Check for updates

RESEARCH ARTICLE

Search strategies to obtain evidence for cancer classification

[version 1; peer review: 1 approved]

Ian Cree[1], Alex Inskip[2], Dilani Lokuhetty[1], Ramon Cierco Jimenez[1], Teresa Lee[1], Gabrielle Goldman-Lévy[1], Christine Giesen[1], Iciar Indave[1], Fiona Campbell[3]

[1]International Agency for Research on Cancer, World Health Organization, Lyon, France
[2]Newcastle University Population Health Sciences Institute, Newcastle, England, UK
[3]School of Social Policy and Society, University of Birmingham, Birmingham, UK

v1 **First published:** 27 Nov 2025, **5**:369
https://doi.org/10.12688/openreseurope.20881.1
Latest published: 27 Nov 2025, **5**:369
https://doi.org/10.12688/openreseurope.20881.1

Abstract

Background

The accuracy of the WHO classification of tumours (WCT), and hence cancer diagnosis, depends on rigorous assessment of the scientific evidence available which describes the defining characteristics of each tumour type. The scale of the task is considerable, given the vast number of publications in the field. The WCT Evidence Gap Map project (WCT EVI MAP; https://wct-evi-map.iarc.who.int/) is designed to answer this need, using Evidence and Gap Map (EGM) methodology to inform the WCT, which is well suited to large datasets.

Methods

As examples to test the strategy, we chose sebaceous carcinoma and cholangiocarcinoma. As a rare tumour, sebaceous carcinoma has a relatively small literature, while the landscape for cholangiocarcinoma is more complex due to recent changes in the classification, the use of outdated terms, and many publications.

Results

The search strategy uses two components - the first describing the tumour type, including alternative terminology, and the second the tumour characteristics (aetiology, pathogenesis, clinical features, epidemiology, histopathology, etc.), with synonyms, as used in the WCT.

Open Peer Review

Approval Status ✓

	1
version 1 27 Nov 2025	✓ view

1. **Angela Crispino**, University of Naples Federico II, Napoli, Italy

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

Conclusions

The results emphasise how relatively small differences in the exact terms and fields searched can produce big differences in the results obtained. It is often important to use a wide range of alternative terms and to cover multiple variations of phrases (via techniques such as proximity searching) to ensure comprehensive results. Limiting results to those characteristics used to classify tumours and using date thresholds helps to improve the relevance of the literature identified.

Plain Language Summary

The World Health Organization (WHO) classifies different types of cancers based on reviews of the scientific literature. Their WHO Classification of Tumours reference book series helps medical professionals make accurate diagnoses. However, reviewing all the available research comprehensively is an enormous task, given the number of studies. To help with this, the WHO launched the WCT Evidence Gap Map (WCT EVI MAP) project. A method called "Evidence and Gap Mapping" is used to organize and assess large amounts of research more efficiently. This method helps identify what is already known and where more research is needed. To test the approach, the researchers modelled their process by looking at two types of cancer: sebaceous carcinoma (a rare cancer with fewer studies) and cholangiocarcinoma (a more complex cancer with many studies and changing terminology). They created search strategies that combined a range of similar terms for the tumour types, along with terms describing their key characteristics, such as how these cancers develop, how they appear under the microscope, and how common they are. They found that small variations in search terms can lead to very different results, so it's important to use a well-worked out search technique using a wide variety of terms. At the same time, the researchers also found that focusing on the most relevant tumour features and using date filters made the search results more useful.

Keywords

Tumour, pathology, classification, evidence gap map, literature search, search strategy.



This article is included in the Horizon Europe gateway.



This article is included in the Health Sciences gateway.

Corresponding author: Ian Cree (creei@iarc.who.int)

Author roles: Cree I: Conceptualization, Investigation, Supervision, Writing – Original Draft Preparation; **Inskip A**: Conceptualization, Data Curation, Formal Analysis, Investigation, Methodology, Writing – Original Draft Preparation; **Lokuhetty D**: Formal Analysis, Project Administration, Writing – Review & Editing; **Cierco Jimenez R**: Formal Analysis, Software, Writing – Original Draft Preparation; **Lee T**: Conceptualization, Methodology, Supervision, Writing – Original Draft Preparation; **Goldman-Lévy G**: Investigation, Writing – Original Draft Preparation, Writing – Review & Editing; **Giesen C**: Conceptualization, Formal Analysis, Investigation, Methodology, Project Administration, Writing – Original Draft Preparation; **Indave I**: Conceptualization, Investigation, Methodology, Project Administration, Resources, Supervision, Writing – Review & Editing; **Campbell F**: Conceptualization, Data Curation, Formal Analysis, Investigation, Methodology, Project Administration, Software, Supervision, Writing – Original Draft Preparation, Writing – Review & Editing

Competing interests: No competing interests were disclosed.

Grant information: This project has received funding from the European Union's Horizon Europe research and innovation programme under grant agreement No 101057127 (Mapping the Evidence for the World Health Organization (WHO) Classification of Tumours: a Living Evidence Gap Map by Tumour Type, WCT EVI MAP).
The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Copyright: © 2025 Cree I *et al*. This is an open access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

How to cite this article: Cree I, Inskip A, Lokuhetty D *et al.* **Search strategies to obtain evidence for cancer classification [version 1; peer review: 1 approved]** Open Research Europe 2025, **5**:369 https://doi.org/10.12688/openreseurope.20881.1

First published: 27 Nov 2025, **5**:369 https://doi.org/10.12688/openreseurope.20881.1

Background

The WHO Classification of Tumours (WCT) is the taxonomy of cancer that underpins cancer diagnosis, and hence treatment and care of cancer patients, as well as research into cancer epidemiology, prevention, diagnosis, and treatment[1–6]. It also informs the WHO International Classification of Diseases for Oncology (ICD-O; https://www.who.int/standards/classifications/other-classifications/international-classification-of-diseases-for-oncology) and the wider International Classification of Diseases (ICD-11; https://icd.who.int/en). However, such global reference tools are not always informed by the best available evidence, with the risk of incorporating misinformation or omitting important study results.

The WCT Evidence Gap Map project (WCT EVI MAP; https://wct-evi-map.iarc.who.int/), developed by the International Collaboration for Cancer Classification and Research (IC3R)[5] and funded by the European Commission (grant number 101057127), uses Evidence and Gap Map (EGM) methodology to inform the WCT and sift the vast literature relevant to tumour classification. This is a novel adaptation of an approach used in social and environmental sciences. EGMs provide an overview of the current available evidence and its quality within a specific topic, theme, or domain to help in the decision-making process. EGM is a type of evidence synthesis methodology within a group of approaches that address broad topic questions[7,8]. The WCT EVI MAP project aims to accelerate progress in cancer diagnosis and management. Integration of evidence-based medicine methods has been shown to have a positive impact in other specialized medical fields[9] and our project represents a first step towards evidence-based pathology.

As an evidence synthesis methodology, EGMs differ from systematic review (SR) by addressing broad research questions and aim to describe the evidence landscape rather than address narrow questions of treatment efficacy[10]. EGMs use a predesigned coding framework/matrix to guide the coding of data and generate an interactive visual table of the findings[11]. This approach is particularly useful in identifying gaps and pockets of low-level evidence in the existing evidence base, aiding decisions on priorities for future research and informing certainty for decision making. The EGM process uses explicit and transparent methods; follows a rigorous process; is accountable, replicable, and updatable; and involves users to ensure relevance and usefulness[12].

We aim to map existing evidence for the WCT, to identify gaps and pockets of low-level evidence, and to synthesise available evidence. The map will compile and publicly present this evidence, increasing the discoverability and use of studies by the WCT and other stakeholders. This will facilitate evidence-based decision-making in pathology and promote the application of higher research standards in the field[6]. There has been intense development of specific software tools to aid the performance of SRs. SRs and EGMs share several steps in their methodology and initial workflow (e.g. systematic search, selection, and critical appraisal), so similar software applications can be used to assist in their production[13].

This paper addresses the need for careful search strategies, a step that is common to all evidence syntheses and that has significant challenges. If the search is too narrow, relevant evidence may be missed, while if it is too broad, it may prove impossible to review the amount of information obtained. Our aim was to identify studies assessing tumours based on the tumour characteristics used in the WCT, and to assess the level of evidence for each based on the study design[14,15]. We therefore describe the steps used to produce the search strategies for the EGM for tumours in the development of the forthcoming 6th edition of the WCT (Figure 1).

Methods

This study has been performed as part of the WCT EVI MAP project. These search strategies have been developed by information specialists in collaboration with topic experts, mainly pathologists, cancer biologists and epidemiologists from the WCT EVI MAP group. Separate search strategies are needed for each EGM to search for all in-remit tumour types and the tumour characteristics as identified in the WCT (Table 1). The results are limited to studies relevant to tumour diagnosis (as described in the protocol's inclusion/exclusion criteria)[15].

For this exercise, comprehensive searches of the PubMed databases have been designed using tailored search strings



Figure 1. Steps used to produce the search strategies for the Evidence Gap Maps of the WHO Classification of Tumours 6th edition.

Table 1. Terms and term combination for the search strategy for the WCT EVI MAP search strategy plan.

Term	Search field
aetiolog*	[tiab]
analysis	[tiab]
Anatomy	[tiab]
Biomarker	[tiab]
burden	[tiab]
"cause of death"	[tiab]
classif*	[tiab]
"clinical features"	[tiab]
clinicopathol*	[tiab]
cytolog*	[tiab]
demograph*	[tiab]
diagnosis	[tiab]
"diagnostic imaging"	[tiab]
epidemiolog*	[tiab]
ethnolog*	[tiab]
etiolog*	[tiab]
"gene expression"	[tiab:~6]
genetic*	[tiab]
"genome sequencing"	[tiab:~6]
histol*	[tiab]
histopathol*	[tiab]
immunohisto*	[tiab]
incidence	[tiab]
mortality	[tiab]
"numerical data"	[tiab]
pathogenesis	[tiab]
pathogeni*	[tiab]
phenotyp*	[tiab]
physiolog*	[tiab]
physiopatholog*	[tiab]
population	[tiab]
prevalence	[tiab]
recurrence	[tiab]
registries	[tiab]
registry	[tiab]
statistic*	[tiab]
survival	[tiab]
transcriptome*	[tiab]
"treatment outcome*"	[tiab]
ultrastructur*	[tiab]
"biomarkers"	[MeSH]
"Cause of Death"	[MeSH]
"Demography"	[MeSH]
"diagnosis, differential"	[MeSH terms:noexp]
"diagnosis"	[MeSH terms:noexp]
"diagnostic errors"	[MeSH terms:noexp]
"early detection of cancer"	[MeSH terms:noexp]
"early diagnosis"	[MeSH terms:noexp]
"Epidemiologic Studies"	[MeSH]
"Epidemiology"	[MeSH]
"Gene Expression Profiling"	[MeSH]
"Global Burden of Disease"	[MeSH]
"Histological Techniques"	[MeSH]
Histology	[MeSH]
"Incidence"	[MeSH]
"molecular diagnostic techniques"	[MeSH terms:noexp]
"Neoplasm Recurrence, Local"	[MeSH]
"Phenotype"	[MeSH]
"prevalence"	[MeSH]
"prognosis"	[MeSH terms:noexp]
Registries	[MeSH]
"risk factors"	[MeSH]
"Transcriptome"	[MeSH]
"Whole Genome Sequencing"	[MeSH]
"analysis"	[MeSH Subheading:noexp]
"anatomy and histology"	[MeSH Subheading:noexp]
"classification"	[MeSH Subheading:noexp]
"cytology"	[MeSH Subheading:noexp]
"diagnosis"	[MeSH Subheading:noexp]
"diagnostic imaging"	[MeSH Subheading:noexp]
"epidemiology"	[MeSH Subheading:noexp]
"ethnology"	[MeSH Subheading:noexp]
"etiology"	[MeSH Subheading:noexp]
"genetics"	[MeSH Subheading:noexp]

Term	Search field
"mortality"	[MeSH Subheading:noexp]
"pathogenicity"	[MeSH Subheading:noexp]
"pathology"	[MeSH Subheading:noexp]
"physiology"	[MeSH Subheading:noexp]
"physiopathology"	[MeSH Subheading:noexp]
"statistics and numerical data"	[MeSH Subheading:noexp]
"Treatment Outcome"	[MeSH:NoExp]
"ultrastructure"	[MeSH Subheading:noexp]

with database-specific subject and keyword terms describing the existing structure of the WCT and current WCT/ICD-O nomenclature (i.e. terms referring to the tumour type and the WCT characteristics). Searches were conducted using English terms; all eligible results had English titles and abstracts. A tailored methodology and step wise approach was developed and applied (Figure 1).

The preliminary draft searches were developed iteratively by information specialists and topic experts. This process was initiated in July 2023 and finalized in May 2024. The first phase was the terminology selection by the identification of representative search terms for both tumour types and tumour characteristics. Subject headings and subheadings in controlled vocabularies (such as MeSH, EMTREE) have been leveraged as much as possible to improve search specificity and sensitivity.

Term combination and strategy building was an iterative process testing the terminology selected. The foundational strategy was built in PubMed for adaptation to Embase. The team defined the terminology selection, i.e., the identification of natural language and controlled vocabulary terms that answer the different components of the research question. Representative terms were identified through the previous edition of the WCT 5th edition (https://tumourclassification.iarc.who.int). Further natural language and controlled terms were added by input from experts on the field, and controlled terms were identified through searches of MeSH terms. Terms were grouped when related.

The WCT EVI MAP project will consider all tumour types listed in the WCT Blue Books 5th edition (WCT5) (https://tumourclassification.iarc.who.int/welcome/). However, given the enormous number of tumour types, exploratory searches were performed to identify key terms and concepts for tumour characteristics that to be included in the final search strategy. As an initial database, PubMed was chosen because all WCT references are indexed there; it has wide coverage of relevant literature; it has global accessibility; and it incorporates medical subject (MeSH) headings and subheadings helpful in identifying studies. The searches were then adapted to Embase, which indexes some journals that are not included by PubMed, to help ensure comprehensive screening of the published literature.

Brief comparison and checking were used to determine whether the initial suggestion of strictly using WCT terminology for the tumour terms in searches was sufficient. This testing found that covering some variations and utilising MeSH where possible was more comprehensive and useful, so this was implemented in the final search strategy.

Term & concept definition

The research focuses on two components based on the study framework: tumour types and tumour characteristics. Terms and concepts that identify and retrieve relevant studies have been chosen based on the research question.

- Year of publication: searches are limited to studies published from 01/01/2018 onwards, i.e. from the start of the 5th edition WCT. For rare tumours (incidence <6/100,000 persons/year) (28687376), a broader time range from 01/01/2013 has been used.

- Language: searches were limited to papers with titles and abstracts published in English, regardless of the full article's publication language.

- Concept validation: after initial pilot searches, the two main components, i.e., tumour type and tumour characteristics, were validated to determine their feasibility for searches.

- Strategy building: For each component the synonymous terms were joined using the OR Boolean logic operator. Different components were joined with the AND Boolean logic operator. The NOT Boolean operator was considered depending on the obtained results. The combination of groups was Component 1 for tumour type AND Component 2 for tumour characteristics. For each tumour type, MeSH terms and text words were used. Proximity operators were used as necessary. Tumour characteristic terms did not vary and were universal for all types of tumours.

- Search builder options. In addition to including both MeSH terms [MeSH] and free text words for tumour types, the following rules were applied for tumour classifiers and study types: MeSH Major Topic [Major] were used occasionally, though not exclusively, in the searches, as this forces PubMed to only search for articles in which the Major Topic has been deemed one of the main focuses of the article by the indexer. Subheadings were used for tumour characteristics [MeSH Subheading]. Narrower MeSH terms below the selected term in the MeSH hierarchy were assessed for inclusion as necessary [MeSH]/[MeSH:NoExp].

- Pilot testing: Pilot searches were performed for studies of a rare tumour with a limited number of published studies, (sebaceous carcinoma), before developing more complex searches for cholangiocarcinoma where recent changes in classification and terminology have occurred. The results

were presented to experts in the WCT EVI MAP team for feedback, validation, and refinement. Based on the experts' input, strategies for all EGM searches were produced. Expert input on this process has also informed the overall search strategies for other tumour types.

Example searches:

1. Sebaceous carcinoma (Table 2): Initial pilot searches were performed for studies addressing sebaceous carcinoma as an example of a rare tumour with limited literature[16]. This type of tumour was selected due to its relative low incidence

Table 2. Comparison of different searches for sebaceous carcinoma.

Search number	Query	Results	Notes
1	"sebaceous carcinoma"[tiab:~6] OR "sebaceous carcinomas"[tiab:~6] OR "sebaceous adenocarcinoma"[tiab:~6] OR "sebaceous adenocarcinomas"[tiab:~6]	1,790	Sebaceous near carcinoma or adenocarcinoma in title or abstract
2	("sebaceous malignant"[tiab:~6] OR "sebaceous malignancy"[tiab:~6] OR "sebaceous malignancies"[tiab: ~6]) NOT ("visceral malign*"[tiab] OR "internal malign*"[tiab])	348	Sebaceous near malignancy in title or abstract (but excluding some unhelpful non-cutaneous terms)
3	"meibomian carcinoma"[tiab:~6] OR "meibomian carcinomas"[tiab:~6] OR "meibomian adenocarcinoma"[tiab:~6] OR "meibomian adenocarcinomas"[tiab:~6]	124	Meibomian near carcinoma or adenocarcinoma in title or abstract
4	"sebaceous carcinoma*"[ot] OR "sebaceous adenocarcinoma*"[ot] OR "sebaceous gland carcinoma*"[ot] OR "meibomian carcinoma*"[ot] OR "meibomian adenocarcinoma*"[ot] OR "meibomian gland carcinoma*"[ot]	283	Searching in 'Other terms' field (can't use phrase search but can use wildcards)
5	"Adenocarcinoma, Sebaceous"[MeSH]	565	Most relevant MeSH heading
6	"Sebaceous Gland Neoplasms"[MeSH Terms:noexp] AND (carcinoma*[tw] OR adenocarcinoma*[tw]) AND (sebaceous[ti] OR carcinoma*[ti] OR adenocarcinoma*[ti] OR tumo*[ti] OR neoplas*[ti] OR cancer*[ti] OR malignan*[ti])	1,024	Partly relevant broad MeSH heading, with additional requirements to make it more likely to be relevant
7	#1 OR #2 OR #3 OR #4 OR #5 OR #6	2,207	Any of the above ways of searching for the sebaceous carcinoma concept
8	aetiolog*[tiab] OR analysis[tiab] OR anatomy[tiab] OR biomarker[tiab] OR burden[tiab] OR "cause of death"[tiab] OR classif*[tiab] OR "clinical features"[tiab] OR clinicopathol*[tiab] OR cytolog*[tiab] OR demograph*[tiab] OR diagnosis[tiab] OR "diagnostic imaging"[tiab] OR epidemiolog*[tiab] OR ethnolog*[tiab] OR etiolog*[tiab] OR "gene expression"[tiab:~6] OR genetic*[tiab] OR "genome sequencing"[tiab:~6] OR histol*[tiab] OR histopathol*[tiab] OR immunohisto*[tiab] OR incidence[tiab] OR mortality[tiab] OR "numerical data"[tiab] OR pathogenesis[tiab] OR pathogeni*[tiab] OR phenotyp*[tiab] OR physiolog*[tiab] OR physiopatholog*[tiab] OR population[tiab] OR prevalence[tiab] OR recurrence[tiab] OR registries[tiab] OR registry[tiab] OR statistic*[tiab] OR survival[tiab] OR transcriptome*[tiab] OR "treatment outcome*"[tiab] OR ultrastructur*[tiab]	14,917,167	Title/abstract search terms for tumour characteristics
9	"biomarkers"[MeSH] OR "Cause of Death"[MeSH] OR "Demography"[MeSH] OR "diagnosis, differential"[MeSH terms:noexp] OR "diagnosis"[MeSH terms:noexp] OR "diagnostic errors"[MeSH terms:noexp] OR "early detection of cancer"[MeSH terms:noexp] OR "early diagnosis"[MeSH terms:noexp] OR Epidemiologic Studies[MeSH] OR "Epidemiology"[MeSH] OR "Gene Expression Profiling"[MeSH] OR "Global Burden of Disease"[MeSH] OR "Histological Techniques"[MeSH] OR "incidence"[MeSH] OR "Incidence"[MeSH] OR "molecular diagnostic techniques"[MeSH terms:noexp] OR "Neoplasm Recurrence, Local"[MeSH] OR "Phenotype"[MeSH] OR "prevalence"[MeSH] OR "prognosis"[MeSH terms:noexp] OR Registries[MeSH] OR "risk factors"[MeSH] OR "Transcriptome"[MeSH] OR "Whole Genome Sequencing"[MeSH]	7,384,065	MeSH search terms for tumour characteristics

Search number	Query	Results	Notes
10	"analysis"[MeSH Subheading:noexp] OR "anatomy and histology"[MeSH Subheading:noexp] OR "classification"[MeSH Subheading:noexp] OR "cytology"[MeSH Subheading:noexp] OR "diagnosis"[MeSH Subheading:noexp] OR "diagnostic imaging"[MeSH Subheading:noexp] OR "epidemiology"[MeSH Subheading:noexp] OR "ethnology"[MeSH Subheading:noexp] OR "etiology"[MeSH Subheading:noexp] OR "genetics"[MeSH Subheading:noexp] OR "mortality"[MeSH Subheading:noexp] OR "pathogenicity"[MeSH Subheading:noexp] OR "pathology"[MeSH Subheading:noexp] OR "physiology"[MeSH Subheading:noexp] OR "physiopathology"[MeSH Subheading:noexp] OR "statistics and numerical data"[MeSH Subheading:noexp] OR "Treatment Outcome"[MeSH:NoExp] OR "ultrastructure"[MeSH Subheading:noexp]	18,746,138	MeSH subheading search terms for tumour characteristics
11	#8 OR #9 OR #10	24,671,064	Full list of terms for diagnosis and classification to cover all WCT characteristics for all disorders.
12	#7 AND #11	2,025	The SC results that also satisfy the WCT characteristics requirements
13	#12 AND (2018:3000/12/12[pdat])	546	Published since 2018
14	#13 NOT (animals[mh] NOT (humans[mh] or model*[tw]))	541	With animal studies removed (unless also tagged as human or mentioning models)

approximately 2 cases per 1 million person-years in White populations, 1 in Asian and Pacific Islander populations, and 0.5 in Black populations in the US) yet high familiarity among pathologists (Dasgupta *et al*. 2009) and, therefore, a lower number of published studies.

2. Cholangiocarcinoma (Table 3): To test our strategy with a tumour type with a more challenging literature landscape for searches, we chose cholangiocarcinoma, as this has had many changes in the classification over the last 10 years. Use of outdated terminology, many synonyms (alternative terminology), and a large number of recently published studies made this a good test of the strategy. We used WCT5 accepted terms and the standard concept 2 terms similar to all tumour types for these searches.

Results

The results of different search strategies are shown in Table 2 and Table 3 for sebaceous carcinoma and cholangiocarcinoma respectively, using PubMed. The results obtained are expressed as the number of publications retrieved.

Search development: Sebaceous carcinoma

Small differences in the exact terms and fields searched can have a substantial effect on the number of results retrieved. Searching for "sebaceous carcinoma"[tiab] (i.e. 'sebaceous carcinoma', the main WCT term, as an exact phrase in the title and/or abstract fields), at the point of testing, retrieved 1,001 results. Many of these would have been relevant. However, the results would not have included some potentially relevant records where the two words weren't directly adjacent, for example articles using the phrase 'sebaceous gland carcinoma', or even just the plural 'sebaceous carcinomas'.

The simplest way to cover variations in phrasing is to combine terms with 'AND' rather than searching as a phrase. Plurals can sometimes be covered using wildcards – in PubMed, the use of the asterisk wildcard, which represents 'any number of any letters, including zero', enables a search term including it to cover word variations, such as plurals.

In contrast to the very specific search, searching for sebaceous[tiab] AND carcinoma*[tiab] (the word sebaceous and the word carcinoma(s) anywhere in the title and/or abstract) returned 2,257 results. This would have covered 'sebaceous gland carcinomas' and other variations. However, those results would also have included non-relevant material where the words 'sebaceous' and 'carcinoma' were separated by several sentences in the abstract, with little semantic connection.

Furthermore, both versions of the search would have missed results that didn't use both of the terms – for example, records referring to 'Meibomian gland carcinoma' (which is listed as acceptable terminology in the WCT 5th edition).

A search for (sebaceous[tiab] OR meibomian[tiab]) AND carcinoma*[tiab] returned 2,332 results – a small increase, mostly due to now including results with 'Meibomian gland carcinoma', 'Meibomian carcinoma'.

Table 3. Searches for cholangiocarcioma.

Search number	Query	Results	Notes
1	"Cholangiocarcinoma"[Majr]	11,403	Cholangiocarcinoma as MeSH major heading
2	cholangiocarcinoma*[ti] OR "bile duct carcinoma*"[ti] OR "cholangiocellular carcinoma*"[ti] OR "cholangiolocellular carcinoma*"[ti] OR "cholangiolocarcinoma*"[ti] OR cholangiocarcinoma*[ot] OR "bile duct carcinoma*"[ot] OR "cholangiocellular carcinoma*"[ot] OR "cholangiolocellular carcinoma*"[ot] OR "cholangiolocarcinoma*"[ot]	14,321	Relevant terms as phrases in title or other terms fields
3	(cholangiocarcinoma*[tiab] OR "bile duct carcinoma*"[tiab] OR "cholangiocellular carcinoma*"[tiab] OR "cholangiolocellular carcinoma*"[tiab] OR "cholangiolocarcinoma*"[tiab]) AND ("Cholangiocarcinoma"[MeSH] OR "Biliary Tract Neoplasms"[Majr:NoExp] OR "Bile Duct Neoplasms"[Majr:NoExp] OR "Liver Neoplasms"[Majr:NoExp])	14,474	Relevant terms in the abstract (title too, but that's already covered above), IF the results also have a fairly relevant MeSH heading (including Cholangiocarcinoma non-major heading, or major heading of broader MeSH terms like 'Bile Duct Neoplams' or 'Liver Neoplams')
4	(cholangiocarcinoma*[tiab] OR "bile duct carcinoma*"[tiab] OR "cholangiocellular carcinoma*"[tiab] OR "cholangiolocellular carcinoma*"[tiab] OR "cholangiolocarcinoma*"[tiab]) NOT medline[sb]	4,260	Relevant terms in abstract, as above, and no MeSH indexing of the records at all.
5	#1 OR #2 OR #3 OR #4	19,959	Any of the above methods of searching for cholangiocarcinoma
6	((distal[tw] OR extrahepatic[tw] OR perihilar[tw]) NOT (intrahepatic[tw] OR Liver[MeSH] OR "Liver Neoplasms"[MeSH:noexp])) OR ((distal[ti] OR extrahepatic[ti] OR perihilar[ti]) NOT intrahepatic[ti])	288,129	Results mentioning types of cholangiocarcinoma that aren't relevant (e.g. distal, extrahepatic) unless also mentioning 'intrahepatic'
7	#5 NOT #6	17,761	Results with the exclusively non-relevant types removed.
8	aetiolog*[tiab] OR analysis[tiab] OR anatomy[tiab] OR biomarker[tiab] OR burden[tiab] OR "cause of death"[tiab] OR classif*[tiab] OR "clinical features"[tiab] OR clinicopathol*[tiab] OR cytolog*[tiab] OR demograph*[tiab] OR diagnosis[tiab] OR "diagnostic imaging"[tiab] OR epidemiolog*[tiab] OR ethnolog*[tiab] OR etiolog*[tiab] OR "gene expression"[tiab:~6] OR genetic*[tiab] OR "genome sequencing"[tiab:~6] OR histol*[tiab] OR histopathol*[tiab] OR immunohisto*[tiab] OR incidence[tiab] OR mortality[tiab] OR "numerical data"[tiab] OR pathogenesis[tiab] OR pathogeni*[tiab] OR phenotyp*[tiab] OR physiolog*[tiab] OR physiopatholog*[tiab] OR population[tiab] OR prevalence[tiab] OR recurrence[tiab] OR registries[tiab] OR registry[tiab] OR statistic*[tiab] OR survival[tiab] OR transcriptome*[tiab] OR "treatment outcome*"[tiab] OR ultrastructur*[tiab]	14,917,167	Title/abstract search terms for tumour characteristics
9	"biomarkers"[MeSH] OR "Cause of Death"[MeSH] OR "Demography"[MeSH] OR "diagnosis, differential"[MeSH terms:noexp] OR "diagnosis"[MeSH terms:noexp] OR "diagnostic errors"[MeSH terms:noexp] OR "early detection of cancer"[MeSH terms:noexp] OR "early diagnosis"[MeSH terms:noexp] OR Epidemiologic Studies[MeSH] OR "Epidemiology"[MeSH] OR "Gene Expression Profiling"[MeSH] OR "Global Burden of Disease"[MeSH] OR "Histological Techniques"[MeSH] OR "incidence"[MeSH] OR "Incidence"[MeSH] OR "molecular diagnostic techniques"[MeSH terms:noexp] OR "Neoplasm Recurrence, Local"[MeSH] OR "Phenotype"[MeSH] OR "prevalence"[MeSH] OR "prognosis"[MeSH terms:noexp] OR Registries[MeSH] OR "risk factors"[MeSH] OR "Transcriptome"[MeSH] OR "Whole Genome Sequencing"[MeSH]	7,384,065	MeSH search terms for tumour characteristics

Search number	Query	Results	Notes
10	"analysis"[MeSH Subheading:noexp] OR "anatomy and histology"[MeSH Subheading:noexp] OR "classification"[MeSH Subheading:noexp] OR "cytology"[MeSH Subheading:noexp] OR "diagnosis"[MeSH Subheading:noexp] OR "diagnostic imaging"[MeSH Subheading:noexp] OR "epidemiology"[MeSH Subheading:noexp] OR "ethnology"[MeSH Subheading:noexp] OR "etiology"[MeSH Subheading:noexp] OR "genetics"[MeSH Subheading:noexp] OR "mortality"[MeSH Subheading:noexp] OR "pathogenicity"[MeSH Subheading:noexp] OR "pathology"[MeSH Subheading:noexp] OR "physiology"[MeSH Subheading:noexp] OR "physiopathology"[MeSH Subheading:noexp] OR "statistics and numerical data"[MeSH Subheading:noexp] OR "Treatment Outcome"[MeSH:NoExp] OR "ultrastructure"[MeSH Subheading:noexp]	18,746,138	MeSH subheading search terms for tumour characteristics
11	#8 OR #9 OR #10	24,671,064	Full list of terms for diagnosis and classification to cover all WCT characteristics for all disorders.
12	#7 AND #11	16,461	The cholangiocarcinoma results that also satisfy the WCT characteristics requirements
13	#12 AND (2018:3000/12/12[pdat])	7,314	Published since 2018
14	#13 NOT (animals[mh] NOT (humans[mh] or model*[tw]))	7,265	With animal studies removed (unless also tagged as human or mentioning models)

However, there are further possible terms, not covered by the WCT 5th edition suggested acceptable terminology, but used by some authors, for example, 'sebaceous adenocarcinoma'. Using MeSH terms can help mitigate this to some extent, but including as many relevant terms as possible helps maximize sensitivity.

The final combination of search terms (see Table 2), was selected to try and cover all likely relevant phrases, whilst also using proximity searching to take a more balanced approach than either words needing to appear as specific phrases (often too restrictive) and words appearing anywhere in the abstract (often too broad). In PubMed, searching for, for example, "sebaceous carcinoma"[tiab:~6] will return results where the words 'sebaceous' and 'carcinoma' appear near (but not necessarily next to) each other – in this case, up to six words apart.

This final combination of terms returned 2,207 results (prior to adding concept 2 terms or date limits).

Finally, component 2, the full list of terms for diagnosis and classification to cover all WCT characteristics, was added to the search (Table 2). Adding this and publication date requirements, plus removal of veterinary studies, located 541 results overall.

Search development: cholangiocarcinoma

For cholangiocarcinoma, the situation is more complex, as the WCT5 separated bile duct carcinomas into intrahepatic and extrahepatic tumours types, with carcinoma of the gallbladder as a further related tumour type. Therefore, for intrahepatic cholangiocarcinoma, publications not relevant to the liver needed to be excluded from the search. However, many papers study multiple tumour types, so papers covering both extrahepatic and intrahepatic cholangiocarcinoma would needed to be included (and would also need including when searching for extrahepatic cholangiocarcinoma). The alternative terminology listed in the WCT is as follows:

Intrahepatic cholangiocarcinoma: Acceptable in WCT: intrahepatic bile duct carcinoma. Not recommended in WCT: peripheral cholangiocarcinoma; cholangiocellular carcinoma; cholangiolocellular carcinoma.

Carcinoma of the extrahepatic bile ducts: Acceptable in WCT: extrahepatic cholangiocarcinoma; bile duct cystadenocarcinoma.

Carcinoma of the gallbladder: Not recommended in WCT: cholangiocarcinoma; gallbladder adenocarcinoma.

Table 3 shows the results for Intrahepatic cholangiocarcinoma. A simple search for "Intrahepatic cholangiocarcinoma"[tiab] returned 5,061 results. Searching the WCT main term directly finds some results but misses results that use a range of other terms, including results where 'intrahepatic' and 'cholangiocarcinoma' are not part of the same phrase. A search for "Intrahepatic"[tiab] AND "cholangiocarcinoma*"[tiab] returns 6,854 results. As expected, separating the terms and adding a

wildcard (for plurals) returns slightly more results. However, it does include results where the two terms are not part of the same phrase, and it does miss results where 'intrahepatic' is implied but not actually mentioned at all. A search for ("Intrahepatic"[tiab] AND "cholangiocarcinoma*"[tiab]) OR "peripheral cholangiocarcinoma"[tiab] OR "cholangiocellular carcinoma"[tiab] OR "cholangiolocellular carcinoma"[tiab] OR "bile duct carcinoma"[tiab] returns 8,618 results. In this example, adding some alternatives, based on the WCT record increases results further, but there is some ambiguity as to whether 'intrahepatic' needs including for any of these terms.

Applying a more comprehensive strategy retrieves significantly more results (17,761 results; Table 3). This may exclude some less relevant results, i.e., those where the terms are only mentioned incidentally in the abstract or those tagged with a non-major MeSH heading without further relevant terms in the title or abstract), but it includes many more by including results that don't necessarily specify 'intrahepatic'. Results that exclusively and specifically refer to extrahepatic cholangiocarcinoma are excluded.

Finally, component 2 is added to the search to cover the full list of terms for diagnosis and classification applicable to WCT characteristics. Using these terms and adding publication date requirements, as well as removing veterinary studies, retrieves 7,265 results overall.

Searches for tumour characteristics

For the individual characteristics used to classify tumours, the issues are compounded by the wide array of synonyms used in the literature. For instance, "Pathogenesis" is the "development of disease" from the cause and is studied with etiology in many papers.

Discussion

The development of a generic search strategy for the WCT EVI MAP project has been challenging for several reasons and serves to illustrate the issues which must be addressed in SR of published literature, particularly in pathology. The characteristics used for classification are described by different disciplines and study types in different ways, leading to significant complexity[14,15]. Furthermore, the widespread use of older terminology in recently published papers means that wider searches are required to find all relevant papers than might be supposed. This includes the use of terms which are not recommended as alternative terminology in the 5th edition WCT. It would be helpful if these terms were routinely included in summary information, either using keywords or within the abstract for all tumour-related publications. The lack of such terms caused some difficulty in developing specific yet consistent search strings for all tumour types, revealing that each tumour type will require a certain degree of bespoke handling. Issues of consistency across all tumour types must be addressed as searches are performed, requiring multi-disciplinary input.

PubMed and other search engines have specific constraints. For example, MeSH terms have a hierarchy which differs from the WCT hierarchy, but can be helpful in gathering relevant evidence, so including these in searches is often helpful to avoid missing relevant references. In addition to the searches for classification characteristics, and each tumour type, we initially performed searches for study designs (e.g. case studies, case-control studies and randomised clinical trial), but these were not defined in all references and would have limited the results, introducing an unnecessary bias. However, it should be noted that the aim of WCT EVI MAP is to map all existing evidence according to their level of evidence[14], which depends largely on the study design. It remains best practice to specify the study design in the summary information for all publications, and this is becoming more important as the biomedical literature continues to expand.

In this paper, we have illustrated our results through PubMed, but best practice in systematic review is to use more than one database, though that does increase the workload. In the WCT EVI MAP project, we have chosen to look at both PubMed and Embase. Most software solutions for handling references handle duplicates. Further assistance comes from tools included with these databases, such as MeSH/EMTREE terms, wildcards (e.g. the use of the asterisk in PubMed), and proximity operators. Although the WCT provides a well-accepted list of preferred tumour names, it is inevitable that some older terms (or new ones) will be used in publications, and it is important that these are included in literature reviews. Use of the alternative terminology listed in the WCT within searches is therefore helpful.

Our decision to limit our searches to the period since the 5th edition of the WCT commenced (2018) was entirely pragmatic to limit the work involved, as we expected that the prior literature would have been adequately reviewed previously. Initial studies show that in several cases, substantial gaps exist in the references cited in the last edition. To address this, we are producing EGMs for each of the volumes published in the 5th edition[15,17] and including all published literature without a date limitation in the database for future work. The combination of these with the new literature searches will be used to ensure optimal citation of references in the forthcoming 6th edition. To date, we have produced EGMs for certain thoracic tumours and breast tumours. We expect that this use of more rigorous searches of the newer literature and the use of redefined levels of evidence[6,14] will allow the WCT to continue to evolve and improve, to the benefit of patients and those interested in advancing their care.

In many cases, comprehensive searches are facilitated by a team approach, including experts in literature search methods as well as those well-versed in the subject matter. We expect some searches to return very large numbers of references, mainly for common tumour types. Evaluating large numbers of studies is a concern for many such projects and where there is a very large literature, adequate review is daunting to the point of impossibility for most research groups. We hope to apply artificial intelligence methods to address this issue in our current project and aid development of EGMs as an aid to revision of the WCT in its next edition[18].

Ethics and consent

Ethical approval and consent were not required.

Disclaimer

The content of this article represents the personal views of the authors and does not represent the views of the authors' employers and associated institutions. Where authors are identified as personnel of the International Agency for Research on Cancer / World Health Organization, the authors alone are responsible for the views expressed in this article and they do not necessarily represent the decisions, policy or views of the International Agency for Research on Cancer / World Health Organization.

Data availability statement

The data that support the findings of this study are available in the article.

Acknowledgements

We are grateful to all those within the WCT EVI MAP teams who have contributed to this search strategy development, especially Nickolas Myles and Ayat Ahmadisaraeilani, and the wider project. The authors would also like to thank Laura Brispot and Benjamin Duclos for their contribution in coordinating this work at IARC.

References

1. Salto-Tellez M, Cree IA: **Cancer taxonomy: pathology beyond pathology.** *Eur J Cancer.* 2019; **115**: 57–60.
PubMed Abstract | Publisher Full Text

2. Nagtegaal ID, Odze RD, Klimstra D, *et al.*: **The 2019 WHO Classification of Tumours of the digestive system.** *Histopathology.* 2020; **76**(2): 182–188.
PubMed Abstract | Publisher Full Text | Free Full Text

3. Cree IA, White VA, Indave BI, *et al.*: **Revising the WHO classification: female genital tract tumours.** *Histopathology.* 2020; **76**(1): 151–156.
PubMed Abstract | Publisher Full Text

4. Cree IA, Indave BI: **Commentary: cancer research quality and tumour classification.** *Tumour Biol.* 2020; **42**(2): 1010428320907544.
PubMed Abstract | Publisher Full Text

5. Cree IA, Indave Ruiz BI, Zavadil J, *et al.*: **The international collaboration for cancer classification and research.** *Int J Cancer.* 2021; **148**(3): 560–571.
PubMed Abstract | Publisher Full Text | Free Full Text

6. Colling R, Indave I, Del Aguilla J, *et al.*: **Moving forward on tumor pathology research reporting: a guide for pathologists from the World Health Organization Classification of Tumors living evidence gap map by tumour type group.** *Mod Pathol.* 2024; **37**(7): 100515.
PubMed Abstract | Publisher Full Text

7. Indave BI, Colling R, Campbell F, *et al.*: **Evidence-levels in pathology for informing the *WHO classification of tumours*.** *Histopathology.* 2022; **81**(4): 420–425.
PubMed Abstract | Publisher Full Text

8. Campbell F, Blank L, Cantrell A, *et al.*: **Factors that influence mental health of university and college students in the UK: a systematic review.** *BMC Public Health.* 2022; **22**(1): 1778.
PubMed Abstract | Publisher Full Text | Free Full Text

9. Connor L, Dean J, McNett M, *et al.*: **Evidence-based practice improves patient outcomes and healthcare system return ofin investment: findings from a scoping review.** *Worldviews Evid Based Nurs.* 2023; **20**(1): 6–15.
PubMed Abstract | Publisher Full Text

10. Snilstveit B, Vojtkova M, Bhavsar A, *et al.*: **Evidence & gap maps: a tool for promoting evidence informed policy and strategic research agendas.** *J Clin Epidemiol.* 2016; **79**: 120–129.
PubMed Abstract | Publisher Full Text

11. Campbell F, Tricco AC, Munn Z, *et al.*: **Mapping reviews, scoping reviews, and evidence and gap maps (EGMs): the same but different- the "Big Picture" review family.** *Syst Rev.* 2023; **12**(1): 45.
PubMed Abstract | Publisher Full Text | Free Full Text

12. White H, Albers B, Gaarder M, *et al.*: **Guidance for producing a Campbell evidence and gap map.** *Campbell Syst Rev.* 2020; **16**(4): e1125.
PubMed Abstract | Publisher Full Text | Free Full Text

13. Cierco Jimenez R, Lee T, Rosillo N, *et al.*: **Machine learning computational tools to assist the performance of Systematic Reviews: a mapping review.** *BMC Med Res Methodol.* 2022; **22**(1): 322.
PubMed Abstract | Publisher Full Text | Free Full Text

14. Colling R, Indave I, Del Aguila J, *et al.*: **A new hierarchy of research evidence for tumor pathology: a Delphi study to define levels of evidence in tumor pathology.** *Mod Pathol.* 2024; **37**(1): 100357.
PubMed Abstract | Publisher Full Text

15. Del Aguila Mejía J, Armon S, Campbell F, *et al.*: **Understanding the use of evidence in the WHO Classification of Tumours: a protocol for an evidence gap map of the classification of tumours of the lung.** *BMJ Open.* 2022; **12**(10): e061240.
PubMed Abstract | Publisher Full Text | Free Full Text

16. Dasgupta T, Wilson LD, Yu JB: **A retrospective review of 1349 cases of sebaceous carcinoma.** *Cancer.* 2009; **115**(1): 158–165.
PubMed Abstract | Publisher Full Text

17. Md Nasir ND, Koh VCY, Cree IA, *et al.*: **Phyllodes Tumour Evidence Gaps Mapped from the 5th edition of the WHO Classification of Tumours of the breast.** *Histopathology.* 2023; **82**(5): 704–712.
PubMed Abstract | Publisher Full Text

18. Giesen C, Goldman-Levy G, Cree IA, *et al.*: **WHO Classification of Tumours: moving towards the sixth edition.** *Histopathology.* 2025; **86**(6): 839–844.
PubMed Abstract | Publisher Full Text

Open Peer Review

Current Peer Review Status: ✔

Version 1

Reviewer Report 10 February 2026

https://doi.org/10.21956/openreseurope.22591.r69132

© **2026 Crispino A.** This is an open access peer review report distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

✔ **Angela Crispino**
University of Naples Federico II, Napoli, Italy

This article describes in a clear and understandable way how literature search strategies can be developed to support cancer classification. The methods are well explained and easy to follow, and the examples of sebaceous carcinoma and cholangiocarcinoma are apt to illustrate the complexity of tumor classification and terminology. The tables are particularly clear and help understand how the search strategies were constructed. Overall, this is a well-written and useful methodological article that achieves its objectives.

Is the work clearly and accurately presented and does it cite the current literature?
Yes

Is the study design appropriate and does the work have academic merit?
Yes

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

If applicable, is the statistical analysis and its interpretation appropriate?
Not applicable

Are all the source data underlying the results available to ensure full reproducibility?
No source data required

Are the conclusions drawn adequately supported by the results?
Yes

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: Digital pathology and cancer research, with a focus on tumour classification

and methodological aspects of literature analysis.

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.