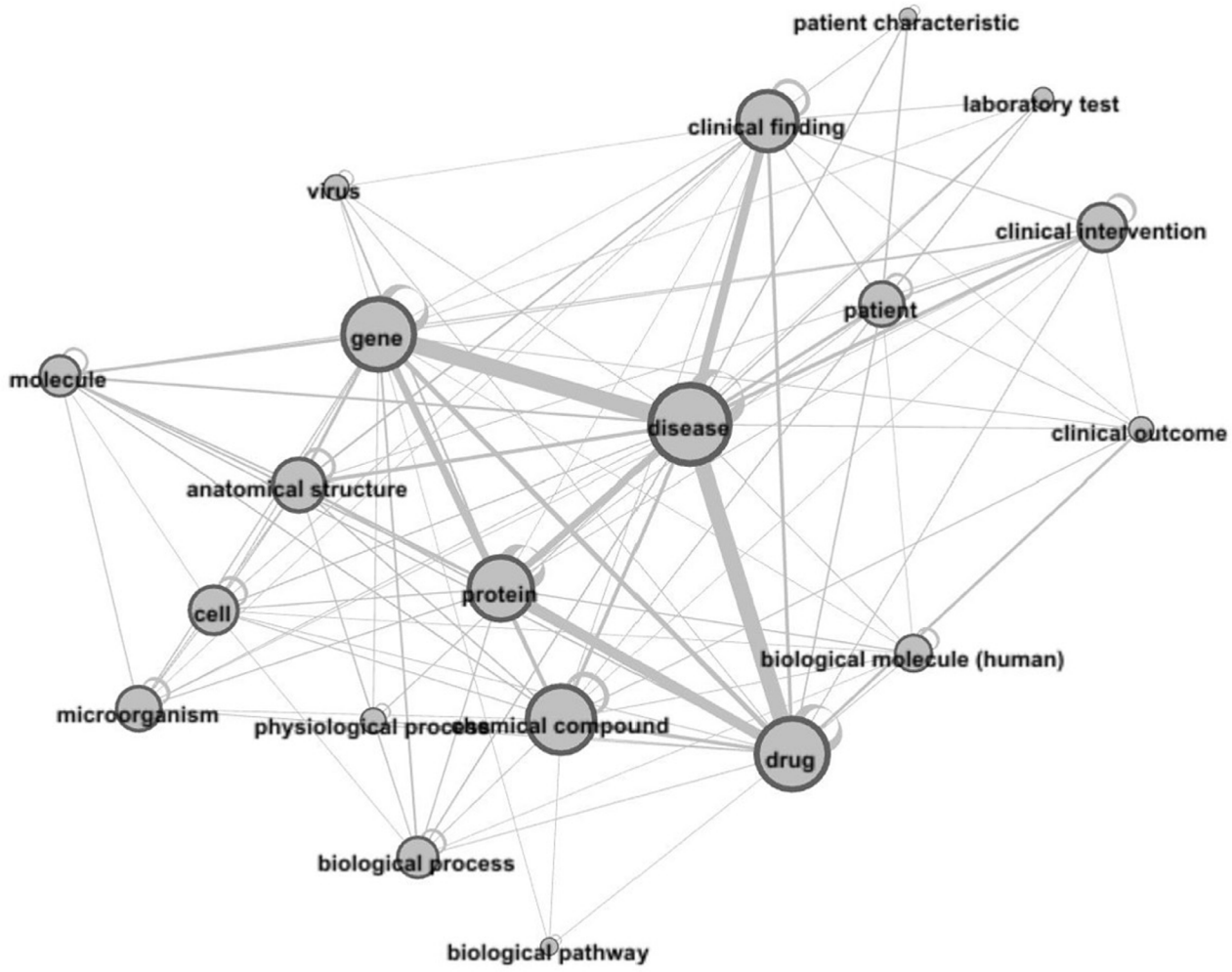

Figure 5. Meta-graph of node classes represented in KG literature. Semi-automated categorisation based on PubMed BERT model and author consensus (see Methods for further details). Edge weights are displayed and correspond to thickness of the edge.

Table 6. List of all source datasets where five or more papers/preprints used the source. Add datasets with counts of 5 or above are listed. DrugBank is the top source, cited almost twice as often as the next most popular source.

Source	Count
DrugBank	46
PubMed	26
UniProt	26
String	21
Sider	19
Omim	18
Reactome	18
KEGG	18
DisGeNet	16
Gene Ontology	15
CheMBL	15
Human Phenotype Ontology	14
Biogrid	12
Comparative toxicogenomics database	12
UMLS	12
PubChem	11
Intact	11
DrugCentral	10
Mondo	10
SemmedDB	9
GTEx	8
GEO	8
Stitch	7
MESH	7
Ensembl	7
Hetionet	7
ClinVar	7
Orphanet	7
PharmgKB	7
Disease Ontology	6
TCGA	6
WikiPathways	6
Interpro	6
MINT	5
Doid	5
GWAS catalog	5
TwoSIDES	5
ChEBI	5
OpenTargets	5
OffSIDES	5
Wikidata	5
Cord-19	5

Table 7. Analysis method breakdown. This highlights diverse techniques in knowledge graph analysis, with graph querying (52 instances), graph embedding (43), and Graph Convolutional Networks (GCNs) (33) being the most frequently used methods.

Analysis method breakdown	Count
querying	52
graph statistics	24
node embedding	20
graph embedding	43
supervised classification methods (non-embedding)	17
unsupervised graph clustering	13
GCN	33
graph attention networks	14
deep learning (other)	17
other	14

used in 52 instances, followed by graph embedding (43) and Graph Convolutional Networks (GCNs) (33).

Table 8 presents the breakdown of validation methods employed in the identified manuscripts. Most papers (120) focus on inside graph validation techniques, such as data splitting, cross-validation, and algorithm diversity to ensure robustness. Less than a third of papers (39) validate findings outside of the graph. Such validation techniques may include *in vitro* testing, or clinical trials. 37 manuscripts combine within and outside graph validation approaches.

Table 9 provides a summary of planned or suggested next steps that authors have stated in Discussion sections of their manuscript The most frequently discussed future directions are data and algorithms, with 99 and 59 counts. Categories aligned with validation (clinical trials, other validation) feature less frequently at 3 and 11 counts.

Table 8. Manuscripts categorized by validation method. Most manuscripts have some form of validation. This is most commonly inside graph validation. Fewer manuscripts validate findings outside of the KG.

Validation method	Count
none	54
inside graph	120
outside graph	39
inside graph, outside graph	37

Table 9. Summary of planned or suggested next steps that authors have stated in Discussion sections of their manuscripts. Suggestions are more likely to focus on proximate improvements in data or algorithms, rather than translatability or applications.

Future plan category	Count
improve data	99
improve algorithm	59
extend use case	42
other validation	21
improve user interface	18
clinical application	11
clinical trials	3
none	25

Table 10 provides a summary of Tables and Figures as related to aims of the scoping review, as recommended by PRISMA-ScR guidelines[18].

Discussion

Findings and implications

KGs are becoming increasingly used in biomedical and healthcare sciences. The most prevalent applications of KGs are in medical science insights and drug repurposing – this is reflected in the common use of drug and protein datasets in KG construction. Research activity is concentrated in North America, China, and Europe. Location of research, together with location of data sources, contributes to bias in dataset curation for openly available datasets[20]. Despite significant commercial interest, the majority of academically published research in this domain continues to be government funded.

The variation in KG utilization suggests that there are potential opportunities in use cases and disease/therapeutic areas where knowledge graphs have seen limited use. KGs tend to either focus on specific diseases or remain highly general, with fewer encompassing entire therapeutic areas or multiple domains. A challenge arises from enrichment of KGs with data for a specific disease; this approach may improve predictions when analysing a disease-relevant question by adding relevant contextual data, but may introduce bias into the graph[21,22].

KGs exhibit substantial heterogeneity in terms of size, as reflected by wide variation in node and edge counts, as well as counts of node and edge classes. This may in part reflect the diverse scope and use cases for knowledge graphs or represent limitations in available storage and compute for analysis but may also represent a lack of known best practice regarding optimal KG size and connectivity.

This study reveals a wide array of data sources used in KG construction, with a preference for open datasets over closed ones. Despite the diversity, there is a clear concentration with some datasets, for example DrugBank, used frequently. Even the ostensible diversity hides a network of dependencies where certain datasets rely on others for primary data sources. For example, Open Targets uses many sources, including UniProt, STRING, Reactome, and ChEMBL[23]. More generally, there are biases stemming from input data. For example,

Table 10. Summary which guides review of Figures and Tables relevant to each review aim.

Aim of scoping review	Relevant Tables and Figures
use-cases	Figure 3, Figure 4
data characteristics	Table 4, Table 5, Table 6; Figure 5
research characteristics	Table 1, Table 2, Table 3, Table 7, Table 9; Figure 2
validation	Table 8

there is a correlation between information available for bio-entities and NIH funding for research into those entities[24].

Figure 5 suggests that on visual inspection, there is a grouping of biomedical node classes (here taken to mean concepts more often found in molecular biology or other 'wet lab' science or bioinformatics) and a grouping of clinical node classes (here used to indicate concepts predominantly patient-related or more commonly encountered in the patient facing context).

This grouping of biomedical and clinical node classes may be due to different data pipelines that would be used to generate node classes in each cluster and the relative ease of capturing some concepts and edges in one type of graph versus the other.

Biomedical node classes often use curated datasets where data has been methodically organized, corrected, annotated, and standardized through mapping and harmonizing to various ontologies and vocabularies[25]. This data is more likely to be available in the public domain through publications, hence the denser connectivity in this cluster.

In contrast, the construction of graphs using clinical node classes is more challenging[26]. First, EHR and public data backends are more individualised. For example, different hospitals may use different coding ontologies, electronic health records, and file storage types and methods from those systems. This lack of common standards makes transfer between health applications challenging. Second, healthcare data tends to be noisy, with large amounts of missingness and censoring, including loss of follow-up, incomplete data recording from patients who present to a care setting, and challenges in capturing phenomic data such as lifestyle and diet. This results in systematic biases in data. Third, there is a data structuring challenge. EHRs contain large amounts of unstructured text, which requires processing and may need either additional, highly expensive manual curation or improved automated curation, for example, through large language models[27,28]. Non-text data such as radiologic and pathologic images would require transformation for ingestion to preserve underlying, complex features and latent variables that might be lost in simple feature extraction. Fourth, regulation may require higher privacy safeguards to ensure that released data does not contain patient details. In the USA, this is governed through the Health Insurance Portability and Accountability Act (HIPAA).

Aside from comparisons between graphs predominantly using biomedical or healthcare node classes, the breadth of use cases depends on graph design, which involves dataset inclusion, schema flexibility, and expansiveness of relevant vocabularies and ontologies. Graphs may also have broad coverage of diseases or be a narrower representation of a single disease or disease group. Though it is likely from current evidence that graph performance will scale in size, performance has also been demonstrated to improve with context[29]. This means that it is likely no single best all-purpose graph. Instead, there is a judgement based on the graph use-case.

The prevalence of graph querying in KG research is surprising given the simplicity of the approach. However, this may be appropriate to the research objectives of those manuscripts if this is for data exploration or transparency in insight generation. Graph machine learning (ML) can perform a wider variety of tasks, for example, link prediction, node classification, or community detection. Graph ML can help mitigate data biases, for example, through graph rewiring or regularization techniques.

Few manuscripts include outside graph validation (Table 8) or explicitly suggest outside graph validation as further work (Table 9). This is concerning but understandable. There may be constraints on carrying out *in vitro* or *in vivo* experimental work or trials due to budgetary restrictions, regulatory constraints, lack of domain expertise, or challenges coordinating cross-disciplinary work. Outside graph validation would help translate insights from KGs into real-world applications.

Future work

There are several areas for further research. First, best practice in KG construction, especially concerning graph size, remains unresolved. Future reviews should further analyze specific use cases of KGs to interrogate tailored outcome measures for those graphs. Identifying factors associated with successful outcomes could inform best graph construction and analysis practices. In addition, it would be helpful to compare the effectiveness of KGs against alternative techniques for data source integration. Second, investigating ways to enhance the integration of -omics data and patient data within KGs could advance personalized medicine and disease understanding. Third, as the paper selection for this review concluded before the rising prevalence of generative AI and large language models (LLMs), an exciting next phase in the field involves understanding how to integrate KGs with these techniques, presenting new possibilities for knowledge representation and utilization[30,31]. Advancements in LLMs will facilitate the incorporation of unstructured data sources such as EHR data. Fourth, given the low percentage of graphs that are open sourced to consider how best to encourage sharing and coordination of KGs in this domain. An example is the Therapeutic Data Commons initiative[32]. One barrier to sharing graphs using clinical data may be regulatory and data privacy concerns.

Ultimately, insights from KGs have yet to realise their potential in delivering clinically actionable findings, especially in patient-facing settings. Adding clinical data to biomedical graphs may be beneficial for drug repurposing and generation of medical science insights. Clinical data itself may be used to better understand disease clustering, patient pathways, and in construction of digital twins. In order to improve the use of clinical data for combination with biomedical data or on its own, the challenges are technical (ontology standardisation, data ingestion) and regulatory (ensuring de-identification, and legal clearance for data sharing). To overcome the

challenge of using proprietary data sources, it is possible to use data that can be accessed on applications, such as MIMIC[33] or use publicly released graphs[26].

Limitations

This study has limitations, including the November 2021 screen cut-off date for manuscript inclusion, which may have excluded more recent developments in KGs. The criteria used for manuscript selection may have inadvertently excluded relevant studies, for example, those that use graphs but do not include this detail in keywords or MESH terms. As there is commercial activity in this area, complementary datasets such as patent screens might provide additional insights into the use of KGs in this area. In conducting this initial scoping review, we established categorizations for use cases, disease areas, and analysis methods through a consensus-driven process involving authors SB and JZ with input from HA and NS. This was followed by an iterative refinement during review, based primarily on ease of manuscript categorisation and identification of areas where manuscripts did not clearly belong to a category. These categorizations are a function of the exploratory nature of this study. Future research may further refine and adapt this framework of categorizations.

Conclusion

In summary, KGs have many possible uses within in biomedicine and healthcare, but their full potential is yet to be realised. The two most popular use cases to date are generation of medical science insights and drug repurposing. There is an opportunity to expand work areas across other use cases and across diseases. Heterogeneity in graph size and context specificity suggests further work is needed to understand optimum graph construction. There are many different techniques used in graph analysis - deploying more sophisticated graph machine learning techniques may improve insights gained from KGs. Validation of findings from graphs through external testing will increase the robustness of conclusions drawn from graphs. While there are many graph-specific factors preventing realisation of utility of KGs in clinical settings, there are many more general barriers to implementation in clinical AI[31,34].

Ethics and consent

Ethical approval and consent were not required.

Data availability

Underlying data

Data used in this article was collected from existing literature. No novel data was collected. All data underlying the results are available as part of the article and no additional source data are required.

Reporting guidelines

Zenodo: PRISMA-ScR Checklist for: 'Scoping review of knowledge graph applications in biomedical and healthcare sciences' https://doi.org/10.5281/zenodo.14781509[19]

Software availability

No software was created.

References

1. Zhang J, Symons J, Agapow P, *et al.*: **Best practices in the real-world data life cycle.** *PLoS Digit Health.* 2022; **1**(1): e0000003.
PubMed Abstract | Publisher Full Text | Free Full Text
2. Krassowski M, Das V, Sahu SK, *et al.*: **State of the field in multi-omics research: from computational needs to data mining and sharing.** *Front Genet.* 2020; **11**: 610798.
PubMed Abstract | Publisher Full Text | Free Full Text
3. Ji S, Pan S, Cambria E, *et al.*: **A survey on knowledge graphs: representation, acquisition, and applications.** *IEEE Trans Neural Netw Learn Syst.* 2022; **33**(2): 494–514.
PubMed Abstract | Publisher Full Text
4. Peng C, Xia F, Naseriparsa M, *et al.*: **Knowledge graphs: opportunities and challenges.** *Artif Intell Rev.* 2023; **56**: 13071–13102.
PubMed Abstract | Publisher Full Text | Free Full Text
5. Patel VL, Evans DA, Groen GJ: **Biomedical knowledge and clinical reasoning**. 1989.
6. Kumar K, Palakal MJ, Mukhopadhyay S, *et al.*: **BioMap: toward the development of a knowledge base of biomedical literature**. In: *Proceedings of the 2004 ACM symposium on Applied computing*. (Association for Computing Machinery, New York, NY USA), 2004; 121–127.
Publisher Full Text
7. Geleta D, Nikolov A, Edwards G, *et al.*: **Biological insights knowledge graph: an integrated knowledge graph to support drug development**. 2021; 2021.10.28.466262.
Publisher Full Text
8. Smith DP, Oechsle O, Rawling MJ, *et al.*: **Expert-augmented computational drug repurposing identified baricitinib as a treatment for COVID-19.** *Front Pharmacol.* 2021; **12**: 709856.
PubMed Abstract | Publisher Full Text | Free Full Text
9. Chen X, Jia S, Xiang Y: **A review: knowledge reasoning over knowledge graph.** *Expert Syst Appl.* 2020; **141**: 112948.
Publisher Full Text
10. Zamini M, Reza H, Rabiei M: **A review of knowledge graph completion.** *Information.* 2022; **13**(8): 396.
Publisher Full Text
11. Nickel M, Murphy K, Tresp V, *et al.*: **A review of relational machine learning for knowledge graphs.** *Proceedings of the IEEE.* 2016; **104**(1): 11–33.
Publisher Full Text
12. Issa S, Adekunle O, Hamdi F, *et al.*: **Knowledge graph completeness: a systematic literature review.** *IEEE Access.* 2021; **9**: 31322–31339.
Publisher Full Text
13. Ye H, Zhang N, Chen H, *et al.*: **Generative knowledge graph construction: a review**. 2023.
Publisher Full Text
14. Zeng X, Tu X, Liu Y, *et al.*: **Toward better drug discovery with knowledge graph.** *Curr Opin Struct Biol.* 2022; **72**: 114–126.
PubMed Abstract | Publisher Full Text
15. Nicholson DN, Greene CS: **Constructing knowledge graphs and their biomedical applications.** *Comput Struct Biotechnol J.* 2020; **18**: 1414–1428.
PubMed Abstract | Publisher Full Text | Free Full Text

16. Chatterjee A, Nardi C, Oberije C, *et al.*: **Knowledge graphs for COVID-19: an exploratory review of the current landscape.** *J Pers Med.* 2021; **11**(4): 300.
PubMed Abstract | Publisher Full Text | Free Full Text

17. Li MM, Huang K, Zitnik M: **Graph representation learning in biomedicine and healthcare.** *Nat Biomed Eng.* 2022; **6**(12): 1353–1369.
PubMed Abstract | Publisher Full Text | Free Full Text

18. Tricco AC, Lillie E, Zarin W, *et al.*: **PRISMA extension for scoping reviews (PRISMA-ScR): checklist and explanation.** *Ann Intern Med.* 2018; **169**(7): 467–473.
PubMed Abstract | Publisher Full Text

19. Budhdeo S: **PRISMA-ScR checklist for: scoping review of knowledge graph applications in biomedical and healthcare sciences.** *Zenodo.* 2025.
http://www.doi.org/10.5281/zenodo.14781509

20. Futoma J, Simons M, Panch T, *et al.*: **The myth of generalisability in clinical research and machine learning in health care.** *Lancet Digit Health.* 2020; **2**(9): e489–e492.
PubMed Abstract | Publisher Full Text | Free Full Text

21. Pujara J, Augustine E, Getoor L: **Sparsity and noise: where knowledge graph embeddings fall short**. In: *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*. (Association for Computational Linguistics, Copenhagen, Denmark), 2017; 1751–1756.
Publisher Full Text

22. Mohamed A, Parambath S, Kaoudi Z, *et al.*: **Popularity agnostic evaluation of knowledge graph embeddings.** In: *Proceedings of the 36th Conference on Uncertainty in Artificial Intelligence (UAI)*. (PMLR), 2020; 1059–1068.
Reference Source

23. Ochoa D, Hercules A, Carmona M, *et al.*: **The next-generation Open Targets Platform: reimagined, redesigned, rebuilt.** *Nucleic Acids Res.* 2023; **51**(D1): D1353–D1359.
PubMed Abstract | Publisher Full Text | Free Full Text

24. Tan F, Zhang T, Yang S, *et al.*: **Discovering booming Bio-entities and their relationship with funds.** *Data Info Manag.* 2021; **5**(3): 312–328.
Publisher Full Text

25. Orchard S, Hermjakob H: **Shared resources, shared costs—leveraging biocuration resources.** *Database (Oxford).* 2015; **2015**: bav009.
PubMed Abstract | Publisher Full Text | Free Full Text

26. Hong C, Rush E, Liu M, *et al.*: **Clinical Knowledge Extraction via Sparse Embedding Regression (KESER) with multi-center large scale Electronic Health Record data.** *NPJ Digit Med.* 2021; **4**(1): 151.
PubMed Abstract | Publisher Full Text | Free Full Text

27. Yang X, Chen A, PourNejatian N, *et al.*: **A Large Language Model for Electronic Health Records.** *NPJ Digit Med.* 2022; **5**(1): 194.
PubMed Abstract | Publisher Full Text | Free Full Text

28. Bean DM, Kraljevic Z, Shek A, *et al.*: **Hospital-wide Natural Language Processing summarising the health data of 1 million patients.** *PLOS Digit Health.* 2023; **2**(5): e0000218.
PubMed Abstract | Publisher Full Text | Free Full Text

29. Li MM, Huang Y, Sumathipala M, *et al.*: **Contextualizing protein representations using deep learning on protein networks and single-cell data**. 2023; 2023.07.18.549602.
Reference Source

30. Zhao J, Zhuo L, Shen Y, *et al.*: **GraphText: graph reasoning in text space.** *arXiv.org.* 2023.
Publisher Full Text

31. Pan JZ, Razniewski S, Kalo JC, *et al.*: **Large Language Models and knowledge graphs: opportunities and challenges.** *arXiv.org.* 2023.
Publisher Full Text

32. Huang K, Fu T, Gao W, *et al.*: **Artificial Intelligence foundation for therapeutic science.** *Nat Chem Biol.* 2022; **18**(10): 1033–1036.
PubMed Abstract | Publisher Full Text | Free Full Text

33. Johnson AEW, Bulgarelli L, Shen L, *et al.*: **MIMIC-IV, a freely accessible Electronic Health Record dataset.** *Sci Data.* 2023; **10**(1): 1.
PubMed Abstract | Publisher Full Text | Free Full Text

34. Zhang J, Budhdeo S, William W, *et al.*: **Moving towards vertically integrated Artificial Intelligence development.** *NPJ Digit Med.* 2022; **5**(1): 143.
PubMed Abstract | Publisher Full Text | Free Full Text

Open Peer Review

Current Peer Review Status: ? ?

Version 1

Reviewer Report 04 March 2025

https://doi.org/10.21956/wellcomeopenres.26035.r119391

?

Bo-Wei Zhao
Southwest University,, Chongqing, China

This paper presents a scoping review of knowledge graph (KG) applications in biomedical and healthcare sciences, analyzing 255 studies to identify key use cases, data characteristics, and validation methodologies. The findings highlight that KGs are most commonly used for medical science insights and drug repurposing, with significant variation in graph size, structure, and analytical techniques. Despite their growing adoption, real-world validation of KG-derived insights remains limited, emphasizing the need for further research on standardization, external validation, and integration with advanced machine learning models.

Based on the analysis of the article provided, here are several specific issues that could be addressed to improve its clarity and scientific rigor:

1.1 The manuscript contains several grammatical and syntactical inconsistencies that affect readability.
1.2 "KGs offered distinct advantages over traditional relational databases, including greater schema flexibility and the capacity to capture nuanced capture of edge characteristics and relationships."
1.3 The manuscript should undergo thorough proofreading to correct redundant phrases, missing articles, and subject-verb agreement errors.

2.1 The methodology follows PRISMA-ScR guidelines, but the rationale for excluding certain knowledge graph use cases (e.g., neuroimaging graph theory) is not well justified.
2.2 The explanation of how node and edge classes were harmonized using the BERT model and k-nearest neighbors requires further clarity—especially regarding parameter tuning and model selection.
2.3 The section discussing meta-graph visualization with Gephi lacks details on preprocessing steps (e.g., thresholding weak connections).

3.1 The manuscript describes node and edge distributions as highly skewed but does not mention whether transformations (e.g., log-scaling) were applied before statistical analysis.
3.2 The discussion on cross-validation for knowledge graph-based machine learning models is minimal, despite its relevance for ensuring robust performance comparisons.
3.3 The analysis lacks hypothesis testing (e.g., statistical significance of observed variations in knowledge graph characteristics across biomedical applications).

4.1 While the review discusses "inside graph" and "outside graph" validation, it does not critically analyze whether the studies included provide sufficient external validation (e.g., real-world clinical implementation).
4.2 The authors should provide more details on whether studies validating drug repurposing insights through preclinical trials were successful or if biases exist in validation methodologies.
4.3 The manuscript does not discuss whether knowledge graphs deployed in clinical settings have demonstrably improved healthcare decision-making.

5.1 While the manuscript cites several prior reviews, it lacks discussion of recent advancements in generative knowledge graph construction and large-scale graph embeddings from 2023/2024.
5.2 Key advancements in graph-based AI models, such as the use of large language models (LLMs) for clinical data extraction and graph reasoning, are not adequately considered.
5.3 The review could benefit from incorporating more recent references (e.g., latest applications of graph-based machine learning in biomedical informatics: 10.1016/j.ins.2024.121360, 10.1039/d4sc06864e, 10.1016/j.csbj.2024.06.032, 10.1093/bib/bbac384, and 10.1039/D4SC03744H).

6.1 The review categorizes use cases into broad themes (e.g., drug repurposing, medical science insights) but does not explain overlap between categories (e.g., drug interactions vs. adverse event prediction).
6.2 Some terms, such as "graph reasoning" and "knowledge graph completion," are used without precise definitions, which could confuse readers unfamiliar with graph-based methodologies.
6.3 The classification of datasets (Table 6) does not indicate whether certain sources contribute disproportionately to specific graph applications, potentially introducing dataset bias.

Are the rationale for, and objectives of, the Systematic Review clearly stated?

Yes

Are sufficient details of the methods and analysis provided to allow replication by others?

Yes

Is the statistical analysis and its interpretation appropriate?

Yes

Are the conclusions drawn adequately supported by the results presented in the review?

Yes

If this is a Living Systematic Review, is the 'living' method appropriate and is the search schedule clearly defined and justified? ('Living Systematic Review' or a variation of this term should be included in the title.)

Not applicable

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: Knowledge graph methods are studied and applied to biomedical and bioinformatics, such as drug-target interactions

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard, however I have significant reservations, as outlined above.

Reviewer Report 26 February 2025

https://doi.org/10.21956/wellcomeopenres.26035.r118658

?

Yuxing Lu
Peking University, Beijing, Beijing, China

This systematic scoping review aims to provide a comprehensive overview of the use of knowledge graphs (KGs) in biomedical and healthcare sciences. The authors performed an extensive literature search on MEDLINE, EMBASE, arXiv, and bioRxiv, ultimately including 255 relevant studies. They identify common use cases for KGs in drug repurposing and medical science insights, with a notable heterogeneity in the size and structure of the graphs used.

Major comments:
1. The authors should clarify why a scoping review, rather than a systematic review or meta-analysis, is needed for this topic.
2. The objectives should also be more focused on how this review will advance the field, detailing the importance of knowledge graph insights in improving real-world healthcare outcomes.
3. For graph analysis, the authors should elaborate on how outliers and non-normal distributions were addressed statistically. Did they apply any normalization methods, or is there a reason for not doing so? A clearer explanation would improve reproducibility.
4. The authors should incorporate a more sophisticated statistical analysis to assess whether variations in graph characteristics (e.g., size, connectivity) influence the conclusions drawn about their utility. For example, they could explore whether certain graph structures are associated with particular use cases or research outcomes.
5. The review should tie conclusions back to the limitations and findings from the studies analyzed to ensure that claims are evidence-based.

Minor comments:
1. Some figures are vague and hard to read, for example Figure 3, Figure 5, and should be updated with high resolution ones.
2. Several sections of the paper are dense with technical terms (e.g., graph embeddings, node

classification), which could be overwhelming for readers from non-computational backgrounds. More definitions and context could be provided, especially in the methods section, to ensure accessibility to a broader audience.
3. There exist some grammar mistakes and the language should be polished.

References

1. Reference Source
2. Bang D, Lim S, Lee S, Kim S: Biomedical knowledge graph learning for drug repurposing by extending guilt-by-association to multiple layers.*Nat Commun*. 2023; **14** (1): 3570 PubMed Abstract | Publisher Full Text

Are the rationale for, and objectives of, the Systematic Review clearly stated?

Partly

Are sufficient details of the methods and analysis provided to allow replication by others?

Yes

Is the statistical analysis and its interpretation appropriate?

Partly

Are the conclusions drawn adequately supported by the results presented in the review?

Partly

If this is a Living Systematic Review, is the 'living' method appropriate and is the search schedule clearly defined and justified? ('Living Systematic Review' or a variation of this term should be included in the title.)

Not applicable

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: Biomedical AI

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard, however I have significant reservations, as outlined above.