

CADTH Health Technology Review

Digital Pathology Using Primary Case Sign-Out

Authors: Rob Edge, Aleksandra Grobelna

ISSN: 2563-6596

Disclaimer: The information in this document is intended to help Canadian health care decision-makers, health care professionals, health systems leaders, and policy-makers make well-informed decisions and thereby improve the quality of health care services. While patients and others may access this document, the document is made available for informational purposes only and no representations or warranties are made with respect to its fitness for any particular purpose. The information in this document should not be used as a substitute for professional medical advice or as a substitute for the application of clinical judgment in respect of the care of a particular patient or other professional judgment in any decision-making process. The Canadian Agency for Drugs and Technologies in Health (CADTH) does not endorse any information, drugs, therapies, treatments, products, processes, or services.

While care has been taken to ensure that the information prepared by CADTH in this document is accurate, complete, and up to date as at the applicable date the material was first published by CADTH, CADTH does not make any guarantees to that effect. CADTH does not guarantee and is not responsible for the quality, currency, propriety, accuracy, or reasonableness of any statements, information, or conclusions contained in any third-party materials used in preparing this document. The views and opinions of third parties published in this document do not necessarily state or reflect those of CADTH.

CADTH is not responsible for any errors, omissions, injury, loss, or damage arising from or relating to the use (or misuse) of any information, statements, or conclusions contained in or implied by the contents of this document or any of the source materials.

This document may contain links to third-party websites. CADTH does not have control over the content of such sites. Use of third-party sites is governed by the third-party website owners' own terms and conditions set out for such sites. CADTH does not make any guarantee with respect to any information contained on such third-party sites and CADTH is not responsible for any injury, loss, or damage suffered as a result of using such third-party sites. CADTH has no responsibility for the collection, use, and disclosure of personal information by third-party sites.

Subject to the aforementioned limitations, the views expressed herein are those of CADTH and do not necessarily represent the views of Canada's federal, provincial, or territorial governments or any third-party supplier of information.

This document is prepared and intended for use in the context of the Canadian health care system. The use of this document outside of Canada is done so at the user's own risk.

This disclaimer and any questions or matters of any nature arising from or relating to the content or use (or misuse) of this document will be governed by and interpreted in accordance with the laws of the Province of Ontario and the laws of Canada applicable therein, and all proceedings shall be subject to the exclusive jurisdiction of the courts of the Province of Ontario, Canada.

The copyright and other intellectual property rights in this document are owned by CADTH and its licensors. These rights are protected by the Canadian *Copyright Act* and other national and international laws and agreements. Users are permitted to make copies of this document for non-commercial purposes only, provided it is not modified when reproduced and appropriate credit is given to CADTH and its licensors.

About CADTH: CADTH is an independent, not-for-profit organization responsible for providing Canada's health care decision-makers with objective evidence to help make informed decisions about the optimal use of drugs, medical devices, diagnostics, and procedures in our health care system.

Funding: CADTH receives funding from Canada's federal, provincial, and territorial governments, with the exception of Quebec.

Questions or requests for information about this report can be directed to Requests@CADTH.ca

Table of Contents

Abbreviations	6
Key Messages	7
Context and Policy Issues	7
Research Questions	8
Methods	8
Literature Search Methods.....	8
Selection Criteria and Methods	8
Exclusion Criteria.....	9
Critical Appraisal of Individual Studies	9
Summary of Evidence.....	9
Quantity of Research Available.....	9
Summary of Study Characteristics.....	10
Summary of Critical Appraisal.....	12
Summary of Findings	13
Limitations.....	15
Conclusions and Implications for Decision- or Policy-Making	16
References	18
Appendix 1: Selection of Included Studies	19
Appendix 2: Characteristics of Included Publications	20
Appendix 3: Critical Appraisal of Included Publications.....	24
Appendix 4: Main Study Findings and Authors' Conclusions	30

List of Tables

Table 1: Selection Criteria.....	9
Table 2: Characteristics of Included Systematic Reviews.....	20
Table 3: Characteristics of Included Primary Clinical Studies	20
Table 4: Strengths and Limitations of Systematic Reviews and Meta-Analyses Using AMSTAR 2 ¹⁰	24
Table 5: Strengths and Limitations of Clinical Studies Using QUADAS-2 ¹¹	24
Table 6: Summary of Findings Included Systematic Reviews	30
Table 7: Summary of Findings of Included Primary Clinical Studies.....	31

List of Figures

Figure 1: Selection of Included Studies 19

Abbreviations

AMSTAR 2	A MeaSurement Tool to Assess systematic Reviews 2
BCSS	breast cancer specific survival
CADTH	Canadian Agency for Drugs and Technologies in Health
CAP	College of American Pathologists
CAP-PLQC	College of American Pathologists - Pathology and Laboratory Quality Center
COI	conflict of interest
DMFS	distant metastasis free survival
LM	light microscopy
PRISMA	Preferred Reporting Items for Systematic Review and Meta-Analysis
PROSPERO	International Prospective Register of Systematic Reviews
QUADAS-2	Quality Assessment of Diagnostic Accuracy Studies 2
SR	systematic review
WSI	whole slide image

Key Messages

- The applicability of the identified findings and the potential impact on patient outcomes for any individual diagnostic setting was unclear.
- In the majority of the identified studies, the authors reported diagnostic accuracy and clinical utility outcomes that supported digital pathology systems as a valuable diagnostic modality, comparable to conventional microscopy. These studies lacked statistical power calculations, making the accuracy of these statements unclear.
- One systematic review and 1 diagnostic study reported clinical utility outcomes of digital pathology. This evidence supported digital pathology using primary case sign-out for accurate prognosis of patient outcomes; however, the clinical utility compared to conventional microscopy was unclear in the identified evidence.
- One systematic review and 13 diagnostic cohort studies reported on the diagnostic accuracy of whole slide image (WSI). The identified outcomes indicated that WSI is a valuable diagnostic modality; however, a large range of diagnostic accuracy in different settings, and a lack of clear statistical power in all studies make comparator conclusions to conventional microscopy unclear.
- One systematic review and 4 diagnostic cohort studies reported diagnostic areas that can present challenges for a digital pathology implementation, the most common being the interpretation and grading of dysplasia.
- One identified systematic review stressed the importance of whole-system validation to identify strengths and weaknesses of specific digital pathology implementations. The range of diagnostic accuracy across studies also indicated that implementation of digital pathology primary case sign-out systems is associated with unclear diagnostic accuracy until appropriately validated.
- No relevant cost-effectiveness evidence for digital pathology using primary case sign-out was identified.

Context and Policy Issues

Digital pathology using primary case sign-out utilizes systems that digitize glass slides of patient specimens to produce a whole slide image (WSI). Traditionally glass slides are evaluated by a pathologist using a conventional light microscope to provide a diagnosis, with most diagnoses requiring multiple slides. WSIs can be rapidly deployed to pathologists for primary case sign-out systems and viewed on a wide variety of digital displays for diagnostics to provide some efficiencies, as well as services to underprivileged and remote areas.¹ Digital pathology using primary case sign-out with WSIs may also have other advantages over glass slides such as ease of archiving, research, teaching, remote expert consultation, improved ergonomics, side-by-side comparisons, larger field of vision, workflow improvements, and quantification of prognostic parameters.^{2,3} Furthermore, algorithm-based pathological diagnostics using WSIs are in development, with current top-performing automated methods comparable to concordance among pathologists.⁴ The clear benefits of this technology, in addition to the logistical pressures of COVID-19, are accelerating adoption of this technology.⁵

Digital pathology systems are considered to be comprised of 2 subsystems: and image acquisition component (i.e., the scanner) and the image viewer.¹ There is a range of Health Canada–approved digital pathology systems available, in addition to validation guidelines

from the College of American Pathologists (CAP) and the Royal College of Pathologists (RCPATH).^{1,6,7} The CAP guidelines state that each pathology laboratory should perform their own validation study, for each clinical use.⁸

This report is an update to a previously published CADTH Reference List report (October 2021).⁹ This report aims to retrieve and review the full-text of this reference list, critically appraise, and summarize the evidence for the clinical utility, diagnostic accuracy, and cost-effectiveness of digital pathology using primary case sign-out.

Research Questions

1. What is the clinical utility for digital pathology using primary case sign-out?
2. What is the diagnostic accuracy of digital pathology using primary case sign-out?
3. What is the cost-effectiveness of digital pathology using primary case sign-out?

Methods

Literature Search Methods

This report makes use of a literature search developed for a previous CADTH report.⁹ For this previous report, a limited literature search was conducted by an information specialist on key resources including MEDLINE, the Cochrane Database of Systematic Reviews, the international health technology assessment (HTA) database, the websites of Canadian and major international health technology agencies, as well as a focused internet search. The search strategy comprised both controlled vocabulary, such as the National Library of Medicine's MeSH (Medical Subject Headings), and keywords. The main search concept was digital pathology. CADTH-developed search filters were applied to limit retrieval to health technology assessments, systematic reviews (SRs), meta-analyses, or network meta-analyses, any types of clinical trials or observational studies and economic studies. Where possible, retrieval was limited to the human population. The search was also limited to English language documents published between January 1, 2016 and October 4, 2021.

Selection Criteria and Methods

One reviewer screened literature search results (titles and abstracts) and selected publications according to the inclusion criteria presented in Table 1. The full text of study publications were not reviewed, but were included in a previously published CADTH Reference List report (October 2021).⁹

In this report a second reviewer screened full-text articles selected for the previously published CADTH Reference List report.⁹ The final selection of full-text articles was again based on the inclusion criteria presented in Table 1.

Exclusion Criteria

Articles were excluded if they did not meet the selection criteria outlined in Table 1, or were published before 2016; however, studies that did not provide any clinical utility evidence (research question 1) were excluded if they were published before 2019. Primary studies retrieved by the search were excluded if they were captured in 1 or more included SRs.

Critical Appraisal of Individual Studies

The included publications were critically appraised by 1 reviewer using the following tools as a guide: A MeaSurement Tool to Assess systematic Reviews 2 (AMSTAR 2)¹⁰ for SRs, and the Quality Assessment of Diagnostic Accuracy Studies 2 (QUADAS-2) checklist¹¹ for diagnostic test accuracy studies. Summary scores were not calculated for the included studies; rather, the strengths and limitations of each included publication were described narratively.

Summary of Evidence

Quantity of Research Available

A total of 38 citations were articles selected for a previous CADTH report (October 2021),⁹ all of which were retrieved for full-text review. Of these potentially relevant articles, 23 publications were excluded for various reasons, and 15 publications met the inclusion criteria and were included in this report. These comprised 2 SRs and 13 diagnostic cohort studies. One SR and 1 diagnostic study reported clinical utility outcomes of digital pathology, while the other SR and all diagnostic cohort studies reported on the diagnostic accuracy of WSI. No studies were identified that examined the cost-effectiveness of digital pathology. Appendix 1 presents the Preferred Reporting Items for Systematic Review and Meta-Analysis (PRISMA)¹² flow chart of the study selection.

Table 1: Selection Criteria

Criteria	Description
Population	Patients suspected of disease requiring histopathology for clinical diagnosis
Intervention	Digital pathology using primary case sign-out in any setting (any digital pathology including WSI, algorithms for dedicated morphometric analysis, algorithms employing artificial intelligence [AI]/ machine learning, natural language processing, and novel microscopic techniques [e.g., multispectral, Fourier transform infrared and other infrared, and second harmonic generation imaging])
Comparator	Standard microscopic evaluation in a lab setting
Outcomes	Q1: Clinical Utility (e.g., benefits and harms, adverse events, safety considerations [i.e., correct patient diagnosis], patient management, patient satisfaction, QoL). Q2: Diagnostic accuracy (e.g., sensitivity, specificity, concordance) Q3: Cost-effectiveness (e.g., cost per QALY gained [i.e., ICER], cost per adverse event avoided)
Study designs	HTA, SRs, randomized controlled trials, non-randomized studies, and economic evaluations

HTA = health technology assessment; ICER = incremental cost-effectiveness ratio; QALY = quality-adjusted life-year; QoL = quality of life; SR = systematic review; WSI = whole slide imaging.

Summary of Study Characteristics

Additional details regarding the characteristics of included publications are provided in Appendix 2.

Study Design

Two SRs met the inclusion criteria presented in Table 1.^{6,13} Araujo et al. did not report any criteria for publication date in the search methodology for diagnostic accuracy studies; however, it only included studies that adhered to the College of American Pathologists Pathology and Laboratory Quality Center (CAP-PLQC) guidelines.^{13,14} These guidelines are recommendations, suggestions, and expert consensus opinion aimed at standardizing validation study methodology.¹⁴ Williams et al. published a SR in 2017 that used a previous systematic electronic literature search for studies published between 1999 and December 2015, which did not specify studies that adhered to CAP-PLQC.^{6,15} This SR met the inclusion criteria because it reports clinical utility outcomes, and was published after 2016.⁶

This report identified and included 13 diagnostic cohort studies that all used a single-gate approach and blinded observers.^{1-5,7,8,16-21} Five of the included studies prospectively examined a diagnostic cohort of current cases,^{3,5,7,18,19} while the remaining 8 studies retrospectively examined a diagnostic cohort of cases.^{1,2,4,8,16,17,20,21} The prospective studies used a consecutive series of current patient cases.^{3,5,7,18,19} One retrospective study randomly selected cases,¹⁶ while 7 used a curated sample of cases intended to be representative.^{1,2,4,8,17,20,21} Davidson et al. used a retrospective representative sample of cases; however, this study also uniquely randomly allocated a large number of pathologist readers to 1 of the 2 diagnostic modalities twice.⁴

Country of Origin

The SRs included in this report originated from Brazil (Araujo et al.)¹³ and the UK (Williams et al.).⁶

The primary clinical studies included in this report were conducted in Italy,^{2,20} Brazil,³ India,^{5,7,8} the US,^{4,16-19} Saudi Arabia,¹ and the UK.²¹ There were no studies identified in this report that originated or were conducted in Canada.^{1,6}

Patient Population

Neither of the included SRs specified a patient population in the systemic search criteria. Araujo et al. described the diagnostic cases as being on slides from dermatologic, central nervous system, gastrointestinal, genitourinary, breast, liver, pediatric organ systems, with subsets from endocrine, head and neck, hematopoietic, hepatobiliary-pancreatic, soft tissue, bone, hematopathology, medical kidney, and transplant biopsies.¹³ Williams et al. did not provide a detailed list of the organ systems from which diagnostic cases originated other than to report that the most common organ system was gastrointestinal, followed by studies that examined a mixed population.⁶

Seven primary diagnostic studies focused on a particular diagnostic area of pathologist expertise.^{1-4,8,20,21} These diagnostic areas included atypical meningiomas,² neuropathology,¹ oral and maxillofacial cases,³ breast cancer,^{4,21} pancreatic solid lesions,²⁰ and prostate core biopsies.⁸ Six primary diagnostic studies had a broader focus on diagnostic accuracy and included cases representing many different organ classes and tissues.^{5,7,16-19}

Interventions and Comparators

Both included SRs examined any digital WSI compared to light microscopy (LM) which was also described as any conventional microscopy by Araujo et al.^{6,13}

Similarly, all primary diagnostic cohort studies also compared WSI to LM.^{1-5,7,8,16-21} While every study provided some details on the scanner used to digitize glass slides, 4 studies provided some detail on the light microscope(s) used,^{1,2,18,20} and 9 provided some details on the hardware and/or software used to examine the WSIs.^{3,4,7,16-21} One study reported on a breast algorithm from Visiopharm (Denmark), without any additional description.⁷ None of the studies described any diagnostic methods as multispectral, Fourier transform infrared, other infrared, or second harmonic generation imaging. All available details on the intervention and comparator hardware and software reported by the primary diagnostic cohort studies are provided in Appendix 2.

The experience and subspecialties of the pathologists reading glass slides or WSI are an essential component of both examined diagnostic modalities and likely impact diagnostic accuracy.^{2,3,8,13} The reporting of the experience and specialties of the reading pathologists was not consistent in the identified studies with 1 study not reporting the experience of the participating pathologists at all.¹⁷ Reading participants were described as expert pathologists,²⁰ senior pathologists,¹⁻³ or residents.¹⁻³ Five studies reported the years of experience of participating pathologists.^{1,4,7,18,19} Additionally, pathologists from various subspecialties were included as readers in 6 studies and as described as neuropathologist,¹ uropathologist⁸ head and neck pathology specialist,⁵ breast pathology specialist,^{5,19,21} gastrointestinal pathology specialist,^{5,16,19} thoracic specialist,⁵ bone and soft tissue specialist,^{5,19} gynecologic specialist,^{5,19} genitourinary specialist,^{5,19} and dermatopathologist.¹⁹ As randomization was conducted in the study by Davidson et al. at the level of the reading pathologist, this study provided additional detail on the experience of the participating pathologists.⁴ The training of pathologists in the use of digital pathology systems, regardless of pathology experience, may also impact diagnostic accuracy of WSI. Three of the included primary diagnostic cohort studies specifically stated that observers had no digital pathology training,^{1,4,21} 5 did not report any information regarding the training of observers,^{2,16-18,20} while 5 reported at least some observer training was completed before initiation of the study.^{3,5,7,8,19}

Outcomes

The 2 SRs reported discordances, which was a focus of the SR by Williams et al.^{6,13} Araujo et al. also summarized a range of intra-observer concordances as reported by included studies.¹³

All included primary diagnostic cohort studies reported intra-observer concordance, that is the degree of agreement between LM and WSI for the same observer.^{1-5,7,8,16-21} Three primary diagnostic cohort studies included measures of inter-observer concordance, reflecting the agreement between different observers for LM and WSI.^{2,4,20} Three studies reported inter and intra-observer concordances using κ , a statistical measure of agreement between observations that ranges from 0 to 1, where 1 represents complete agreement and 0 represents agreement that would be expected by random chance.^{3,20,21} Larghi et al. also reported diagnostic accuracy outcomes by using a historical definitive diagnosis as the gold standard compared to new observations using LM and WSI.²⁰ Additional outcomes reported in this identified body of evidence may have implications for the implementation of digital pathology using primary case sign-out and may include deferral rate,^{5,7,17} diagnostic turnaround time,^{7,8,17,19,20} and slide rescan rate.^{5,7,17,19} Borowsky et al. uniquely provided an

overall discrepancy rate as well as a discrepancy rate broken down by tissue type.¹⁷ Rakha et al. provided an analysis of the association of histological grade as determined by LM and WSI with 2 clinical utility outcomes, breast cancer specific survival (BCSS), and distant metastasis free survival (DMFS).²¹ Ammendola et al. also provided data on the prognostic accuracy for the recurrence of atypical meningiomas.²

Summary of Critical Appraisal

The 2 SRs included in this report had many methodological strengths. A significant difference between the 2 SRs is that Williams et al. relied on a prior SR for literature inclusion,¹⁵ and while the authors provided methodology for the systematic literature search, literature search selection, duplicate literature screening, and data extraction, similar to Araujo et al,¹³ it did not conduct a critical appraisal or report the risk of bias of identified body of evidence.⁶ Both SRs provided a defined research objective and registered the protocol with PROSPERO.^{6,13} Additionally, the SR of Araujo et al. followed PRISMA guidelines, and had a statement of no conflicts of interest (COI).¹³ Williams et al. reported that 1 author is on the advisory board and conducts collaborative projects with a WSI device manufacturer.⁶ Both SRs conducted minimal quantitative analysis of the identified evidence and described findings narratively, and Williams et al. synthesized clinical utility evidence regarding the potential impact of discordances.^{6,13} An unclear risk of bias associated with case selection was reported by Araujo et al. in the included studies. Additionally, a high-risk of bias associated with the threshold definitions used for diagnostic concordance, otherwise the identified body of evidence identified by Araujo et al. was evaluated as at low concern for bias.¹³

Critical appraisal of the included primary diagnostic studies revealed some common strengths and limitations throughout this body of evidence. The blinding of observers,^{1-5,7,8,16,17,19-21} consistent evaluation of cases,^{1-5,7,8,16-21} defined outcomes,^{1-5,7,8,16-21} and the role of investigators^{1-5,7,17-20} was well described in most, if not all of the studies, which minimized the potential impact of measurement bias in body of evidence. In all but 1 study, there were no clear instances of inappropriate case exclusion.^{1-5,7,8,17-21} Critical appraisal identified an unclear risk of selection bias in this body of evidence where 3 studies excluded cases before slide scanning,^{7,16,17} 6 studies selected representative cases,^{1,4,8,17,20,21} and/or used a single representative slide for each case.^{1,2,4,8,20,21} Random case selection was described in 2 studies; however, in the context of these diagnostic cohort study designs was not akin to the randomization of patients in a randomized controlled trial.^{7,16} One study design was unique where observers were randomized twice to LM or WSI diagnostic interventions for representative cases. Therefore the observers in this study could be randomized to 1 diagnostic modality followed by the other, or randomized to the same modality twice.⁴ Five of the 13 studies were prospective, in that the cases were live patient cases evaluated by both diagnostic modalities.^{3,5,7,18,19} Four of these prospective studies did not select cases and instead evaluated a consecutive cohort of patients, which would minimize the potential for selection bias.^{3,5,18,19} None of the included studies provided any sample size justification,^{1-5,7,8,16-21} including a study that the authors described as a noninferiority study.¹⁷ The training of the observers with regard to pathologist experience and specialty was reported in 12 studies,^{1-5,7,8,16,18-21} however 4 studies did not report training on the digital pathology system.^{1,2,17,20} Every study reported a washout period (i.e., the time between observations by alternate diagnostic modalities to prevent the observer recalling the diagnosis as determined by the previous modality), which had a considerable range; 2 days,^{8,18} 2 weeks,^{5,16} 3 to 6 weeks,² 1 month,^{3,17} 8 weeks,¹ 3 months,^{7,20,21} 13 weeks,¹⁹ and 9 months.⁴ The applicability of the findings within this body of evidence had strengths including

that observers used a variety of hardware for WSI diagnosis in 7 studies,^{1,3-5,7,18,20} and a variety of LM in 2 studies,^{1,20} which may represent a more realistic remote diagnostic setting. Eleven studies also provided helpful insights from the author's perspective on the limitations of their studies.^{1-5,7,17-21} Within this body of evidence there were 5 studies that reported a potential COI,¹⁷⁻¹⁹ or did not provide a COI statement.^{1,16}

Additional details regarding the strengths and limitations of included publications are provided in Appendix 3.

Summary of Findings

Appendix 4 presents the main study findings and authors' conclusions.

Clinical Utility of Digital Pathology Using Primary Case Sign-out

Williams et al. conducted an SR focused on outcomes of discordance and the potential clinical impact of the discordances in the identified evidence body. The authors summarized 335 discordances, out of a total 8,069 diagnoses (4% discordance). The largest category of discordances was missed diagnoses of malignant, dysplastic, or atypical conditions where malignant tissues were diagnosed as benign. Of a total 109 discordances in this category, 101 of the preferred diagnoses agreed with conventional microscopy over WSI. Over all categories 335 discordances were examined, 28 of which (0.35% of total diagnoses), had the potential to cause moderate to severe patient harm. It was also reported that of the 335 discordances, 169 (50.4%) were determined to be of appreciable diagnostic difficulty and recognized inter-observer variation.⁶

Rakha et al. conducted a large diagnostic study of breast cancer cases (n = 1,675) that reported diagnostic accuracy in addition to a survival analysis by examining histological grade association as determined by LM and WSI, with BCSS and with DMFS. Grading with either LM or WSI, regardless of the observer, demonstrated strong association with both clinical outcomes. Individual WSI graded components demonstrated statistically significant differences for BCSS and DMFS. LM graded histological components showed stronger association with BCSS and DMFS than WSI with the exception of tubule formation; however, these differences were not statistically significant.²¹

Ammendola et al. examined the prognostic accuracy of WSI and LM for atypical meningioma recurrence following surgical resection. High mitotic index was the histological parameter with the most predictive power for recurrence using either WSI or LM. The observed greater predictive accuracy of WSI for high mitotic index, brain invasion, and sheeting as compared to LM did not reach statistical significance.²

Diagnostic Accuracy of Digital Pathology Using Primary Case Sign-out

Concordance and Diagnostic Accuracy

All included studies reported diagnostic concordance outcomes, except for Williams et al.^{1-5,7,8,13,16-21} Araujo et al. conducted an SR which identified 13 studies that reported on the concordance of WSI as compared to LM. The intra-observer concordance ranged from 87% to 98.3%, with a κ coefficient range from 0.8 to 0.98 indicating excellent agreement.¹³ In a diagnostic cohort study published in 2021, Ramaswamy et al. conducted a retrospective validation on breast cancer cases followed by a prospective analysis of a wider range of histological subspecialties and found a major intra-observer concordance between WSI and LM of 100%, when minor discordances were included the intra-observer concordance was

98.9%. The authors also briefly reported that a breast algorithm assessment had between 97.2% and 100% concordance for different breast biomarkers.⁷ In another analysis of a wide range of pathologies, 3 observers demonstrated a major intra-observer concordance of 100%, with a minor discordance of 1.1%.⁵ Two prospective validation studies on wide ranging pathologies were conducted at Memorial Sloan Kettering Cancer Center, published in 2019¹⁹ and 2020.¹⁸ The first study found an intra-observer diagnostic concordance of 99.3%, and an intra-observer grade concordance of 94.1% among 8 observers.¹⁹ This study was followed by a study that found a major intra-observer concordance of 100% among 12 observers with a minor discordance rate of 1.1%.¹⁸ Samuelson et al. used validation methodology for WSI in compliance with the CAP guidance for remote sign-out validation and observed a major intra-observer concordance between 5 observers untrained in WSI of 94.7%, and an overall concordance of 83.62% with LM when examining a wide range of pathologies.¹⁶ A study by Borowsky et al. examined surgical pathology for primary diagnosis and found an intra-observer concordance of 96.1% between WSI and LM. The largest major discrepancy rate difference between LM and WSI compared to the definitive diagnosis was observed for skin diagnoses, where WSI exceeded LM by 2.3%. LM had a larger major discrepancy for salivary gland diagnoses by 1.14%.¹⁷ Ammendola et al. determined the diagnostic accuracy of WSI as compared to LM for grading atypical meningioma, and found greater inter-observer concordance between senior pathologists than between residents for both diagnostic modalities, and a higher inter-observer concordance using WSI than LM for all histological components except for mitotic index. Intra-observer concordance for atypical meningioma was 89%. Histological components with the highest intra-observer concordance were sheeting and small cells (96%), while the lowest intra-observer concordance was observed for high mitotic index (78%) where all observers classified more cases as having high mitotic index by WSI than by LM.² Araujo et al., examined diagnostic accuracy of WSI for oral and maxillofacial pathology found an intra-observer agreement between WSI and LM where κ ranged from 0.85 to 0.98 which indicated excellent agreement.³ A study examining prostate core biopsies reported a major intra-observer concordance of 100%, with a minor discordance of 1.2%.⁸ Neuropathology cases, examined by Alassiri et al. demonstrated an intra-observer concordance of 82.1% that included 10% major discordances and 7.9% minor discordances between WSI and LM. The authors concluded that formally trained neuropathologists would provide more accurate diagnoses using WSI.¹ A well-conducted retrospective study from Davidson et al. twice randomly assigned 208 pathologists to either WSI or LM to grade breast cancer cases and found an intra-observer grade concordance of 73% when LM was assigned twice, 68% when WSI was assigned twice, and 63% when observers were switched from once diagnostic modality to the other. None of the intra-observer concordance differences were statistically significant; however, significant differences were observed for inter-observer concordance. The inter-observer concordance for Nottingham grading of breast cancer was 68% in the first assignment and 69% in the second assignment to LM, whereas the inter-observer concordance was 60% in the first assignment and 62% in the second assignment to WSI. The authors concluded that WSI may be associated with increased variability between pathologists in assignment of Nottingham grade for invasive breast carcinomas.⁴ An intra-observer agreement of 68% between WSI and LM for the exact grade of breast cancer was also reported by Rakha et al. This study found moderate overall concordance of grade between WSI and LM; however, 1 histological component, pleomorphism, was only of fair agreement ($\kappa = 0.27$).²¹ In another retrospective study, there were no statistically significant differences between the intra- or inter-observer for WSI and LM for diagnostic classification or histological components of pancreatic solid lesions. This study by Larghi et al. also reported diagnostic performance measures, which were also not significantly different between

WSI and LM. The sensitivity and specificity of LM was 0.92 and 0.96, respectively, while the sensitivity was 0.93 and the specificity was 0.88 for WSI.²⁰

Discordances

Four studies provided some additional information on discordances between WSI and LM.^{3,6,13,21} Studies identified in the SR by Araujo et al. reported that in instances of discordance a minority (37.3%) agreed with WSI over conventional microscopy to the preferred diagnoses.¹³ Both SRs provided narrative conclusions that some areas of pathological diagnoses present diagnostic difficulties.^{6,13} Williams et al. conclude that their analysis of the discordances reveals specific areas that present problematic diagnostic challenges for WSI and that awareness of these areas is important. Furthermore, to create accurate awareness of these areas, Williams et al. recommend that diagnostic departments conduct in-house validations for WSI to evaluate the strengths and weaknesses of their specific systems for primary case sign-out diagnosis.⁶ A prospective study by Araujo et al. observed that most discordances were found on dysplasia grading, and differentiation between severe dysplasia and microinvasive oral squamous cell carcinoma.³ A study examining breast carcinoma identified a major discordance of 1.5% of which significantly more WSI diagnoses were of the lower grade than the LM diagnoses ($P < 0.00001$).²¹

Deferral Rate

Three studies reported a deferral rate for WSI. Two studies reported deferral rates for WSI but not for the LM gold standard, both for a wide range of pathologies of 0.34%,⁷ and 4.5%.⁵ Borowsky et al. reported a deferral rate for WSI of 3.5% and 3.3% for LM; however, the statistical significance was not reported.¹⁷

Delayed Diagnosis

With regard to the implementation of digital pathology primary case sign-out systems there were 2 statistically significant findings of an increased time to diagnose with WSI.^{8,20} Three additional studies also observed an increased WSI diagnostic time that was not statistically significant; however, it is unclear if those studies were sufficiently powered to detect differences in these outcomes.^{7,17,19}

Rescan Rate

When slides are scanned for WSI systems they may have to be rescanned for a variety of reasons which could have decreased the efficiency of digital pathology. Rescan rates of 0.33%,⁷ 0.67%,¹⁷ 2.3%,⁵ and 7%.¹⁹ were reported by 4 studies.

Cost-Effectiveness of Digital Pathology Using Primary Case Sign-out

No cost-effectiveness evidence for digital pathology using primary case sign-out was identified.

Limitations

One limitation of this report is that some studies did not examine digital pathology primary, case sign-out in a remote setting; however the study design intentions were to examine digital pathology for primary diagnosis in a potential remote scenario and therefore these studies were included. A lack of prospective studies looking at clinical utility outcomes also

limited the ability to draw conclusions regarding important patient centred outcomes when diagnosed by digital pathology using primary case sign-out. The applicability of the findings from diagnostic accuracy studies of the body of evidence is unclear, as it is not associated with clinical utility and contains significant variation in study design, intervention, and population. None of the identified studies were conducted in Canada and the applicability to the Canadian health care setting is unclear. However, narrative introductions of 2 studies cited literature to report that Canada is 1 of a limited number of examples that utilize WSI for large-scale primary diagnostic purposes.^{1,6}

Conclusions and Implications for Decision- or Policy-Making

Three studies, 1 SR and 2 diagnostic cohort studies, reported clinical utility outcomes.^{2,6,21} The SR found that 0.35% of disagreements between WSI and LM diagnostic modalities had the potential to cause moderate to severe patient harm, the largest category of these discrepancies was the missed diagnosis of malignant, dysplastic, or atypical conditions. LM was the preferred diagnostic modality for 94% of discrepancies in this category, indicating to the authors that the diagnosis of dysplasia may be a pitfall of digital pathology.⁶ The 2 diagnostic cohort studies found that LM and WSI offer significant diagnostic predictive power for atypical meningioma recurrence,² and significant association with breast cancer survival.²¹ Neither diagnostic cohort study demonstrated a significant difference between the 2 diagnostic modalities in prognostic accuracy, however it is not clear that either study was sufficiently powered to do so.^{2,21} This evidence supported digital pathology using primary case sign-out for accurate prognosis of patient outcomes, however the clinical utility compared to conventional microscopy was unclear in the identified evidence.

Diagnostic accuracy was examined in 1 SR and 13 diagnostic cohort studies.^{1-5,7,8,13,16-21} The SR was evaluated as having few limitations, and evaluated a body of evidence that consisted of 13 diagnostic cohort studies, none of which were also included in this report. The SR evaluated the included studies as having minor concerns of bias and reported a concordance between WSI and LM of between 87% and 98.3%. The majority of discordances (62.7%) agreed with LM as the preferred diagnosis. Specific findings within certain areas of pathology were identified as being more challenging for WSI diagnosis including dermatopathology, pediatric pathology, neuropathology, and gastrointestinal pathology.¹³ Thirteen studies examined the diagnostic accuracy of digital pathology and met the inclusion criteria of this report. A wide range of pathologies, pathologist specialties, pathologist experience, and digital pathology platforms were examined in this evidence, but all compared the diagnostic accuracy of WSI as compared to LM.^{1-5,7,8,16-21} The body of evidence overall was at potential risk of selection bias, although 4 prospective diagnostic cohort studies avoided this and had few relevant concerns of potential biases in the reported methodology.^{3,5,18,19} The breadth of diagnostic settings examined in these studies was reflected in the wide range of reported intra-observer concordances and author expectations of intra-observer concordances between WSI and LM (Appendix 4). All 13 identified studies reported intra-observer concordance, and the authors of 11 of these studies reported that the intra-observer concordances supported WSI as a valuable diagnostic modality, comparable to LM.^{1-5,7,8,17-20} This included a mean overall intra-observer concordance that ranged from 82.1% in a setting of neuropathological diagnoses,¹ to a mean overall intra-observer concordance of 98.9%

in 2 studies in a setting of diverse pathological diagnoses.^{5,18} The authors of a diagnostic validation study on a variety of pathologies expressed concern regarding the range of intra-observer concordance (75.5% to 92.2%), and reported that perhaps validation studies should aim for a range of diagnostic concordance rather than a fixed mean.¹⁶ Similar to the SR of Williams, 4 diagnostic cohort studies reported that the area of most discordance involved dysplasia grading and atypical diagnosis.^{2-4,21} Other outcomes that may potentially inform the implementation of a digital pathology system identified in this evidence are the rescan rate,^{5,7,17,19} delayed diagnoses,^{7,8,17,19,20} and referral rate.^{5,7,17}

The authors of 1 SR stated that it is “important that diagnostic departments perform their own whole-system validations for WSI, to evaluate the strengths and weaknesses of the combination of hardware and software components they propose to use for primary diagnosis.” (p. 1717)⁶ This report identified 8 diagnostic cohort studies that were conducted specifically to validate a digital pathology primary case sign-out system before full implementation.^{1,3,5,7,8,16,18,19} These studies reported a range of validation methodology, adhered to different validation standards, articulated implementation concerns, and provided concordance data across different diagnostic settings.

Lastly, no relevant cost-effectiveness evidence for digital pathology using primary case sign-out was identified.

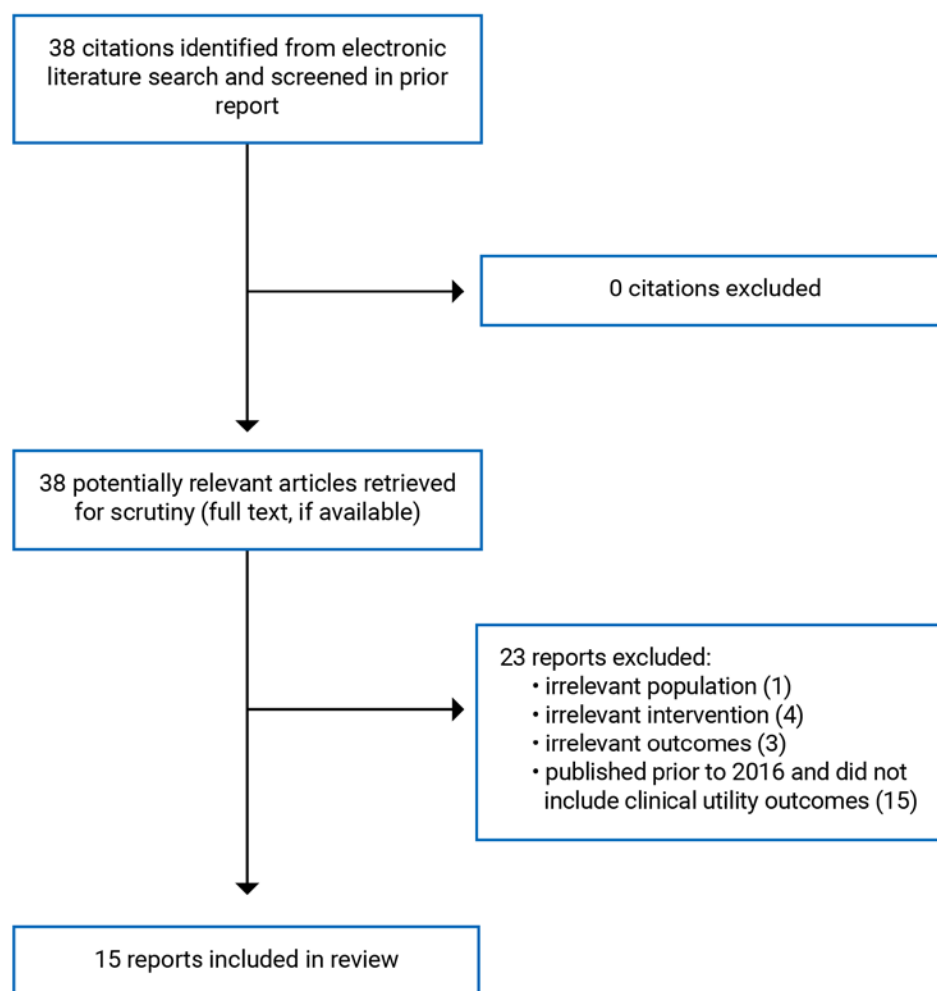
This report identified a range of diagnostic accuracy among studies that suggested implementation of digital pathology primary case sign-out systems is associated with unclear diagnostic accuracy until appropriately validated.

References

1. Alassiri A, Almutrafi A, Alsufiani F, et al. Whole slide imaging compared with light microscopy for primary diagnosis in surgical neuropathology: a validation study. *Ann Saudi Med.* 2020;40(1):36-41. [PubMed](#)
2. Ammendola S, Bariani E, Eccher A, et al. The histopathological diagnosis of atypical meningioma: glass slide versus whole slide imaging for grading assessment. *Virchows Arch.* 2021;478(4):747-756. [PubMed](#)
3. Araujo ALD, do Amaral-Silva GK, Perez-de-Oliveira ME, et al. Fully digital pathology laboratory routine and remote reporting of oral and maxillofacial diagnosis during the COVID-19 pandemic: a validation study. *Virchows Arch.* 2021;479(3):585-595. [PubMed](#)
4. Davidson TM, Rendi MH, Frederick PD, et al. Breast cancer prognostic factors in the digital era: comparison of Nottingham grade using whole slide images and glass slides. *J Pathol Inform.* 2019;10:11. [PubMed](#)
5. Rao V, Kumar R, Rajaganesan S, et al. Remote reporting from home for primary diagnosis in surgical pathology: a tertiary oncology center experience during the COVID-19 pandemic. *J Pathol Inform.* 2021;12:3. [PubMed](#)
6. Williams BJ, DaCosta P, Goacher E, Treanor D. A systematic analysis of discordant diagnoses in digital pathology compared with light microscopy. *Arch Pathol Lab Med.* 2017;141(12):1712-1718. [PubMed](#)
7. Ramaswamy V, Tejaswini BN, Uthaiiah SB. Remote reporting during a pandemic using digital pathology solution: experience from a tertiary care cancer center. *J Pathol Inform.* 2021;12:20. [PubMed](#)
8. Rao V, Subramanian P, Sali AP, Menon S, Desai SB. Validation of whole slide imaging for primary surgical pathology diagnosis of prostate biopsies. *Indian J Pathol Microbiol.* 2021;64(1):78-83. [PubMed](#)
9. Hill S, Grobelna A. Digital pathology using primary case sign-out. (CADTH Rapid response report: reference list). Ottawa (ON): CADTH; 2021: <https://www.cadth.ca/sites/default/files/pdf/htis/2021/RA1193%20Digital%20Pathology%20Final.pdf>. Accessed 2021 Oct 20.
10. Shea BJ, Reeves BC, Wells G, et al. AMSTAR 2: a critical appraisal tool for systematic reviews that include randomised or non-randomised studies of healthcare interventions, or both. *BMJ.* 2017;358:j4008. [PubMed](#)
11. Whiting PF, Rutjes AW, Westwood ME, et al. QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. *Ann Intern Med.* 2011;155(8):529-536. [PubMed](#)
12. Liberati A, Altman DG, Tetzlaff J, et al. The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate health care interventions: explanation and elaboration. *J Clin Epidemiol.* 2009;62(10):e1-e34. [PubMed](#)
13. Araújo ALD, Arboleda LPA, Palmier NR, et al. The performance of digital microscopy for primary diagnosis in human pathology: a systematic review. *Virchows Arch.* 2019;474(3):269-287. [PubMed](#)
14. Pantanowitz L, Sinard JH, Henricks WH, et al. Validating whole slide imaging for diagnostic purposes in pathology: guideline from the College of American Pathologists Pathology and Laboratory Quality Center. *Arch Pathol Lab Med.* 2013;137(12):1710-1722. [PubMed](#)
15. Goacher E, Randell R, Williams B, Treanor D. The diagnostic concordance of whole slide imaging and light microscopy: a systematic review. *Arch Pathol Lab Med.* 2017;141(1):151-161. [PubMed](#)
16. Samuelson MI, Chen SJ, Boukhar SA, et al. Rapid validation of whole-slide imaging for primary histopathology diagnosis. *Am J Clin Pathol.* 2021;155(5):638-648. [PubMed](#)
17. Borowsky AD, Glassy EF, Wallace WD, et al. Digital whole slide imaging compared with light microscopy for primary diagnosis in surgical pathology. *Arch Pathol Lab Med.* 2020;144(10):1245-1253. [PubMed](#)
18. Hanna MG, Reuter VE, Ardon O, et al. Validation of a digital pathology system including remote review during the COVID-19 pandemic. *Mod Pathol.* 2020;33(11):2115-2127. [PubMed](#)
19. Hanna MG, Reuter VE, Hameed MR, et al. Whole slide imaging equivalency and efficiency study: experience at a large academic center. *Mod Pathol.* 2019;32(7):916-928. [PubMed](#)
20. Larghi A, Fornelli A, Lega S, et al. Concordance, intra- and inter-observer agreements between light microscopy and whole slide imaging for samples acquired by EUS in pancreatic solid lesions. *Dig Liver Dis.* 2019;51(11):1574-1579. [PubMed](#)
21. Rakha EA, Aleskandarani M, Toss MS, et al. Breast cancer histologic grading using digital microscopy: concordance and outcome association. *J Clin Pathol.* 2018;71(8):680-686. [PubMed](#)

Appendix 1: Selection of Included Studies

Figure 1: Selection of Included Studies



Appendix 2: Characteristics of Included Publications

Note that this appendix has not been copy-edited.

Table 2: Characteristics of Included Systematic Reviews

Study citation, country, funding source	Study designs and numbers of primary studies included	Population characteristics	Intervention and comparator(s)	Outcomes
Araujo, 2019, Brazil ¹³ Funding: CAPES/ PROEX, CNPq, FAPESP	Diagnostic cohort studies (n = 13)	Slides from organ systems: dermatologic, CNS, gastrointestinal, genitourinary, breast, liver, pediatric. Subsets included endocrine, head and neck, hematopoietic organ, hepatobiliary-pancreatic organ, soft tissue, bone, hematopathology, medical kidney and transplant biopsies.	WSI Comparator: any conventional microscopy	Concordance: intra-observer Discordance analysis
Williams, 2017, UK ⁶ Funding: partial funding from Sectra AB (Linköping, Sweden), Leica Biosystems (Vista, CA), FFEI Ltd (Hemel Hempstead, Hertfordshire, England)	This study used the systematic review of Goacher, 2017 ¹⁵ to examine instances of discordance from the WSI validation literature 38 diagnostic studies: crossover (n = 6), prospective cohort (n = 19), retrospective cohort (n = 13)	Slides from unreported organ systems, the most common being gastrointestinal (n = 7), and mixed (n = 10).	WSI Comparator: LM	Discordance between WSI and LM instances: potential for harm, preferred diagnostic medium, attribution of discordance

CAPES/PROEX = Coordination for the Improvement of Higher Education Personnel; CNPq = National Council for Scientific and Technological Development; CNS = central nervous system; FAPESP = Sao Paulo Research Foundation; LM = light microscopy; WSI = whole slide image.

Table 3: Characteristics of Included Primary Clinical Studies

Study citation, country, funding source	Study design	Population characteristics	Intervention and comparator(s)	Outcomes
Ammendola, 2021, Italy ² Funding: University of Verona	Diagnostic cohort Case samples (n = 35) selected randomly evaluated by 2 senior pathologists and 2 residents	Atypical meningiomas, a single representative slide per case	WSI: NR Scanner: NanoZoomer S360 Digital slide scanner (Hamamatsu Photonics) LM: Nikon Eclipse 80i light microscope with a x 10/22 mm micrometer eyepiece	Concordance: intra-rater and inter-rater Prognostic accuracy for recurrence

Study citation, country, funding source	Study design	Population characteristics	Intervention and comparator(s)	Outcomes
Araujo, 2021, Brazil ³ Funding: CAPES/PROEX, CNPq, FAPESP	Diagnostic consecutive cohort Case samples evaluated by 1 pathologist and 3 trainees	Oral and maxillofacial cases (n = 162)	WSI: Various consumer grade workstations Scanner: Aperio Digital Pathology System (Leica Biosystems, Wetzlar, Germany) LM: NR	Concordance: intra-rater and inter-rater
Ramaswamy, 2021, India ⁷ Funding: None	Diagnostic cohort Retrospective case samples were selected randomly and evaluated by 3 pathologists for validation. Followed by 886 prospective cases	Retrospective cases from breast (n = 100) Prospective cases from breast, head and neck, gastrointestinal, female reproductive organs, urogenital and male reproductive system, soft tissue and bone, lung, mediastinum, pleura, lymph nodes, CNS, skin, ear, endocrine organs (n = 886, slides = 2,142)	WSI: Various consumer grade workstations, Breast algorithm (Visiopharm (Denmark)) Scanner: An FDA-approved Philips UFS 300 (Ultrafast scanner 300) scanner with Image Management System (IMS) software LM:NR	Concordance, deferral rate, turnaround time, rescan rate
Rao, 2021(1), India ⁸ Funding: None	Diagnostic cohort Representative case samples for training (n = 10) and for validation (n = 60) evaluated by 3 pathologists	Prostate core biopsies representing benign and malignant prostate pathology (n = 70)	WSI: NR Scanner: Panoramic MIDI II scanner (3DHISTECH; Budapest, Hungary) LM: NR	Concordance: intra-rater, read times
Rao, 2021(2), India ⁵ Funding: None	Diagnostic cohort Live case samples for training (n = 10) and for validation in real-time environment (n = 594) evaluated by 18 pathologists	Head and neck, breast, gastrointestinal thoracic pathology gynecologic pathology, genitourinary, bone and soft tissue (n = 594)	WSI: NR remote workstations Scanner: VENTANA DP200 whole-slide scanner (Hemel Hempstead, UK) LM: NR	Concordance, deferrals, rescan rate
Samuelson, 2021, US ¹⁶ Funding: NR	Validation study using diagnostic cohort Case samples were selected randomly for each evaluating pathologist (n = 5) from a large dataset of established LM-based primary diagnoses	Gastrointestinal, gynecologic, head and neck, breast, genitourinary, and dermatologic pathologies (n = 171)	WSI: CaseViewer 2.3.0 (3DHitech). Scanner: P1000 Panoramic scanner (3DHitech) LM: NR	Concordance: intra-rater

Study citation, country, funding source	Study design	Population characteristics	Intervention and comparator(s)	Outcomes
Alassiri, 2020, Saudi Arabia ¹ Funding: None	Validation study using diagnostic cohort Case samples (one representative per case) selected from recent cases (n = 60) for reading by pathologists (n = 4)	A broad range of neuropathological diagnoses (n = 60)	WSI: NR Scanner: Aperio scanner (ScanScope AT Turbo) LM: Pathologist's personal LM	Concordance: intra-rater
Borowsky, 2020, US ¹⁷ Funding: Leica Biosystems Imaging, Inc., Beckman Coulter, Inc., and UC Davis	Diagnostic consecutive cohort study Case samples were selected randomly for each reading pathologist (n > 15) from a large dataset of established LM-based primary diagnoses	Dataset was enriched for difficult diagnostic categories. Breast, prostate, lung/bronchus/larynx/oral cavity/nasopharynx, colorectal, GE junction, stomach, skin, lymph node, bladder, gynecological, liver/bile duct neoplasm, endocrine, brain/CNS, kidney neoplastic, salivary gland, hernial/peritoneal, gallbladder, appendix, soft tissue tumours, anus/perianal (n = 2045 cases, 5,849 slides)	WSI: Dell (Round Rock, TX) workstations with medical-grade monitor Scanner: Aperio AT2 DX system (Leica Biosystems, Inc., Vista, California) LM: NR	Concordance: intra-rater, discrepancy rates by organ type, rescan rate, diagnostic times, deferral rate
Hanna, 2020, US ¹⁸ Funding: partial funding from Paige.AI and PathPresenter	Validation study using diagnostic cohort Case samples were selected randomly for each reading pathologist (n = 12), evaluated on random days representing a day's workload of primary diagnoses	Cases (n = 2,119) from genitourinary, dermatopathology, breast, gastrointestinal, head and neck, bone and soft tissue, gynecologic, neuropathology	WSI: consumer grade workstations Scanner: Aperio GT450 whole slide scanner (Leica Biosystems, Buffalo Grove, Illinois, US). LM: Olympus BX43 (Olympus)	Concordance
Davidson, 2019, US ⁴ Funding: NIH/NCI, Ventana Medical Systems, Inc.	Diagnostic cohort Pathologists (n = 208) randomly assigned to a characterized slide set (WSI or glass), then followed by a second randomization to WSI or glass slide of the same slide set.	Breast cancer cases (n = 22) - full spectrum of breast pathology spanning the Nottingham grade scale	WSI: HD View SL custom viewer Scanner: iScan Coreo Au™ (Ventana Medical Systems, Inc.) LM: NR	Concordance: intra-rater and inter-rater

Study citation, country, funding source	Study design	Population characteristics	Intervention and comparator(s)	Outcomes
Hanna, 2019, US ¹⁹ Funding: Paige.AI and NR	Validation study using diagnostic cohort Active case samples were selected randomly for each reading pathologist (n = 8), evaluated on random days representing a day's workload of primary diagnoses	Cases (WSI = 199, LM = 204) of genitourinary, dermatopathology, breast, gastrointestinal, bone and soft tissue, gynecologic, neuropathology	WSI: MSK Slide Viewer (custom) Scanner: Leica Aperio AT2 (Leica Biosystems, Buffalo Grove, Illinois, US) LM: NR	Concordance defined as not having a significant impact on clinical management Rescan rate Diagnostic time
Larghi, 2019, Italy ²⁰ Funding: None	Validation study using diagnostic cohort Representative cases selected and evaluated by 5 expert pathologists	Pancreatic solid lesion cases (n = 60)	WSI: Aperio ImageScope (Leica Biosystems, Buffalo Grove, IL) software. Scanner: Aperio ScanScope XTscanner (Leica Biosystems, Buffalo Grove, IL) LM: Pathologist's personal LM	Concordance: intra-rater and inter-rater
Rakha, 2018, UK ²¹ Funding: None	Diagnostic consecutive cohort Consecutive cases evaluated by 1 pathologist	Invasive primary operable breast cancer patients (n = 1,675) with long-term clinical follow-up (median = 135 months)	WSI: 3D Histech Panoramic Viewer (3DHISTECH Ltd., Budapest, Hungary) Scanner: 3D Histech Panoramic 250 Flash II scanner (3DHISTECH Ltd., Budapest, Hungary) LM: NR	Concordance: intra-rater Prognostic analysis for BCSS and DMFS

BCSS = breast cancer specific survival; CNS = central nervous system; DMFS = distant metastasis free survival; GE = gastroesophageal; NIH/NCI = national institutes of health/national cancer institute; NR = not reported.

Appendix 3: Critical Appraisal of Included Publications

Note that this appendix has not been copy-edited.

Table 4: Strengths and Limitations of Systematic Reviews and Meta-Analyses Using AMSTAR 2¹⁰

Strengths	Limitations
Systematic Reviews	
Araujo, 2019¹³	
<ul style="list-style-type: none"> • Defined research objective • Literature search selection/inclusion/exclusion methodology clear • Follows PRISMA guidelines and registered protocol with PROSPERO • Literature screened in duplicate • Critical appraisal using validated criteria of included studies in duplicate • Risk of bias of body of evidence assessed • Data extraction methodology described • Statement of no conflict of interest 	<ul style="list-style-type: none"> • Narrative summary only of included evidence • Limited information on included study characteristics
Williams, 2017⁶	
<ul style="list-style-type: none"> • Defined research objective • Literature search selection/inclusion/exclusion methodology clear • Registered protocol with PROSPERO • Literature screened in duplicate • Data extraction methodology described reviewed in triplicate 	<ul style="list-style-type: none"> • Stated conflict of interest • No critical appraisal of included evidence • Narrative summary only of included evidence

AMSTAR 2 = A MeaSurement Tool to Assess systematic Reviews 2; PRISMA = Preferred Reporting Items for Systematic Reviews and Meta-Analysis; PROSPERO = International Prospective Register of Systematic Reviews; NR = not reported; NA = not applicable.

Table 5: Strengths and Limitations of Clinical Studies Using QUADAS-2¹¹

Strengths	Limitations
Ammendola, 2021²	
Risk of Bias <ul style="list-style-type: none"> • No inappropriate case exclusion • All cases evaluated similarly • Blinded pathologists • Role of investigators clear • Wash out period (3 to 6 weeks) • Outcomes well defined • Statement of no COI • Discussion of study limitations Applicability <ul style="list-style-type: none"> • Some data on pathologist training level (senior pathologists and residents) 	Risk of Bias <ul style="list-style-type: none"> • Retrospective cases • Limited data on telepathology training • Single representative slide/case • No statistical power calculation • No slide deidentifying methodology reported Applicability <ul style="list-style-type: none"> • All assessments used same LM • No description of WSI viewer • Diagnostic study only - no clinical outcome data

Strengths	Limitations
Araujo, 2021³	
Risk of Bias <ul style="list-style-type: none"> • Consecutive cases • Prospective cases • No inappropriate case exclusion • All cases evaluated similarly • Blinded pathologists • Role of investigators clear • Wash out period (1 month) • Outcomes well defined • Statement of no COI • Discussion of study limitations Applicability <ul style="list-style-type: none"> • Training level of diagnostic investigators reported • Variety of computer hardware used for remote evaluation 	Risk of Bias <ul style="list-style-type: none"> • One pathologist and 3 trainees as evaluators Applicability <ul style="list-style-type: none"> • Diagnostic study only - no clinical outcome data • No description of LM
Ramaswamy, 2021⁷	
Risk of Bias <ul style="list-style-type: none"> • Randomly selected cases • Prospective validation component • All cases evaluated similarly • Blinded pathologists • Role of investigators clear • Wash out period (3 months) • Outcomes well defined • Statement of no COI • Discussion of study limitations Applicability <ul style="list-style-type: none"> • Training level of diagnostic investigators reported • Variety of computer hardware used for remote evaluation 	Risk of Bias <ul style="list-style-type: none"> • Retrospective component • Cases excluded upon pre-scan QC • No statistical power calculation • No slide deidentifying methodology reported Applicability <ul style="list-style-type: none"> • Diagnostic study only - no clinical outcome data • No description of LM

Strengths	Limitations
Rao, 2021(1)⁸	
Risk of Bias <ul style="list-style-type: none"> • No inappropriate case exclusion • All cases evaluated similarly • Blinded pathologists • Role of investigators clear • Wash out period (4 weeks for validation component) • Outcomes well defined • Statement of no COI Applicability <ul style="list-style-type: none"> • Training level of diagnostic investigators reported 	Risk of Bias <ul style="list-style-type: none"> • Retrospective cases • Representative case selection • Single representative slide/case • No discussion on limitations • No statistical power calculation • Wash out period (2 days for prospective component) • No slide deidentifying methodology reported Applicability <ul style="list-style-type: none"> • Diagnostic study only - no clinical outcome data • No description of remote hardware • No description of LM
Rao, 2021(2)⁵	
Risk of Bias <ul style="list-style-type: none"> • Consecutive cases • Prospective cases • No inappropriate case exclusion • All cases evaluated similarly • Blinded pathologists • Role of investigators clear • Wash out period (2 weeks) • Outcomes well defined • Statement of no COI • Discussion of study limitations Applicability <ul style="list-style-type: none"> • Training level of diagnostic investigators reported • Variety of computer hardware used for remote evaluation 	Risk of Bias <ul style="list-style-type: none"> • No statistical power calculation • No slide deidentifying methodology reported Applicability <ul style="list-style-type: none"> • Diagnostic study only - no clinical outcome data • No description of LM
Samuelson, 2021¹⁶	
Risk of Bias <ul style="list-style-type: none"> • Random selection of cases enrolled • All cases evaluated similarly • Blinded pathologists • Role of investigators clear • Wash out period (2 weeks) • Outcomes well defined Applicability <ul style="list-style-type: none"> • Training level of diagnostic investigators reported 	Risk of Bias <ul style="list-style-type: none"> • Cases excluded upon post-scan QC • Retrospective cases • No discussion on limitations • No statistical power calculation • No COI statement Applicability <ul style="list-style-type: none"> • Diagnostic study only - no clinical outcome data • All assessments used same WSI viewer • No description of LM

Strengths	Limitations
Alassiri, 2020¹	
Risk of Bias <ul style="list-style-type: none"> • No inappropriate case exclusion • All cases evaluated similarly • Blinded pathologists • Role of investigators clear • Wash out period (8 weeks) • Outcomes well defined • Discussion of study limitations Applicability <ul style="list-style-type: none"> • Training level of diagnostic investigators reported • Variety of computer hardware used for remote evaluation • Variety of LM used 	Risk of Bias <ul style="list-style-type: none"> • Retrospective cases • Representative case selection • Limited data on telepathology training • Single representative slide/case • No statistical power calculation • No COI statement Applicability <ul style="list-style-type: none"> • Diagnostic study only - no clinical outcome data
Borowsky, 2020¹⁷	
Risk of Bias <ul style="list-style-type: none"> • No inappropriate case exclusion • All cases evaluated similarly • Blinded pathologists • Role of investigators clear • Wash out period (31 days) • Outcomes well defined • Discussion of study limitations Applicability <ul style="list-style-type: none"> • Hardware described 	Risk of Bias <ul style="list-style-type: none"> • Cases excluded upon pre-scan QC • Retrospective cases • Representative case selection • Limited data on telepathology training • No statistical power calculation • Statement of potential COI Applicability <ul style="list-style-type: none"> • Diagnostic study only - no clinical outcome data • No description of LM
Hanna, 2020¹⁸	
Risk of Bias <ul style="list-style-type: none"> • Consecutive cases • Prospective cases • No inappropriate case exclusion • All cases evaluated similarly • Role of investigators clear • Outcomes well defined • Discussion of study limitations Applicability <ul style="list-style-type: none"> • Training level of diagnostic investigators reported • Variety of computer hardware used for remote evaluation 	Risk of Bias <ul style="list-style-type: none"> • No statistical power calculation • Blinding unclear • Statement of potential COI • No slide deidentifying methodology reported • Short washout period (mean 2 days) Applicability <ul style="list-style-type: none"> • Diagnostic study only - no clinical outcome data • All assessments used same LM

Strengths	Limitations
Davidson, 2019⁴	
Risk of Bias <ul style="list-style-type: none"> • Randomized assignment of pathologists to WSI or LM twice • No inappropriate case exclusion • All cases evaluated similarly • Blinded pathologists • Role of investigators clear • Wash out period (9 months) • Outcomes well defined • Statement of no COI • Discussion of study limitations Applicability <ul style="list-style-type: none"> • Training level of diagnostic investigators reported • Variety of computer hardware used for remote evaluation 	Risk of Bias <ul style="list-style-type: none"> • Retrospective cases • Representative case selection • Less than 80% of pathologists completed readings • Single representative slide/case • No statistical power calculation • No slide deidentifying methodology reported Applicability <ul style="list-style-type: none"> • Diagnostic study only - no clinical outcome data • No description of LM
Hanna, 2019¹⁹	
Risk of Bias <ul style="list-style-type: none"> • Consecutive cases • Prospective cases • No inappropriate case exclusion • All cases evaluated similarly • Blinded pathologists • Role of investigators clear • Wash out period (13 weeks) • Outcomes well defined • Discussion of study limitations Applicability <ul style="list-style-type: none"> • Training level of diagnostic investigators reported 	Risk of Bias <ul style="list-style-type: none"> • No statistical power calculation • Statement of potential COI Applicability <ul style="list-style-type: none"> • Diagnostic study only - no clinical outcome data • All assessments used same WSI viewer • No description of LM

Strengths	Limitations
Larghi, 2019²⁰	
Risk of Bias <ul style="list-style-type: none"> • No inappropriate case exclusion • All cases evaluated similarly • Blinded pathologists • Role of investigators clear • Wash out period (3 months) • Outcomes well defined • Statement of no COI • Discussion of study limitations Applicability <ul style="list-style-type: none"> • Training level of diagnostic investigators reported • Variety of computer hardware used for remote evaluation • Variety of LM used 	Risk of Bias <ul style="list-style-type: none"> • Retrospective cases • Representative case selection • Limited data on telepathology training • Single representative slide/case • No statistical power calculation Applicability <ul style="list-style-type: none"> • Diagnostic study only - no clinical outcome data
Rakha, 2018²¹	
Risk of Bias <ul style="list-style-type: none"> • No inappropriate case exclusion • All cases evaluated similarly • Blinded pathologists • Wash out period (3 months) • Outcomes well defined • Statement of no COI • Discussion of study limitations Applicability <ul style="list-style-type: none"> • Training level of diagnostic investigators reported 	Risk of Bias <ul style="list-style-type: none"> • Retrospective cases • Representative case selection • Single representative slide/case • No statistical power calculation • No slide deidentifying methodology reported Applicability <ul style="list-style-type: none"> • All assessments used same WSI viewer • No description of LM

COI = conflict of interest; LM = light microscope; QC = quality control; WSI = whole slide image.

Appendix 4: Main Study Findings and Authors' Conclusions

Note that this appendix has not been copy-edited.

Table 6: Summary of Findings Included Systematic Reviews

Main study findings	Authors' conclusion
Systematic reviews	
Araujo, 2019¹³	
<p>Intra-observer concordance</p> <p>Range 87% to 98.3%</p> <p>κ coefficient range 0.8 to 0.98</p> <p>Discordance</p> <p>61.5% of studies provided a preferred diagnosis for disagreements. Among a total of 99 disagreements, 37 (37.3%) of preferred diagnoses agreed with WSI over conventional microscopy.</p> <p>Critical Appraisal:</p> <p>Unclear risk of bias in 15.4% of studies due to unclear case selection criteria. Two other studies (15.4%) were at high-risk of bias with regard to the thresholds classifying diagnostic concordance. Otherwise, the identified evidence was evaluated as low concern for bias.</p>	<p>"In general, this systematic review showed a high concordance between diagnoses achieved by using WSI and conventional light microscope (CLM), summarizes difficulties related to specific findings of certain areas of pathology— including dermatopathology, pediatric pathology, neuropathology, and gastrointestinal pathology— and demonstrated that WSI can be used to render primary diagnoses in several subspecialties of human pathology." (p270)</p>
Williams, 2017⁶	
<p>Discordances</p> <p>Discordance occurrences: 335/8069 (4%)</p> <p>Among a total of 335 disagreements, 44 (13%) of preferred diagnoses agreed with WSI over conventional microscopy.</p> <p>Among a total of 335 disagreements, 28 (8.4% or 0.35% of total reads) had the potential to cause moderate/severe patient harm.</p> <p>The largest category of discordance was missed diagnosis of malignant/dysplastic/atypical conditions where malignant tissue was diagnosed as benign.</p> <p>Among a total of 109 disagreements regarding the diagnosis of malignant/dysplastic/atypical conditions, 101 of preferred diagnoses agree with conventional microscopy over WSI.</p> <p>Most discordances (169/335) had appreciable diagnostic difficulty and recognized inter-observer variation.</p>	<p>"Systematic analysis of concordance studies reveals specific areas that may be problematic on whole slide imaging. It is important that pathologists are aware of these areas to ensure patient safety." (p1712)</p> <p>"...we believe it is important that diagnostic departments perform their own whole-system validations for WSI, to evaluate the strengths and weaknesses of the combination of hardware and software components they propose to use for primary diagnosis." (p1717)</p>

LM = light microscopy; WSI = whole slide imaging.

Table 7: Summary of Findings of Included Primary Clinical Studies

Main study findings	Authors' conclusion
Ammendola, 2021 ²	
<p>Inter-observer concordance for senior pathologists (n = 2)</p> <p>Atypical meningioma: LM = 63%; WSI = 74%</p> <p>Atypical for major criteria: LM = 86%; WSI = 86%</p> <p>Atypical for minor criteria: LM = 60%; WSI = 77%</p> <p>Brain invasion: LM = 97%; WSI = 97%</p> <p>High mitotic index: LM = 86%; WSI = 80%</p> <p>Hypercellularity: LM = 77%; WSI = 86%</p> <p>Sheeting: LM = 74%; WSI = 77%</p> <p>Macronucleoli: LM = 49%; WSI = 51%</p> <p>Small cells: LM = 49%; WSI = 49%</p> <p>Spontaneous necrosis: LM = 51%; WSI = 54%</p> <p>Inter-observer concordance for residents (n = 2)</p> <p>Atypical meningioma: LM = 54%; WSI = 60%</p> <p>Atypical for major criteria: LM = 69%; WSI = 80%</p> <p>Atypical for minor criteria: LM = 46%; WSI = 63%</p> <p>Brain invasion: LM = 83%; WSI = 89%</p> <p>High mitotic index: LM = 80%; WSI = 69%</p> <p>Hypercellularity: LM = 74%; WSI = 86%</p> <p>Sheeting: LM = 57%; WSI = 66%</p> <p>Macronucleoli: LM = 37%; WSI = 40%</p> <p>Small cells: LM = 34%; WSI = 34%</p> <p>Spontaneous necrosis: LM = 26%; WSI = 31%</p> <p>Intra-observer concordance (median %) all observers (n = 4); LM vs WSI</p> <p>Atypical meningioma: 89%</p> <p>Brain invasion: 94%</p> <p>High mitotic index: 78%</p> <p>Hypercellularity: 93%</p> <p>Sheeting: 96%</p> <p>Macronucleoli: 89%</p> <p>Small cells: 96%</p> <p>Spontaneous necrosis: 94%</p> <p>Predictive accuracy (P > 0.05)</p> <p>All 35 cases underwent complete surgical resection and 25 (71%) developed a recurrent tumour.</p> <p>High mitotic index was the histological parameter most associated with recurrence.</p> <p>There was no statistically significant difference between LM and WSI for predictive power for recurrence.</p>	<p>"In conclusion, this study shows that atypical meningioma may be safely diagnosed using WSI. The transition to this modality could simplify and standardize the assessment of mitotic index, without the need of normalization according to the microscope used. Although the inter-observer reproducibility of minor atypical criteria remains unsatisfactory, in this study, it was slightly higher using WSI compared to glass slides. Finally, the similar predictive value of all histopathological features when using the two different modalities further highlights the reliability of the diagnosis of atypical meningioma with WSI." (p755)</p> <p>"... the predictive accuracy of all histopathological parameters for recurrence was not significantly different between the two viewing modes." (p 753)</p>

Main study findings	Authors' conclusion
Araujo, 2021 ³	
<p>Intra-observer concordance all observers (n = 4)</p> <p>κ coefficient range (95% CI): 0.85 to 0.98 (0.81 to 0.98)</p> <p>Differentiation between dysplasia grading and differentiation between severe dysplasia and microinvasive OSCC had the most discordance among less trained readers</p>	<ul style="list-style-type: none"> • “Flipping is a great advantage of WSI (rotation of the image with a single click). • The wide view provided by a scanned image, automated focus, and easy navigation within different magnifications allows fast recognition of regions of interest, overcoming light, focus, and magnification handling issues, and characteristics of LM. • Pathologists should be cautious to not miss important histological structures on WSI when their confidence increases. By relying on the wide view provided by WSI, pathologists may feel secure to give a diagnosis at a lower magnification, being prone to error—not a technology limitation. • Training time (experience) and calibration in pathology are crucial for good performance. • Reported pitfalls when using a digital environment were as follows: <ul style="list-style-type: none"> ◦ Technology-related pitfalls: lag screen mirroring, lack of details of inflammatory cells, and need for a higher magnification to assess dysplasia. ◦ Case-related pitfalls: bad quality clinical photo, challenging/borderline case, clinical information, and hypothesis do not relate with the histological characteristics, lack of clinical photo/information, lack of radiographs, misleading clinical diagnosis/hypothesis, necrosis, nonrepresentative biopsy/small amount of tissue, need for special staining, the subjectivity of dysplasia analysis. ◦ Technical processing-related pitfalls: artifact, fixation, the thickness of tissue section, inclusion, staining, and cases that required a deeper tissue sectioning.” (p9)

Main study findings	Authors' conclusion
Ramaswamy, 2021⁷	
Intra-observer concordance all observers (n = 3) Major concordance (mean): 100% All concordance (mean): 98.9% Deferral rate 3/886 (0.34%) deferred for microscopy Turnaround time 97.3% met the turnaround time 2.7% required additional sampling or discussion Rescan rate 0.33% samples required rescanning	<p>"Our retrospective validation study showed that major intraobserver diagnostic concordance between WSIs on laptops and medical-grade monitors was 100%. Prospective validation with all three modalities also showed major diagnostic concordance of 100%." (p9)</p> <p>"Digital pathology is an excellent technology, which is well integrated with the workflow. Along with a team approach, it proves that remote reporting and sign-out is noninferior to on-site reporting and is comparable to WSIs on medical-grade monitors and light microscopy. Such studies on remote reporting opens the door for the use of digital pathology for interinstitutional consultation and collaboration. Regulatory bodies have approved remote reporting and can refine guidelines for validation and user acceptability." (p10)</p>
Rao, 2021(1)⁸	
Intra-observer concordance all observers (n = 3) Concordance: 98.8% Major discordance: 0.0% Minor discordance: 1.2% Time to diagnose (median seconds (IQR)); LM, WSI Pathologist 1 (P = 0.794) 60 (50 to 90) 60 (50 to 87.5) Pathologist 2 (P = 0.01) 39 (28.25 to 51) 32 (23.25 to 44) Pathologist 3 (P < 0.001) 25 (20 to 40) 63 (43.75 to 83)	<p>"Overall findings contribute to the growing evidence that histologic interpretation of routinely reported parameters on digital slides is comparable with routine microscopic evaluation even in a setting of specialty practice, with a number of immediate applications inherent to WSI." (p82)</p>
Rao, 2021(2)⁵	
Intra-observer concordance all observers (n = 3) Major concordance (mean): 100% All concordance (mean): 98.9% Deferral rate (n (%)) 27/594 (4.5%) Rescan rate (n (%)) 33/1426 (2.3%)	<p>"Careful re-assessment of existing infrastructure and need-based repurposing helped in quick adoption of DP and efficient management of our laboratory workflow. This study also validates a DP system and digital workflow for primary diagnosis from remote site with absolute concordance and proves the efficiency of the workflow. It reinforces the noninferiority of WSI when compared with microscopy even in a remote setting and provides evidence for safe and efficient diagnostic services when carried out in a risk-mitigated environment." (p8)</p>

Main study findings	Authors' conclusion
Samuelson, 2021¹⁶	
Intra-observer concordance (n = 5) Concordance (mean [range]): 83.62% (71.8% to 96.9%) Major concordance (mean [range]): 94.72% (93.7% to 96.9%)	<p>"We described a method for rapid validation of digital pathology for primary digital diagnosis using minimum resources that fully complies with CAP recommendations. In a broader sense, there continues to be a need to evolve better and standardized methods for anatomic pathology validation and measurement of diagnostic performance of digital WSI." (p10)</p>
Alassiri, 2020¹	
Intra-observer concordance (n = 4) Concordance (mean [range]): 82.1% (71.7% to 88.3%) Major discordance (mean [range]): 10% (3.3% to 16.7%) Minor discordance (mean [range]): 7.9% (3.3% to 11.7%)	<p>"WSI as a diagnostic modality is not inferior to LM and gradual transitioning into digital pathology is possible with close monitoring and sufficient training. The pre-analytical phase should be well controlled with quality H&E slides. However, to ensure the best results, only formally trained neuropathologists should handle the digital neuropathology service." (p40)</p>
Borowsky, 2020¹⁷	
Intra-observer concordance (n = 4) Concordance (overall): 96.1% Major discrepancy rate difference WSI - LM Overall: 0.44% (95% CI, -0.15% to 1.03%) Anus/perianal: 1.16% Appendix: 0.00% Bladder: 0.93% Brain/neuro: 0.55% Breast: 0.76% Colorectal: 0.00% Endocrine: -0.53% Gastroesophageal Junction: 0.54% Gallbladder: 0.00% Gynecological: 1.10% Hernia/peritoneal: 0.00%	<p>"This study demonstrated that clinical diagnoses made by pathologists via WSI using the Leica Biosystems Aperio AT2 DX system are not inferior to the traditional LM method for a large collection of pathology cases with diverse tissues/organs and sample types." (p1251)</p>

Main study findings	Authors' conclusion
<p>Kidney: -0.56%</p> <p>Liver/bile duct: 1.06%</p> <p>Lung: 1.55%</p> <p>Lymph node: -0.78%</p> <p>Prostate: -0.44%</p> <p>Salivary gland: -1.14%</p> <p>Skin: 2.30%</p> <p>Soft tissue: -0.60%</p> <p>Stomach: 1.06%</p> <p>Rescan rate, n (%): 39/5849 (0.67%)</p> <p>Read time (minutes per case diagnosis)</p> <p>WSI: 5.20; LM: 4.95</p> <p>Deferral rate, n (%)</p> <p>WSI: 271/7781 (3.5%), LM: 258/7781 (3.3%)</p>	
Hanna, 2020¹⁸	
<p>Intra-observer concordance (n = 12)</p> <p>Major concordance (mean (range)): 100%</p> <p>Minor concordance (mean [range]): 98.9%</p>	<p>"The validation successfully demonstrated operational feasibility of supporting remote review and reporting of pathology specimens and verification of remote access performance and usability for remote primary diagnostic signout." (p9)</p>
Davidson, 2019⁴	
<p>Nottingham grade Intra-observer concordance P = 0.22</p> <p>LM both phases (n = 49) (mean [95% CI]): 73% (68% to 78%)</p> <p>WSI both phases (n = 41) (mean [95% CI]): 68% (61% to 75%)</p> <p>LM to WSI (n = 45) (mean [95% CI]): 61% (55% to 67%)</p> <p>WSI to LM (n = 37) (mean [95% CI]): 66% (59% to 68%)</p> <p>Combined (n = 82) (mean [95% CI]): 63% (59% to 68%)</p> <p>Nottingham grade Inter-observer concordance P < 0.001</p> <p>LM phase I (n = 115) (mean [95% CI]): 68% (66% to 70%)</p> <p>WSI phase I (n = 93) (mean [95% CI]): 60% (57% to 62%)</p> <p>LM phase II (n = 86) (mean [95% CI]): 69% (67% to 71%)</p> <p>WSI phase II (n = 86) (mean [95% CI]): 62% (60% to 64%)</p>	<p>"Pathologists' intraobserver agreement (reproducibility) is similar for Nottingham grade using glass slides or WSI. However, slightly lower agreement between pathologists suggests that verification of grade using digital WSI may be more challenging." (p1)</p> <p>"While digitized pathology slides offer multiple advantages, use of the WSI digital format may be associated with increased variability among pathologists in assigning the Nottingham grade for invasive breast carcinomas. Advances in digital technology resolution, development of digital image analysis aids, and training in digital WSI interpretation may help address current limitations in grade assessment and be important for provision of the highest quality of clinical care." (p8)</p>

Main study findings	Authors' conclusion
Hanna, 2019¹⁹	
Intra-observer concordance (n = 8) Diagnostic: 99.3% Grade: 94.1% Margin: 100% LVI/PNI: 83.3% pT: 97.3% pN: 97.1% Efficiency WSI vs LM (P > 0.05) 19 seconds longer per slide by WSI 177 seconds longer per case by WSI Rescan rate, n (%): 148/2091 (7%)	"This investigation serves to further validate whole slide images being non-inferior to glass slides from the standpoint of diagnostic concordance, but importantly demonstrates loss of efficiency in the diagnostic turnaround time in a true clinical environment, requiring improvements in other aspects of the pathology workflow to support full adoption of digital pathology." (p12)
Larghi, 2019²⁰	
Diagnostic Performance (P > 0.05): LM, WSI Sensitivity: 0.92 (0.87 to 0.95), 0.93 (0.89 to 0.95) Specificity: 0.96 (0.80 to 0.99), 0.88 (0.69 to 0.97) PPV: 0.99 (0.97 to 0.99), 0.99 (0.97 to 0.99) NPV: 0.51(0.41 to 0.61), 0.52 (0.41 to 0.63) Diagnostic Accuracy: 0.92 (0.88 to 0.94), 0.92 (0.88 to 0.94) Intra-observer Agreement (κ (95% CI)) (P > 0.05): LM vs WSI Diagnostic classification: 0.87 (0.81 to 0.93) Core tissue: 0.68 (0.59 to 0.77) # Of lesional cells: 0.67 (0.56 to 0.77) % Lesional cells: 0.77 (0.71 to 0.83) Inter-observer Agreement (κ (95% CI)) (P > 0.05): LM, WSI Diagnostic classification: 0.79 (0.71 to 0.88), 0.78 (0.69 to 0.87) Core tissue: 0.59 (0.45 to 0.72), 0.53 (0.40 to 0.66) # Of lesional cells: 0.62 (0.52 to 0.71), 0.53 (0.43 to 0.63) % Lesional cells: 0.40 (0.30 to 0.50), 0.38 (0.28 to 0.47) Efficiency (seconds/diagnosis) (P < 0.001): LM, WSI Median (range): 84 (30 to 150), 108 (54 to 240)	"In conclusion, our results show a high concordance between light microscopy and whole slide imaging, as well as a substantial inter-observer agreement and a complete intra-observer agreement regarding diagnostic classification on EUS-guided cell-block or histological acquired biopsy samples from patients with pancreatic solid lesions. Methods to decrease WSI reading time and make it more cost-effective to use digital images will be required for wider adoption of this technique in clinical practice." (p1578)

Main study findings	Authors' conclusion
Rakha, 2018 ²¹	
<p>Intra-observer Agreement (κ (95% CI)) Parameters: LM vs WSI Grade: 0.51 (0.47 to 0.54) Mitosis scores: 0.46 (0.43 to 0.50) Tubules scores: 0.48 (0.44 to 0.52) Pleomorphism scores: 0.27 (0.24 to 0.31) Parameters: WSI 2 readings Grade: 0.65 (0.60 to 0.68) Mitosis scores: 0.60 (0.56 to 0.63) Tubules scores: 0.64 (0.60 to 0.68) Pleomorphism scores: 0.56 (0.52 to 0.59) Histology association with BCSS (HR [95% CI]) (P < 0.001): LM, WSI* Grade: 2.4 (2.0 to 3.0), 1.9(1.6 to 2.3) Tubules: 1.9 (1.5 to 2.4), 2.8(1.9 to 4) Pleomorphism: 2.7 (2 to 3.7), 1.8(1.5 to 2.2) Mitosis: 1.7 (1.5 to 1.9), 1.5(1.3 to 1.7) * from first read Histology association with DMFS (HR [95% CI]) (P < 0.001): LM, WSI* Grade: 2.1(1.8 to 2.5), 1.8 (1.5 to 2.1) Tubules: 1.7(1.4 to 2.1), 2.6 (1.9 to 3.6) Pleomorphism: 2.2 (1.7 to 2.9), 1.6 (1.3 to 1.8) Mitosis: 1.6 (1.4 to 1.8), 1.4 (1.3 to 1.6) * from first read Discordances (P < 0.00001) Major discordance rate of 1.5% where significantly more WSI diagnoses were of the lower grade as compared to LM.</p>	<p>"WSI grading showed moderate concordance with LM grading comparable to concordance rate reported among different pathologists who graded breast cancer using conventional microscopy. Exact grade agreement between WSI and LM grading was reached in 68% of cases." (p8)</p> <p>"This study demonstrates that grading using WSI is not only reproducible but also provides significant survival information comparable to glass slides." (p10)</p> <p>"Virtual microscopy is a reliable and reproducible method for assessing BC histologic grade. Regardless of the observer or assessment platform, histologic grade is a significant predictor of outcome. Continuing advances in imaging technology could potentially provide improved performance of WSI BC grading and in particular mitotic count assessment." (p1)</p>

BCSS = breast cancer specific survival; CAP = College of American Pathologists; CI = confidence interval; DMFS = distant metastasis free survival; DP = digital pathology; HR = hazard ratio; IQR = interquartile range; LM = light microscope; LVI/PNI = lymphovascular invasion/perineural invasion; OSCC = oral squamous cell carcinoma; pT = pathological stage; pN = nodal staging; WSI = whole slide image.