REVIEW

Developing consensus in Histopathology: the role of the Delphi method

Dilek Taze,[1,2,3] Collette Hartley,[2,4] Ann W Morgan,[1,2,3,4] Aruna Chakrabarty,[1,3] Sarah L Mackie[3,5] & Kathryn J Griffin[1,2,3]

[1]St James' University Hospital NHS Trust, Leeds Institute [2]of Cardiovascular and Metabolic Medicine, School of Medicine and [5]of Rheumatic and Musculoskeletal Medicine, University of Leeds, NIHR Leeds [3]Biomedical Research Centre, and [4]Medtech and In Vitro Diagnostics Co-operative, Leeds Teaching Hospitals NHS Trust, Leeds, UK

Developing consensus in Histopathology: the role of the Delphi method

The Delphi method is a well-established research tool, used for consensus building across a number of fields. Despite its widespread use, and popularity in many medical specialities, there is a paucity of literature on the use of the Delphi method in Histopathology. This literature review seeks to critique the Delphi methodology and explore its potential applications to histopathology-based clinical and research questions. We review those published studies that have utilized the Delphi methodology in Histopathology settings and specifically outline the advantages and limitations of this technique, highlighting situations where its application can be most effective.

Keywords: consensus, Delphi, histopathology

Introduction

The Delphi method was first developed in the 1950s and 60s as a military forecasting method to predict the effect of emerging technologies on warfare.[1] It has since evolved to be a valuable consensus-building tool in a varied array of disciplines from business to healthcare. This, now widely utilized, methodology has the sole purpose of generating consensus using an expert panel in situations where robust evidence is lacking, limited, or contradictory. Despite its acceptance in other medical disciplines, the use of the Delphi methodology in Histopathology has not previously been examined in the literature, and this review seeks to explore and critique the various facets of the Delphi method as they relate to Histopathology.

Address for correspondence: Dr. Kathryn Griffin, Leeds Institute of Cardiovascular and Metabolic Medicine, School of Medicine, University of Leeds, Leeds, UK. e-mail: k.j.griffin@leeds.ac.uk

Materials and methods

Pubmed and Google Scholar databases were searched using Boolean combinations of the search terms "Histopathology", "Delphi" and "consensus". The search was limited to a date range from 2015 to the present, resulting in a combined total of 1659 articles. The title and/or abstract of each of these articles were screened for relevance and the following inclusion criteria applied; papers available in full text format, written in English, and with use of the Delphi methodology solely in a histopathology setting. Fifty articles met the inclusion criteria and the references of these were searched, identifying 4 additional relevant papers. After initial assessment, papers using a "hybrid" consensus methodology (such as Delphi combined with a nominal group technique) were excluded to ensure that the review focussed exclusively on the Delphi methodology as described by Jones and Hunter.[2] Papers were also excluded if the specific number of participating pathologists in a

multidisciplinary Delphi was not stated, if the pathology contribution was undefined or unclear, or if the study related primarily to Histopathology education. Seventeen studies were appraised in detail; no Cochrane review articles were identified.

Outline of the Delphi method

The Delphi method uses an expert panel to generate consensus when empirical evidence is lacking or limited. There are four main principles which underpin the Delphi method, these are; the use of an expert panel, the anonymity of participants, utilizing multiple rounds of questionnaires to assess participant opinion, and finally the provision of "feedback" to participants between each subsequent round.[3]

A typical Delphi process starts with review of the current literature to identify gaps in knowledge or understanding, followed by the recruitment of a panel of participants deemed as "expert" in the field; that is professionals who are knowledgeable or experienced in the area of study.[4] The researchers design a structured questionnaire comprising concise "statements" which are derived from the literature, from best practice or which pertain to areas of uncertainty. This is sent to the expert panel members who are required to anonymously rate their level of agreement of each statement using, most commonly, a Likert-type scale.[5] The individual responses are then collated and analysed by the researchers to determine which statements achieved consensus. The participants receive feedback detailing their own score as well as the position of the whole group. This provides an opportunity for participants to reconsider their ratings, especially if they form a minority opinion. Statements that did not reach consensus are clarified and/or reworded, and a revised questionnaire is then circulated. This process of rounds of surveying and feedback is repeated until complete consensus is reached (rarely), until the predetermined number of rounds is completed or, until no additional statements gain consensus. Fundamentally, there should be no communication between the panel members at any point during this process.

Delphi methods can be adapted to incorporate in-person, teleconference or online meetings to allow for structured debate and discussion amongst the panel members in between survey rounds. These are typically referred to as "modified Delphi" methods because they differ from that originally described, but crucially they still maintain the anonymity of participants with regards to their individual responses.[4]

Table 1 summarizes the key considerations for optimal Delphi design and highlights some of our recommendations to researchers.

The Delphi in histopathology

The Delphi method has been used in Histopathology for achieving consensus on a variety of topics ranging from determining best practice for diagnostic processes,[6,7] to the formulation of standardized reporting criteria.[8–12] The expert panel utilized in these studies also varies, with some composed exclusively of histopathologists and others made up of multidisciplinary consensus groups of which histopathologists are just one component. Whilst the number of survey rounds is generally similar in the studies explored in this paper (usually 2–4 survey rounds are performed) the panel size varies and demonstrates geographical diversity. The characteristics of these studies are summarized in Table 2.

Defining terminology and nomenclature

Certain histological features can provide information on prognosis, guide treatment strategies or influence response to therapy. Inconsistencies in nomenclature can create confusion for clinicians interpreting pathology reports, especially if unfamiliar words are used interchangeably by different pathologists. Standardizing terminology and establishing reproducible definitions is therefore of paramount importance, not only for day-to-day practice, but also for predicting the biological behaviour of a lesion on the basis of its morphology.

Carr *et al.*[13] used a modified Delphi method to reach consensus on terminology which may be misleading or confusing in the diagnosis of pseudomyxoma peritonei and associated appendiceal neoplasia. This was partly achieved by identifying synonyms or histological terms which were being used interchangeably in the pathological diagnosis of this condition. The panel aimed to reach consensus and provide guidance on the standardized use of these terminologies. As an example, the participants agreed that the term "goblet cell tumour" should be introduced as the preferred term for the synonym "goblet cell carcinoid" to prevent confusion with the alternative diagnosis of an appendiceal carcinoid or neuroendocrine tumour.

Similarly, Mariette *et al.*[14] reported marked inconsistencies in the finding of a relationship between signet-ring-cell type histology in gastric carcinoma

Table 1. Considerations and recommendation for the optimal design of a Delphi study based upon the findings from this literature review

Aspect of the Delphi	Considerations and recommendations
Review of the literature	Identify the specific areas where empirical evidence is lacking or limited.
Recruitment of the expert participants	Determine the definition of 'expert' for the subject of interest. Will the participants comprise histopathologists only or a multidisciplinary approach? Have factors for increasing diversity been considered i.e., differing levels of experience, geographical variation, specialist vs generalist?
Type of Delphi	Traditional (strictly no communication between participants) or Modified (some form of facilitated communication between participants). If a Modified Delphi is selected, how will factors of anonymity and dominance be accounted for?
Formulation of statements	There are no clear evidence-based guidelines on how to formulate statements. We recommend that each statement is concise and assesses one piece of information only.
Number of survey rounds	This can be predetermined or repeated until complete consensus is achieved. We recommend two to four survey rounds to minimize the risk of participant fatigue.
Rating scale	Likert-type rating scales are favoured in the literature for ease of statistical analysis. We recommend a nine-point Likert scale with a "no opinion" option and a free-text comment box.
Feedback to participants	At the end of each survey round, the results can be communicated to the participants in quantitative or qualitative form, or both. We recommend that the results are graphically represented to the participants with feedback detailing individual scores as well as the relationship to the whole group. Any comments made should also be included anonymously.
Consensus	This should be determined before the study is initiated (a priori) A consensus level of 70% is typical.

and poor prognosis. They outline that this was, at least in part, due to a lack of standardized definitions. The authors compared the terminology used in two classification systems and found that tumours with signet ring cells were variably described as "diffuse type", "poorly cohesive" or "signet ring type", thus leading to inconsistencies in how they were classified. A multidisciplinary Delphi was then undertaken to determine how gastric carcinoma subtypes should best be defined and subsequently classified. The participants reached consensus that the most recent edition of the WHO classification system[15] should be used to report gastric adenocarcinomas, and only carcinomas with more than 90% poorly cohesive cells with signet cell morphology should be labelled as signet ring carcinoma.

Both studies identify the importance of unifying histopathological terminology, however, their reported methodologies were incomplete with regards the Delphi process utilized. Specifically, Mariette et al.[14] did not detail how many rounds of surveys were undertaken, the type of rating scale used and if consensus was defined a priori. In addition, the authors described their study as a traditional Delphi, yet the participants engaged in a series of web-based discussions more in keeping with a modified Delphi approach. Carr et al.[13] provide slightly more methodological detail, but provided the participants with response "options" which they then voted on, rather than using a Likert-type rating scale. Likert-type scales are favoured in Delphi studies because they facilitate calculation of the data median and interquartile ranges.[5] The results can be graphically represented to show the distribution of responses for each statement and the participants can see their own position compared to the whole group.[2] This provides an opportunity for the participants to reconsider their initial ranking if they wish to do so, especially if they are deemed to be an outlier for a given statement.[2,16] Researchers can expand their use by choosing to include "open text boxes" thereby enabling participants the option to add in comments which may inform participant feedback and future rounds of surveys.

In contrast, Seoane et al.[17] provide a very good example of methodological detail in their multidisciplinary

Table 2. Delphi in Histopathology – Summary of literature review

Author	Type of Delphi utilized*	No. of participants (no. of pathologists)	No. of countries†	No. of rounds	Rating scale utilized	Consensus defined a priori?
Burgues et al.[6]	Modified	46	1 (Spain)	2	9-point Likert scale	Yes
Teoh et al.[7]	Classic	200 (8)	Not specified (Intercontinental)	2	9-point Likert scale + "unable to score" option	Yes
Tejera et al.[8]	Modified	8	1 (Spain)	2	9-point Likert scale	Yes
Raya et al.[9]	Modified	37	1 (Spain)	2	4-point Likert scale	Yes
Haddad et al.[10]	Modified	14	Not specified	3	Not specified	Yes
Klimstra et al.[11]	Classic	28 (12)	7	1	Yes/no voting system	Yes
Fisher et al.[12]	Classic	50 (6)	Not mentioned	1	0–10 rating scale	Yes
Carr et al.[13]	Modified	71 (34)	15	4	Not specified	No
Mariette et al.[14]	Classic (disputed)	26 (12)	8	Not specified	Not specified	No
Seoane et al.[17]	Classic	21 (3)	Not specified (Europe and Latin America only)	2	7-point Likert scale	Yes
Tsekrekos et al.[18]	Classic	6	Not specified (Western countries only)	4	4-point Likert scale + free text	Yes
Saliba et al.[19]	Classic	15	12	3	5-point Likert scale + free text	Yes
Lord et al.[21]	Modified	15	8	2	5-point Likert scale	Yes
Kojima et al.[22]	Not specified	8	1 (Japan)	3	6-point Likert scale	Yes
Carney et al.[23]	Modified	3	1 (USA)	Not specified	Not specified	Not specified
Dufraing et al.[24]	Modified	10	10	2	Not specified	Yes
Simpson et al.[29]	Classic	73 (7)	14	3	5-point Likert scale	Yes

*This refers to whether the authors conducted their research using a traditional Delphi process without allowing for communication between the participants (classic Delphi), or if they adapted the method to incorporate a form of in-person or online meetings at some stage (modified Delphi).
†This refers to the geographical locations of the participants.

Delphi study which aimed to reach consensus on the nomenclature for actinic cheilitis (a potentially malignant lesion of the lip). A list of all the statements sent to participants was provided in the paper, alongside the response rates and median agreement scores. However, despite this robust approach, there was a lack of pathology representation in the study (3/21 participants) and it failed to reach consensus on the key issue of whether epithelial dysplasia is a requirement for diagnosis, thus highlighting the importance of panel composition in Delphi studies.

Evaluating classification systems

Classification systems are routinely used in Histopathology across a range of disease entities. It is not uncommon for different versions of a given classification system to exist, especially if there is no agreed international standard. Tumour Regression Grading (TRG) systems provide important prognostic information to clinicians by quantifying viable tumour cells post-therapy, yet there are several reporting systems in routine use with no single system recognized as

the "gold standard". Both Tsekerkos *et al.*[18] and Saliba *et al.*[19] used the Delphi method to reach consensus on the most appropriate TRG system for the assessment of residual malignancy in upper gastrointestinal tumours post-treatment. Both studies achieved consensus and reported that a 4-tiered TRG scoring system for the primary tumour, combined with a 3-tiered TRG system for regional metastatic nodal disease, should be taken as the optimum international standard. Whilst both studies recruited an international panel of experts with a similar number of survey iterations, they differed in the provision of statistical feedback to participants between rounds. Saliba *et al.* omitted this step to "not influence the individual's response." This contradicts what is described in the gold standard methodology of the Delphi method, whereby feedback to participants is critical to allow participants to reflect and potentially re-rank their agreement to a particular statement.[5]

Although both studies reached similar conclusions regarding the optimum TRG system, Tsekrekos *et al.* acknowledged that, whilst reaching consensus is the ultimate aim of the Delphi method, there is no universally accepted criterion as to what constitutes consensus. Indeed, a systematic review investigating how consensus is determined and utilized in Delphi studies found that there is considerable variation.[20] Some authors suggest that a consensus level of 70% is typical, but this is not supported by robust evidence and cut-off criteria have been adopted with variable rigidity (see Table 2).[5] For example, in a Delphi study by Burgues *et al.*,[6] investigating the specimen handling and histological analysis of breast specimens post-neoadjuvant systemic therapy, consensus was defined a priori as 70%. All the statements reached consensus in the first round yet, for reasons not provided, the statements in the consensus range of 70–80% were still subject to a second round of voting.

For the purposes of scientific credibility, it is important that consensus is predetermined and cut-off values explicitly stated before the study is initiated. One histopathological study stands out in terms of its methodological rigour and fulfilment of the main Delphi principles. Lord *et al.*[21] sought to gain consensus on how tumour deposits (TDs) in colorectal cancer are defined and classified within the TNM staging system. They undertook an international modified Delphi study with 15 expert histopathologists and their process of recruiting an expert panel, defining consensus a priori and undertaking iterative rounds of survey with feedback to participants, is clearly outlined. The participants reached consensus that the current positioning of TDs in the TNM staging was not prognostically reflective. For example, at present TDs are categorized within the nodal (N) category as 'N1c' although by definition they are not nodal and actually confer worse prognostic outcomes than lymph node involvement.[21] Nonetheless, consensus could not be reached to remove TDs from the N category altogether. The authors therefore proposed the need for a new, comprehensive staging system to appropriately capture important prognostic information which is currently missed or incomplete. Perhaps the only limitation of this paper is the exclusive recruitment of experts who had published in this subject area, with no representation of general pathologists who routinely use the TNM staging system. We hypothesise that the general pathology perspective, from individuals who use the system on a daily basis, may have altered the opinions expressed in this survey. Regardless, this excellent study highlights how Delphi methodology can highlight areas for future research and be used to generate recommendations for practice.

Resolving diagnostic disconcordance

Diagnostic Histopathology is inherently subjective and whilst this is accepted and mitigated against (for example through engagement in educational schemes and continued professional development activities), there remain challenging areas with consistently high interobserver variability. Such areas include the reporting of lymphovascular space invasion, as investigated by Kojima *et al.*,[22] and in the diagnosis of melanocytic lesions, as reported by Carney *et al.*[23] Both groups utilized the Delphi method as part of a multi-stage research study. This is acceptable, however, even when just one component of a study, it is vital that the methodology is clearly outlined to ensure it is fit-for-purpose. Carney *et al.*[23] do not specify the basic elements of their Delphi method (number of rounds, type of rating scale, how consensus was determined etc) and the 3 study participants were known to one another and reviewed "borderline" cases at a multi-header microscope, removing all anonymity. Anonymity prevents dominant characters from exerting undue influence, and hence is an important factor in the Delphi method.[5] Even when the voting process is anonymous, the participants may know or predict each other's responses, especially in a relatively small expert panel, in a small subspecialty, or if a participant is known to hold a particular view. We recommend that multi-stage research studies incorporating a Delphi method

clearly explain why, and at which stage, consensus methodology was required, and how anonymity was maintained.

Another area with high interobserver variation is in the estimation of neoplastic cell percentage (NCP) in tumour samples. Dufraing et al.[24] utilized an international modified Delphi method comprising 10 pathologists with experience in molecular pathology to reach recommendations for the practice of determining NCP in colorectal carcinoma. Common to all three of these studies exploring diagnostic disconcordance, is that the panel sizes comprised 10 or fewer participants. Dufraing et al.[24] highlight that no published guidelines exist for determining the optimum panel size. This is true, yet Delphi studies do not need to be statistically representative, and quality, rather than quantity, of the experts is deemed paramount.[25] Some studies also highlight that diversity within groups can have a positive impact by creating a pool of different perspectives, knowledge base and ideas.[26] In our opinion, diversity can be achieved in Histopathology Delphi studies in several ways; by having an international expert panel, recruiting pathologists with differing levels of experience, and involving both general and specialist pathologists.

Standardizing histological assessment and reporting

Datasets are evidence-based documents that outline the minimum information that histopathologists should include in their report for specific entities. The purpose of datasets is to act as an aide memoir and ensure that the most important histological features for a given pathology are captured in the report. Determining the most pertinent items to be recorded in a pathology "minimum dataset" can be challenging in areas lacking robust empirical evidence, and this process can be facilitated with the use of consensus building methods. For example, Haddad et al.[10] recognized that the assessment and reporting of tumour budding in colorectal carcinomas varies widely, despite its established adverse prognostic implications. They undertook a modified Delphi study with 14 international gastroenterology pathologists. Although the rating scale used was not stated, this study did provide a "no opinion" option for the expert panel. This prevents the participants from feeling compelled to select a neutral option (i.e. "undecided") or even not rate the statement altogether, in cases where they feel that they do not have adequate knowledge to form a judgement. Tejera et al.[8] also

used a modified Delphi methodology to establish which histologic variables should be included in their reporting protocol for cutaneous melanoma. By the end of the process, they had reached consensus on 30 of 36 variables and thus provide guidelines for the reporting of cases submitted to the Spanish National Registry.[8] Finally, Raya et al.[9] utilized a modified Delphi comprising 37 expert haematopathologists. The participants reached consensus and developed dataset recommendations concerning the clinical information, ancillary testing and morphological features for the diagnosis of primary myelofibrosis after a 2-round modified Delphi study. The authors state that the dataset items formulated as a result of their work will help to standardize the reporting of bone marrow biopsies in Spain for this challenging entity.

In our opinion, simply stating that a "modified Delphi" was performed without elaborating on the details of how the traditional Delphi model was adapted, as was the case with Raya et al.,[9] prevents the reader from assessing its methodological rigour. Modified Delphi studies provide opportunity for experts to discuss and debate statements that do not reach consensus, paving way for new ideas to be generated. Nonetheless, the "modifications" must be clearly detailed, and the factors of anonymity and dominance accounted for as much as possible. Without this, studies risk compromising the fundamental principles of the Delphi study.

The standardizing of feature assessment and reporting have also been the focus of a number of multidisciplinary consensus groups in Histopathology. For example, Klimstra et al.[11] sought to determine the common features of neuroendocrine tumours and subsequently develop a minimum pathology dataset. This study had a standardized design in that the participants were all present in a conference room, required to vote "yes" or "no" only, and the results were immediately tabulated. Unique to this study is the use of a binary voting system instead of the favoured Likert-type scales described the literature.[16,26] The purpose of rating scales in Delphi studies is two-fold: to measure the extent a participant agrees with a statement and to determine the degree in which the participants agree with each other.[2] In contrast, binary voting systems do not offer these insights. It is, however, worth highlighting that there is no preferential type of Likert scale (five-point, seven-point etc) described in the literature, although nine-point Likert scales are most frequently used.[27]

Lip salivary gland histopathology is used in the primary diagnosis of Sjogren's syndrome and has the potential to stratify disease and determine inclusion

in clinical trials. Fisher *et al.*[12] used a multidisciplinary Delphi study to confirm consensus and highlight uncertainty within this challenging field. One topic showing strong consensus amongst participants, was the need for further guidance on the identification of lymphocytic germinal centres, as these may be an important feature to predict prognosis. The authors concluded that a weakness in their study was the reliance on expert opinion when the evidence base is lacking. The Delphi method is undertaken when empirical evidence for a given subject is limited, and this acts an intrinsic motivation for seeking expert opinion to reach consensus on contentious issues. We therefore question the authors' understanding of the fundamental principles governing the Delphi method and challenge the notion that relying on expert opinion is a weakness.

Representation of pathologists in multidisciplinary Delphi studies

In Medicine, healthcare professionals rarely work in isolation, and Histopathology is no exception. As outlined above, multidisciplinary Delphi studies can bring together international experts across a range of specialities to reach consensus on all aspects of a patient's journey from diagnosis to treatment. Two of the largest Delphi studies identified in our search are multidisciplinary in nature and are discussed below and summarized in Table 2.

Teoh *et al.*[7] utilized a modified Delphi method to develop consensus statements for best practice standards for the en-bloc resection of bladder tumours (ERBT). Two-hundred specialists from across the globe participated in the process, of which only 8 were pathologists. Whilst there is no denying that pathologists were under-represented in this panel, this is not necessarily surprising given that ERBT is a surgical procedure. However, 9 out of 103 statements pertained to Histopathology (6 of which reached consensus) and all of these statements were rated by over 135 participants, considerably more than the number of participating pathologists. Upon closer review, it is apparent that many of these statements related to specimen handling and the pathology features that "should" be included in the report. It is not unreasonable for urologists to rate such statements given that these pathological factors, for example tumour stage and margin involvement, are associated with prognosis and can influence management options. The authors acknowledged that not all specialists would have adequate knowledge to offer an opinion on a given statement and sought to mitigate this by incorporating an "unable to score" option, rather than "neither agree nor disagree" in their rating scale. We feel that this greatly added to the already robust methodology presented in this study. In addition, the statements were grouped together under common domains and were clear and succinct. Unfortunately, the formulation of Delphi statements is another area with minimal published guidance, therefore evaluating the quality and ease of use of these statements is challenging and is largely opinion based.[28]

Simpson *et al.*[29] recruited 73 participants, including 7 pathologists, to reach consensus on the clinicopathological diagnostic criteria for erosive Lichen Planus (ELP). The methodological criteria for a gold-standard Delphi were fulfilled in this study, however the main limitation was the lack of histopathology representation. For example, the histopathologists determined that certain epidermal changes (such as "sawtooth" acanthosis) are important features for the diagnosis of ELP, however, the rest of the group (non-pathologists) did not support this view and the statement did not reach consensus. It would seem intuitive that that histopathologists, with the relevant training and expertise in their field, are the ones best suited to offer opinions on statements relating to histology. However, as illustrated by this example, there is strength in numbers, and we hypothesise that an "unable to score" option would have made non-consensus less likely. This example illustrates the importance of appropriate panel selection and statement design to not only ensure a balanced representation across different stakeholders, but to capture the viewpoints specific to the participants area of expertise.

Conclusion

The Delphi method is a robust consensus building research tool which has been used in a small number of histopathological studies across a wide variety of indications. Whilst there is a paucity of literature exploring the use of Delphi in Histopathology, this review shows that, where applied robustly, Delphi has the potential to establish reproducible definitions, define best practices for specimen handling, resolve areas of diagnostic inconsistencies and determine the pertinent information to be included in pathology reports. We particularly advocate the use of the Delphi method in instances where there is ambiguity or redundancy in nomenclature, and when core pathology features need to be identified and/or agreed upon.

Whether exclusively histopathologist led, or as part of a multidisciplinary panel, Delphi studies can bring together geographically dispersed individuals with an array of expertise, however, the purpose for each stakeholder group needs to be clearly defined and adequate and/or proportionate representation ensured.

Acknowledgements

Conflicts of interest

KJG is an NIHR-funded Clinical Lecturer. AWM, SLM and AC are supported by the NIHR Leeds Biomedical Research Centre and AWM is additionally supported by the NIHR Leeds Medtech and In Vitro Diagnostics Co-operative, an NIHR Senior Investigator award and an MRC TARGET Partnership grant. SLM reports consultancy on behalf of her institution for Roche/Chugai, Sanofi, AbbVie and AstraZeneca. AWM has received research grant or educational funding or undertaken consultancy on behalf of her institution for the following pharmaceutical companies in the last 5 years: AstraZeneca, Kiniska Pharmaceuticals, Regeneron, Roche/Chugai, Sanofi and Vifor. The views expressed in this article are those of the authors and not necessarily those of the Leeds Hospitals Charity, NIHR or the Department of Health and Social Care.

Author contributions

KG and SM were responsible for study conception and design. DT and CH performed the literature search and were responsible for sourcing and organizing the original data. DT and KG applied the inclusion and exclusion criteria and performed analysis of the included studies. Expert histopathology opinion was provided by AC. DT and KG wrote the original manuscript with review and editing performed by SM, AM and AC. All authors have reviewed and approved the final version of the manuscript.

Data availability statement

Data sharing not applicable to this article as no datasets were generated or analysed during the current study

References

1. Dalkey N, Helmer O. An experimental application of the Delphi method to the use of experts. *Manag. Sci.* 1963; **9**; 458–467.
2. Jones J, Hunter D. Consensus methods for medical and health services research. *BMJ* 1995; **311**(7001); 376–380.
3. Humphrey-Murto S, Wood TJ, Gonsalves C, Mascioli K, Varpio L. The Delphi method. *Acad. Med.* 2020; **95**(1); 168.
4. Humphrey-Murto S, Varpio L, Gonsalves C, Wood TJ. Using consensus group methods such as Delphi and Nominal Group in medical education research. *Med. Teach.* 2017; **39**(1); 14–19.
5. Taylor E. We agree, don't we? The Delphi method for health environments research. *HERD* 2020; **13**(1); 11–23.
6. Burgués O, López-García M, Pérez-Míes B *et al.* The ever-evolving role of pathologists in the management of breast cancer with neoadjuvant treatment: recommendations based on the Spanish clinical experience. *Clin. Transl. Oncol.* 2018; **20**(3); 382–391.
7. Teoh JYC, MacLennan S, Chan VWS *et al.* An international collaborative consensus statement on en bloc resection of bladder tumour incorporating two systematic reviews, a two-round Delphi survey, and a consensus meeting. *Eur. Urol.* 2020; **78**(4); 546–569.
8. Tejera-Vaquerizo A, Fernández-Figueras MT, Santos-Briz Á *et al.* Protocol for the histologic diagnosis of cutaneous melanoma: consensus statement of the Spanish Society of Pathology and the Spanish Academy of Dermatology and Venereology (AEDV) for the national cutaneous melanoma registry. *Rev. Esp. Patol.* 2021; **54**(1); 29–40.
9. Raya JM, Montes-Moreno S, Acevedo A *et al.* Pathology reporting of bone marrow biopsy in myelofibrosis; application of the Delphi consensus process to the development of a standardised diagnostic report. *J. Clin. Pathol.* 2014; **67**(7); 620–625.
10. Haddad TS, Lugli A, Aherne S *et al.* Improving tumor budding reporting in colorectal cancer: a Delphi consensus study. *Virchows Arch.* 2021; **479**(3); 459–469.
11. Klimstra DS, Modlin IR, Adsay NV *et al.* Pathology reporting of neuroendocrine tumors: application of the Delphic consensus process to the development of a minimum pathology data set. *Am. J. Surg. Pathol.* 2010; **34**(3); 300–313.
12. Fisher BA, Jonsson R, Daniels T *et al.* Standardisation of labial salivary gland histopathology in clinical trials in primary Sjögren's syndrome. *Ann. Rheum. Dis.* 2017; **76**(7); 1161–1168.
13. Carr NJ, Cecil TD, Mohamed F *et al.* A Consensus for classification and pathologic reporting of pseudomyxoma peritonei and associated appendiceal neoplasia: the results of the Peritoneal Surface Oncology Group International (PSOGI) modified Delphi process. *Am. J. Surg. Pathol.* 2016; **40**(1); 14–26.
14. Mariette C, Carneiro F, Grabsch H *et al.* Consensus on the pathological definition and classification of poorly cohesive gastric carcinoma. *Gastric Cancer* 2019; **22**(1); 1–9.
15. Bosman FT, Carneiro F, Hurban RG, Theise ND. *WHO classification of tumours of the digestive system*. 4th ed. Lyon: IARC Press, 2010; 48–58.
16. Hsu C-C, Sandford BA. The Delphi technique: making sense of consensus. *Pract. Assess. Res. Eval.* 2007; **12**; 10.
17. Seoane J, Warnakulasuriya S, Bagán JV *et al.* Assembling a consensus on actinic cheilitis: a Delphi study. *J. Oral Pathol. Med.* 2021; **50**(10); 962–970.
18. Tsekrekos A, Detlefsen S, Riddell R *et al.* Histopathologic tumor regression grading in patients with gastric carcinoma

submitted to neoadjuvant treatment: results of a Delphi survey. *Hum. Pathol.* 2019; **84**; 26–34.

19. Saliba G, Detlefsen S, Carneiro F *et al*. Tumor regression grading after neoadjuvant treatment of esophageal and gastroesophageal junction adenocarcinoma: results of an international Delphi consensus survey. *Hum. Pathol.* 2021; **108**; 60–67.

20. Diamond IR, Grant RC, Feldman BM *et al*. Defining consensus: a systematic review recommends methodologic criteria for reporting of Delphi studies. *J. Clin. Epidemiol.* 2014; **67**(4); 401–409.

21. Lord A, Brown G, Abulafi M *et al*. Histopathological diagnosis of tumour deposits in colorectal cancer: a Delphi consensus study. *Histopathology* 2021; **79**(2); 168–175.

22. Kojima M, Shimazaki H, Iwaya K *et al*. Pathological diagnostic criterion of blood and lymphatic vessel invasion in colorectal cancer: a framework for developing an objective pathological diagnostic system using the Delphi method, from the Pathology Working Group of the Japanese Society for Cancer of the Colon and Rectum. *J. Clin. Pathol.* 2013; **66**(7); 551–558.

23. Carney PA, Reisch LM, Piepkorn MW *et al*. Achieving consensus for the histopathologic diagnosis of melanocytic lesions: use of the modified Delphi method. *J. Cutan. Pathol.* 2016; **43**(10); 830–837.

24. Dufraing K, van Krieken JH, De Hertogh G *et al*. Neoplastic cell percentage estimation in tissue samples for molecular oncology: recommendations from a modified Delphi study. *Histopathology* 2019; **75**(3); 312–319.

25. Powell C. The Delphi technique: myths and realities. *J. Adv. Nurs.* 2003; **41**(4); 376–382.

26. Murphy MK, Black NA, Lamping DL *et al*. Consensus development methods, and their use in clinical guideline development. *Health Technol. Assess.* 1998; **2**(3); i–88.

27. Lange T, Kopkow C, Lützner J *et al*. Comparison of different rating scales for the use in Delphi studies: different scales lead to different consensus and show different test-retest reliability. *BMC Med. Res. Methodol.* 2020; **20**(1); 28.

28. Niederberger M, Spranger J. Delphi technique in health sciences: a map. *Front. Public Health* 2020; **8**; 457.

29. Simpson RC, Thomas KS, Leighton P, Murphy R. Diagnostic criteria for erosive lichen planus affecting the vulva: an international electronic-Delphi consensus exercise. *Br. J. Dermatol.* 2013; **169**(2); 337–343.