Mod Pathol 37 (2024) 100357

MODERN PATHOLOGY

USCAP UNITED STATES AND CANADIAN ACADEMY OF PATHOLOGY
Creating a Better Pathologist

Journal homepage: https://modernpathology.org/

Research Article

A New Hierarchy of Research Evidence for Tumor Pathology: A Delphi Study to Define Levels of Evidence in Tumor Pathology

Richard Colling[a,b,*], Iciar Indave[c], Javier del Aguila[c], Ramon Cierco Jimenez[c], Fiona Campbell[d], Magdalena Chechlińska[e], Magdalena Kowalewska[f], Stefan Holdenrieder[g], Inga Trulson[g], Karolina Worf[g], Marina Pollán[h,i], Elena Plans-Beriso[h], Beatriz Pérez-Gómez[h,i], Oana Craciun[h], Ester García-Ovejero[h], Irmina Maria Michałek[j], Kateryna Maslova[e], Grzegorz Rymkiewicz[j], Joanna Didkowska[k], Puay Hoon Tan[l], Nur Diyana Md Nasir[l], Nickolas Myles[c], Gabrielle Goldman-Lévy[c], Dilani Lokuhetty[c], Ian A. Cree[c]

[a] *Nuffield Department of Surgical Sciences, University of Oxford, John Radcliffe Hospital, Oxford, UK;* [b] *Department of Cellular Pathology, Oxford University Hospitals NHS Foundation Trust, Oxford, UK;* [c] *International Agency for Research on Cancer (IARC), World Health Organization, Lyon, France;* [d] *Population Health Sciences Institute, Newcastle University, Newcastle Upon Tyne, UK;* [e] *Department of Cancer Biology, Maria Sklodowska-Curie National Research Institute of Oncology (MSCI), Warsaw, Poland;* [f] *Department of Molecular and Translational Oncology, Maria Sklodowska-Curie National Research Institute of Oncology (MSCI), Warsaw, Poland;* [g] *Institute of Laboratory Medicine, German Heart Centre Munich (DHM), Munich, Germany;* [h] *National Center for Epidemiology, Instituto de Salud Carlos III (ISC III), Madrid, Spain;* [i] *Consortium for Biomedical Research in Epidemiology and Public Health (CIBERESP), Madrid, Spain;* [j] *Department of Cancer Pathology, Maria Sklodowska-Curie National Research Institute of Oncology (MSCI), Warsaw, Poland;* [k] *Polish National Cancer Registry, Department of Epidemiology and Cancer Prevention, Maria Sklodowska-Curie National Research Institute of Oncology (MSCI), Warsaw, Poland;* [l] *Luma Medical Centre, Singapore*

ARTICLE INFO

Article history:
Received 4 August 2023
Revised 3 October 2023
Accepted 15 October 2023
Available online 20 October 2023

Keywords:
tumor
pathology
hierarchy
histopathology
evidence based
classification

ABSTRACT

The hierarchy of evidence is a fundamental concept in evidence-based medicine, but existing models can be challenging to apply in laboratory-based health care disciplines, such as pathology, where the types of evidence and contexts are significantly different from interventional medicine. This project aimed to define a comprehensive and complementary framework of new levels of evidence for evaluating research in tumor pathology—introducing a novel Hierarchy of Research Evidence for Tumor Pathology collaboratively designed by pathologists with help from epidemiologists, public health professionals, oncologists, and scientists, specifically tailored for use by pathologists—and to aid in the production of the World Health Organization Classification of Tumors (WCT) evidence gap maps. To achieve this, we adopted a modified Delphi approach, encompassing iterative online surveys, expert oversight, and external peer review, to establish the criteria for evidence in tumor pathology, determine the optimal structure for the new hierarchy, and ascertain the levels of confidence for each type of evidence. Over a span of 4 months and 3 survey rounds, we collected 1104 survey responses, culminating in a 3-day hybrid meeting in 2023, where a new hierarchy was unanimously agreed upon. The hierarchy is organized into 5 research theme groupings closely aligned with the subheadings of the WCT, and it consists of 5 levels of evidence—level P1

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.

* Corresponding author.
E-mail address: richard.colling@nds.ox.ac.uk (R. Colling).

https://doi.org/10.1016/j.modpat.2023.100357

representing evidence types that merit the greatest level of confidence and level P5 reflecting the greatest risk of bias. For the first time, an international collaboration of pathology experts, supported by the International Agency for Research on Cancer, has successfully united to establish a standardized approach for evaluating evidence in tumor pathology. We intend to implement this novel Hierarchy of Research Evidence for Tumor Pathology to map the available evidence, thereby enriching and informing the WCT effectively.

Introduction

Evidence-based medicine (EBM) aims to guide clinical decision-making by integrating the best available evidence with clinical expertise and patient values.[1] Evidence-based decision-making is fundamental in modern health care, ensuring that clinical practices are guided by reliable evidence to achieve the best possible outcomes for patients. In tumor pathology, where accurate diagnosis and precise classification play a pivotal role in determining treatment strategies, the need for evidence-based principles cannot be overstated. Part of the EBM process is ranking evidence in a hierarchy. The typical pyramid of evidence is a well-known concept; however, applying this paradigm to diagnostic specialties, especially in laboratory medicine, can be challenging because of the diverse types of evidence and unique contexts involved in pathology practice. As a result, the adoption of evidence-based principles in pathology has been slow, and there is a lack of consensus regarding what constitutes good evidence in the field.[2,3]

To attempt to boost the use of EBM in tumor pathology, several international collaborative forums have been established by the World Health Organization's International Agency for Research on Cancer (IARC) since 2019.[4] One such forum is the International Collaboration for Cancer Classification and Research, which aims to promote evidence-based practice and standards for cancer classification and research. Part of this is the plan to systematically find and rank all published evidence related to the World Health Organization Classification of Tumors (WCT).[5-7] The WCT serves as a fundamental basis for cancer classification, forming the foundation for all research and clinical management in oncology globally.[8] The project titled "Mapping the Evidence for the World Health Organization Classification of Tumors: A Living Evidence Gap Map by Tumor Type" (WCT EVI MAP)[9] aims to produce evidence gap maps, visually highlighting and ranking existing evidence to identify pockets of low-level or absent data in the evidence base.[5,6,9-11] We realised that to effectively rank evidence, a consensus on what constitutes "good" evidence in tumor pathology was needed. Thus, the decision was made to develop a new Hierarchy of Research Evidence for Tumor Pathology (HETP), by which we mean a hierarchy of the research evidence to support the classification.[2] In accordance with other similar hierarchies of evidence, the idea was to focus on the study design, which is inherently related to the risk of bias, to develop a hierarchy that would not require individual study appraisal.

A robust HETP tailored to histopathology is paramount to address the unique challenges and diverse evidence types encountered in this discipline. A new hierarchy could have the potential to significantly impact patient care, research advancements, and resource allocation in oncology globally. This study aimed to reach a consensus among experts in the field of tumor pathology, with input from those working in EBM, to develop this new HETP. The goal was to create a complementary, adaptable hierarchy that reflects existing models but is specifically tailored to the unique context of tumor pathology. In this study, we outline how the process was undertaken, including the results of a Delphi study and the evolving structure, to present the final HETP, which we intend to use in producing the WCT evidence gap maps.

Materials and Methods

Approach

The study aimed to establish a ranking order for various types of evidence cited in the WCT and develop an appropriate HETP in the context of tumor pathology. The objective was to rank evidence, such as individual published studies, based on their quality, indicating the inherent risk of bias derived from the study's design. The most robust evidence, with lower risk of bias, would be ranked higher than less robust evidence, following the model used in EBM hierarchies.[12]

To achieve this, our approach involved collaborating with a core group of experts in the field and actively seeking input from a wide range of allied experts. We tackled the fundamental question through a series of focused steps:

1. Determining whether an overarching hierarchy was sufficient or if separate subhierarchies were needed based on different research and clinical areas.
2. Identifying appropriate groupings for the subhierarchies, if necessary.
3. Establishing the number of levels that the hierarchy should comprise.
4. Defining which types of publication/work should be included in the hierarchy, clarifying what constitutes "evidence," and identifying exclusions.
5. Determining how different studies should be ranked together and at what level.

We deliberately excluded interventions/drugs from the new hierarchy. We recognize that these aspects are well addressed in existing frameworks and not immediately relevant for the WCT.

Study Design

We used an adapted form of the e-Delphi method in this study. The Delphi method is a structured approach to achieve a consensus among a group of experts through iterative rounds of assessments. In each round, experts review previous results and reassess their views based on the group's collective feedback. The process is illustrated in Figure 1. This process continues until consensus is achieved, minimizing the influence of a few

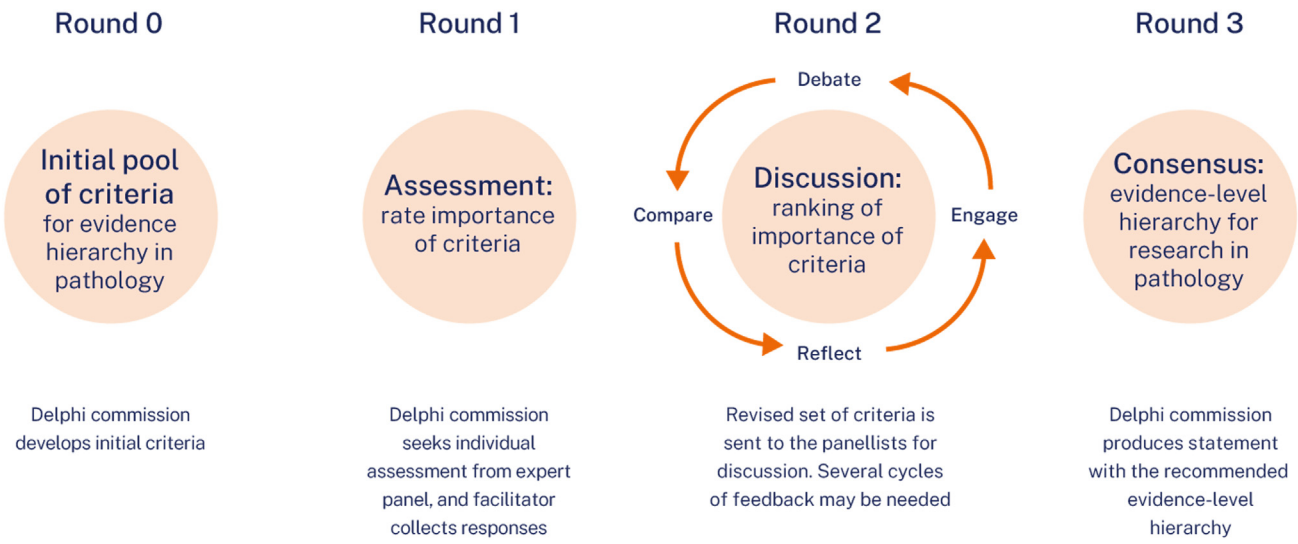

Figure 1.
An overview of the original plan for the rounds within the Delphi part of this work.

dominant views.[13,14] Our study closely followed the Delphi method with some modifications outlined below.

Participants

The experts invited to participate were authors and/or editors of the WCT, representing a multidisciplinary group of ~2000 pathologists (the vast majority), oncologists, surgeons, scientists, public health specialists, and other allied health professionals. Members of the WCT EVI MAP team were excluded from taking part in the surveys. All Delphi rounds were conducted anonymously, ensuring no collection of personal or identifying data. Participants were presented with a participant information sheet and explicitly asked for consent to participate. The same cohort of experts were invited to participate in all 3 rounds. Ethics approval was secured.

Rounds

Three rounds of surveys were conducted in English using an online format hosted on the Jisc platform. Experts received an invitation email from the IARC containing a survey link. The surveys remained active for 2 weeks following the invitation or were extended if the target sample size had not been met. Each survey started with an introduction and an outline of the project's scope, provided the full participant information sheet, collected consent, and gathered demographic data. Similar information was made available online for later or separate viewing if desired. Non-consenting participants were unable to proceed with the survey at that round but could participate later. Participants were not mandated to take part in future rounds, and as surveys were anonymous, it was not possible to know if the same participants took part in subsequent rounds. The completion of the 3 rounds provided sufficient data to inform the subsequent 3-day meeting.

WCT EVI MAP Team Meeting

Following the 3 Delphi rounds, the WCT EVI MAP team convened a 3-day hybrid meeting in Lyon, France, in April 2023. The meeting attendees included the Steering Group (see below) and the 12 Advisory Board members (see below). All discussions were live streamed using the Zoom platform, and online attendees participated alongside in-person guests. The Advisory Board members joined 2 sessions, on days 1 and 2, to discuss the Delphi results and finalize the new hierarchy. Chaired by the outgoing head of the WCT (I.C.), the meeting also included the incoming head for the sixth edition (D.L.).

Sample Size

Each survey aimed to achieve a sample size of 300 responses, ~15% of the expert population, which was felt an acceptable and representative response rate.

Consensus

The objective was to achieve agreement of 75% on any survey question at each round. Questions without consensus were reposed to experts in the next round, considering peer responses.

Oversight

A study Steering Group was set up, chaired by the chief editor of the WCT (I.C.). Oxford (R.C.) led on running the study, drafted the surveys and complete initial analyses on the results. The Steering Group finalised survey questions, reviewed, and analyzed the data after each round and made final executive decisions on the hierarchy's structure. The final hierarchy was agreed upon by a central panel, including the Steering Group and the wider WCT EVI MAP project team, during the 3-day hybrid meeting in Lyon, France. The external WCT EVI MAP Advisory Board (below) provided input and feedback midway through this meeting. Thus, the final hierarchy represents a consensus between the WCT authors, the WCT EVI MAP expert project group, and input from external advisors and stakeholders.

Advisory Board

The Advisory Board consisted of international experts from the Editorial Board members of the WCT and International

WCT topics:	Etiology		Clinical features, Radiology, Localisation, Macroscopy, Histopathology, Immunohistochemistry, Cytology	Diagnostic molecular pathology Diagnostic immunohistochemistry	Prognosis, Staging	Prediction
	Epidemiology, Prevalence, Incidence, Risk factors	Mechanisms & Pathogenesis				
Research focus:	Population	Cellular, molecular & genetic biology	Pathological characterisation and diagnosis of the tumor	Diagnostic tests (accuracy) and reproducibility (precision)	Prognostics	Predictive biomarkers
Example questions:	How common is this tumour? What factors influence the probability of developing this tumour? Who is at risk?	What are the molecular alterations in this tumor? What mechanisms cause this tumor?	What are the key features of this tumor? How to we define the tumor?	How well does this test confirm the diagnosis?	What features influence five-year survival?	Can this marker predict treatment?
Level P1	*Systematic reviews**	*Systematic reviews**	*Systematic reviews**	*Systematic reviews**	*Systematic reviews**	*Systematic reviews**
Level P2	*Population-based descriptive studies* *Prospective cohort studies* *Randomised-controlled trials*	*Prospective cohort studies* *Studies derived from randomised-controlled trials*	*Case-control studies* *Diagnostic agreement/reproducibility studies* *Diagnostic test accuracy studies* *Prospective cohort studies* *Studies derived from randomised-controlled trials*	*Case-control studies* *Diagnostic agreement/reproducibility studies* *Diagnostic test accuracy studies* *Prospective cohort studies* *Randomised-controlled trials (of diagnostic tests)* *Studies derived from randomised-controlled trials*	*Prospective cohort studies* *Studies derived from randomised-controlled trials*	*Prospective cohort studies* *Randomised-controlled trials* *Studies derived from randomised-controlled trials*
Level P3	*Case-control studies* *Retrospective cohort studies* *Other cross-sectional studies*	*Case-control studies* *Mechanistic clinical studies* *Other cross-sectional studies* *Retrospective cohort studies*	*Consensus studies* *Other cross-sectional studies* *Retrospective cohort studies*	*Case series* *Consensus studies* *Other cross-sectional studies* *Retrospective cohort studies*	*Case-control studies* *Observational trials* *Other cross-sectional studies* *Retrospective cohort studies*	*Case-control studies* *Retrospective cohort studies*
Level P4	*Case series*	*Animal studies* *Case series* *Mechanistic laboratory studies* *Molecular database studies* *Observational trials*	*Case series*	*Clinical laboratory test validation studies*	*Diagnostic test accuracy studies* *Population-based descriptive studies*	*Diagnostic test accuracy studies*
Level P5	*Case reports*	*Case reports*	*Case reports*	*Not used*	*Case reports* *Case series*	*Case reports* *Case series* *Mechanistic clinical studies*

Figure 2.
The newly proposed Hierarchy of Research Evidence for Tumor Pathology. The hierarchy is arranged at the top level to map to the various sections of the World Health Organization Classification of Tumors, that is, matching chapter headings/subheadings as far as possible to cover all topics discussed in the series. The various headings are grouped into similar research topics of interest, such as tumor characteristics, prognostics, and so forth. Example subject-based questions that clinicians may ask are given below. The levels presented below are in order of robustness, that is, level P1 having the greatest confidence and level P5 the lowest. Each P level for each subsection contains the list of evidence types that rank at that level (in alphabetical order, no intralevel ranking implied). Most types of evidence can be thought of as "studies" but not all. Studies of rare tumors (defined here as those with an incidence of <1/million population/annum) are automatically upgraded 1 level. This would include case reports that may be the only unique diagnosis of its type in the literature. Large sample size level P2 studies should be upgraded to level P1. This table should be used in conjunction with the glossary (Supplementary File 2) as the terms used may be different from those that some readers may be used to. *Systematic reviews predominantly including level P2 studies. Systematic reviews of predominantly level P3-P5 studies are placed in level P2. Rapid reviews are placed in level P2.

Collaboration for Cancer Classification and Research. A total of 12 members were invited and accepted the invitation. They represented 10 different countries (1 from Africa, 1 from South America, 4 from Asia-Pacific, 2 from North America, and 4 from Europe) and had professional backgrounds in histopathology (8 members), epidemiology (3 members), and oncology (1 member).

Results

At each Delphi study round, an invitation was sent to the IARC emailing list for the WCT series authorship group. The demographics of the participants at each round are included in Supplementary File 1. The Delphi study started using an extensive list of different terms for various types of potential evidence, study designs, or publication types. This attempt aimed to avoid bias or missing any types of work from the hierarchy. However, it quickly became apparent that this approach highlighted different understandings of the terms across respondents. Therefore, in round 2, we started to develop a glossary of terms and combined some similar terms under an umbrella. The terminology used for the evidence types evolved at each round and at the 3-day meeting into the final agreed format (Supplementary File 2).

Delphi Round 1

The first survey invitations were distributed on November 3, 2022, and the survey closed on November 20, 2022. The WCT authors (n = 2070) were invited to take part, and 464 completed the survey, yielding a response rate of ~22%. In this round, the focus was on reaching consensus on the basic hierarchy structure and on determining the types of studies/research publication that should be included in the hierarchy as evidence. Most respondents (75%) felt that just more than one overarching hierarchy was needed, leading the Steering Group to create a hierarchy with subhierarchies (columns), each focusing on a different topic area based around the chapter headings/subheadings or other sections of the WCT series (the structure of this is evident in Fig. 2). The WCT headings were used as far as possible, although it was not

Table
Evaluation of various evidence types during Delphi rounds

Round	Include in hierarchy (counts as evidence)	Exclude from hierarchy (does not count as evidence)
1	Case-control studies	Conference abstract
	Case series	Conference posters
	Consensus studies	Editorials
	Cross-sectional/observational studies	Narrative/nonsystematic reviews
	Cohort studies	Opinion pieces
	Diagnostic test studies	Oral platform abstracts
	Evidence synthesis papers	Patent descriptors
	Laboratory in vitro studies	Personal/author/expert opinion written in text
	Randomized controlled trials	Study protocols
	Systematic reviews	"Unpublished data" as quote
2	Cancer registry data	Clinical trial protocols
	Clinical laboratory test studies	Patient interview studies
	Genome database entries	Preprint versions of papers
		Surgical technique papers
3	Animal studies	Commentaries
	Case reports	Guidelines[a]
	Molecular biology databases	Letters
	Noncontrolled or nonrandomized trials	Surveys

The types of work considered in this study and the decisions in the subsequent Delphi rounds regarding their inclusion or exclusion from the evidence hierarchy. The terminology of the listed types of work evolved during the Delphi study, and the Steering Group added and merged some categories at various points in the process to deal with areas where agreement could not be reached. Some terms change later in the study, see the glossary (Supplementary File 2) for full and final definitions.
[a] Unless systematically developed, then counted as a systematic review.

always possible to match these exactly. The participants were presented with a list of 31 study types or publications (potential "evidence types") complied by the Steering Group. The participants were then invited to give their opinion on if each of the types of work should be included in the hierarchy (ie, counted as evidence for the purposes of the HETP). The Steering Group list aimed to be exhaustive and include different terminologies for the same types of work to avoid missing any potential evidence types—however, the participants were also invited to make their own suggestions of potential evidence types, which could in turn be voted on in subsequent rounds. In this round, consensus was reached on 10 types of work that should be included in the new hierarchy and 10 types of work that should be excluded (Table, round 1). The remaining types of work where agreement was not reached were rolled forward to round 2 for reevaluation. Table outlines all the types of work (from the Steering Group and participant suggestions, 35 in total) that were evaluated in the Delphi and indicates at what stage consensus was reached to include or exclude them.

Delphi Round 2

Invitations were sent on December 19, 2022, and the survey closed on January 8, 2023. Of 2020 email invitations (some of the round 1 invitees replied to the email and asked not to be invited again), 353 completed the survey, resulting in a response rate of 17%. In this round, participants reappraised their opinions on nonconsensus types of work from round 1 in light of peer responses, leading to consensus to include 3 types of work and exclude 4 types (Table, round 2). Round 2 also allowed participants to rank where each type of evidence should sit within the hierarchy. Participants were presented with a blank hierarchy structure (the same outline as seen in Supplementary File 2 but lacking the study types) and asked to place each study type into a position within the hierarchy. The most frequent choice was chosen. Where agreement was not reached, the study type was rolled over to the next round. These initial rankings are shown in black text in Supplementary File 3. It was during round 2 that the idea that studies of rare disease should be given a higher ranking was conceived by the Steering Group as large sample sizes in these studies are virtually impossible and so tend to be case report/series in nature.

Delphi Round 3

Invitations were sent to 2013 authors (again, some asked not to be invited again or some had left the mailing list) on February 2, 2023, and 406 had completed the survey by February 26, 2023, giving a response rate of 20%. In this round, the experts reappraised their opinions on the remaining nonconsensus types of work, resulting in consensus to include 4 types of work and exclude 4 types (Table, round 3). Round 3 also presented the initial rankings from round 2 and gave the opportunity to revise this, but the consensus was that no revisions were needed. Participants then had the opportunity to rank or exclude the final remaining types of work. The final rankings from this round are shown in red text in Supplementary File 3.

WCT EVI MAP Meeting

By the end of the 3-day hybrid meeting, the Steering Group reached a consensus for a new HETP (Fig. 2) and a glossary of evidence types (Supplementary File 2). The meeting involved an in-depth discussion on each subhierarchy (column) with all members of the WCT EVI MAP project team present. Each study type was discussed, and a final position in the hierarchy was decided, taking into account other existing hierarchies, such as the Oxford Centre for Evidence-Based Medicine Levels of Evidence.[12] The aim was to complement existing hierarchies while adapting to the unique needs of tumor pathology. For instance, evidence on tumor etiology and pathogenesis included animal studies and other mechanistic laboratory studies, acknowledging the

challenges of conducting direct experimental human research because of ethical considerations.

The meeting was multidisciplinary, bringing together epidemiologists, pathologists, molecular biologists, clinicians, and laboratory medicine and public health specialists. One point that became evident during the meeting was the need to distinguish between sections related to the term "etiology" in the WCT, a term which has different meanings for some members of the group. This led to the distinction between etiology and risk factors (investigated through epidemiologic population studies) and etiology and pathogenesis (studies of the cause of disease). The hierarchy structure reflects this distinction (Fig. 2).

The Advisory Board of external experts provided feedback, and some amendments were made based on their input. Following the meeting, final amendments and the glossary were decided via email. The final version of the hierarchy is shown in Figure 2. For the glossary, see Supplementary File 2.

Discussion

The concept of "levels" or a "hierarchy" of evidence has been present since the early days of the EBM movement. Early attempts in the late 1970s from Canada ranked evidence, considering randomized controlled trials as level 1, followed by cohort and case-control studies at level 2, and expert opinions at the lowest level 3.[15] Many others have followed suit, with more complex hierarchies existing. However, research in pathology and tumor classification has primarily focused on different techniques used in the laboratory to assess various aspects relevant to the diagnosis and definition of tumor types.[12,15] Some attempts have been made to classify pathology evidence, such as the one used by the UK Royal College of Pathologists.[16] However, these efforts have not fully explored the views of pathologists regarding the applicability of such hierarchies in both clinical diagnostic and research pathology and are somewhat limited in scope. Furthermore, many of these existing hierarchies are based on frameworks developed by patient-facing groups of clinicians and lack input from pathologists. They often oversimplify the highly varied types of evidence available in tumor pathology. In our experience during the WCT EVI MAP project, we found it challenging to translate existing hierarchies to accommodate the wide and complex range of evidence cited in the WCT as most hierarchies lacked the necessary clarity and detail concerning the types of study design included.

Our approach to developing a new HETP used a modified Delphi technique, allowing us to gather the perspectives of a wide-ranging group of clinicians and scientists in tumor pathology. These views informed a core group of experts forming the Steering Group, who played a crucial role in shaping the final hierarchy. At all stages, every effort was made to follow the consensus views of the WCT authors, but with some executive decision being made by the Steering Group. The outcome of the Delphi exercise and the Steering Group's discussion led to the development of a comprehensive evidence hierarchy applicable (we believe) to almost any published tumor pathology work.

An early decision (based on consensus in round 1) in the study was to create a hierarchy with separate sections focusing on discrete topic areas rather than a unified scale that could be applied in all circumstances. A unified scale may appear easier to apply in practice and, for most busy pathologists, this would look more straightforward. However, a problem with a unified hierarchy is that it loses resolution and can consequently become more difficult to apply in practice. When piloting the mapping exercise mentioned in the introduction (WCT EVI MAP), the authors found that applying a more simplified hierarchy was challenging and by default ranked pathology papers lower than we felt was fair. This is what prompted work to develop a more nuanced HETP. In addition, by trying to consider every possible type of evidence, peer-reviewed and published or unpublished, we attempted a consensus view for each, resulting in a more inclusive and sophisticated classification system. Inevitably, this results in a more complex system that might not always be the best to apply in all circumstances; however, the aim was not to replace, but to complement, existing hierarchies and offer an alternative where preferred.

The multidisciplinary nature of our meeting, which brought together epidemiologists, pathologists, molecular biologists, clinicians, and laboratory medicine and public health specialists, ensured that a broad range of perspectives enriched the discussions. This of course brought challenges. Notably, in addressing the term etiology, we recognized the need to distinguish between epidemiologic studies exploring risk factors and more mechanistic studies investigating the cause and pathogenesis of diseases. This differentiation allowed for a more nuanced approach to the hierarchy, accounting for the unique considerations of each domain. This is reflected in the HETP design and in some ways is a compromise between what pathologists and epidemiologists may think of as etiology.

The final results of the Delphi, modified by the Steering Group, culminated in identifying a consensus on what counts as evidence in tumor pathology and how these should be ranked depending on the question being asked. Up to 5 levels were needed to cover the ranking preferences, and these are named "P" levels (for "pathology") to distinguish them from other clinical hierarchies that use levels. Similar to most other hierarchies, the focus primarily is on the study design. Inherently, this is based on the idea that some studies have a greater risk of bias than others - this was the guiding (but not only) principle on which we designed the HETP. The intention was to generate a tool, which gives a high-level perspective on a group of studies published on a topic, rather than attempting to individually appraise studies. However, the Steering Group recognized that in some situations, a minor degree of appraisal is helpful. Unlike other hierarchies, some allowance is made in specific circumstances. For example, studies on rare diseases are upgraded by a level, or large level P2 studies are upgraded to level P1). We have also produced an explanatory glossary, which defines the types of evidence in the hierarchy, something unique in these types of hierarchy publications. The way the hierarchy is presented allows one to search for evidence on a particular clinical question and then identify where each of those studies sits within the hierarchy. One may choose to focus on the highest-level study found for making clinical decisions or weigh up the overall quality of the body of evidence on a particular topic. Researchers may also use this process to identify what work is needed to improve the evidence base. It would be up to the individual to decide what to do with the information from the hierarchy. For example, a working pathologist might use the information just to find the best evidence to make a clinical decision. However, an author of a national guideline may need to demonstrate how the evidence supporting the guideline was evaluated, with which hierarchy, and state in the guideline what the level of evidence is.

The hierarchy is not designed to take an individual study and place it directly into a particular level. The intent is that the user starts with a clinical question or topic. Although some of the levels for different topics do contain the same study types, there are several study types listed at different levels depending on the context. This is because, for example, diagnostic test accuracy

studies are clearly very important when trying to evaluate if a particular antibody stain can diagnose a certain tumor subtype, but less important (although not totally irrelevant) for determining if that antibody can predict outcome, but entirely irrelevant when assessing the prevalence of a tumor. When one starts with a clinical question and finds the evidence, it is easy to determine where to place it within the hierarchy. When one starts with just a single study, it is not immediately clear where it sits in the hierarchy. Of course, this could become clear when reading the paper, where the clinical question may become obvious, but this is not always the case. An individual study may answer more than one question, and the same study could be placed at different levels depending on the context in which it is being evaluated. Therefore, the HETP should be used in the context of what the evidence is attempting to inform rather than to just evaluate various *ad hoc* publications.

One key area of potential confusion that is worth highlighting briefly is that of the terminology around "diagnostics." A diagnosis may be made based on the reference (or sometimes called "gold") standard definitive criteria (eg, a collective of descriptive clinical, imaging, and laboratory characteristics) or using a surrogate (proxy) to reference standard (a "diagnostic test"). In patient-facing settings, the diagnosis is rarely made using the ideal and definitive (often impractical and costly) reference standard in clinic, and requesting surrogate tests (blood, tissue, and radiology) to narrow down a clinical differential diagnosis is the routine approach. Consequently, much focus in EBM is on assessing the accuracy of these 'diagnostic tests', and so, this is often what most clinicians think about when we talk about "diagnostics". From a simplified EBM reasoning perspective, the histopathologic diagnosis can be viewed as such a diagnostic test (or a series of cell-based and tissue-based tests). However, for pathologists, this distinction is more arbitrary; most of the diagnostic work is a reference standard—tumors are diagnosed by examining the defining, characteristic features (eg, morphology and immunoprofile) that are described in the WCT. Sometimes, pathologists do use immunohistochemistry (IHC) or molecular assays in the true EBM sense of a diagnostic test, but the processes are blended in practice and thought of as a concept—the "final diagnosis." Nonetheless, studies trying to answer questions about defining tumor characteristics are different from those trying to assess the accuracy of a diagnostic test, and so it was important to reflect this distinction in the HETP. IHC is a subtle example of this problem, sometimes pathologists use it as part of the defining features of a tumor (being part of the reference standard), sometimes as a true add-on diagnostic test (eg, to determine the likely primary site of a metastasis), and sometimes this distinction is overlapping. This is reflected in the hierarchy with IHC being listed in a descriptive context and in a diagnostic test context.

A special explanation for studies using molecular technology is also needed. In the HETP, we distinguish between studies evaluating molecular diagnostic tests and studies focusing on the molecular pathogenesis of tumors. The latter category at a first glance might not obviously fall within one of the groups in the hierarchy or glossary. Although these are not epidemiologic studies, for most molecular investigations, the fundamental study design can still be classified as such (eg, following a cohort or case-control methodology). There are also many exploratory molecular pathology studies that fall under what we term "mechanistic" designs. Finally, there are studies of molecular alterations in just 1 patient or tumor, which can often be considered a case report.

The Delphi method is certainly not perfect, and although it is designed to allow experts to freely express their views,[6,7] the method is not without its limitations.[17] Experts' judgments can be influenced by personal and professional factors that can lead to bias. A large sample size should limit the effect of this, although that assumption is not universally accepted.[18] Although we set a reasonable threshold[13,14] of 75% agreement for survey questions, some may debate the level of consensus required. Nevertheless, the Delphi method is not supposed to be solely reliant on rounds of questions, and an expert oversight is expected. To ensure the robust decision-making, we maintained oversight through the Steering Group and sought feedback from an independent and external Advisory Board. We hope the modified Delphi approach we used in this study helped minimize any bias in the process.

A further limitation of the Delphi study is that although we achieved a target of 15% of the WCT authorship (which is probably quite high for a survey of mainly pathologists), it would have been desirable to have had a higher response rate of course. The demographics at each round showed a range of backgrounds that are likely representative of the WCT authorship (and indeed the readership), but the potential for some bias could not be eliminated. A final point to mention is that we cannot know the number of participants who took part in all 3 rounds because of the online study design to maintain anonymity, and so, we cannot be sure of the effect of some participants not being present at all 3 rounds. Some evidence does exist to suggest that separate groups do tend to agree with each other and that may provide some, although it limited, reassurance.[18]

A final point to highlight is that the HETP is intended to focus very much on tumor pathology. Although some aspects of the HETP could be generalized and applied in nontumor contexts, there may be many types of study that may not have been considered for inclusion, and some of the ranking may not be appropriate in that context. A future development could be to modify the HETP to include nontumor pathology publications.

Conclusion

In this study, we believe we have used a compromise of approaches to reach a consensus on how evidence should be ranked in tumor pathology. We have developed a new HETP that we believe can be universally applied across all areas of tumor pathology. By providing a robust and adaptable framework for evaluating evidence, our hierarchy aims to contribute to improved patient outcomes and the overall advancement of knowledge in the field of tumor pathology.

We strongly encourage others to adopt this comprehensive hierarchy, in conjunction with other complementary models, when evaluating evidence in systematic reviews and guidelines. Although we recognize that a Delphi methodology may not be perfect and our outcomes may not be indisputable, it is probably the best and most pragmatic approach available for attempting such an ambition task as this.

Moving forward, we plan to implement the HETP in the next phase of the WCT EVI MAP project, wherein we will systematically gather and map all available evidence to the chapters of the entire WCT series. This will begin with a pilot validation of the HETP. These evidence gap maps will play a pivotal role in evaluating the evidence for the upcoming sixth edition of the WCT series and will serve as a valuable resource for academics and funders in the field of tumor pathology to focus their research efforts in the future.

Acknowledgments

The authors thank the University of Oxford for sponsoring, hosting, and supporting this work; Laura Brispot for her significant contribution in coordinating this project at the International Agency for Research on Cancer and organizing the email invitations; and the International Agency for Research on Cancer for supporting and leading the project.

Author Contributions

The study was conceived by R.C., I.I., and I.C. The surveys were run and analyzed by R.C. with contributions from all authors. The manuscript was drafted by R.C., and all authors reviewed and approved the final version.

Data Availability

The data sets used and/or analyzed during the current study are available from the corresponding author on reasonable request.

Funding

The overall project, International Agency for Research on Cancer, and beneficiaries (German Heart Centre Munich, Maria Sklodowska-Curie National Research Institute of Oncology, and Instituto de Salud Carlos III) are funded by the European Commission (HORIZON grant no. 101057127). R.C. and F.C. are funded by UK Research and Innovation. S.H. has received research funding or honoraria from Roche, BMS, Merck, Sysmex, Thermo, Volition, Trillium, Medica, and Instand and is a founder of SFZ BioCoDE and CEBIO. P.H.T. has received honoraria from AstraZeneca.

Declaration of Competing Interest

The authors report no relevant conflicts of interest.

Ethics Approval and Consent to Participate

This study was conducted with approval from the University of Oxford Medical Sciences Interdivisional Research Ethics Committee (approval reference: R83130/RE001). All data were collected anonymously; however, consent to participate was obtained from each participant during each survey. Attendees at the World Health Organization Classification of Tumors EVI MAP team meeting in Lyon, as well as participation in the Steering Group and Advisory Board, were voluntary, and consent was implied. The study was performed in accordance with the principles outlined in the Declaration of Helsinki.

Supplementary Material

The online version contains supplementary material available at https://doi.org/10.1016/j.modpat.2023.100357

References

1. Sackett DL, Rosenberg WMC, Gray JAM, Haynes RB, Richardson WS. Evidence based medicine: what it is and what it isn't. *BMJ*. 1996;312:71–72.
2. Indave BI, Colling R, Campbell F, Tan PH, Cree IA. Evidence-levels in pathology for informing the WHO classification of tumours. *Histopathology*. 2022;81:420–425.
3. Cree IA, Indave BI. Commentary: cancer research quality and tumour classification. *Tumour Biol*. 2020;42:1010428320907544.
4. IARC. About IARC. Accessed October 2, 2023. https://www.iarc.who.int/cards_page/about-iarc/
5. 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:e061240.
6. 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;82:704–712.
7. Cree IA, Indave Ruiz BI, Zavadil J, et al. The International Collaboration for Cancer Classification and Research. *Int J Cancer*. 2021;148:560–571.
8. IARC. WHO classification of tumours online. Accessed October 2, 2023. https://tumourclassification.iarc.who.int/welcome/
9. IARC. WCT EVI MAP project overview. Accessed October 2, 2023. https://wct-evi-map.iarc.who.int/project-overview/
10. White H, Albers B, Gaarder M, et al. Guidance for producing a Campbell evidence and gap map. *Campbell Syst Rev*. 2020;16:e1125.
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;12:45.
12. OCEBM Levels of Evidence Working Group. The Oxford levels of evidence 2. Accessed October 2, 2023. https://www.cebm.ox.ac.uk/resources/levels-of-evidence/ocebm-levels-of-evidence
13. Hasson F, Keeney S, McKenna H. Research guidelines for the Delphi survey technique. *J Adv Nurs*. 2000;32:1008–1015.
14. Nair R, Aggarwal R, Khanna D. Methods of formal consensus in classification/diagnostic criteria and guideline development. *Semin Arthritis Rheum*. 2011;41:95–105.
15. Burns PB, Rohrich RJ, Chung KC. The levels of evidence and their role in evidence-based medicine. *Plast Reconstr Surg*. 2011;128:305–310.
16. Marchevsky AM, Wick MR. Evidence levels for publications in pathology and laboratory medicine. *Am J Clin Pathol*. 2010;133:366–367.
17. Barrett D, Heale R. What are Delphi studies? *Evid Based Nurs*. 2020;23:68–69.
18. Hasson F, Keeney S. Enhancing rigour in the Delphi technique research. *Technol Forecast Soc Change*. 2011;78:1695–1704.