Systematic Review of the Use of Telepathology During Intraoperative Consultation

Robin L. Dietz, MD[○], *Douglas J. Hartman, MD, and Liron Pantanowitz, MD*

From the Department of Pathology, UPMC, Pittsburgh, PA.

ABSTRACT

Objective: *To compare studies that used telepathology systems vs conventional microscopy for intraoperative consultation (frozen-section) diagnosis.*

Methods: *A total of 56 telepathology studies with 13,996 cases in aggregate were identified through database searches.*

Results: *The concordance of telepathology with the reference standard was generally excellent, with a weighted mean of 96.9%. In comparison, we identified seven studies using conventional intraoperative consultation that showed a weighted mean concordance of 98.3%. Evaluation of the risk of bias showed that most of these studies were low risk.*

Conclusions: *Despite limitations such as variation in reporting and publication bias, this systematic review provides strong support for the safety of using telepathology for intraoperative consultations.*

The term *telepathology* (TP), first used by Weinstein[1] in 1986, is defined as using telecommunications technology to make remote diagnoses on a computer screen rather than directly through the lens of a microscope. Almost any type of specimen viewed with light microscopy (LM) can also be evaluated using TP. TP can also refer to sending nonimaging file types to make diagnoses, but for the purposes of this review, we refer to TP in the context of digital pathology to support remote intraoperative consultation (IOC).

There are five broad technological categories for TP methods[2,3]: (1) static, (2) dynamic, (3) robotic, (4) whole-slide image, and (5) hybrid. *Static images* are still digital images taken from preselected areas. Their image quality depends heavily on the experience of the photographer transmitting the images, who must select the most representative area and magnification for the pathologist receiving the images. *Dynamic telepathology* is the transmission of a video stream from a microscope to a receiving pathologist. Similar to TP with static images, dynamic TP requires a host to control the microscope on site. The receiving pathologist has more control with this modality as he or she is able to communicate with the host and specify in near real time which areas to view and focus on. *Robotic microscopy* (RM) gives control of the microscope stage, objectives, and focus to the receiving pathologist and also transmits a live dynamic image feed. This eliminates the need to give verbal orders to a host. *Whole-slide images* (WSIs) are large images composed of many smaller static images stitched together. While this modality also eliminates the need for a host, if multiple focus planes (Z-stacks) are desired, the slide must be scanned using Z-stacking, which increases the scanning time and WSI file size. *Hybrid methods* (also known as multimodality[3]) use combinations of either RM/WSI or dynamic/WSI.

One of the principal applications for TP has been the remote interpretation of frozen sections or IOC. Studies measuring the accuracy of TP for IOC span as far back to 1991.[4] A meta-analysis on the topic of TP in IOC has not been performed since the review by Wellnitz et al[5] in 2000. They identified 13 published studies at the time involving around 1,300 cases and found that the accuracy of TP for IOC had a slightly lower diagnostic accuracy (0.91) than conventional (manual glass slide examination with a traditional LM) IOC (0.98).[5] They attributed these differences to higher rates of deferred cases and false-negative diagnoses using TP. The specificity of both TP and conventional methods was not significantly different.[5] The study by Wellnitz et al[5] focused on static and dynamic TP methods, and since then, robotic and WSI studies have emerged as the most popular methods tested in published studies.

Recently, Bashshur et al[5] performed a systematic review of the use of TP in different areas of pathology, including hematopathology, cytopathology, and IOC. Similar to Wellnitz et al,[5] they found a consistent trend of concordance between TP and conventional LM. Their analysis found that there was more agreement in cancer detection for higher-grade cancers and that experience of the pathologist in information technology (IT) and higher-resolution images correlated with diagnostic accuracy and diagnostic confidence, respectively.[6]

There is a need to perform a systematic review on the accuracy of TP, specifically for IOC, for several reasons. IOCs are unique compared with primary diagnosis due to the variable quality of slide preparation and the need to arrive at a diagnosis under strict time pressures. Also, a systematic review dedicated to TP for IOC has not been performed in nearly two decades, and since then, there have been technological advances in pathology equipment and more generally in computational power. The aim of this review was to perform an updated evaluation of the literature focusing on the use of TP for IOC. The concordance of TP to conventional methods was the primary outcome measure, and time to diagnosis was a secondary outcome measure.

Materials and Methods

This systematic review was conducted following the Preferred Reporting Items for Systematic Reviews and Meta-Analysis (PRISMA) guidelines.[7]

Search Strategy

No prior related review on TP was identified in the International Prospective Register of Systematic Reviews database of registered systematic reviews in health care. A search was accordingly conducted using the following databases: MEDLINE by PubMed platform, Scopus, and the Cochrane Library. Search terms included the following: [ALL (telepathology) AND ALL (frozen section)], [ALL (telepathology) AND ALL (intraoperative consultation)], [ALL (digital pathology) AND ALL (frozen section)], [ALL (digital pathology) AND ALL (intraoperative consultation)], [ALL (virtual pathology) AND ALL (frozen section)], ALL (virtual pathology) AND ALL (intraoperative consultation)], and [ALL (telepathology)]. As a result, 1,291 articles from PubMed, 452 articles from Scopus, and 15 articles from the Cochrane Library were identified. A Google search did not contribute any additional articles. A manual search of cited references was also performed. We sought the assistance of a librarian to perform the database search.

Article Screening and Eligibility Criteria

Articles were screened according to the PRISMA flow diagram shown in ❚Figure 1❚.[7] No limits were placed on the date range, and two abstracts from the gray literature were considered. The two abstracts met our criteria for full article review, but one of the abstracts subsequently published an article with the same data that were used in our analysis.[8] The other abstract only reported κ values and results from the Cochran-Mantel-Haenzel test, and it was not included in our analysis.[9] EndNote X9.1 (Clarivate Analytics) was used to manage our database searches and references, as well as identify duplicate studies. Only articles about the use of TP for frozen-section diagnosis were included. Articles that used exclusively cytopathology preparations (fine-needle aspirations or smears) or routine surgical pathology cases were excluded. When an article combined IOC data with other data (eg, cytopathology), the study was included only if there was separate reporting of the frozen-section data that permitted data extraction.

Data Extraction

The data extraction was conducted by the primary researcher (R.L.D.) and verified by two researchers (D.J.H. and L.P.). The following information was recorded: author, publication year, location of sending and receiving institutions, TP method used (eg, static, dynamic), equipment used, study type (retrospective or prospective), study design type ❚Figure 2❚, tissue type involved, and percent concordance. The percent concordances were used to calculate weighted means between different categories such as study type and TP method. When available, we also recorded case selection criteria,
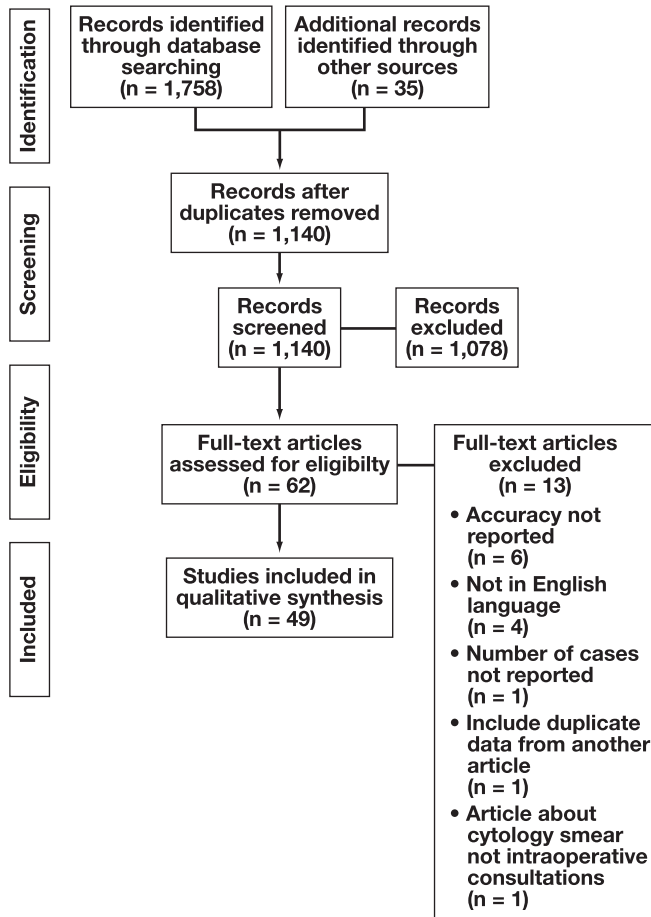
Figure 1 Preferred Reporting Items for Systematic Reviews and Meta-Analysis (PRISMA) flow diagram detailing the study selection process. For the final review, data was extracted from the full texts of 49 articles.

Figure 2 Study design classification in published TP for IOC articles. Prospective studies included study types A and B, and retrospective studies included study types B through F. **A**, Original LM diagnosis to original TP diagnosis. **B**, Original LM diagnosis to original IOC diagnosis. **C**, Original LM diagnosis to study TP diagnosis. **D**, Original IOC diagnosis to study TP diagnosis. **E**, Study LM diagnosis to study TP diagnosis. **F**, Study IOC diagnosis to study TP diagnosis. IOC, intraoperative consultation; LM, light microscopy; TP, telepathology.

accuracy of conventional methods, turnaround times (TATs) of both TP and conventional methods, deferral rates, software used, transmission speed, screen resolution, cost, and technical failures.

The design of studies we evaluated varied, but generally studies compared the concordance of TP with respect to the final LM diagnosis in prospective studies and either the conventional IOC diagnosis or the final LM diagnosis in retrospective studies. The terms *prospective* and *retrospective* were defined with regard to the temporal relation between the collection of index test data and the time of the surgery from which tissue was acquired. This binary distinction between study types is similar to that used by Wellnitz et al[5] in their meta-analysis. Prospective studies are those performed on cases that were reported back to the surgeon and used for patient treatment. Retrospective studies are those that involved pulling cases from slide archives and analyzed outside of routine clinical workflow.
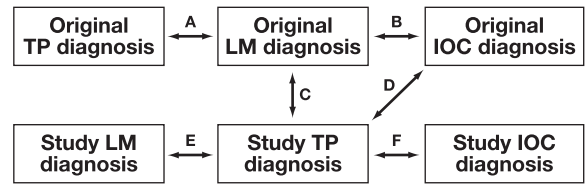
Some of the terminology in the literature used to describe TP methods was inconsistent with the terminology we have defined above. For instance, if a study claimed that it used a "dynamic" TP system with a robotic stage,[10] we classified this publication as using a robotic TP system. For our study, "dynamic" pathology systems were those that used a video stream without a robotic stage. Although some instruments offer hybrid functionality, we classified the study according to the method that was actually used for remotely viewing cases.

Discrepancies included "major discrepancies" or those that would have resulted in a change in therapy. "Minor discrepancies" were differences between diagnoses that did not result in a change in therapy or that were merely semantic. We chose to count cases with minor discrepancies as concordant and major discrepancies as discordant.

An area of variability among studies that caused difficulty in data extraction was deferrals. Some studies classified deferrals as major discrepancies.[11] Some studies counted deferrals as concordant if the original IOC deferred that case.[12,13] Many studies did not mention deferrals or report how deferrals were calculated,[10,14-25] or they chose to completely exclude deferrals from their analysis.[26-30]

When detailed information was not provided, we elected to extract concordance rates the authors provided. When a study required us to calculate concordance rates, we opted to count cases that were deferred or that had minor discrepancies as concordant cases. LM of formalin-fixed (permanent) slides was considered the primary reference standard when available; otherwise, conventional IOC using LM was considered the reference standard. The TP diagnosis was the index test in all studies. A 2-week

washout period was considered adequate for retrospective reviews.

Studies were further categorized by the type of tissue evaluated. If the study used only one type of tissue, it was classified as such; otherwise, it was classified as "mixed." Types of tissue included in the mixed category were abdominal wall, artery, bile duct, bone, breast, bronchus, chest wall, diaphragm, duodenum, esophagus, larynx, liver, lung, lymph node, nose, ovary, pancreas, parathyroid, peritoneum, sinus, soft tissue, stomach, thyroid, tongue, tonsil, ureter, uterus, urinary bladder, and uvula.

Detailed explanations about training of pathologists to use TP systems were rarely reported, so we did not evaluate this parameter. We also did not take into consideration the viewing devices (eg, monitors, tablets) used to make diagnoses. Besides, increased display resolution has not been shown to be associated with additional diagnostic benefit,[31] and studies of commercial vs medical-grade monitors have shown similar performance.[32]

Quality Assessment

Different study types are inherently fraught with bias. Studies that are performed prospectively have the risk of self-affirmation and timing bias. The pathologist signing out the final LM diagnosis on a case may be less likely to disagree with the original IOC diagnosis, and the final diagnoses in these cases are not blinded. The quality of included studies was assessed by one reviewer using the Quality Assessment of Diagnostic Accuracy Studies 2 (QUADAS-2) tool modified to address specific concerns of bias in the included studies ❙Table 1❙.

Quantitative Synthesis

The studies included in our analysis showed high levels of diversity in their study design (Figure 2), organ systems evaluated, hardware and software systems used, index test conditions, deferral rate reporting, and outcome measures. As a result, we felt that estimating effect size in the meta-analysis was not justified.[35] The weighted means were calculated for the total and for each organ system by normalizing the number of cases as weights such that they summed up to 1. Weighted means were performed to correct for the effects of study size on the means.

Results

Using the search strategy detailed above, 1,758 articles were identified through database searching. A total of 1,140 articles were screened after duplicates were removed. The number of articles excluded at each stage can be found in Figure 1. A total of 49[4,8,10-30,36-61] articles were included. Of note, 34 of these articles were found from manual reference searching, while 15 were identified through the search strategy. Six articles[8,17,18,21,27,47] that compared more than one type of TP system were separated into individual studies for the numerical analysis, resulting in a total of 56 studies for analysis.

Details of each study are provided in ❙Table 2❙, and the percent concordance range (PCR) and weighted means organized by organ system are shown in ❙Table 3❙. The weighted mean for the entire study set was 96.9% and ranged from 88.2% for lymph node cases to 99.4% for parathyroid cases. The PCR for the included studies ranged from 54% to 100%. The largest PCR range was in the central nervous system (CNS) (54%-98%), which also had the largest number of cases of any tissue type outside of the mixed type. The weighted mean for seven studies of conventional IOC[62-68] with 120,996 cases was 98.3%.

We also compared the weighted means of the concordance rates prior to 2000, 91.1%, and from 2000 to present, 97.2%. This distinction was made since it represents the time since the last meta-analysis was conducted by Wellnitz et al,[5] and it also marked a turning point from the use of mostly static and dynamic TP methods to primarily robotic and WSI methods.

Graphs highlighting different aspects of study characteristics related to concordance are shown in ❙Figure 3❙. Forty-six (82%) of 56 studies had a concordance above 90%, and 36 (64%) of 56 had concordances above 95%. There was no clear correlation between study size and concordance rates (Figure 3A). A total of 27 (48%) of 56 studies were prospective, and 29 (52%) of 56 were retrospective (Figure 3B). The unweighted means for these study types were 92.52% and 94.95%, respectively. The most common TP method tested was RM (n = 24), followed by WSI (n = 13). The unweighted means for these methods were 93.95% and 97.3%, respectively (Figure 3C). Dynamic systems (n = 9) had a mean of 93.1%, and static imaging had a slightly lower mean at 88.3%. The only study involving a hybrid TP system showed a concordance of 98.3%. Figure 3D shows the percent concordance vs year, color coded by TP method.

The available average TAT was extracted from seven articles,[20-22,27,39,53,61] incorporating a total of 3,990 cases. The weighted average of the TAT, with weights normalized to 1 to account for study size, was 17.75 minutes per case. While other articles reported TAT,[29,40,52,59] they were not in a format that was easily extractable or reported other time metrics (eg, total viewing time).

The results from the studies reviewed come from 14 countries and use a myriad of hardware and software systems. Some

Table 1

Modified QUADAS-2[a]

Item	Yes	No	Unclear
Domain 1: Patient Selection			
A. Risk of bias[b]			
Was a consecutive or random sample of the patients enrolled?	( )	( )	( )
Did the study avoid inappropriate exclusions?	( )	( )	( )
Could the selection of patients have introduced bias?	( )	( )	( )
B. Concerns regarding applicability			
Is there concern that the included patients do not match the review question?	( )	( )	( )
Domain 2: Index Test(s)			
A. Risk of bias			
Were the index test results interpreted without knowledge of the results of the reference standard?	( )	( )	( )
Were clinical details provided for each case?[c]	( )	( )	( )
Are participants trained in using the index test?[c]	( )	( )	( )
Were deferrals reported?	( )	( )	( )
Could conduct/interpretation have introduced bias?	( )	( )	( )
B. Concerns regarding applicability			
Is there concern that the index test, its conduct, or interpretation differ from the review question?	( )	( )	( )
Domain 3: Reference Standard			
A. Risk of bias			
Is the reference standard likely to correctly classify the target condition?	( )	( )	( )
Were the reference standard results interpreted without knowledge of the results of the index test?	( )	( )	( )
Were clinical details provided for each case?[c]	( )	( )	( )
Could the reference standard, its conduct, or its interpretation have introduced bias?	( )	( )	( )
B. Concerns regarding applicability			
Is there concern that the target condition as defined by the reference standard does not match the review question?	( )	( )	( )
Domain 4: Flow and Timing			
A. Risk of bias			
Was there an appropriate interval between index test(s) and the reference standard?[d]	( )	( )	( )
Did all patients receive a reference standard?	( )	( )	( )
Did patients receive the same reference standard?	( )	( )	( )
Were all patients included in the analysis?	( )	( )	( )
Could the patient flow have introduced bias?	( )	( )	( )

QUADAS-2, Quality Assessment of Diagnostic Accuracy Studies 2.

[a]This modified version of QUADAS-2 was adapted from the original and based on recent work on a similar topic.[33,34] The QUADAS-2 assessment was used as a tool to assess the quality of diagnostic accuracy studies.[34]

[b]Domain in which a signaling question (Was a case-control design avoided?) was omitted.

[c]Additional signaling question added.

[d]The minimal appropriate time interval was considered 2 weeks.

of the image acquisition hardware used in the past 10 years includes Nikon COOLSCOPE,[15,21] Motic EasyScan,[29] Aperio (CS2,[58,61] LV1,[8] and ScanScope[18,19]), Navigo,[60] MikroScanD2,[8] Remote Medical Technologies,[30] Mirax,[20,57] Hamamatsu NanoZoomer,[16] and Trestle MedMicro.[17] Older image acquisition hardware devices include CLARO VASSALO scanner,[23] Leica TPS2,[27] Zeiss Axioplan,[25] TelePath,[48] Telemed A200[39], and Pathtran 1000 system.[38] Homegrown systems used a variety of cameras paired with software, such as Olympus DP-71 camera with Olympus cellSens software,[24] Sony Handyman digital video camera with iChat AV software,[54] Nikon E1000M microscope camera with Zem software,[52] and Sony 3CCD camera with Paint Shop Pro.[43]

Quality Assessment and Risk of Bias

A summary of the quality assessment of the 49 articles reviewed with the QUADAS-2 tool can be found in **Table 4**. Bar graphs of the results organized by risk of bias and applicability are shown in **Figure 4**. The risk of bias was assessed across four QUADAS-2 domains: flow and timing, reference standard, index test, and patient selection. The reference standard domain showed the lowest risk of bias, with seven (14%) unclear articles and 42 articles showing low risk of bias. The patient selection domain showed the highest risk of bias of any domain in the quality assessment, with 16 unclear articles, two articles with high risk of bias, and 31 articles with low risk of bias.

Applicability was assessed across three QUADAS-2 domains: reference standard, index test, and patient selection. Similar to the risk of bias assessment, the reference standard showed the lowest concerns regarding applicability, with five (10%) unclear articles and 44 (90%) articles with low concern. The patient selection domain again showed the highest concern, with nine (18%) unclear articles, three (6%) articles with high concern, and 37 (76%) articles with low concern.

Table 2
Study Characteristics[a]

First Author	Year	Design Type	Country	Study Type	Tissue Type	No. of Cases	Concordance, %
Dynamic							
Baak[44]	2000	D	Netherlands	Retrospective	Mixed	128	93.8
Dawson[45]	2000	A	United States	Prospective	Skin	66	93.9
Hufnagl[47]	2000	B, C	Germany	Retrospective	Mixed	125	96.4
Ongürü[12]	2000	C	Turkey	Retrospective	Mixed	98	85.7
Horbinski[17]	2007	A	United States	Prospective	CNS	40	81
McKenna[54]	2007	D	United States	Retrospective	Skin	20	100
Liang[55]	2008	A	China	Prospective	Mixed	50	92
Vitkovski[30]	2015	C	United States	Prospective	Lung	103	98
Vosoughi[24]	2018	A	United States	Retrospective	Mixed	333	97.4
Hybrid							
Pradhan[18]	2016	D	United States	Retrospective	GU	20	98.3
Robotic							
Nordrum[4]	1991	A	Norway	Prospective	Mixed	17	100
Shimosato[36]	1992	NS	Japan	Prospective	Mixed	16	68.8
Nordrum[39]	1995	A	Norway	Prospective	Mixed	100	96
Winokur[13]	1998	E	United States	Retrospective	Mixed	64	96.9
Della Mea[46]	2000	D	Italy	Retrospective	Mixed	60	100
Hufnagl[47]	2000	B, C	Germany	Retrospective	Breast	125	96.4
Hufnagl[47]	2000	B, C	Germany	Retrospective	Breast	53	100
Winokur[48]	2000	E	United States	Retrospective	Mixed	99	94.9
Demichelis[49]	2001	F	Italy	Retrospective	Mixed	70	98.6
Chorneyko[26]	2002	D, F	Canada	Retrospective	Mixed	80	86.6
Kaplan[50]	2002	D	United States	Retrospective	Mixed	120	100
Nehal[51]	2002	F	United States	Retrospective	Skin	60	100
Hutarew[52]	2003	A	Austria	Prospective	Mixed	413	99.4
Moser[25]	2003	C	Austria	Retrospective	Mixed	270	91.5
Hitchcock[28]	2005	A	United States	Prospective	Breast	195	95.3
Hutarew[53]	2006	A	Austria	Prospective	CNS	343	97.9
Horbinski[17]	2007	A	United States	Prospective	CNS	362	78.2
Evans[27]	2009	A	Canada	Prospective	CNS	350	97.5
Evans[27]	2009	A	Canada	Prospective	CNS	633	98
Horbinski[56]	2009	A	United States	Prospective	CNS	262	78.2
Slodkowska[21]	2009	A	Poland	Prospective	Lung	81	97.5
Gifford[15]	2012	A	Australia	Prospective	Lymph node	52	88.2
Hufnagl[59]	2016	A	Germany	Prospective	Breast	35	94.3
Chandraratnam[8]	2018	C	Australia	Retrospective	Parathyroid	76	99.1
Chandraratnam[8]	2018	C	Australia	Retrospective	Parathyroid	76	99.6
Static							
Becker[10]	1993	D	United States	Retrospective	CNS	52	54
Oberholzer[37]	1993	A	Switzerland	Prospective	Mixed	16	75
Fujita[38]	1995	A	Japan	Prospective	Mixed	59	94.9
Oberholzer[40]	1995	A	Switzerland	Prospective	Mixed	53	90.3
Adachi[41]	1996	A	Japan	Prospective	Mixed	117	93.2
Weinstein[42]	1997	C, D, F	United States	Retrospective	Skin	48	95.8
Della Mea[43]	1999	D	Italy	Retrospective	Mixed	151	96.7
Stauch[11]	2000	A	Germany	Retrospective	Breast	200	97
Frierson[14]	2007	D	United States	Retrospective	Mixed	100	98
Whole-slide image							
Tsuchihashi[23]	2008	NS	Japan	Retrospective	Unknown	15	100
Slodkowska[21]	2009	A	Poland	Prospective	Lung	33	100
Fallon[57]	2010	D	United States	Retrospective	Ovary	52	96
Ramey[19]	2011	D	United States	Retrospective	Mixed	67	89
Gould[16]	2012	A	Canada	Retrospective	CNS	30	96.7
Ribback[20]	2014	C	Germany	Prospective	Mixed	1,204	98.6
Têtu[22]	2014	A	Canada	Prospective	Mixed	1,329	98
Bauer[58]	2015	D	United States	Retrospective	Mixed	70	98.5
Pradhan[18]	2016	D	United States	Retrospective	GU	20	100
Cima[60]	2018	F	Italy	Retrospective	Mixed	121	94.2
French[61]	2018	A	England	Prospective	Lung	31	96.7
Huang[29]	2018	A	China	Prospective	Mixed	5,233	99.8

CNS, central nervous system; GU, genitourinary; NS, not specified.
[a]The design type column refers to the information derived from Figure 2.

Table 3

Percentage of Concordance Range (PCR) and Weighted Means Organized by Organ System

	No. (%) of Cases	PCR, %	Weighted Mean
Telepathology			
Breast	733 (5)	94.3-100	96.4
Central nervous system	2,072 (15)	54-98	90.5
Genitourinary	40 (<1)	98.3-100	99.15
Lung	248 (2)	96.7-100	97.9
Lymph node	52 (<1)	88.2	88.2
Mixed	10,438 (75)	68.8-100	98.2
Ovary	52 (<1)	96	96
Parathyroid	152 (1)	99.1-99.6	99.4
Skin	194 (1)	93.9-100	96.9
Unknown	15 (<1)	100	100
Total	13,996 (100)	54-100	96.9
Conventional microscopy			
Mixed	120,479 (>99)	92.9-98.6	98.3
Ovary	517 (<1)	87.8-99.3	90.9
Total	120,996 (100)	87.8-99.3	98.3

Discussion

Herein we present an updated systematic review focused on the use of TP in IOC. For the 49 articles included in our review incorporating 13,996 cases handled by TP, we identified a mean concordance of 96.9% between TP and LM per study. Most (64%) these studies showed concordance rates above 95%. The TAT for IOC performed by TP was infrequently reported (n = 7) but averaged 17 minutes per case. By organ system (Table 3), we were surprised to find that the CNS cases had a relatively low weighted mean (90.5%) compared with organ systems. This is also disturbing since the CNS was the most studied organ system after the "mixed" cases category. One explanation is that CNS cases are inherently more challenging than other organ systems. Another possibility is that the weighted means will fall in other organ system categories as more and larger studies are published.

The meta-analysis previously published by Wellnitz et al[5] on the use of TP for IOC reported an overall diagnostic accuracy of 91%. The weighted mean of the concordance rates we found up to 2000 was very similar at 91.1%. From 2000 onward, the weighted mean of the concordance rates was 97.2%. Although statistical testing for significance was not performed, this is comparable to the 98.3% weighted mean of concordance rates found from conventional IOC studies. We suspect the improvement in concordance rates was primarily due to improvements in image transmission speeds and quality, as well as increased utilization of RM and WSI rather than static images. Our findings are similar to prior reviews by Wellnitz et al[5] and Bashshur et al,[6] which argued that there is an acceptable loss of overall concordance and increase in overall TAT when using TP for IOC vs conventional (LM) IOC.

The change in popular TP methods is made evident by graphing concordance vs year of publication (Figure 3D). This shows that prior to 2000, static TP was the method most often tested. From 2000 to 2007, RM was more common, whereas after 2007, WSI was the most popular method. Although the now outdated static TP systems appeared to perform worse than newer nonstatic TP systems, we could not find a clear advantage of these newer TP methods over one another. Thus, we feel pathologists should come to an agreement with the referring site about which TP system to use for IOC depending on their need, preference, IT infrastructure, and funding.

The transmission speeds reported by authors have greatly increased from 64 kilobytes per second in 1991 to 100 megabytes per second in 2018, which has allowed for larger, higher-quality images to be quickly transmitted. Transmission speeds in the gigabyte per second range are now becoming more common. A frequent technical problem was internet slowness and outright internet failures, which contributed to longer TATs and even cancelling the IOC.[27,55] The transmission speed may be related to accuracy since pathologists are inclined to view fewer images before making a diagnosis when there is considerable lag in image transmission. The most common technical problem in studies performed in the past 10 years was failure of the scanner to digitize slides.[19,60,61] Cima et al[60] reported that the failure to digitize slides occurred in cases with high fat content, fibrosis, and necrosis. Despite occasional technical issues, most authors were generally satisfied with the performance of their TP systems. No adverse medical-legal outcomes were reported.

Deferrals were inconsistently reported across studies and ranged from none reported to 30%.[47] The highest deferral rate was attributed to a lack of appropriate overview
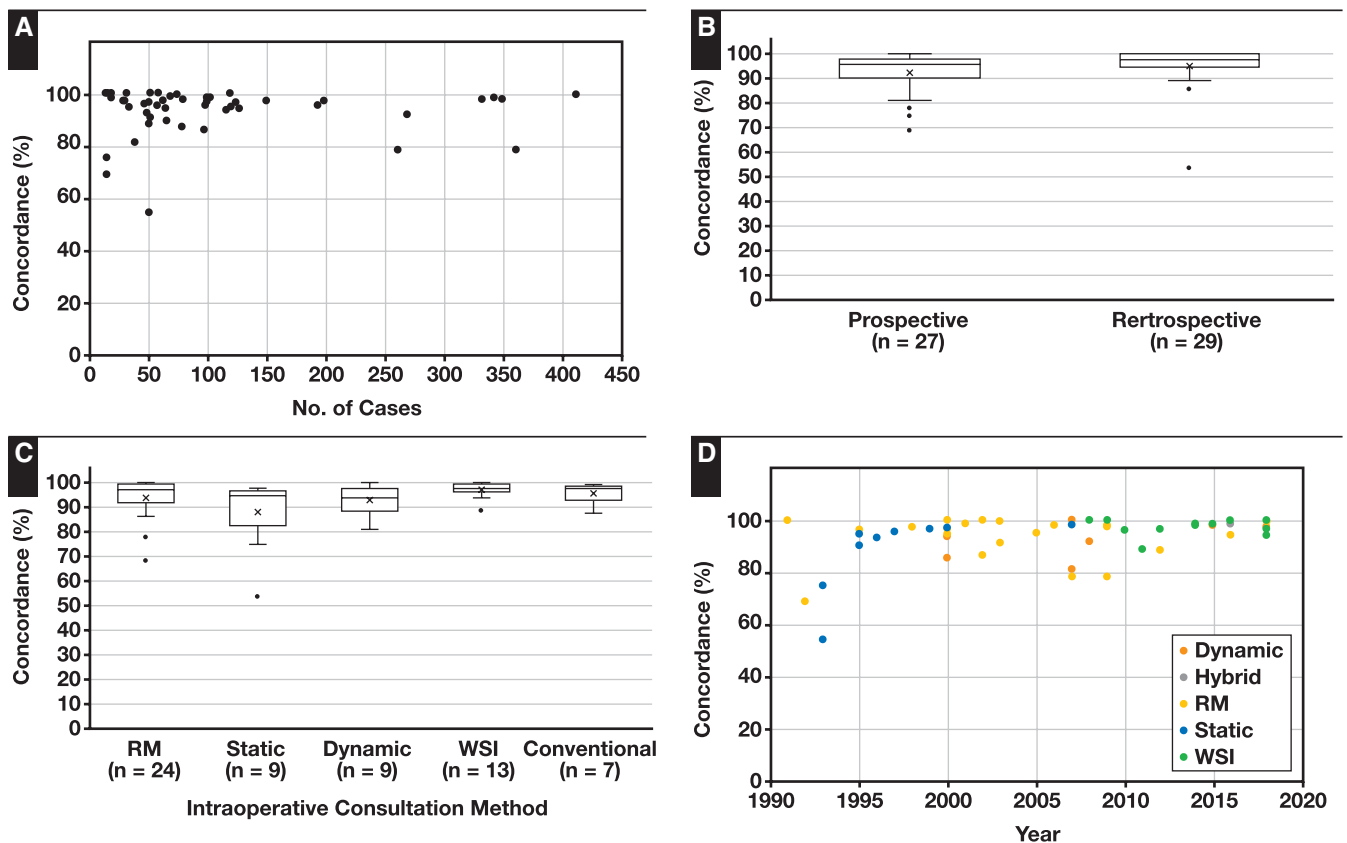
Figure 3 The means calculated in the scatterplots are weighted by the number of cases. RM, robotic microscopy; WSI, whole-slide image. **A**, Concordance vs the number of cases. Most studies have concordance rates more than 90%. Many studies did not reach the College of American Pathologists' validation requirement of 80 cases. Three studies with total numbers of cases more than 450 were omitted from the graph with study sizes of 633, 1,204, and 1,329 and concordance rates of 98%, 98.6%, and 98%, respectively. **B**, Box-and-whisker plot of concordance vs study type. The means of both prospective (92.5%) and retrospective (94.9%) studies are shown as an "X," and the median is the bar within the box. **C**, Box-and-whisker plot of concordance vs intraoperative consultation method. The means of RM (93.9%), static (88.3%), dynamic (93.1%), WSI (97.3%), and conventional microscopy (95.9%) are shown. The single study with a hybrid method (concordance = 98.3%) was omitted from this graph. **D**, Concordance vs year of publication. Points are color coded by method. Eras of popular telepathology methods are more apparent. Static images predominated prior to 2000, RM from 2000 to 2007, and WSI from 2007 to present.

images.[47] Winokur et al[13] attributed uncertainty to image quality, the nature of the specimen, and insufficient expertise. Other authors attributed deferrals to uncertainty in diagnosis,[28] and these were more common in benign or reactive conditions, lymphomas, and unusual cases.[17] Several authors noted that deferral rates were similar to or lower than conventional IOC.[12,17]

The precise reasons for discrepancies varied between studies, but several problems were common among studies. Preanalytical factors such as poor tissue sampling, staining, and slide preparation as well as lack of knowledge of gross findings were common causes of discrepancies reported.[8,11,14,16,25,28,42,48,50] Wellnitz et al[5] noted that in prospective studies the gross examination is likely

to be performed by a nonpathologist, which increases the risk of inappropriate tissue selection for microscopic examination. Tissue selection is only needed in larger, more complex specimens and will likely not affect smaller biopsy specimens in most cases.

Technical error involving unfocused images and lack of images of diagnostic areas also was frequently reported.[11,12,49] Several studies cited possible lack of familiarity of the TP system as contributing to diagnostic discrepancies.[26,44,45] Similar to conventional IOC, small foci of cancer were missed with higher frequency than larger tumor foci.[46,52] Borderline cases, benign/reactive conditions, lymphomas, and rare diagnoses also caused difficulty in TP diagnoses.[17,56-58] In a few studies, it should

Table 4

Modified QUADAS-2 Assessment of Individual Studies

First Author	Risk of Bias				Applicability Concerns		
	Patient Selection	Index Test	Reference Standard	Flow and Timing	Patient Selection	Index Test	Reference Standard
Nordrum[4]	Unclear	Low	Low	Low	Low	Low	Low
Shimosato[36]	Unclear	Unclear	Unclear	Unclear	Unclear	Unclear	Unclear
Becker[10]	Low	Unclear	Low	Low	Low	Low	Low
Oberholzer[37]	Low	Low	Low	Low	Low	Low	Low
Fujita[38]	Low	Low	Low	Low	Low	Low	Low
Nordrum[39]	Unclear	Low	Unclear	Low	Unclear	Low	Low
Oberholzer[40]	Low	Low	Low	Low	Low	Low	Low
Adachi[41]	Unclear	Unclear	Low	Low	Unclear	Low	Low
Weinstein[42]	Unclear	Unclear	Low	Low	Low	Low	Low
Winokur[13]	Unclear	Unclear	Unclear	Unclear	Unclear	High	Unclear
Della Mea[43]	Unclear	Unclear	Low	Low	Unclear	Low	Low
Baak[44]	Low	Low	Low	Unclear	Low	Low	Low
Dawson[45]	Unclear	Low	Low	Low	Low	Low	Low
Della Mea[46]	Low	Unclear	Low	Low	Low	Low	Low
Hufnagl[47]	Unclear	Low	Low	Unclear	Low	Low	Low
Ongürü[12]	Low	Low	Low	Low	Low	Low	Low
Stauch[11]	Low	Low	Low	Low	Low	Low	Low
Winokur[48]	Low	Low	Low	Low	Low	Low	Low
Demichelis[49]	Low	Low	Low	Low	Low	Low	Low
Chorneyko[26]	Low	Unclear	Low	Low	Low	Low	Low
Kaplan[50]	Unclear	Low	Low	Low	Low	Low	Low
Nehal[51]	Low	Low	Low	Low	Low	Low	Unclear
Hutarew[52]	Low	Low	Low	Low	Low	Low	Low
Moser[25]	Unclear	Unclear	Low	Low	Unclear	Low	Low
Hitchcock[28]	Low	Low	Low	Low	Low	Low	Low
Hutarew[53]	Low	Low	Low	Low	Low	Low	Low
Frierson[14]	Low	Low	Low	Low	Low	Low	Low
Horbinski[17]	Low	Low	Low	Low	Low	Low	Low
McKenna[54]	Low	Unclear	Unclear	High	High	High	Low
Liang[55]	High	Low	Low	Unclear	High	Low	Low
Tsuchihashi[23]	Unclear	Unclear	Unclear	High	Unclear	Unclear	Unclear
Evans[27]	Low	Low	Low	Low	Low	Low	Low
Horbinski[56]	Low	Low	Low	Low	Low	Low	Low
Slodkowska[21]	Unclear	Low	Low	Low	Low	Low	Low
Fallon[57]	Low	Low	Low	Low	Low	Low	Low
Ramey[19]	Low	Low	Low	Low	Low	Low	Low
Gifford[15]	Low	Low	Low	Low	Low	Low	Low
Gould[16]	Unclear	Unclear	Low	Unclear	Unclear	Unclear	Low
Ribback[20]	Low	Low	Low	Low	Low	Low	Low
Têtu[22]	Low	Low	Low	Unclear	Low	Low	Low
Bauer[58]	Low	Low	Low	Low	Low	Low	Low
Vitkovski[30]	High	Low	Low	Unclear	High	Unclear	Low
Hufnagl[59]	Low	Low	Low	Low	Low	Low	Low
Pradhan[18]	Unclear	Unclear	Unclear	Unclear	Unclear	Low	Low
Chandraratnam[8]	Low	Low	Low	Low	Low	Low	Low
Cima[60]	Low	Low	Low	Low	Low	Low	Low
French[61]	Low	Low	Low	Low	Low	Low	Low
Huang[29]	Unclear	Low	Unclear	Low	Low	Low	Unclear
Vosoughi[24]	Low	Low	Low	Low	Low	Low	Low

QUADAS-2, Quality Assessment of Diagnostic Accuracy Studies 2.

be noted that TP diagnosis from a specialist resulted in a better diagnosis than what would have been reported by a general pathologist using conventional microscopy.[30,55]

There were several limitations to this systematic review. We found that the reporting area with the highest risk of bias and concern for applicability was in the patient selection domain (Table 4 and Figure 4). Pathologists who publish studies validating TP systems should specify exactly how the patients were selected and how randomization was performed. The studies we found with high risk in this domain chose which patients to include based on their opinion without predetermined criteria or only
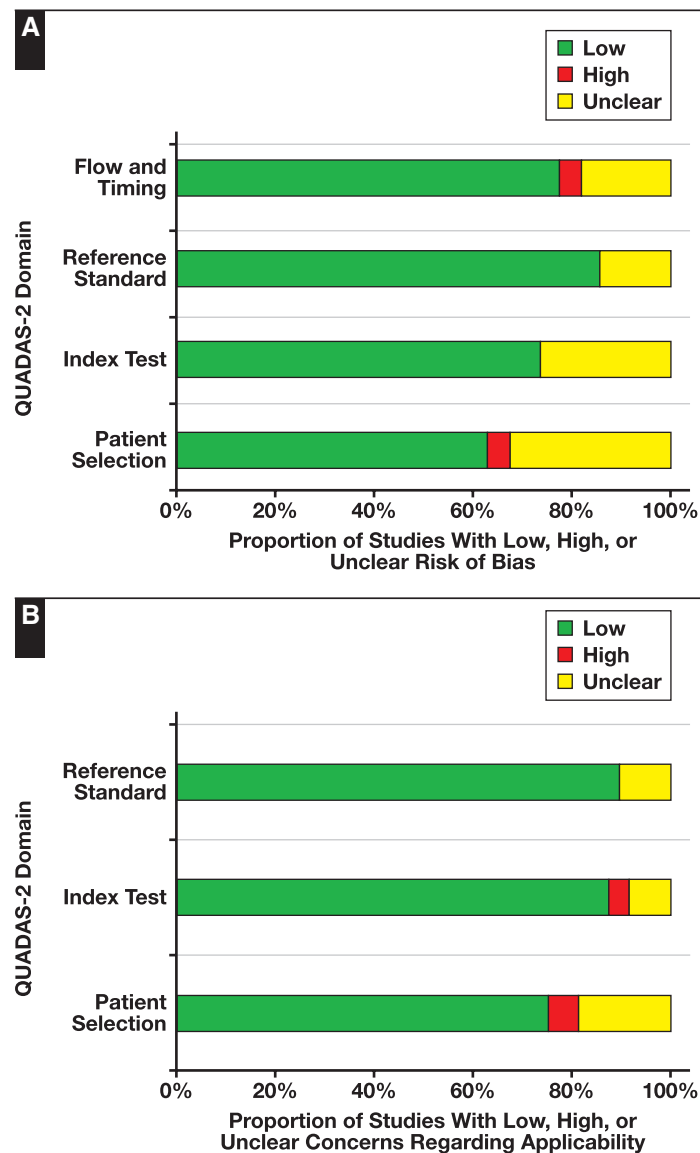
Figure 4 Quality Assessment of Diagnostic Accuracy Studies 2 (QUADAS-2) quality assessment results. **A**, Proportion of studies reviewed (n = 49) with low, high, or unclear risk of bias. **B**, Proportion of studies reviewed (n = 49) with low, high, or unclear concerns regarding applicability. Graphs adopted from the template for graphical display from the QUADAS-2 resource page.[33]

included cases that were screened for high risk of malignancy. Our advice to improve future publications in this field would be to be specific on the criteria used to select cases included in the validation.

There was unfortunately too much diversity in the study design and study reporting to conduct a meta-analysis. Future publications about validation of TP systems should better specify how patients were selected and how randomization was performed, in addition to providing detailed information about deferrals and discrepancies.

Increasing subspecialization in pathology and consolidated health systems are likely to increase the demand

for TP. These two factors work together, but they are somewhat separate. Consolidated health systems involve workload balancing to reduce the total number of employees. New technologies, such as TP, enable the system to provide partial services at different geographic sites at an efficient cost (eg, splitting pathologists' time). Subspecialization also encourages the use of this technology because specific organ frozen sections can be evaluated by subspecialized experts.

In summary, this systematic review shows that TP systems exhibit comparable performance to IOC performed by conventional LM with respect to concordance

and TAT. The small gap between TP and LM concordance rates will likely decrease as technology continues to improve and digital pathology becomes an essential part of pathology training and practice.

References

1. Weinstein RS. Prospects for telepathology. *Hum Pathol.* 1986;17:433-434.

2. Dietz RL, Hartman DJ, Zheng L, et al. Review of the use of telepathology for intraoperative consultation. *Expert Rev Med Devices.* 2018;15:883-890.

3. Pantanowitz L, Dickinson K, Evans AJ, et al. American Telemedicine Association clinical guidelines for telepathology. *J Pathol Inform.* 2014;5:39.

4. Nordrum I, Engum B, Rinde E, et al. Remote frozen section service: a telepathology project in northern Norway. *Hum Pathol.* 1991;22:514-518.

5. Wellnitz U, Binder B, Fritz P, et al. Reliability of telepathology for frozen section service. *Anal Cell Pathol.* 2000;21:213-222.

6. Bashshur RL, Krupinski EA, Weinstein RS, et al. The empirical foundations of telepathology: evidence of feasibility and intermediate effects. *Telemed J E Health.* 2017;23:155-191.

7. Moher D, Liberati A, Tetzlaff J, et al; PRISMA Group. Preferred Reporting Items for Systematic Reviews and Meta-Analyses: the PRISMA statement. *J Clin Epidemiol.* 2009;62:1006-1012.

8. Chandraratnam E, Santos LD, Chou S, et al. Parathyroid frozen section interpretation via desktop telepathology systems: a validation study. *J Pathol Inform.* 2018;9:41.

9. Barasch SL, Li Z, Stewart J III. Telepathology for frozen section analysis: a validation of remote meeting technologies (RMT) software. https://www.uwhealth.org/files/uwhealth/docs/pdf6/telepathology_telehealth.pdf. Accessed December 12, 2018.

10. Becker RL Jr, Specht CS, Jones R, et al. Use of remote video microscopy (telepathology) as an adjunct to neurosurgical frozen section consultation. *Hum Pathol.* 1993;24:909-911.

11. Stauch G, Schweppe KW, Kayser K. Diagnostic errors in interactive telepathology. *Anal Cell Pathol.* 2000;21:201-206.

12. Ongürü O, Celasun B. Intra-hospital use of a telepathology system. *Pathol Oncol Res.* 2000;6:197-201.

13. Winokur TS, McClellan S, Siegal GP, et al. An initial trial of a prototype telepathology system featuring static imaging with discrete control of the remote microscope. *Am J Clin Pathol.* 1998;110:43-49.

14. Frierson HF Jr, Galgano MT. Frozen-section diagnosis by wireless telepathology and ultra portable computer: use in pathology resident/faculty consultation. *Hum Pathol.* 2007;38:1330-1334.

15. Gifford AJ, Colebatch AJ, Litkouhi S, et al. Remote frozen section examination of breast sentinel lymph nodes by telepathology. *ANZ J Surg.* 2012;82:803-808.

16. Gould PV, Saikali S. A comparison of digitized frozen section and smear preparations for intraoperative neurotelepathology. *Anal Cell Pathol (Amst).* 2012;35:85-91.

17. Horbinski C, Fine JL, Medina-Flores R, et al. Telepathology for intraoperative neuropathologic consultations at an academic medical center: a 5-year report. *J Neuropathol Exp Neurol.* 2007;66:750-759.

18. Pradhan D, Monaco SE, Parwani AV, et al. Evaluation of panoramic digital images using Panoptiq for frozen section diagnosis. *J Pathol Inform.* 2016;7:26.

19. Ramey J, Fung KM, Hassell LA. Use of mobile high-resolution device for remote frozen section evaluation of whole slide images. *J Pathol Inform.* 2011;2:41.

20. Ribback S, Flessa S, Gromoll-Bergmann K, et al. Virtual slide telepathology with scanner systems for intraoperative frozen-section consultation. *Pathol Res Pract.* 2014;210:377-382.

21. Słodkowska J, Pankowski J, Siemiatkowska K, et al. Use of the virtual slide and the dynamic real-time telepathology systems for a consultation and the frozen section intra-operative diagnosis in thoracic/pulmonary pathology. *Folia Histochem Cytobiol.* 2009;47:679-684.

22. Têtu B, Perron É, Louahlia S, et al. The Eastern Québec Telepathology Network: a three-year experience of clinical diagnostic services. *Diagn Pathol.* 2014;9(suppl 1):S1.

23. Tsuchihashi Y, Takamatsu T, Hashimoto Y, et al. Use of virtual slide system for quick frozen intra-operative telepathology diagnosis in Kyoto, Japan. *Diagn Pathol.* 2008;3(suppl 1):S6.

24. Vosoughi A, Smith PT, Zeitouni JA, et al. Frozen section evaluation via dynamic real-time nonrobotic telepathology system in a university cancer center by resident/faculty cooperation team. *Hum Pathol.* 2018;78:144-150.

25. Moser PL, Lorenz IH, Sögner P, et al. The accuracy of telediagnosis of frozen sections is inferior to that of conventional diagnosis of frozen sections and paraffin-embedded sections. *J Telemed Telecare.* 2003;9:130-134.

26. Chorneyko K, Giesler R, Sabatino D, et al. Telepathology for routine light microscopic and frozen section diagnosis. *Am J Clin Pathol.* 2002;117:783-790.

27. Evans AJ, Chetty R, Clarke BA, et al. Primary frozen section diagnosis by robotic microscopy and virtual slide telepathology: the University Health Network experience. *Semin Diagn Pathol.* 2009;26:165-176.

28. Hitchcock CL, Hitchcock LE. Three years of experience with routine use of telepathology in assessment of excisional and aspirate biopsies of breast lesions. *Croat Med J.* 2005;46:449-457.

29. Huang Y, Lei Y, Wang Q, et al. Telepathology consultation for frozen section diagnosis in China. *Diagn Pathol.* 2018;13:29.

30. Vitkovski T, Bhuiya T, Esposito M. Utility of telepathology as a consultation tool between an off-site surgical pathology suite and affiliated hospitals in the frozen section diagnosis of lung neoplasms. *J Pathol Inform.* 2015;6:55.

31. Randell R, Ambepitiya T, Mello-Thoms C, et al. Effect of display resolution on time to diagnosis with virtual pathology slides in a systematic search task. *J Digit Imaging.* 2015;28:68-76.

32. Norgan AP, Suman VJ, Brown CL, et al. Comparison of a medical-grade monitor vs commercial off-the-shelf display for mitotic figure enumeration and small object (*Helicobacter pylori*) detection. *Am J Clin Pathol.* 2018;149:181-185.

33. Goacher E, Randell R, Williams B, et al. The diagnostic concordance of whole slide imaging and light microscopy: a systematic review. *Arch Pathol Lab Med.* 2017;141:151-161.

34. Whiting PF, Rutjes AW, Westwood ME, et al; QUADAS-2 Group. QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. *Ann Intern Med.* 2011;155:529-536.

35. Higgins J, Green S. *Cochrane Handbook for Systematic Reviews of Interventions.* Version 5.1.0. The Cochrane Collaboration; 2011. http://handbook-5-1.cochrane.org. Accessed January 12, 2019.

36. Shimosato Y, Yagi Y, Yamagishi K, et al. Experience and present status of telepathology in the National Cancer Center Hospital, Tokyo. *Zentralbl Pathol.* 1992;138:413-417.

37. Oberholzer M, Fischer HR, Christen H, et al. Telepathology with an integrated services digital network—a new tool for image transfer in surgical pathology: a preliminary report. *Hum Pathol.* 1993;24:1078-1085.

38. Fujita M, Suzuki Y, Takahashi M, et al. The validity of intraoperative frozen section diagnosis based on video-microscopy (telepathology). *Gen Diagn Pathol.* 1995;141:105-110.

39. Nordrum I, Eide TJ. Remote frozen section service in Norway. *Arch Anat Cytol Pathol.* 1995;43:253-256.

40. Oberholzer M, Fischer HR, Christen H, et al. Telepathology: frozen section diagnosis at a distance. *Virchows Arch.* 1995;426:3-9.

41. Adachi H, Inoue J, Nozu T, et al. Frozen-section services by telepathology: experience of 100 cases in the San-in District, Japan. *Pathol Int.* 1996;46:436-441.

42. Weinstein LJ, Epstein JI, Edlow D, et al. Static image analysis of skin specimens: the application of telepathology to frozen section evaluation. *Hum Pathol.* 1997;28:30-35.

43. Della Mea V, Cataldi P, Boi S, et al. Image sampling in static telepathology for frozen section diagnosis. *J Clin Pathol.* 1999;52:761-765.

44. Baak JP, van Diest PJ, Meijer GA. Experience with a dynamic inexpensive video-conferencing system for frozen section telepathology. *Anal Cell Pathol.* 2000;21:169-175.

45. Dawson PJ, Johnson JG, Edgemon LJ, et al. Outpatient frozen sections by telepathology in a Veterans Administration medical center. *Hum Pathol.* 2000;31:786-788.

46. Della Mea V, Cataldi P, Pertoldi B, et al. Combining dynamic and static robotic telepathology: a report on 184 consecutive cases of frozen sections, histology and cytology. *Anal Cell Pathol.* 2000;20:33-39.

47. Hufnagl P, Bayer G, Oberbamscheidt P, et al. Comparison of different telepathology solutions for primary frozen section diagnostic. *Anal Cell Pathol.* 2000;21:161-167.

48. Winokur TS, McClellan S, Siegal GP, et al. A prospective trial of telepathology for intraoperative consultation (frozen sections). *Hum Pathol.* 2000;31:781-785.

49. Demichelis F, Barbareschi M, Boi S, et al. Robotic telepathology for intraoperative remote diagnosis using a still-imaging-based system. *Am J Clin Pathol.* 2001;116:744-752.

50. Kaplan KJ, Burgess JR, Sandberg GD, et al. Use of robotic telepathology for frozen-section diagnosis: a retrospective trial of a telepathology system for intraoperative consultation. *Mod Pathol.* 2002;15:1197-1204.

51. Nehal KS, Busam KJ, Halpern AC. Use of dynamic telepathology in Mohs surgery: a feasibility study. *Dermatol Surg.* 2002;28:422-426.

52. Hutarew G, Dandachi N, Strasser F, et al. Two-year evaluation of telepathology. *J Telemed Telecare.* 2003;9:194-199.

53. Hutarew G, Schlicker HU, Idriceanu C, et al. Four years experience with teleneuropathology. *J Telemed Telecare.* 2006;12:387-391.

54. McKenna JK, Florell SR. Cost-effective dynamic telepathology in the Mohs surgery laboratory utilizing iChat AV videoconferencing software. *Dermatol Surg.* 2007;33:62-68.

55. Liang WY, Hsu CY, Lai CR, et al. Low-cost telepathology system for intraoperative frozen-section consultation: our experience and review of the literature. *Hum Pathol.* 2008;39:56-62.

56. Horbinski C, Hamilton RL. Application of telepathology for neuropathologic intraoperative consultations. *Brain Pathol.* 2009;19:317-322.

57. Fallon MA, Wilbur DC, Prasad M. Ovarian frozen section diagnosis: use of whole-slide imaging shows excellent correlation between virtual slide and original interpretations in a large series of cases. *Arch Pathol Lab Med.* 2010;134:1020-1023.

58. Bauer TW, Slaw RJ. Validating whole-slide imaging for consultation diagnoses in surgical pathology. *Arch Pathol Lab Med.* 2014;138:1459-1465.

59. Hufnagl P, Guski H, Hering J, et al. Comparing conventional and telepathological diagnosis in routine frozen section service. *Diagn Pathol.* 2016;2:1-19.

60. Cima L, Brunelli M, Parwani A, et al. Validation of remote digital frozen sections for cancer and transplant intraoperative services. *J Pathol Inform.* 2018;9:34.

61. French JMR, Betney DT, Abah U, et al. Digital pathology is a practical alternative to on-site intraoperative frozen section diagnosis in thoracic surgery. *Histopathology.* 2019;74:902-907.

62. Ferreiro JA, Myers JL, Bostwick DG. Accuracy of frozen section diagnosis in surgical pathology: review of a 1-year experience with 24,880 cases at Mayo Clinic Rochester. *Mayo Clin Proc.* 1995;70:1137-1141.

63. Hashmi AA, Naz S, Edhi MM, et al. Accuracy of intraoperative frozen section for the evaluation of ovarian neoplasms: an institutional experience. *World J Surg Oncol.* 2016;14:91.

64. Mahe E, Ara S, Bishara M, et al. Intraoperative pathology consultation: error, cause and impact. *Can J Surg.* 2013;56:E13-E18.

65. Novis DA, Gephardt GN, Zarbo RJ; College of American Pathologists. Interinstitutional comparison of frozen section consultation in small hospitals: a College of American Pathologists Q-Probes study of 18,532 frozen section consultation diagnoses in 233 small hospitals. *Arch Pathol Lab Med.* 1996;120:1087-1093.

66. Sams SB, Wisell JA. Discordance between intraoperative consultation by frozen section and final diagnosis. *Int J Surg Pathol.* 2017;25:41-50.

67. Wasinghon P, Suthippintawong C, Tuipae S. The accuracy of intraoperative frozen sections in the diagnosis of ovarian tumors. *J Med Assoc Thai.* 2008;91:1791-1795.

68. White VA, Trotter MJ. Intraoperative consultation/final diagnosis correlation: relationship to tissue type and pathologic process. *Arch Pathol Lab Med.* 2008;132:29-36.