UK National Screening Committee

Evidence summary on the use of digital pathology in breast and cervical cancer screening

External review against programme appraisal criteria for the UK National Screening Committee

Version: 3.0

Author: Solutions for Public Health

Date: November 2021

The UK National Screening Committee secretariat is hosted by the Department of Health and Social Care

About the UK National Screening Committee (UK NSC)

The UK NSC advises ministers and the NHS in the 4 UK countries about all aspects of population screening and supports implementation of screening programmes.

Conditions are reviewed against evidence review criteria according to the UK NSC's evidence review process.

Read a complete list of UK NSC recommendations.

UK National Screening Committee, Southside, 39 Victoria Street, London, SW1H 0EU

www.gov.uk/uknsc

Blog: https://nationalscreening.blog.gov.uk/

For queries relating to this document, please contact: https://view-health-screening-recommendations.service.gov.uk/helpdesk/

Contents

Plain English summary

In pathology glass slides are usually checked by a pathologist using a microscope. In digital pathology the glass slides are scanned to create a digital image. The pathologist then checks the digital image on a computer screen or mobile device.

Digital pathology is not currently used for diagnosis in national cancer screening programmes. Instead, it is mainly used for education, training and maintaining standards. Using digital pathology in screening could improve the distribution of cases to pathologists. It could also mean faster access to past cases for comparison. But there are some concerns that it would be more difficult to identify certain types of cancer using digital images.

In 2017, a systematic review was published looking at studies up to 2015. This review found that diagnoses made using glass slides or digital images were comparable. But there were limitations to the studies included in the systematic review.

In 2020, the UK NSC searched for new evidence on the use of digital pathology in an evidence map. The conclusion of this work was to conduct a further UK NSC evidence summary of digital pathology in breast cancer and cervical cancer.

This evidence summary considers new evidence published since 2015. It aimed to answer 3 questions to see whether:

- digital pathology images are good enough to identify screened people with cancer
- health professionals and the public are comfortable with the idea of using digital pathology in screening
- using digital pathology in screening is cost-effective

The conclusion of this evidence summary was that there is not enough evidence to recommend the use of digital pathology in breast cancer or cervical cancer screening programmes. This is because:

- few studies have explored the accuracy of digital pathology images in screening cases
- most studies included a small number of cases
- health professionals and the public’s comfort with the use of digital pathology in cancer screening is unclear
- there is a lack of evidence on the cost-effectiveness of using digital pathology in screening programmes

Executive summary

Purpose of the review

This document reviews the evidence on the use of digital pathology in breast and cervical cancer screening against the UK National Screening Committee (NSC) criteria about the test accuracy, the acceptability of the technology to health professionals and the public, and its cost-effectiveness. The scope of this evidence summary is limited to the pathologist's interpretation of digital images and does not include automated image analysis of digital slides.

Background

Digital pathology is a technology that allows pathologists to review glass histopathology slides digitally on a computer screen, rather than with a light microscope. The key technology enabling digital pathology is 'whole slide imaging', a technology which creates a digital image of the entire glass slide with a scanning device to provide a high-resolution image that can be stored and viewed on a computer screen or mobile device for later review.

The adoption of digital pathology is still at an early stage and digitisation of histopathology is only available in a very limited number of hospitals in the UK. Digital pathology is not currently part of the screening pathway in nationally implemented UK adult cancer screening programmes. It is currently used by UK screening programmes only for education, audit, and maintaining standards by training and continuing professional development. It is reported that the use of digital pathology would benefit screening by allowing for more streamlined distribution of screening cases to pathologists and faster access to archived cases for comparison. However, there is a concern about whether the digital images would allow the accurate identification of precancers, atypia and early-stage cancers which involve the identification of subtler morphologies.

No guidance was identified that was specifically about the use of digital pathology in the breast cancer or cervical cancer screening pathways. However, generic guidelines or best practice recommendations on the use of digital pathology are available.

A 2017 systematic review examined the published literature from 1999 to March 2015 on the concordance of diagnoses rendered by whole slide imaging compared with those rendered by light microscopy. The authors concluded that the review found evidence to support a high level of diagnostic concordance. However, they noted that the review findings were predominantly based on small studies that were inadequately powered to provide data of non-inferiority and were of varying quality, and therefore recommended that further validation studies are still needed.

Focus of the review

In 2020, the UK NSC was asked by the National Coordinating Committee for Breast Pathology and by the Royal College of Pathologists to consider the evidence regarding the use of whole slide imaging for the preoperative diagnosis of tissue specimens from the NHS Breast Cancer Screening Programme. The UK NSC agreed that work should be undertaken to consider the use of digital pathology in breast cancer screening and also in bowel and cervical cancer screening, in the form of a preliminary evidence map. The map evaluated the type and amount of evidence related to the use of digital pathology to help determine if further work was required in this area. The evidence map concluded that there was sufficient evidence to justify commissioning a more sustained review on the use of digital pathology in breast and cervical cancer screening in line with the UK NSC evidence review process. The volume and type of evidence related to the use of digital pathology for bowel cancer screening was insufficient to justify an evidence review at this stage.

Recommendation under review

The aim of this evidence summary is to review the use of digital pathology in breast and cervical cancer screening, focusing on the test accuracy, the acceptability of the technology to health professionals and the public, and its cost-effectiveness. The interest for cervical cancer screening is in the digital pathology of histology slides, not liquid-based cytology.

Findings and gaps in the evidence of this review

Sixteen[*] publications were included in the evidence summary. The evidence for each question is summarised below.

Criteria 4 and 5 — '*There should be a simple, safe, precise and validated screening test' and 'The distribution of test values in the target population should be known and a suitable cut-off level defined and agreed'.*

Question 1 — What is the evidence on the accuracy of digital pathology (whole slide imaging) in breast and cervical cancer screening?

Ten publications on the accuracy of digital pathology compared to light microscopy in breast cancer were included. Six related to the primary diagnosis of breast cancer and 4 to the grading of breast cancer. Four publications were from the UK. The studies included between 22 and 1,675 breast cancer cases. One of these publications used cases detected from the NHS Breast Cancer

[*] One publication was included in 2 of the 3 review questions

Screening Programme (between 50 and 1,077 cases in the 3 studies reported within this single publication). There were differences in the way that the individual studies were conducted and in the way that the results were reported. Generally, studies relating to the primary diagnosis of breast cancer reported their results as the percent concordance/agreement between digital pathology using whole slide images and glass slides read by light microscopy. Studies relating to the grading of cancer generally reported their results as kappa scores for agreement between digital pathology using whole slide images and glass slides read by light microscopy. In studies reporting concordance/agreement this was generally over 90% and in studies reporting kappa score this was generally reported to show moderate or substantial agreement. However, agreement scores were lower in some individual studies. The kappa statistic was developed to account for the possibility of chance agreement to increase the reliability of a result. There are advantages and disadvantages to both percent agreement and kappa score. Both can provide useful information and the extent to which chance is likely to influence the results influences the value of each. In interpreting the results, whether presented as percent agreement or kappa score, detail on the definition of the 'agreement' that is being assessed and the nature and implications of any disagreements observed is important and these details were not always adequately reported by the studies. Although all studies used light microscopy as the reference standard, the process varied between studies. In some, the reference standard was archived reports, in others it was a consensus diagnosis by a group of pathologists or review by the same pathologist either immediately or after washout periods that ranged from one week to 9 months. Other differences between the studies included the number of pathologists involved and whether they had received training in digital pathology. One of the potential concerns about the use of digital pathology in cancer screening is whether the digital images would allow the accurate identification of subtler morphologies such as precancers, atypia and early-stage cancers. Case agreement rates by diagnostic category were reported in one study which reported that agreement with the reference standard (consensus diagnosis by 3 experienced pathologists by light microscopy) was statistically significantly higher for glass slides than digital whole slide images for most of the diagnostic categories. However, there was limited evidence identified on this aspect.

Only one publication on the accuracy of digital pathology compared to light microscopy in relation to the primary diagnosis of cervical cancer was included. This small study included 157 cases and reported high levels of agreement between digital pathology and light microscopy.

Overall, based on the findings of this evidence summary, criteria 4 and 5 are not met.

Criterion 12 — '*There should be evidence that the complete screening programme (test, diagnostic procedures, treatment/ intervention) is clinically, socially and ethically acceptable to health professionals and the public'.*

Question 2 — Is the use of digital pathology in breast and cervical cancer screening clinically, socially and ethically acceptable to health professionals and the public?

Six UK publications relating to the acceptability of the use of digital pathology were included. Benefits mentioned across several studies included the ease and efficiency of sharing images for teaching, second opinions and multi-disciplinary team meetings. Barriers mentioned included financial costs, slide and equipment quality and lack of training and standardisation. However, the studies included a small number of pathologists and were not specific to the acceptability of the use of digital pathology in breast or cervical cancer screening. It was not clear how representative the views expressed were of the wider community of pathologists in the UK. No evidence was identified on the acceptability of using digital pathology in cancer screening to the public. Overall, based on the findings of this evidence summary, criterion 12 is not met.

Criterion 14 — '*The opportunity cost of the screening programme (including testing, diagnosis and treatment, administration, training and quality assurance) should be economically balanced in relation to expenditure on medical care as a whole (ie. value for money). Assessment against this criterion should have regard to evidence from cost benefit and/or cost-effectiveness analyses and have regard to the effective use of available resource'.*

Question 3 – Are there any health economic/ cost-effectiveness analyses and/ or models on the use of digital pathology in breast and cervical cancer screening compared to the use of light microscopy? If so, what do they show?

No studies on the cost-effectiveness of digital pathology compared to light microscopy were identified. Therefore, criterion 14 is not met.

Recommendations on screening

The volume, quality and direction of new evidence is insufficient at present to determine the accuracy, acceptability and cost-effectiveness of digital pathology in breast and cervical cancer screening. The included studies on the accuracy of digital pathology in breast cancer generally reported a high degree of agreement between digital pathology and light microscopy, although with limited information to interpret the clinical significance of the results in some studies. The evidence specifically relating to the accuracy of digital pathology in cases detected by screening is limited. The interpretation of the evidence base is also affected by the small number of cases involved, differences in the designs and process used in the studies and differences in the way the results were reported. Further, larger, high-quality studies with improved methodological consistency assessing the accuracy of digital pathology using cases detected through breast cancer screening programmes are needed. The evidence available on the accuracy of digital pathology in cervical cancer was limited to a single small study, restricting any conclusions that can be drawn on the

use of digital pathology in this area. Evidence relating to the acceptability of digital pathology amongst pathologists in the UK was mixed with both positives and concerns identified. However, these studies were not specifically exploring the acceptability of digital pathology in screening programmes. In addition, they included a small number of participants, so it is unclear how representative the views expressed by participants in these studies are to the wider UK population of pathologists, health professionals or the public. Studies exploring the acceptability of the use of digital pathology within screening programmes within a wider population of stakeholders are needed. Studies on the cost-effectiveness of digital pathology compared to light microscopy are also needed, as no evidence was found addressing this specific point. An ongoing Health Technology Assessment (HTA) primary study is due to conclude in October 2021 and, once the results are released, is expected to produce useful results on the performance of digital pathology in cancer screening. This HTA also has a qualitative component which aims to explore views and experiences of pathologists and laboratory staff migrating from light microscopy to digital pathology. In addition, part of the HTA consists of a health economic study which is expected to assess the incremental costs associated with digitalisation compared to light microscopy-based pathology.

Limitations

This evidence summary was conducted according to the UK NSC evidence review process over a condensed period of time. The review only looked for peer-reviewed scientific work and does not include work published elsewhere (grey literature). Studies not available in the English language, abstracts and poster presentations were not eligible for inclusion. Given that these are accepted methodological adjustments for a rapid review, and that the searches for this evidence summary covered relevant literature since January 2015 (when the searches for the 2017 systematic review were carried out), these limitations should not have led to the exclusion of any pivotal studies.

Evidence uncertainties

The volume, quality and direction of new evidence is insufficient to meet the UK NSC criteria about the test accuracy, acceptability and cost-effectiveness of digital pathology at present. Further work specifically exploring the accuracy, acceptability and cost-effectiveness of digital pathology in breast cancer screening is warranted. There is insufficient evidence to recommended further work on the use of digital pathology in cervical cancer screening at this time. The release of the results of the ongoing HTA report on a multi-centred validation of digital whole slide imaging for routine diagnosis could be the point at which further consideration is needed on whether a re-evaluation of the evidence is required on this topic.

Introduction and approach

This document reviews the evidence on the use of digital pathology in breast and cervical cancer screening against the UK National Screening Committee (NSC) criteria about the test accuracy, the acceptability of the technology to health professionals and the public, and its cost-effectiveness.

Background

Digital pathology is a technology that allows pathologists to review glass histopathology slides digitally on a computer screen, rather than with a light microscope which is considered the reference standard conventionally used for diagnosis in pathology[1]. Digital pathology can be broadly defined as the creation, management, sharing and interpretation of pathology information (including slides and data) in a digital environment[2]. The key technology enabling digital pathology is 'whole slide imaging', a technology which creates a digital image of the entire glass slide with a scanning device to provide a high-resolution image that can be stored and viewed on a computer screen or mobile device for later review[2,3]. As the digital images are up to 10 gigapixels in size, specialised software is needed to compress the image to reduce the size to an accessible and optimal size for viewing and analysis. Modern whole slide imaging scanners can scan from one to 400 slides at a time, typically using a microscope lens with magnification of 20x or 40x[4].

Digital pathology is cited as having many advantages compared to traditional light microscopy. These include[1,3]:

- easier access as it enables electronic transfer of slides from the laboratory to the pathologist and improved workflow in the laboratory which can help to minimise the impact of local shortages of pathologists
- images can be simultaneously shared between multiple sites and pathologists thereby reducing time taken to gain second opinions
- individual pathologists can examine multiple slides allowing side-by-side comparisons of different magnifications of the same case
- slides can be annotated and potentially subjected to standardised-image analysis software
- images can be shared and stored virtually reducing the risk of slide degradation and physical damage

However, the adoption of digital pathology is still at an early stage and digitisation of histopathology is only available in a very limited number of hospitals in the UK[3,5]. Limitations and barriers to the implementation of digital pathology include[1,3,6,7]:

- a limited evidence base to inform its use

- evidence published suggests an increased time to diagnosis and reduced diagnostic confidence compared with light microscopy
- more difficult assessment of dysplasia grading, weddellite calcification (calcium oxalate), mitotic counts and viral inclusions on digital systems
- inadequate infrastructure for digital pathology in many laboratories
- costs associated with implementation of whole slide imaging
- lack of regulatory approval
- pathologists' reluctance to use the technology or lack of knowledge of its advantages and limitations

Digital pathology in cancer screening

In the UK, digital pathology is not currently part of the screening pathway in nationally implemented adult cancer screening programmes. It is currently used by UK screening programmes only for education, audit, and maintaining standards by training and continuing professional development[8]. For example, many UK screening programmes use digital pathology based external quality assessment schemes where pathologists review digital images of training cases online to support quality and standards in screening[7,8]. It is reported that digital pathology would benefit screening by allowing for more streamlined distribution of screening cases to pathologists and faster access to archived cases for comparison, but there is a concern about its ability to accurately identify precancers, atypia and early-stage cancers which involve the identification of subtler morphologies[7].

There is particular interest in relation to the use of whole slide imaging in breast cancer screening. This evidence summary also considers the use of whole slide imaging in cervical cancer screening. However, the interest for cervical cancer screening is in the digital pathology of histology slides, not liquid-based cytology. Digital imaging of cervical liquid-based cytology is currently confined to imaging for archive purposes[†].

Some practitioners consider digital pathology not as a new test to be introduced in the screening pathway but as an alternative diagnostic platform that will replicate or replace the microscope[9]. Currently, pathologist review of glass slides using light microscopy is considered the reference (gold) standard[1]. The 2018 best practice recommendations for implementing digital pathology from the Royal College of Pathologists state that glass slides should be considered the primary reference image that is retained for the patient record until further work has established the equivalence of glass and digital images[3].

[†] UK National Screening Committee, personal communication May, 2021

If digital pathology using whole slide images were to be adopted as the reference standard in the future, the validation and internal and external quality assurance of systems, laboratories and pathologists that are in place for the use of light microscopy would still be required. For example, in a related context, when direct digital mammography systems were adopted for routine mammography in the NHS Breast Screening Programme, guidance for routine quality control was developed to ensure that equipment was working within agreed standards and parameters and was functioning safely[10].

The Royal College of Pathologists' position is that digital pathology is still a new technology in which experience is limited, so a cautious approach to adoption is warranted. However, they state that the technology can be used safely for primary diagnosis if the Royal College of Pathology validation guidelines are followed, which allow pathologists to gain experience in digital pathology while deferring to the microscope when there is uncertainty[3,8]. Preliminary discussions between Public Health England, the UK NSC and the Royal College of Pathologists on digital pathology took place in 2016-17. At that time, the systematic review by Goacher *et al.*[1] was available and recommended that further studies be conducted to address a limited evidence base. Those initial discussions also prompted a Health Technology Assessment (HTA) primary study.

This HTA is currently ongoing and it is expected to produce useful results on the performance of digital pathology in cancer screening[11]. This HTA is a multi-centre validation study comparing interpretation of slides by pathologists using light microscopy and digital pathology for routine diagnosis in Coventry, Belfast, Lincoln, Oxford and Nottingham. The order in which the pathologists view the images (light microscopy and digital pathology) will be randomised. The study will include 2,000 complete histopathology samples including 600 samples each of breast, gastrointestinal (both including 200 cancer screening samples) and skin, and 200 renal samples. For each of the 4 tissue areas, 4 pathologists will examine the same cases using both light microscopy and digital pathology with a 6-week washout period between viewings. The 4 pathologists will collectively decide the correct diagnosis (ground truth) for all cases using multi-head light microscopy, taking into account the original report (reference diagnosis) as required. An independent pathologist will judge whether any differences observed are major (would alter the patient's treatment) or minor (would not alter the treatment). In addition to diagnostic accuracy, the study will examine perceptions and experiences of pathologists prior to and during the progress of the study, assessing acceptability, timing and attitudes towards digital pathology. This HTA will also include a health economic study which will assess the incremental costs associated with digitalisation compared to light microscopy-based pathology. The study is expected to conclude in October 2021.

Current guidance

No guidance was identified that was specifically about the use of digital pathology in the breast cancer or cervical cancer screening pathways. However, generic guidelines or best practice recommendations on the use of digital pathology are available.

In 2013, the College of American Pathologists produced guidelines on the validation of whole slide imaging for diagnostic purposes in pathology[12]. They consist of 12 guideline statements which include specific recommendations for validation studies, such as at least 60 routine cases per application, training in whole slide imaging for participants and, for intraobserver studies, a washout period of at least 2 weeks between viewing digital and glass slides. These guidelines were updated in 2021, reaffirming these recommendations[13]. In 2017, the College of American Pathologists also provided guidance on the major issues to be considered by pathologists contemplating the introduction of whole slide imaging into clinical practice based on the personal experience of early adopters of the technology[14].

In the UK, the Royal College of Pathologists published a digital pathology strategy setting out their objectives in this area in 2019, supporting the use of digital pathology in diagnosis, research, education and training[15]. In 2018, they published best practice recommendations advising on the technical and practical aspects of implementing digital pathology[3]. These include an overview of digital pathology technology, published evidence and personal experience on diagnostic use and practical advice for implementation. The recommendations include general principles for the validation and verification of digital pathology, which are less resource intense than the American guidelines. The recommended validation process for pathologists to follow involves (1) basic skills training, (2) practice with feedback, (3) validation of a training set of at least 20 retrospective cases, (4) validation of live cases over one to 3 months, (5) validation statement documenting discordance/concordance rates, (6) ongoing monitoring. One of the example validation sets provided in the document appendix relates to breast pathology. These validation sets list the scope of tests slides (such as specimen and tissue types, stains, diagnoses and tasks) and potential pitfalls.

In March 2020, the Royal College of Pathologists published guidance for remote reporting of digital pathology slides during periods of exceptional service pressure[16,17]. This guidance outlines recommendations around validation, training, equipment and reporting for the temporary remote reporting of digital slides in times of clinical and service necessity, such as the Covid-19 pandemic. The document appendix includes potential areas of digital diagnostic difficulties in general and for sub-specialities including breast and gynaecological.

Additional UK generic guidance identified was on the use of digital pathology from the Leeds Teaching Hospitals NHS Trust and the University of Leeds. This includes best practice guidance

for organisations that are interested in implementing digital pathology for routine diagnosis[4] and achieving International Organisation for Standardisation (ISO) accreditation[18].

Current policy context and previous reviews

In 2017, a systematic review by Goacher *et al.*[1], examined the published literature from 1999 to March 2015 on the concordance of pathologic diagnoses rendered by whole slide imaging compared with those rendered by light microscopy, and identified 38 studies for inclusion[1]. The studies included a case mix of organ systems and ranged in size from 20 to 524 cases (mean 140). Countries of studies were not reported. It was not reported whether any of the studies evaluated diagnoses of specimens from screening programmes. The authors concluded that the review found evidence to support a high level of diagnostic concordance with a weighted mean diagnostic concordance of 92.4% and a weighted mean kappa coefficient[‡] of 0.75 between whole slide imaging and light microscopy. The percentage of concordance range was reported to be 91.0% to 99.0% from the 6 breast studies and 94.0% for the 2 gynaecological studies[§]. However, they noted that the review findings were predominantly based on small studies that were inadequately powered to provide data of non-inferiority and were of varying quality, and therefore recommended that further validation studies are still needed.

In 2020, the UK NSC was asked by the National Coordinating Committee for Breast Pathology and by the Royal College of Pathologists to consider the evidence regarding the use of whole slide imaging for the preoperative diagnosis of tissue specimens from the NHS Breast Cancer Screening Programme. The UK NSC agreed that work should be undertaken to consider the use of digital pathology in breast cancer screening and also in bowel and cervical cancer screening, in the form of a preliminary evidence map to evaluate the type and amount of evidence related to the use of digital pathology to help determine if further work was required in this area. Solutions for Public Health was appointed to carry out the preliminary evidence map.

The evidence map evaluated the volume and type of evidence exploring the use of digital pathology in breast cancer screening, bowel cancer screening and cervical cancer screening. The search period was restricted to January 2015 (informed by the searches carried out for the systematic review by Goacher *et al.*[1]) to 19 October 2020. The map also searched for guidelines and/or recommendations on the use of digital pathology in breast, bowel of cervical cancer screening. The evidence map concluded that there was sufficient evidence to justify

---

[‡] kappa < 0 = less than chance agreement; kappa 0.01–0.20 = slight agreement; kappa 0.21–0.40 = fair agreement; kappa 0.41–0.60 = moderate agreement; kappa 0.61–0.80 = substantial agreement; kappa 0.81–0.99 = almost perfect agreement

[§] One of the 2 gynaecological studies included in the Goacher *et al.* systematic review was Ordi *et al*. 2015, which has also been included in this evidence summary. All 6 of the breast studies included in the Goacher *et al.* systematic review were published prior to January 2015 and were therefore not eligible for inclusion in this evidence summary

commissioning a more sustained review on the use of digital pathology in breast and cervical cancer screening in line with the UK NSC evidence review process. The volume and type of evidence related to the use of digital pathology for bowel cancer screening was insufficient to justify an evidence review at this stage.

Objectives

The aim of this evidence summary is to review the use of digital pathology in breast and cervical cancer screening, focusing on the test accuracy, the acceptability of the technology to health professionals and the public, and its cost-effectiveness.

The scope of this evidence summary is limited to the pathologist's interpretation of whole slide images and does not include automated image analysis of whole slide images. However, a brief summary acknowledging the studies of automated image analysis returned by the searches is provided in a short narrative synthesis to the key question on test accuracy.

Table 1. Key questions for the evidence summary, and relationship to UK NSC screening criteria

	Criterion	Key questions	Studies included
	THE TEST		
4	There should be a simple, safe, precise and validated screening test.	What is the evidence on the accuracy of digital pathology (whole slide imaging) in breast and cervical cancer screening?	11
5	The distribution of test values in the target population should be known and a suitable cut-off level defined and agreed.		
	THE SCREENING PROGRAMME		
12	There should be evidence that the complete screening programme (test, diagnostic procedures, treatment/ intervention) is clinically, socially and ethically acceptable to health professionals and the public.	Is the use of digital pathology in breast and cervical cancer screening clinically, socially and ethically acceptable to health professionals and the public?	6**
14	The opportunity cost of the screening programme (including testing, diagnosis and treatment, administration, training and quality assurance) should be economically balanced in relation to expenditure on medical care as a whole (ie. value for money). Assessment against this criterion should have regard to evidence from cost benefit and/or cost-effectiveness analyses and have regard to the effective use of available resource.	Are there any health economic/ cost-effectiveness analyses and/ or models on the use of digital pathology in breast and cervical cancer screening compared to the use of light microscopy? If so, what do they show?	0

** One paper was included in both question 1 and question 2

Methods

The current evidence summary was conducted by Solutions for Public Health, in keeping with the UK National Screening Committee evidence review process. Database searches were conducted on 19 October 2020 and 20 May 2021 to identify studies relevant to the questions detailed in Table 1.

Eligibility for inclusion in the review

The following review process was followed:

1. Each title and abstract was reviewed against the inclusion/exclusion criteria by one reviewer. Where the applicability of the inclusion criteria was unclear, the article was included at this stage in order to ensure that all potentially relevant studies were captured
2. Full-text articles required for the full-text review stage were acquired
3. Each full-text article was reviewed against the inclusion/exclusion criteria by one reviewer, who determined whether the article was relevant to one or more of the review questions
4. Any queries at the abstract or full-text review stage were resolved through discussion with a second reviewer
5. The review was quality assured by a second senior reviewer, not involved with the writing of the review.

Eligibility criteria for each question are presented in Table 2 below. Further details relating to the eligibility of studies are provided in Appendix 2. This includes decisions made regarding the prioritisation of studies for this evidence summary based on the results of the evidence map.

A total of 1,342 unique references were identified and sifted by an information scientist by title and abstract for potential relevance. A reviewer assessed 196 titles and abstracts for further appraisal and possible inclusion in the final review. Overall, 53 studies were identified as possibly relevant during title and abstract sifting and were further assessed at full text. Appendix 2 contains a full PRISMA flow diagram (Figure 1), along with a table of the included publications and details of which questions these publications were identified as being relevant to (Table 7).

Table 2. Inclusion and exclusion criteria for the key questions

Key question	Inclusion criteria							Exclusion criteria
	Population	**Target condition**	**Intervention/ Index test**	**Reference standard**	**Comparator**	**Outcome**	**Study type**	
What is the evidence on the accuracy of digital pathology (whole slide imaging) in breast and cervical cancer screening?	Women offered breast and cervical cancer screening	Breast cancer and cervical cancer	Digital pathology (whole slide imaging)	Light microscopy	Light microscopy	• concordance of whole slide imaging and light microscopy • kappa statistic • non-inferiority • time to diagnosis • diagnostic confidence • sensitivity • specificity • positive and negative predictive values • likelihood ratios • area under the curve	Studies from the UK screening programme to be prioritised Studies in randomly assigned or consecutively enrolled populations to be prioritised Studies in the English language published since January 2015	Case reports, conference abstracts, comment/ editorials/ letters
Is the use of digital pathology in breast and cervical cancer screening	Adult population	Breast cancer and cervical cancer	Use of digital pathology (whole slide imaging) in breast and cervical	N/A	Current use of light microscopy No comparator	Perceptions, views and/or attitudes and/ or experiences of pathologists, health professionals, patients and	Qualitative, quantitative and mixed methods studies. Randomised controlled	Opinion-based papers

clinically, socially and ethically acceptable to health professionals and the public?			cancer screening			members of the public regarding the use of digital pathology in breast and cervical cancer screening	trials, cohort studies and systematic reviews of any of the above Studies in the English language published since January 2015	
Are there any health economic/ cost-effectiveness analyses and/ or models on the use of digital pathology in breast and cervical cancer screening compared to the use of light microscopy? If so, what do they show?	Adult population	Breast cancer and cervical cancer	Use of digital pathology (whole slide imaging) in breast and cervical cancer screening	N/A	Light microscopy	• total cost of screening using digital pathology • incremental cost • incremental life-year saved • incremental cost-effectiveness ratio (ICER) • number of lives saved • cost per life saved • any other outcome as outlined by the study	Economic evaluations, such as studies comparing at least 2 alternative interventions in terms of costs and outcomes. Cost-minimisation, cost-effectiveness, cost-utility, cost-benefit and cost-consequence analyses can all be considered. Reviews of economic	N/A

			evaluations can also be included Studies in the English language published since January 2015

Appraisal for quality/risk of bias tool

The following tools were used to assess the quality and risk of bias of each study included in the evidence summary:

- diagnostic accuracy studies: Quality Assessment of Diagnostic Accuracy Studies (QUADAS-2) tool
- qualitative studies: CASP critical appraisal checklist for qualitative research

Databases/sources searched

Systematic searches of 3 databases (Medline, Embase and Cochrane Library) were conducted to identify studies relevant to the questions detailed in Table 1.

The searches for the evidence map were conducted on 19 October 2020. Two searches were carried out, one specifically for breast, bowel or cervical digital pathology studies and one broader search for digital pathology validation studies. Checks for additional relevant papers were made using the PubMed related papers function and checking the reference lists of relevant systematic reviews. An additional search for guidelines was conducted on the TRIPdatabase, NICE Evidence Search and using an internet search engine. The searches were updated on 20 May 2021 to check for any additional studies published since the evidence map search.

The search strategies are presented in Appendix 1.

Question level synthesis

Criteria 4 and 5 — Test accuracy

4. There should be a simple, safe, precise and validated screening test.

5. The distribution of test values in the target population should be known and suitable cut-off level defined and agreed.

Question 1 — What is the evidence on the accuracy of digital pathology (whole slide imaging) in breast and cervical cancer screening?

A 2017 systematic review[1] found evidence of a high level of diagnostic concordance between whole slide imaging and light microscopy (weighted mean diagnostic concordance of 92.4% and a weighted mean kappa coefficient of 0.75). However, there was a low volume of evidence and the studies were small and of variable quality.

The 2020 evidence map explored the evidence base for the use of digital pathology in breast, bowel and cervical cancer screening. It included studies comparing digital pathology to light microscopy for the primary diagnosis and grading of cancer and studies assessing the interobserver concordance for whole slide imaging interpretation between pathologists. The 2020 evidence map also included sections briefly describing other studies identified that were potentially of interest but were not formally included in the evidence map. These included non-UK studies that included a case mix of samples for different organ types or specialties or evaluated the performance of digital pathology for the prognosis or classification of cancers. Studies evaluating automated systems for analysing whole slide imaging compared to pathologist interpretation of glass slides/whole slide images were also mentioned but were not formally included in the evidence map.

Eligibility for inclusion in the review

Studies in which a pathologist's interpretation of whole slide images is compared to a pathologist's interpretation of glass slides for the primary diagnosis or grading of breast or cervical cancer have been prioritised for inclusion. This is because these studies would be more informative in relation to the use of digital pathology in a screening programme and more reflective of current practice. Studies assessing variation in whole slide imaging interpretation, in the absence of a comparison with light microscopy, are not included in this evidence summary but are briefly described in the final section of the narrative below.

Studies evaluating some form of automated image analysis of whole slide images have not been included in this evidence summary, although the volume and type of evidence related to this specific but separate area of digital pathology is briefly summarised for information. This is because this is still a very immature area of digital pathology as recognised by the Royal College of Pathology. The College recognises that the interpretation of microscopic images is a complex and holistic process involving the interpretation of subtle image features using clinical knowledge and experience and recommends that image analysis systems be properly evaluated, and their strengths and weaknesses understood, before introduction to clinical use[3].

Studies that included a mixture of different tissue types with no separate reporting for breast or cervical cancer outcomes were excluded. Studies assessing the use of digital pathology in liquid-based cytology were excluded.

Description of the evidence

Of the 53 papers reviewed at full text, 42 related to this question. Eleven studies met the criteria for inclusion. Appendix 2 contains a full PRISMA flow diagram (Figure 1), along with a table of the included publications and details of which questions these publications were identified as being relevant to (Table 7). The excluded studies are listed in Table 8 in Appendix 2.

The 11 studies that met the inclusion criteria for this question comprised:

- 6 papers (3 UK, 3 non-UK) comparing digital pathology to light microscopy relating to the primary diagnosis of breast cancer (Babwale *et al*. 2021[19], Borowsky *et al*. 2020[20], Elmore *et al*. 2017[21], Mukhopadhyay *et al*. 2018[22], Williams *et al*. 2018[23], Williams *et al*. 2020[7])
- 4 papers (one UK, 3 non-UK) comparing digital pathology to light microscopy relating to the grading of breast cancer (Al-Janabi *et al*. 2016[24], Davidson *et al*. 2019[25], Rakha *et al*. 2018[26], Wilbur *et al*. 2015[27])
- one paper (non-UK) comparing digital pathology to light microscopy relating to the primary diagnosis of cervical cancer (Ordi *et al*. 2015[28])

The 6 papers relating to the primary diagnosis of breast cancer included data from 8 different studies. There were between 50 and 1,077 cases in 3 studies detected from the UK screening programme (all described in Williams *et al.* (2020)7); 2 studies with between 47 and 694 cases from UK centres (detection route unclear) and 3 studies with between 240 and 304 cases from the USA (detection route unclear). In addition, 225 cases from Lithuania (detection route unclear) were included within Williams *et al.* (2020)[7]. The number of pathologists, where stated was between one and 4 in all but one of the studies, which included more than 80 pathologists.

The 4 papers relating to the grading of breast cancer included data from 5 different studies. It is not clear if any of the cases included in these studies were detected by screening. The single UK

study included 1,675 cases. The 3 non-UK studies from the USA, Netherlands and Denmark included between 22 and 184 cases. The number of pathologists was between one and 3 in all but one of the studies, which included 82 pathologists.

The one study relating to the primary diagnosis of cervical cancer (Ordi *et al.* 2015[28]) included 157 cases referred to colposcopy because of an abnormal Pap smear from one centre in Spain.

Discussion of findings

A study-level summary of data extracted from each included publication is presented in the 'summary and appraisal of individual studies' in Appendix 3 where publications are stratified by question and are presented in alphabetical order. In the discussion below, studies relating to the primary diagnosis of breast cancer are presented first, followed by studies relating to the grading of breast cancer and the primary diagnosis of cervical cancer.

Digital pathology compared to light microscopy for the primary diagnosis of breast cancer

The 6 papers (reporting on 8 studies) comparing digital pathology (whole slide images) to light microscopy (glass slides) relating to the primary diagnosis of breast cancer were all validation studies seeking to establish the diagnostic accuracy of digital whole slide images. Table 3 presents results relating to diagnostic concordance (agreement) or discordance (disagreement) between assessments made using digital whole slide images or light microscopy. In Table 3, studies using data from UK Breast Cancer Screening Programmes are presented first, followed by other UK data and then non-UK data. Confidence intervals are provided where reported. The paper by Williams *et al.* (2020)[7] reported 3 separate studies with different sites within each study. These are presented as separate rows in Table 3. Further details about all the studies are presented in Appendix 3.

All studies used light microscopy as the reference standard. Five analyses compared digital whole slide images to the original diagnosis made using light microscopy from archive reports, or to a consensus diagnosis made using light microscopy. In these analyses, the glass slides may or may not have been read by the same pathologist as the digital whole slide image. Six analyses compared digital whole slide images to review of glass slides on light microscopy by the same pathologist (intraobserver agreement), either immediately or after a washout period. The studies reported slightly different definitions of concordance, some reporting any difference in reporting and others focusing on differences that might affect patient management. In some studies, the definition of the concordance or agreement reported was not clear. The terminology and definitions as reported by the study authors are provided as footnotes in Table 3 and in Appendix 3.

In 4 studies, participating pathologists had received training and/or access to test sets (Babwale *et al*. (2021)[19], Mukhopadhyay *et al*. (2018)[22], Williams *et al.* (2018)[23], Williams *et al.* (2020)[7] (study 2). In 4 studies it was not clear if the pathologists had received any training in digital pathology (Borowsky *et al.* (2020)[20], Elmore *et al.* (2017)[21], Williams *et al.* (2020)[7] (studies 1 and 3).

Table 3: Summary of diagnostic concordance (agreement) for the primary diagnosis of breast cancer

Test (magnification)	Reference standard	Population	Number of pathologists	Slides available	Result	Study
Digital whole slide images (x40)	Archived light microscopy reports	250 cases, detected through the UK Breast Cancer Screening Programme	2	Complete cases	Clinical concordance[a]: 99.6% Clinically significantly discordant cases: 1/250 (0.4%)	Williams *et al.* (2020)[7] study 1 (Coventry UK)
Digital whole slide images (x40)	Immediate review of glass slide by light microscopy	896 cases, detected through the UK Breast Cancer Screening Programme	4	Complete cases	Clinical concordance[a]: 99.0% Clinically significantly discordant cases: 9/896 (1.0%)	Williams *et al.* (2020)[7] study 2 (Leeds, UK)
Digital whole slide images (x40)	Immediate review of glass slide by light microscopy	181 cases, detected through the UK Breast Cancer Screening Programme	1	Complete cases	Clinical concordance[a]: 99.4% Clinically significantly discordant cases: 1/181 (0.6%)	Williams *et al.* (2020)[7] study 2 (Lincolnshire, UK)
Digital whole slide images (x40)	Review of glass slides by light microscopy with 2-week washout	50 diagnostically challenging cases, detected through the UK Breast Cancer Screening Programme	3	Single representative slide selected for each case	Agreement[b]: 87% Kappa score[c]: 0.80 (95%CI 0.70 to 0.90)	Williams *et al.* (2020)[7] study 3 (Leeds, UK)
Digital whole slide images (x20)	Authorised reports on glass slides by light microscopy	47 cases from one UK centre (detection route unclear)	Not stated	As required	Concordance[d]: 95.7% Clinically significantly discordant cases: 0/47	Babwale *et al*. (2021)[19] (Wales, UK)
Digital whole slide images (x40)	Immediate review of glass slide by light microscopy	694 cases from 3 UK pathologists' total workload (detection route unclear)	3	All case slides	Complete clinical concordance[e]: 98.8% Clinically significantly discordant cases: 8/694 (1.2%) Complete concordance: 96.2% Any observable difference: 3.8%	Williams *et al.* (2018)[23] (Leeds, UK)

Test (magnification)	Reference standard	Population	Number of pathologists	Slides available	Result	Study
Digital whole slide images (x20)	Original diagnosis by light microscopy	304 routine care cases from the archives of 5 US sites (detection route unclear)	Not stated	All case slides or a single representative slide where appropriate	Major discrepancy rate[f]: 4.29%	Borowsky *et al.* (2020)[20] (US)
Digital whole slide images (x40)	Glass slides reviewed by light microscopy with 9-month minimum washout	240 randomly selected cases from US pathology registries (detection route unclear)	82	Not stated	Concordance in diagnosis for individual pathologists in 2 study phases (intraobserver, digital vs glass)[g]: 77% (95%CI 75 to 78)	Elmore *et al.* (2017)[21] (US)
Digital whole slide images (x40) Glass slides by light microscopy	Consensus diagnosis by 3 experienced pathologists by light microscopy	240 randomly selected cases from US pathology registries (detection route unclear)	93	Not stated	Case agreement rates against reference standard by diagnostic category (digital vs glass[g]: • benign without atypia: 82% (95%CI 79 to 85) vs 87% (95%CI 85 to 89), p<0.01 • atypia: 43% (95%CI 39 to 47) vs 48% (95%CI 44 to 52), p=0.08 • ductal carcinoma *in situ*: 79% (95%CI 77 to 82) vs 84% (95%CI 82 to 86), p<0.01 • invasive carcinoma: 93% (95%CI 90 to 95) vs 96% (95%CI 94 to 97), p=0.04	Elmore *et al.* (2017)[21] (US)
Digital whole slide images (x40)	Glass slides reviewed by light microscopy with ≥4-week minimum washout	299 consecutive routine surgical cases from 4 US institutions	4	1-16 slides per case	Major discordance rate[h]: 4.2%	Mukhopadhyay *et al*. (2018)[22] (US)
Digital whole slide images (x20)	Archived light microscopy reports	225 cases from 1 Lithuanian centre (detection route unclear)	1	Complete cases	Clinical concordance rate[b]: 96.0% Clinically significantly discordant cases: 9/225 (4.0%)	Williams *et al.* (2020)[7] study 1 (Vilnius Lithuania)

Abbreviations: CI - confidence interval; UK – United Kingdom; US – United States

a In Williams *et al.* (2020) studies 1 and 2, a clinically significant discordance = any material difference in the diagnosis, regardless of whether or not this would have affected patient prognosis or treatment
b In Williams *et al.* (2020) study 3, agreement was for breast lesion classification (not further defined). The authors also reported intraobserver agreement within format (digital and digital or glass and glass)
c kappa < 0 = less than chance agreement; kappa 0.01–0.20 = slight agreement; kappa 0.21–0.40 = fair agreement; kappa 0.41–0.60 = moderate agreement; kappa 0.61–0.80 = substantial agreement; kappa 0.81–0.99 = almost perfect agreement
d In Babwale *et al.* (2021), concordance = agreement in diagnosis between digital and glass slide reports. Discordant cases were divided into discordance of clinical significance or discordance of no clinical significance and no impact on patient management
e In Williams *et al.* (2018), no definitions were provided for 'complete concordance' or 'complete clinical concordance'
f In Borowsky *et al.* (2020), major discrepancy rate = different diagnosis associated with different patient management. The authors also reported the major discrepancy rate for glass slides compared to the original diagnosis (see Appendix 3)
g In Elmore *et al.* (2017), the study authors referred to agreement in diagnosis and reproducibility. However, this was not further defined
h In Mukhopadhyay *et al*. (2018), major discordance = a difference in diagnosis associated with a difference in patient management. The authors also reported major discordance rate for glass slides compared to the reference standard (see Appendix 3)

Generally, across studies, the concordance between digital pathology using whole slide images and glass slides read by light microscopy relating to the primary diagnosis of breast cancer was over 90%. The highest concordance scores reported were from 2 studies reported in Williams *et al.* (2020)[7] that analysed data from cases detected through the UK Breast Cancer Screening Programme; over 99% in each of the study sites. In these studies, the corresponding discordance of less than 1% related to any material difference in the diagnosis, regardless of whether or not this would have affected patient prognosis or treatment. In the third study reported by Williams *et al.* (2020)[7], which focused on diagnostically challenging cases detected through the UK Breast Cancer Screening Programme, the kappa score of 0.80 suggested substantial agreement. Comparison of these results with the other studies reported is complicated by differences in the populations, definitions and processes used. However, the 2 other UK studies also both reported a concordance of over 95% with few discordant cases that were described as being clinically significant. The non-UK studies generally reported a similar level of agreement with the exception of Elmore *et al.* (2017)21 which reported a concordance in diagnosis of 77%. The concordance figure reported was not fully defined, so it is not clear how comparable this result is to the other studies. Elmore *et al.* (2017)[21] also reported case agreement rates against the reference standard by diagnostic category for primary diagnosis. These ranged from less than 50% for atypia to over 90% for invasive carcinoma. Agreement with the reference standard (consensus diagnosis by 3 experienced pathologists by light microscopy) was statistically significantly higher for glass slides than digital whole slide images for most of the diagnostic categories.

Two studies also reported a range of additional outcomes. Elmore *et al.* (2017)21 also reported predictive values of initial digital or glass interpretation compared to confirmation by a reference panel, confidence and time to diagnosis. Williams *et al.* (2018)23 also reported diagnostic confidence.

In Elmore *et al.* (2017)21, predictive values of initial digital or glass interpretation were calculated by combining case agreement results with the prevalence of diagnostic outcomes in US women aged 50-59 years who received breast biopsies after screening. Results were reported as the likelihood that the initial interpretation using digital whole slide images or glass slides would be confirmed:

- benign without atypia
  - digital: 95.7% (95%CI 95.0 to 96.4)
  - glass: 97.1% (95%CI 96.7 to 97.4)
- atypia
  - digital: 27.8% (95%CI 23.9 to 32.5)
  - glass: 37.8% (95%CI 33.6 to 42.7)
- ductal carcinoma *in situ*
  - digital: 57.1% (95%CI 50.6 to 64.8)
  - glass: 69.6% (95%CI 64.4 to 75.3)
- invasive carcinoma

- digital: 97.2% (95%CI 95.6 to 98.6)
- glass: 97.7% (95%CI 96.5 to 98.7)

The authors reported that the estimated predictive values were statistically significantly lower for digital compared to glass for atypia (p=0.002) and ductal carcinoma *in situ* (p=0.007). No p value was reported for benign without atypia and invasive carcinoma.

In Elmore *et al.* (2017)[21] confidence in the interpretation of cases was reported in a number of ways:

- confidence in interpretive format: digital 78.6% vs glass 81.7%, p=0.22
- percentage of interpretations marked as borderline: digital 24.6% vs glass 26.1%, p=0.35
- images rated as challenging cases: digital 38.5% vs glass 30.0%, p=0.003
- pathologists desiring a second opinion: digital 42.5% vs glass 35.5%, p=0.03

Williams *et al.* (2018)[23] also reported diagnostic confidence, assessed on a scale from 0 (not at all confident) to 7 (very confident). Mean confidence in reading the digital whole slide images was between 6.70 and 6.90 for the 3 pathologists, with a range of scores from 4 to 7 for 2 of the pathologists and from 0 to 7 for the third pathologist. The mean confidence scores for reading glass slides by light microscopy were between 6.80 and 6.99 with a range of scores from 4 to 7 for 2 of the pathologists and from 6 to 7 for the third pathologist.

No absolute values for time to diagnosis were reported by Elmore *et al.* (2017)21. However, the percentage of the pathologists who spent 20 hours participating in the study (the maximum time allowed) was statistically significantly higher for reading digital whole slide images than glass slides by light microscopy (76% vs. 51%, p=0.01, n=208).

Digital pathology compared to light microscopy for the grading of breast cancer

The 4 papers comparing digital pathology (whole slide images) to light microscopy (glass slides) relating to the grading of breast cancer were all validation studies seeking to establish the diagnostic accuracy of digital whole slide images. Table 4 presents results relating to the agreement in grading made using digital whole slide images or light microscopy. Two studies reported agreement in grading using the Nottingham criteria, or similar grading criteria from the College of American Pathologists (Davidson *et al.* 2019[25], Rakha *et al.* 2018[26]). Two studies reported agreement in human epidermal growth factor receptor (HER2) score (Al-Janabi *et al.* 2016[24], Wilbur *et al.* 2015[27]). Both assessed prognostic factors. The grade considers the aggressiveness of the tumour, the HER2 score is used to assess which patients will be responsive to particular therapies[24].

In Table 4, the study using UK data is presented first, followed by studies using non-UK data. Confidence intervals are provided where reported. The paper by Wilbur *et al.* (2015)[27] reported 2

separate studies. These are presented as separate rows in Table 4. Further details about all the studies are presented in Appendix 3.

All studies used light microscopy as the reference standard. One study compared digital whole slide images to the original diagnosis from patient notes made using glass slides read by light microscopy (Rakha *et al.* 2018[26]). The glass slide reports may or may not have been read by the same pathologist as the digital whole slide image. Three studies compared digital whole slide images to review of glass slides on light microscopy by the same pathologist (intraobserver agreement) after a washout period (Davidson *et al.* 2019[25], Wilbur *et al.* 2015[27] study 1 and study 2). One study used both approaches as the reference standard, depending on the availability of records (Al-Janabi *et al.* 2016[24]). The level of agreement was often presented as a kappa score, where a kappa score of < 0 is less than chance agreement, 0.01–0.20 is slight agreement, 0.21–0.40 is fair agreement, 0.41–0.60 is moderate agreement, 0.61–0.80 is substantial agreement and 0.81–0.99 is almost perfect agreement. Level of agreement was also presented as a Cramner's V score in Rakha *et al.* (2018)[26]. This measures the association between 2 variables where 0 is no association and 1 is complete association. Wilbur *et al.* (2015)[27] created dichotomous categories for positive results (HER2 score of 2+ and 3+) and negative results (HER2 score of 0 and 1+) and then reported the positive and negative percentage agreement.

Wilbur *et al.* (2015)[27] stated that all the participating pathologists had received training in digital whole slide images with standard case sets. Davidson *et al.* (2019)[25] stated that no training sets were provided to pathologists. In the other 2 studies (Al-Janabi *et al.* 2016[24], Rakha *et al.* 2018[26]), it was not clear if the pathologists had received any training in digital pathology.

Table 4: Summary of diagnostic concordance results for the grading of breast cancer

Test (magnification)	Reference standard	Population	Number of pathologists	Slides available	Result	Study
Digital whole slide images (x20)	Patient note reports using glass slides	1,675 cases of early-stage invasive breast cancer (detection route unclear)	1	Single representative slide selected for each case	Exact agreement of cancer grading[a]: 68% Kappa score: 0.51 (95%CI 0.47 to 0.54) Cramner's V: 0.58 Discordance: 32%	Rakha *et al.* (2018)[26] (Nottingham, UK)
Digital whole slide images (x40)	Glass slides review by light microscopy from records or re-assessed after a minimum washout period of 6-weeks	96 archive cases of invasive breast cancer (detection route unclear)	2	Not stated	Agreement of HER2 score: 73.1% (95%CI 63.9 to 82.3) Kappa score: 0.588	Al-Janabi *et al.* (2016)[24] (Netherlands)
Digital whole slide images (x40)	Glass slides review by light microscopy after a minimum washout period of 9-months	22 invasive breast cancer cases from registries in 2 US states (detection route unclear)	82	Single representative slide selected for each case	Agreement in Nottingham grade for digital and glass reviews by the same pathologist: 63% (95%CI 59 to 68) Kappa score: 0.38 (95%CI 0.30 to 0.46)[b]	Davidson *et al.* (2019)[25] (US)
Digital whole slide images (magnification not stated)	Glass slides review by light microscopy after a minimum washout period of 7 days	180 breast cancer cases from a Danish tissue bank (detection route unclear)	3	Not stated	Overall agreement in HER2 score: 87.2% (95%CI 84.1 to 89.8) Positive percentage agreement: 94.9% (95%CI 91.3 to 97.1)	Wilbur *et al.* (2015)[27], study 1[c] (Denmark)

					Negative percentage agreement: 81.3% (95%CI 76.6 to 85.3)	
Digital whole slide images (magnification not stated)	Glass slides review by light microscopy after a minimum washout period of 7 days	184 breast cancer cases from a Danish tissue bank (detection route unclear)	3	Not stated	Overall agreement in HER2 score: 88.8% (95%CI 85.7 to 91.7) Positive percentage agreement: 95.7% (95%CI 93.1 to 98.0) Negative percentage agreement: 82.8% (95%CI 77.9 to 87.3)	Wilbur *et al.* (2015)[27], study 2[c] (Denmark)

Abbreviations: CI - confidence interval; HER2 - human epidermal growth factor receptor; UK – United Kingdom; US – United States

a In Rakha *et al.* (2018) discordance was described as largely being between adjacent levels of grade. The authors also reported agreement, kappa scores and Cramner V for the grading of component: tubular score, nuclear pleomorphism and mitotic score and intraobserver agreement for 2 separate readings of digital whole slide images (see Appendix 3)
b Davidson *et al.* (2019) also reported agreement and kappa scores for the grading of component: tubular score, nuclear pleomorphism and mitotic score (see Appendix 3)
c In Wilbur *et al.* (2015) study 1, glass slides were read before digital whole slide images. In study 2, the order in which digital whole slide images or glass slides were read was randomised. Pathologists were blinded to any previous results.

It is not clear if any of the studies on the grading of breast cancer included studies detected by screening. In the UK study by Rakha *et al.* (2018)[26], the kappa score of 0.51 reported equates to moderate agreement, although the Cramner V score of 0.58 reported was stated to be substantial agreement by the study authors. In the 2 non-UK studies that also reported a kappa score, this was 0.588 (moderate agreement) in Al-Janabi *et al.* (2016)[24] and 0.38 (fair agreement) in Davidson *et al.* (2019)[25]. Wilbur *et al*. (2015)[27] did not report a kappa score but reported overall agreement of 87% and 89%. As with the studies about the primary diagnosis of breast cancer, the comparison of studies is complicated by differences between the studies. The lowest agreement reported was by Davidson *et al.* (2019)[25]. This study included a large number of pathologists, approximately half of which had experience of using digital pathology in their professional practice. The study authors stated that pathologists were not provided with training sets, written instructions or standardised diagnostic criteria.

Additional outcomes reported by Al-Janabi *et al.* (2016)24 were test metrics for scoring the non-amplified (normal) and amplified categories for HER2 score[††] and time spent in scoring. Sensitivity was 95.45%, specificity was 100%, positive predictive value was 100% and negative predicative value was 97.14%. Time spent in scoring (reported for 30 cases) was 81.7 seconds for digital whole slide images and 86.9 seconds for glass slides by light microscopy (no statistical comparison of digital vs glass reported).

Quality of the evidence base

The quality of the included studies was appraised using an adapted QUADAS-2 checklist (Table 5). The main area of high or unclear risk of bias across multiple studies related to the study populations. In half of the studies, the cases were either not randomly or consecutively selected, or it was unclear how cases were selected. With the exception of the study by Williams *et al.* (2020)[7], it was unclear if any of the cases used in the studies were detected by screening.

In Williams *et al.* (2020)[7] (study 2) and Williams *et al.* (2018)[23], the glass slides were reviewed immediately after the digital whole slide images. The reference standard results were therefore not interpreted without knowledge of index test results (high risk of bias). In relation to the 'flow and timing' domain, the College of American Pathologists guidelines recommend a period of at least 2 weeks between viewing digital and glass slides[12]. No washout period appears to have been applied in Williams *et al.* (2020)[7] (study 2) and Williams *et al.* (2018)[23] suggesting that there was not an appropriate interval between index test and reference standard (high risk of bias). In Wilbur

[††] In Al-Janabi *et al.* (2016), samples with an average of <4 spots/nucleus were considered non-amplified (normal); samples with between 4 and 6 spots/nucleus were considered equivocal; samples with ≥6 spots/nucleus were considered amplified

et al. (2015)[27] the minimum washout period was 7 days, although the range was 7 to 51 days. The proportion of intervals between reads that were appropriate is therefore unclear.

Most of the studies reporting concordance relating to the primary diagnosis of breast cancer reported their results as percent agreement or disagreement. Only one of the studies relating to the primary diagnosis of breast cancer, Williams *et al.* (2020)[7] (study 3), also reported the result as a kappa score. Kappa score was more commonly reported in the studies focusing on the grading of breast cancer. The kappa statistic was developed to account for the possibility of chance agreement to increase the reliability of a result[29]. It often therefore appears to indicate a lower level of agreement than percent agreement, although there are standard thresholds for the interpretation of kappa score, in terms of whether it represents for example, a moderate or substantial agreement. There are advantages and disadvantages to both percent agreement and kappa score. Both can provide useful information and the extent to which chance is likely to influence the results influences the value of each[29]. In interpreting the results, whether presented as percent agreement or kappa score, detail on the definition of the ‘agreement’ that is being assessed and the nature of any disagreements observed is important, and these details were not always adequately reported by the studies. Another limitation for the interpretation of the results is the absence of any confidence intervals around the results in many of the studies.

Table 5. QUADAS-2 scores summary indicating the areas of low, unclear or high risk of bias

Risk of bias	Al-Janabi *et al.* (2016)[24]	Babwale *et al.* (2021)[19]	Borowsky *et al.* (2020)[20]	Davidson *et al.* (2019)[25]	Elmore *et al.* (2017)[21]	Mukhopad-hyay *et al.* (2018)[22]	Rakha *et al.* (2018)[26]	Wilbur *et al.* (2015)[27‡‡]	Williams *et al.* (2018)[23]	Williams *et al.* (2020)[7] study 1	Williams *et al.* (2020)[7] study 2	Williams *et al.* (2020)[7] study 3
Domain I: Patient selection												
Consecutive or random sample of population enrolled?	High	Unclear	Low	High	Low	Low	Low	High	Low	Unclear	Low	High
Case-control design avoided?	N/A	Low	Low	N/A	Low	Low	N/A	N/A	Low	Low	Low	Low
Inappropriate exclusions avoided?	Low	Low	Low	Low	Low	Low	Low	Low	Low	Low	Low	Low
Domain II: Index Test												
Index test results interpreted without knowledge of reference standard results?	Low	Low	Low	Low	Low	Low	Low	Low	Low	Low	Low	Low
Threshold pre-specified?	Low	Low	Low	Low	Low	Low	Low	Low	Low	Low	Low	Low
Domain III: Reference standard												
Reference standard likely to correctly classify condition?	Low	Low	Low	Low	Low	Low	Low	Low	Low	Low	Low	Low
Reference standard results interpreted without knowledge of index test results?	Low	Low	Low	Low	Low	Low	Low	Low	High	Low	High	Low
Domain IV: Test strategy flow and timing												
Appropriate interval between index test and reference standard?	Low	Low	Low	Low	Low	Low	Low	Unclear	High	Low	High	Low

‡‡ The quality scores were the same for study 1 and study 2 reported by Wilbur *et al.* (2015)

Risk of bias	**Al-Janabi *et al.* (2016)[24]**	**Babwale *et al.* (2021)[19]**	**Borowsky *et al.* (2020)[20]**	**Davidson *et al.* (2019)[25]**	**Elmore *et al.* (2017)[21]**	**Mukhopad-hyay *et al.* (2018)[22]**	**Rakha *et al.* (2018)[26]**	**Wilbur *et al.* (2015)[27‡‡]**	**Williams *et al.* (2018)[23]**	**Williams *et al.* (2020)[7] study 1**	**Williams *et al.* (2020)[7] study 2**	**Williams *et al.* (2020)[7] study 3**
Did all participants receive same reference standard?	Low	Low	Low	Low	Low	Low	Low	Low	Low	Low	Low	Low
All patients included in analysis?	Low	Low	Low	Low	Low	Low	Low	Low	Unclear	Low	Low	Low
Domain V: Applicability												
Applicable to UK screening population of interest?	Unclear	Unclear	Unclear	Unclear	Unclear	Unclear	Unclear	Unclear	Unclear	Low	Low	Low
Applicable to UK screening test of interest?	Low	Low	Low	Low	Low	Low	Low	Low	Low	Low	Low	Low
Target condition measured by reference test applicable to UK screening condition of interest?	Low	Low	Low	Low	Low	Low	Low	Low	Low	Low	Low	Low

Overall, most of the studies identified included a small number of cases and a small number of pathologists. No studies reported a power calculation relating to the number of cases included in the study. It is not clear if the studies were adequately powered. The quality of the included studies was generally fairly high when assessed against the QUADAS-2 tool. However, the areas of potential bias identified in some studies do potentially reduce confidence in the results.

The consistency of the results across studies was difficult to assess due to differences in the study populations, designs of the studies and the reporting of results. For example, differences in the magnification used to view digital slide images, the processes used in the way that the pathologists viewed the cases or in how the light microscopy reference standard was assessed, the definitions of what was considered agreement and differences in the way in which the results were analysed. Most of the studies did not report confidence intervals around the concordance or agreement scores. Where they were reported, they were wide in some studies.

The applicability of the evidence base to screening programmes in the UK was also unclear. Only one paper used data from the UK Breast Cancer Screening Programme. The applicability of some studies to clinical practice was also unclear. For example, in some studies a single representative slide was selected for each case, in some studies pathologists either had access to the full case slides or could request any slides or stains required.

Digital pathology compared to light microscopy for the primary diagnosis of cervical cancer

The one study comparing digital pathology (whole slide images) to light microscopy (glass slides) relating to the primary diagnosis of cervical cancer (Ordi *et al.* 2015[28]) evaluated the validity of digital whole slide images in 157 biopsies or excisions (number of slides not clear) of the uterine cervix from patients referred to colposcopy in one centre in Spain because of an abnormal Pap smear. The concordance between one pathologist who reviewed all cases using digital whole slide imaging and a second pathologist who reviewed the same cases on glass slides by light microscopy (the reference standard) was reported:

- complete agreement: 86.6% (95%CI 80.3 to 91.5)
- kappa score: 0.832 (95%CI 0.757 to 0.906)
- major discrepancies: 8/157 (5.1%)
- minor discrepancies: 13/157 (8.3%)

Major discrepancies were defined as differences with clinical and/or prognostic implications for the patients. Minor discrepancies were defined as mild differences which would not have any clinical or prognostic implications. The kappa score reported equates to almost perfect agreement.

This single study included a small number of cases and involved 2 pathologists, one pathologist reviewing cases using a digital format and one using a glass slide format. There were no areas of concern when the study was assessed using the QUADAS-2 tool and the study was deemed to be at low risk of bias overall. However, due to the design of the study it is not possible to determine if the differences in complete agreement observed are due to the different viewing format or due to differences between the 2 pathologists. This study was set in Spain. It is not clear if the results of this study would be consistent with or applicable to UK practice.

Further details of the study are presented in Appendix 3.

Additional studies identified of potential interest (not formally included in the evidence summary)

Two UK studies were identified reporting concordance within a modality format (for example agreement between separate readings using digital whole slide images) (Dessauvagie *et al.* 2018[30], van Seijen *et al.* 2021[31]). These differ from the studies that were formally included in this evidence summary which explored concordance between digital whole slide images and light microscopy. The studies below are therefore briefly summarised for information. Additional results on intraobserver agreement within a modality format from 2 UK studies that were included in this evidence summary are also briefly described (Rakha *et al.* 2018[26], Williams *et al.* 2020[7]).

A study coordinated from the UK, Dessauvagie *et al.* (2018)[30], reported the results of a diagnostic audit of 69 fibroepithelial lesions categorised as B3 (lesion of uncertain malignant potential) on core needle biopsies and subsequent excisions retrieved from archives of the Leeds Teaching Hospital NHS Trust from 2009 to 2018. It was not stated if any of these were detected by screening. Eight pathologists from 4 tertiary pathology institutions in 3 unspecified countries assessed the whole slide images. The pathologists ranged in experience from recently qualified (less than 10 years specialist experience) to specialised breast pathologists (more than 10 years specialist experience with a dominant practice in breast pathology). For core needle biopsies, the mean kappa score for agreement of diagnosis using whole slide images between pathologists was 0.36 (fair agreement; no confidence intervals reported; range 0.15 to 0.55). For specialist pathologists, the mean kappa value[§§] was 0.44 (moderate agreement; no confidence intervals reported; range 0.33 to 0.51) and 0.35 for generalist pathologists (fair agreement; no confidence intervals reported; range 0.22 to 0.51). For excision biopsies, the mean kappa scores for agreement of diagnosis using whole slide images between pathologists was 0.49 (moderate agreement; no confidence intervals reported; range 0.32 to 0.74) for all pathologists, 0.54 for specialists (moderate agreement; no confidence intervals reported; range 0.44 to 0.74), and 0.44 for generalists (moderate agreement; no confidence intervals reported; range 0.23 to 0.63).

One study evaluated agreement in grading for 425 breast ductal carcinoma *in situ* samples using only whole slide images between 9 British, Dutch and American pathologists (van Seijen *et al.* 2021[31]). It is not clear if these samples were from screening programmes. For grading (low, intermediate or high), the chance-corrected kappa score for association between pathologists was 0.50 (95%CI 0.44 to 0.56) indicating a moderate association. For a subgroup analysis of pathologists using UK pathology guidelines this was 0.58 (95%CI 0.56 to 0.61) (moderate agreement).

In addition, Rakha *et al.* (2018)[26] and Williams *et al.* (2020)[7] study 3 also reported intraobserver agreement within modality format (see included studies above and in Appendix 3 for study details). Rakha *et al*. (2018) reported the results for 2 separate readings of digital whole slide images by one pathologist with a 3-month washout period. The kappa score for agreement of grade was 0.65 (95% 0.60 to 0.68) and the Cramner's V statistic was 0.65, both suggesting substantial agreement. Williams *et al.* (2020) reported the results for 2 separate readings of digital whole slide images and glass slides by light microscopy by the same pathologist with a 2-week washout period. For digital whole slide images, the agreement was 87% with a kappa score of 0.80 (95%CI 0.72 to 0.87). For glass slides, the agreement was 85% with a kappa score of 0.78 (95%CI 0.57 to 0.81). Both of these kappa scores suggest substantial agreement.

[§§] kappa < 0 = less than chance agreement; kappa 0.01–0.20 = slight agreement; kappa 0.21–0.40 = fair agreement; kappa 0.41–0.60 = moderate agreement; kappa 0.61–0.80 = substantial agreement; kappa 0.81–0.99 = almost perfect agreement

In addition, 3 non-UK studies included in this evidence summary also reported concordance within a modality for the same pathologist and between different pathologists (Davidson *et al.* 2019[25], Elmore *et al.* 2017[21], Wilbur *et al.* 2015[27]). Please see Appendix 3 for further details of these results.

More than 50 additional validation studies were found concerning the evaluation of automated systems for analysing whole slide imaging compared to pathologist interpretation of glass slides/whole slide images. These artificial intelligence studies related mostly to the classification of breast cancer. Four studies were primarily concerned with the diagnosis of breast cancer[32,33,34,35] and 3 were concerned with cervical cancer histology samples[36 ; 37; 38]. Approximately a third of the studies identified were from the USA or Canada, approximately a quarter were from Asia or Europe and approximately 10% were from the UK.

Further considerations

As described in the introduction, pathologist review of glass slides using light microscopy is considered the current reference (gold) standard[1]. The studies included in this evidence summary all used pathologist review of glass slides as the reference standard. However, they differed in whether one, or more than one, pathologists were involved in the process of establishing the diagnosis. Likewise, in clinical practice, a diagnosis may be made by one pathologist or a second opinion or consensus diagnosis may be sought for some cases. These processes by which pathologists make a diagnosis, or establish ground truth, would be unlikely to change if whole slide images were to be considered an equivalent viewing mechanism to light microscopy and introduced into the screening programme. It is unclear at present whether the intention would be to entirely replace the use of light microscopes or whether these would remain as an option in more difficult cases if it was felt that the digital image was insufficient to make a diagnosis. If the option of reviewing a glass slide were to be removed before the digital images of glass slides are considered equivalent to the original glass slide in terms of the quality of the image, this might be considered a situation where the current reference standard was being replaced by a less accurate reference standard. However, if the digital images produced from glass slides are considered equivalent to the original glass slide in terms of the quality of the image for diagnostic purposes, the use of digital pathology might represent an alternative viewing means rather than a change to the reference standard.

Uncertainty about the future position of light microscopy in the diagnostic pathway is also of relevance to potential issues relating to the spectrum of disease and the potential presence of subtler morphologies in screening cases. If light microscopy remained an option in more difficult cases, the importance of the specific consideration of evidence from screen detected cases may be lessened. However, it would still be important to understand the

judgements and decisions that are made about screen detected cases using digital pathology. For example, whether there is a change in the spectrum of cases detected using digital pathology compared to glass slides using light microscopy.

One of the suggested advantages for the use of digital pathology is the potential for increased use of automated image analysis. If automated image analysis were to be introduced into the screening pathway to replace pathologist review of an image, this would represent a difference to the process by which a diagnosis is established and would require a revisitation of potential methodological issues surrounding the replacement of a reference standard. However, the use of automated image analysis is outside the scope of this evidence summary.

Glass slides are produced and scanned to create the whole slide images. At present, the Royal College of Pathologists considers glass slides as the primary reference image and recommends that they are retained for at least 10 years[3]. The current recommendation is that digital images are retained for a period of 2 laboratory inspection cycles[3]. However, this is based on the current recommendation that the glass slide is retained. Governance issues around the place of glass slides and digital images within the validation and internal and external quality assurance of systems, laboratories and pathologists will need to be considered if digital pathology is to be introduced into the screening pathway.

Summary of Findings Relevant to Criteria 4 and 5: Criteria not met[***]

The included studies on the accuracy of digital pathology in breast cancer generally showed a high degree of agreement between digital pathology and light microscopy. However, the evidence specifically relating to the accuracy of digital pathology in cases detected by screening is limited. The interpretation of the evidence base is also limited by the small number of cases involved and differences in the designs and process used in the studies. Further, larger, high-quality studies with improved methodological consistency assessing the accuracy of digital pathology using cases detected through breast cancer screening programmes are needed. An ongoing HTA primary study is due to conclude in October 2021 and, once published, is expected to produce useful results on the performance of digital pathology in cancer screening.

[***] **Met** -for example, this should be applied in circumstances in which there is a sufficient volume of evidence of sufficient quality to judge an outcome or effect which is unlikely to be changed by further research or systematic review.
Not Met - for example, this should be applied in circumstances where there is insufficient evidence to clearly judge an outcome or effect or where there is sufficient evidence of poor performance.
Uncertain -for example, this should be applied in circumstances in which the constraints of an evidence summary prevent a reliable answer to the question. An example of this may be when the need for a systematic review and meta-analysis is identified by the rapid review.

The evidence available on the accuracy of digital pathology in cervical cancer was limited to a single small study, restricting any conclusions that can be drawn on the use of digital pathology in this area.

Overall, criteria 4 and 5 are not met.

Criterion 12 — Acceptability of the use of digital pathology to health professionals and the public

12. There should be evidence that the complete screening programme (test, diagnostic procedures, treatment/ intervention) is clinically, socially and ethically acceptable to health professionals and the public

Question 2 — Is the use of digital pathology in breast and cervical cancer screening clinically, socially and ethically acceptable to health professionals and the public?

The 2020 evidence map described evidence relating to the acceptability of digital pathology from 5 UK studies.

Eligibility for inclusion in the review

Qualitative, quantitative and mixed methods studies assessing the perceptions, views and/or attitudes, and/or experiences of pathologists, health professionals, patients and members of the public regarding the use of digital pathology in breast and cervical cancer screening were all eligible for inclusion. As UK studies were identified, these were prioritised for inclusion. Though studies on the acceptability of digital pathology in non-UK countries were not formally included in this evidence summary, they are synthesised in a short narrative for information. Studies relating to the acceptability of digital pathology in breast and cervical cancer screening were prioritised. However, UK studies exploring the acceptability of digital pathology more broadly were also included. Studies on the acceptability of digital pathology in liquid-based cytology were not eligible for inclusion.

Description of the evidence

Of the 53 papers reviewed at full text, 11 related to this question. Six studies met the criteria for inclusion. Appendix 2 contains a full PRISMA flow diagram (Figure 1), along with a table of the included publications and details of which questions these publications were identified as being relevant to (Table 7). The excluded studies are listed in Table 8 in Appendix 2.

The 6 included UK studies for this question comprised:

- one validation study on digital pathology in breast cancer (using screening data) that included some feedback from pathologists (Williams *et al.* 2020[7])
- one survey on the use of digital pathology in breast cancer (Dessauvagie *et al.* 2018[30])
- 2 surveys on the use of digital pathology in clinical settings (case mix) (Browning *et al.* 2020[39], Williams *et al.* 2018[40])

- one focus group study on the benefits and challenges of transitioning to digital pathology (case mix) (Turnquist *et al.* 2019[41])
- one validation study on digital pathology (case mix) that included information on lessons learnt (Williams *et al.* 2019[42])

The paper by Williams *et al.* (2020)7 involved cases detected through the NHS Breast Screening Programme. However, no studies specifically explored the acceptability of digital pathology within screening programmes. No studies on the acceptability of digital pathology in cervical cancer, outside of liquid-based cytology, were identified. No evidence was identified on the acceptability of using digital pathology in screening to the public.

The number of pathologists providing feedback ranged from 7 to 24, where this was known. Three studies were based in one centre and 2 included pathologists from 4 centres. One study, (Williams *et al.* 2018)[40] sought views from 41 pathology departments. These views were expressed either at a departmental or at an institutional level.

Discussion of findings

A study-level summary of data extracted from each included publication is presented in the 'summary and appraisal of individual studies' in Appendix 3. In Appendix 3, publications are stratified by question and are presented in alphabetical order.

In the discussion below, studies relating to the acceptability of digital pathology in breast cancer are presented first, followed by studies about the use of digital pathology more broadly.

Acceptability of digital pathology in breast cancer

The paper by Williams *et al.* (2020)[7] described benefits of digital pathology reported by participating pathologists at 4 clinical sites (3 UK sites and one in Lithuania). The number of pathologists who provided feedback is not clear. The validation studies that the pathologists were participating in primarily used cases from the NHS Breast Screening Programme. Dessauvagie *et al.* (2018)[30] reported the results of a feedback survey on pathologists' experiences, comfort and confidence in using digital pathology. The 8 pathologists from 4 centres (based in 3 unspecified countries) who provided feedback were taking part in a study on interobserver variability using breast cancer cases from the archives of one UK centre.

In Dessauvagie *et al.* (2018)[30], 7 pathologists (88%) reported feeling comfortable utilising the digital platform for the audit and that they were satisfied with the quality of the digital images provided. Six pathologists (75%) felt comfortable using digital pathology for diagnosis in routine practice. Four of the pathologists (50%) reported using digital pathology in routine diagnostic work

and all (100%) reported use in some aspect of their clinical practice. Williams *et al.* (2020)[7] provided more detail on the benefits of using digital pathology. These included:

- loss of need for glass slide transport and transfer delays
- rapid and convenient availability of images for sharing and second opinion
- rapid access to previous biopsies for comparison with resection/repeat biopsies
- perceived increased efficiency in the diagnosis of large volume biopsies/multi-slide and multi-level cases
- occupational health benefits from being able to complete work digitally
- enhanced opportunities to demonstrate pathology in multi-disciplinary team meetings
- useful for teaching a larger cohort of trainees across a wide geographical area
- feasibility of applying artificial intelligence-based tools in the routine setting of breast pathology reporting

In Dessauvagie *et al.* (2018)[30], barriers with digital pathology reported included slow visual scanning of slides at low power, worse resolution compared to conventional microscopy, poor viewing screen quality and it being harder and more time consuming to identify and quantify mitosis. Williams *et al.* (2020)[7] did not report on barriers to the use of digital pathology.

Acceptability of digital pathology (case mix studies)

Four UK studies, Browning *et al.* (2020)[39], Turnquist *et al*. 2019[41], Williams *et al.* (2018)[40] and Williams *et al.* (2019)[42], considered the acceptability and potential barriers of digital pathology. These studies included a mix of specialties or organ types with results not reported separately by specialty or organ type.

Two surveys were completed by pathologists (number not reported) from 41 UK NHS and academic pathology departments in 2017 (Williams *et al.* 201840) and by 18 pathologists (of 34 approached) from one UK centre (Oxford) in 2020 (Browning *et al.* 202039). The latter 2020 survey included consideration of the impact of the COVID-19 pandemic.

In Williams *et al.* (2018)[40], 60% (23/39) of institutions had access to a digital pathology scanner, 60% (24/40) had access to a digital pathology workstation and 58.5% (24/41) listed the investigation and use of digital pathology as a high or essential priority at their institution. In Browning *et al.* (2020)[39], 9 of the 18 pathologists (50%) had used digital pathology prior to the COVID-19 pandemic and 14 (78%) had used it during the pandemic. Fourteen of the 18 pathologists (78%) agreed that digital pathology is a positive step for their specialty team and 16 (89%) agreed that they would likely continue reporting digitally beyond the pandemic.

In Williams *et al.* (2018)[40], the most popular applications of digital pathology in current use were undergraduate and postgraduate teaching and research and quality assurance. Thirty-one per cent

(n not stated) used digital pathology for primary diagnosis and 36% (n not stated) used digital pathology for secondary diagnosis. The majority of respondents agreed or strongly agreed that digital pathology would improve:

- collaboration (97%)
- efficiency (70%)
- turnaround times (56%)
- staff time (56%)

In Browning *et al.* (2020)[39], 3 of the 18 pathologists (17%) used digital pathology to report all clinical cases and 6 (33%) some clinical cases. Other uses of digital pathology included:

- quick review of a case to determine if immunohistochemistry/ special stains were needed (83%)
- second opinions (67%)
- to demonstrate images in a multi-disciplinary team meeting (56%)
- to prepare/ review a case prior to a multi-disciplinary team meeting (78%)

In Williams *et al.* (2018)[40], areas of concern raised by the 41 responding UK NHS and academic pathology departments included laboratory safety (37%) and costs (20%), with 83% agreeing or strongly agreeing that initial financial cost was a barrier to wider digital pathology usage at their institution. Other barriers to wider use included time cost (49%), not being perceived as useful in the department (33%) and lack of training (31%). In Browning *et al.* (2020)[39], concerns raised by 7 of the 18 pathologists (39%) related to occasional out of focus slides. Specific challenges faced during the upscaling of digital pathology due to the COVID-19 pandemic included setting up workstations, including internet access and equipment, for remote working.

The focus group study (Turnquist *et al.* 2019[41]) included 3 pathologists and 3 biomedical scientists (a general laboratory manager, a quality manager and an IT manager) in one UK hospital (Oxford) who had been involved in a digital pathology pilot. Additional comments were also sought from a breast pathologist who was not present at the focus group. Benefits to digital pathology identified included improvements to:

- collaboration
- training and teaching (access to an archive of cases)
- cost savings
- research
- growth of speciality
- multi-disciplinary team meetings
- patient-centred care (through patients having more access to their images, fostering greater patient involvement and communication between patient and doctor)

Barriers to the implementation of digital pathology included concerns about:

- standardisation
- validation
- national implementation
- storage and backups
- variations in training (due to differences in digital pathology implementation)
- technical issues
- cost-effectiveness
- workload
- privacy/ legality

The final UK study included, (Williams *et al.* 2019[42]), explored lessons learned with 24 pathologists at one UK centre (Leeds) who had completed training in digital immunohistochemistry using 1,480 digital immunohistochemistry slides. Feedback was reported separately for pathologists who had a high satisfaction score (6 or 7 out of 7) and pathologists who had low confidence or satisfaction (less than 6 out of 7). The mean satisfaction score with digital slides was 5.91 (range 2 to 7). The mean confidence score with digital slides was 6.1 (range 2 to 7). Advantages for digital slides over glass slides from pathologists with high satisfaction included:

- digital as quick and easy as glass slides
- easier to spot areas of concern at low power
- positive results are spotted more quickly
- easier to assess a multi-slide case
- easier and quicker to use
- digital slides seem more crisp

Feedback from pathologists with lower confidence or satisfaction included:

- headache from screening large volumes of tissue for rare positive cells
- took longer to scroll through all the tissue at high power than on light microscope
- need higher magnification scanning for some stains
- *H pylori* blurry and difficult to spot

The authors reported that for the stains that were found to be difficult or time consuming to assess using 20x digital slides, rescanning at 40x improved the confidence of the pathologists to make a digital assessment.

Quality of the evidence base

Three of the 6 studies were formally assessed with the CASP critical appraisal checklist for qualitative research (Appendix 3, Tables 23, 25 and 28). The other studies included some feedback on acceptability but were designed as validation studies. These studies were not formally assessed with the checklist for qualitative studies. However, the key areas covered by the

checklist were considered for the feedback aspects of these studies. The main areas of concern across the 6 studies related to how representative or widespread the views expressed were. This was due to the small number of pathologists providing feedback but also the fact that many of them were participating in a specific study or training programme on digital pathology.

The contexts and areas explored by the studies varied. However, there was some overlap in the advantages and limitations to digital pathology described. Studies from the UK were identified, although only one was conducted in the context of cases from a UK screening programme. It is not clear if the views expressed in this study or any of the other studies related specifically to the use of digital pathology in screening. Two of the 6 studies were in the context of a study about breast cancer. The remaining 4 studies were not specific to any discipline. No studies about the acceptability of digital pathology in cervical cancer were identified.

Additional studies identified of potential interest (not formally included in the evidence summary)

Six non-UK studies were also identified with some information relating to the acceptability of digital pathology (Elmore *et al.* 2020[43], Hanna *et al.* 2020[44], Hanna *et al.* 2019[45], Lundström *et al.* 2016[46], Stathaniankos *et al.* 2019[47], Unternaehrer *et al.* 2020[48]). None of these studies were based on a specific speciality and there was no indication that any were exploring the acceptability of the use of digital pathology in screening. The studies below are briefly summarised for information and have not been formally included in this evidence summary, as UK-based studies were prioritised instead.

Three studies were from the USA. Elmore et al. (2020)43 reported the results of a 2019 survey with 76 students on 9 pathology trainee programmes (of 159 trainees invited to participate). Exposure to digital pathology varied and was higher for students who started training more recently. Comfort with using whole slide images for interpretation was higher for trainees with exposure to whole slide images in medical school (29% vs 12%, p=0.06). Most trainees agreed that whole slide images can be used to make accurate primary diagnoses (92%, 95%CI 6 to 98) or for obtaining second opinions (93%, 95%CI 88 to 99). Hanna *et al.* (2019)45 was a validation study with 8 pathologists in one centre from various specialities including breast and gynaecologic (not reported separately). Six pathologists responded to a survey on their experience, with a median of 4.5 years of experience using digital pathology (range 2 to 10). Most of the questions related to the specific digital systems and process used. However, the median overall experience of using digital whole slide images was neutral on a 5-point scale ranging from very poor to very good. Comfort level with the option of using whole slide images for primary diagnosis varied with the median response being uncomfortable on a 5-point scale ranging from very uncomfortable to very comfortable. Hanna *et al.* (2020)[44] investigated the validation of a digital pathology system during the COVID-19 pandemic. Twelve pathologists from 9 specialities including breast and

gynaecologic (not reported separately) remotely reviewed whole slide images. A survey on their experiences was completed by 10 pathologists with a median of 5 years of experience using digital pathology (range 3 to 10). Satisfaction with the performance in navigating the whole slide images ranged from neutral to very good on a 5-point scale from very poor to very good. Comfort level with the option of using whole slide images with the availability of glass slides was good or very good for 90% of respondents. Five pathologists (50%) rated their level of comfort in using digital pathology for primary diagnosis without the availability of glass slides as comfortable.

The remaining 3 studies were from Europe. Unternaehrer *et al.* (2020)[48] reported the results of a national survey in Switzerland assessing pathologists' experiences of digital pathology. The estimated response rate was 39.5%, of which 89% had experience of digital whole slide images, mainly in education (61%) and primary diagnostics (20%). Of the 134 respondents, 66% reported feeling comfortable with the idea of making a primary diagnosis digitally and 10% reported that they would not be comfortable. Most (54%) of respondents believed that more standards and regulations are necessary for the clinical employment of digital pathology. Stathaniankos *et al.* (2019)[47] described the experience of one centre in the Netherlands that implemented a fully digital workflow for primary diagnostics in 2015. This included a survey of 23 pathologists from different specialities with at least 6 months experience of the digital system. Most respondents felt very confident (43.5%) or rather confident (30.4%) in working digitally. The most frequent complaint about the quality of scanned slides related to difficulties in discriminating microorganisms and mitosis. Neither the age nor experience of the pathologists was correlated with confidence. Most respondents also reported digital microscopy to be ergonomic (47.8%) or very ergonomic (17.4%) with fewer injuries related to digital microscopy than light microscopy. Lundström *et al.* (2016)[46] reported a summary of the 2015 Nordic symposium on digital pathology which included 190 attendees from 15 countries. Of 83 attendees invited to complete a survey on the use of digital pathology, 49 responded. Respondents included pathologists (35%), pathologists in training (14%), lab technologists (2%), clinical management (10%), IT staff (19%) and other (20%). Respondents felt that the perceived added safety risks from digitalisation was lower relative to other risks within the diagnostic pipeline (results only reported graphically). The conclusion of the symposium discussion was a broad consensus that continued efforts to scrutinise digital pathology were needed, with differing views on the degree to which this would affect the pace of adoption.

Summary of Findings Relevant to Criterion 12: Criterion not met[†††]

UK studies relating to the acceptability of digital pathology were identified. Benefits mentioned across several studies included the ease and efficiency of sharing images for teaching, second opinions and multi-disciplinary team meetings. Barriers mentioned included financial costs, slide and equipment quality and lack of training and standardisation. However, the studies included a small number of pathologists and were not specific to the acceptability of the use of digital pathology in breast or cervical cancer screening. It was not clear how representative the views expressed were of the wider community of pathologists in the UK. No evidence was identified on the acceptability of using digital pathology in screening to the public. Therefore, criterion 12 is not met.

[†††] **Met** -for example, this should be applied in circumstances in which there is a sufficient volume of evidence of sufficient quality to judge an outcome or effect which is unlikely to be changed by further research or systematic review.
Not Met - for example, this should be applied in circumstances where there is insufficient evidence to clearly judge an outcome or effect or where there is sufficient evidence of poor performance.
Uncertain -for example, this should be applied in circumstances in which the constraints of an evidence summary prevent a reliable answer to the question. An example of this may be when the need for a systematic review and meta-analysis is identified by the rapid review.

Criterion 14 — Cost-effectiveness of digital pathology compared to light microscopy

14. The opportunity cost of the screening programme (including testing, diagnosis and treatment, administration, training and quality assurance) should be economically balanced in relation to expenditure on medical care as a whole (value for money). Assessment against this criterion should have regard to evidence form cost benefit and/or cost-effectiveness analysis and have regard to the effective use of available resource.

Question 3 — Are there any health economic/ cost-effectiveness analyses and/or models on the use of digital pathology in breast and cervical cancer screening compared to the use of light microscopy? If so, what do they show?

No studies on the cost-effectiveness of digital pathology in breast and cervical cancer screening, or on the use of digital pathology more broadly, compared to the use of light microscopy were identified.

Summary of Findings Relevant to Criterion 14: Criterion not met[‡‡‡]

As no evidence on the cost-effectiveness of digital pathology in breast and cervical cancer screening, or on the use of digital pathology more broadly, compared to the use of light microscopy was identified, it is not possible to draw conclusions in this area. Therefore criterion 14 is not met.

[‡‡‡] **Met** -for example, this should be applied in circumstances in which there is a sufficient volume of evidence of sufficient quality to judge an outcome or effect which is unlikely to be changed by further research or systematic review.
Not Met - for example, this should be applied in circumstances where there is insufficient evidence to clearly judge an outcome or effect or where there is sufficient evidence of poor performance.
Uncertain -for example, this should be applied in circumstances in which the constraints of an evidence summary prevent a reliable answer to the question. An example of this may be when the need for a systematic review and meta-analysis is identified by the rapid review.

Review summary

Conclusions and implications for policy

This evidence summary reviews the evidence on the use of digital pathology in breast and cervical cancer screening against the UK NSC criteria about the test accuracy, the acceptability of the technology to health professionals and the public, and its cost-effectiveness.

The volume, quality and direction of new evidence is insufficient at present to determine the accuracy, acceptability and cost-effectiveness of digital pathology in breast and cervical cancer screening.

The included studies on the accuracy of digital pathology in breast cancer generally reported a high degree of agreement between digital pathology and light microscopy, although with limited information to interpret the clinical significance of the results in some studies. The evidence specifically relating to the accuracy of digital pathology in cases detected by screening is limited. The interpretation of the evidence base is also affected by the small number of cases involved, by the differences in the designs and process used in the studies, and by the differences in the way that results were reported. Further, larger, high quality studies with improved methodological consistency, assessing the accuracy of digital pathology using cases detected through breast cancer screening programmes, are needed. An ongoing Health Technology Assessment primary study is due to conclude in October 2021 and, once the results are released, is expected to produce useful results on the performance of digital pathology in cancer screening. The evidence available on the accuracy of digital pathology in cervical cancer was limited to a single small study, restricting any conclusions that can be drawn.

Evidence relating to the acceptability of digital pathology amongst pathologists in the UK was mixed with both positives and concerns identified. However, these studies were not specifically exploring the acceptability of digital pathology in screening programmes. It is also unclear how representative the views expressed by participants in these studies are to the wider UK population of pathologists or health professionals. No evidence was identified on the acceptability of using digital pathology in screening to the public. Studies exploring the acceptability of the use of digital pathology within screening programmes within a wider population of stakeholders are needed. The ongoing HTA should provide some additional evidence in this area as its protocol outlines a qualitative component which aims to explore views and experiences of pathologists and laboratory staff migrating from light microscopy to digital pathology.

Studies on the cost-effectiveness of digital pathology compared to light microscopy are also needed, given that no evidence addressing this specific area was identified in the available peer reviewed literature. As previously mentioned, part of the ongoing HTA also consists of a health economic study which is expected to assess the incremental costs associated with digitalisation compared to light microscopy-based pathology.

In conclusion, there is insufficient evidence to recommended further work on the use of digital pathology in cervical cancer screening at this time. Further work specifically exploring the accuracy, acceptability and cost-effectiveness of digital pathology in breast cancer screening is warranted. However, the publication of the ongoing HTA report on a multi-centred validation of digital whole slide imaging for routine diagnosis could be the point at which further consideration is needed on whether additional work is required on this topic.

Limitations

This evidence summary was conducted according to the UK NSC evidence review process over a condensed period of time. The review only looked for peer-reviewed scientific work and does not include work published elsewhere (grey literature). Studies not available in the English language, abstracts and poster presentations were not eligible for inclusion. Given that these are accepted methodological adjustments for a rapid review, and that the searches for this evidence summary covered relevant literature since January 2015 (when the searches for the systematic review by Goacher *et al*. were carried out), these limitations should not have led to the exclusion of any pivotal studies.

Appendix 1 — Search strategy

Electronic databases

The search strategy included searches of the databases shown in Table 6. The searches were originally run on the 19 October 2020 and were updated on 20 May 2021.

In May 2021, the Endnote library of results from the 2020 search was also re-checked for cost studies (using cost and economic as search terms) and experience/perceptions studies (using view, attitude, opinion, perspective, perception, satisf., focus group, interview, questionnaire and survey as search terms). This was done to reflect the addition of specific key questions on acceptability and costs to the commissioning document for the evidence summary and resulted in 122 records for re-screening.

Table 6. Summary of electronic database searches and dates

Database	Platform	Searched on date	Date range of search
MEDLINE, MEDLINE In-Process, MEDLINE Daily, Epub Ahead of Print	Ovid SP	19th October 2020 20th May 2021	January 2015 to May 2021
Embase	Ovid SP	19th October 2020 20th May 2021	January 2015 to May 2021
The Cochrane Library, including: - Cochrane Database of Systematic Reviews (CDSR) - Cochrane Central Register of Controlled Trials (CENTRAL) - Database of Abstracts of Reviews of Effects (DARE)	Wiley Online	19th October 2020 20th May 2021	January 2015 to May 2021

Search Terms

SOURCES SEARCHED: Medline, Embase and Cochrane Library

DATES OF SEARCH: January 2015 to 20 May 2021

SEARCH STRATEGIES

Medline search 1			Embase search 1		
1	exp Breast Neoplasms/	304880	1	exp breast cancer/	494900
2	Uterine Cervical Neoplasms/	77239	2	exp uterine cervix cancer/	103361
3	exp Colorectal Neoplasms/	209938	3	exp colon cancer/	28373
4	((breast or cervi* or colon* or rect* or colorect* or colo-rect* or bowel) adj3 (cancer* or neoplas* or carcinoma*	649108	4	((breast or cervi* or colon* or rect* or colorect* or colo-rect* or bowel) adj3 (cancer* or neoplas* or carcinoma*	927202

	or malignan* or tumor* or tumour* or lesion?)).ti,ab,kw.			or malignan* or tumor* or tumour* or lesion?)).ti,ab,kw.	
5	1 or 2 or 3 or 4	785064	5	1 or 2 or 3 or 4	1099422
6	Image Interpretation, Computer-Assisted/	46558	6	((whole slide imag* or wsi) and (patholog* or histopatholog* or cytopatholog* or cytodiagnos* or biops*)).ti,ab,kw.	1652
7	Pathology, Clinical/	5628	7	(digital* adj2 (patholog* or histopatholog* or cytopatholog* or cytodiagnos*)).ti,ab,kw.	2235
8	cytodiagnosis/ or exp biopsy/	301169	8	6 or 7	3291
9	7 or 8	306080	9	limit 8 to "reviews (maximizes specificity)"	12
10	6 and 9	1082	10	5 and 8	680
11	((whole slide imag* or wsi) and (patholog* or histopatholog* or cytopatholog* or cytodiagnos* or biops*)).ti,ab,kw.	994	11	9 or 10	692
12	(digital* adj2 (patholog* or histopatholog* or cytopatholog* or cytodiagnos*)).ti,ab,kw.	1264	12	limit 11 to (english language and yr="2015 -Current")	561
13	10 or 11 or 12	2855	13	(editorial or letter or note or conference*).pt. or case report.ti.	7836648
14	limit 13 to ("systematic review" or "reviews (maximizes specificity)")	19	14	12 not 13	329
15	5 and 13	609			
16	14 or 15	627			
17	limit 16 to (english language and yr="2015 -Current")	384			
18	(comment or editorial or letter).pt. or case report.ti.	2210064			
19	17 not 18	382			

Medline search 2			**Embase search 2**		
1	Observer Variation/	43463	1	diagnostic accuracy/	264720
2	"Reproducibility of Results"/	415876	2	reproducibility/ or observer variation/	241611
3	(validat* or accuracy or accurate or discordan* or concordan* or nonconcordan* or non-concordan*).ti,ab,kw.	1305270	3	(validat* or accuracy or accurate or discordan* or concordan* or nonconcordan* or non-concordan*).ti,ab,kw.	1774178
4	(observer or intraobserver or interobserver).ti,ab,kw.	54823	4	(observer or intraobserver or interobserver).ti,ab,kw.	72778
5	1 or 2 or 3 or 4	1613461	5	1 or 2 or 3 or 4	2100617
6	Image Interpretation, Computer-Assisted/	46558	6	((whole slide imag* or wsi) and (patholog* or histopatholog* or cytopatholog* or cytodiagnos* or biops*)).ti,ab,kw.	1652
7	Pathology, Clinical/	5628	7	(digital* adj2 (patholog* or histopatholog* or cytopatholog* or cytodiagnos*)).ti,ab,kw.	2235
8	cytodiagnosis/ or exp biopsy/	301169	8	6 or 7	3291

9	7 or 8	306080	9	limit 8 to "reviews (maximizes specificity)"	12
10	6 and 9	1082	10	5 and 8	1698
11	((whole slide imag* or wsi) and (patholog* or histopatholog* or cytopatholog* or cytodiagnos* or biops*)).ti,ab,kw.	994	11	9 or 10	1701
12	(digital* adj2 (patholog* or histopatholog* or cytopatholog* or cytodiagnos*)).ti,ab,kw.	1264	12	limit 11 to (english language and yr="2015 -Current")	1370
13	10 or 11 or 12	2885	13	(editorial or letter or note or conference*).pt. or case report.ti.	7836648
14	limit 13 to ("systematic review" or "reviews (maximizes specificity)")	19	14	12 not 13	733
15	5 and 13	1410			
16	14 or 15	1419			
17	limit 16 to (english language and yr="2015 -Current")	936			
18	(comment or editorial or letter).pt. or case report.ti.	2210064			
19	17 not 18	929			

Cochrane search	
#6	MeSH descriptor: [Image Interpretation, Computer-Assisted] explode all trees
#7	MeSH descriptor: [Pathology, Clinical] explode all trees
#8	MeSH descriptor: [Cytodiagnosis] this term only
#9	MeSH descriptor: [Biopsy] explode all trees
#10	#7 or #8 or #9
#11	#6 and #10
#12	((("whole slide imag*" or wsi) and (patholog* or histopatholog* or cytopatholog* or cytodiagnos* or biops*))):ti,ab,kw OR ((digital* NEAR/2 (patholog* or histopatholog* or cytopatholog* or cytodiagnos*))):ti,ab,kw
#13	#11 or #12

Results by database

Medline	**1,311**
Embase	**1,062**
Cochrane Library	**91**
PubMed related articles	**12**
Total	**2,476**

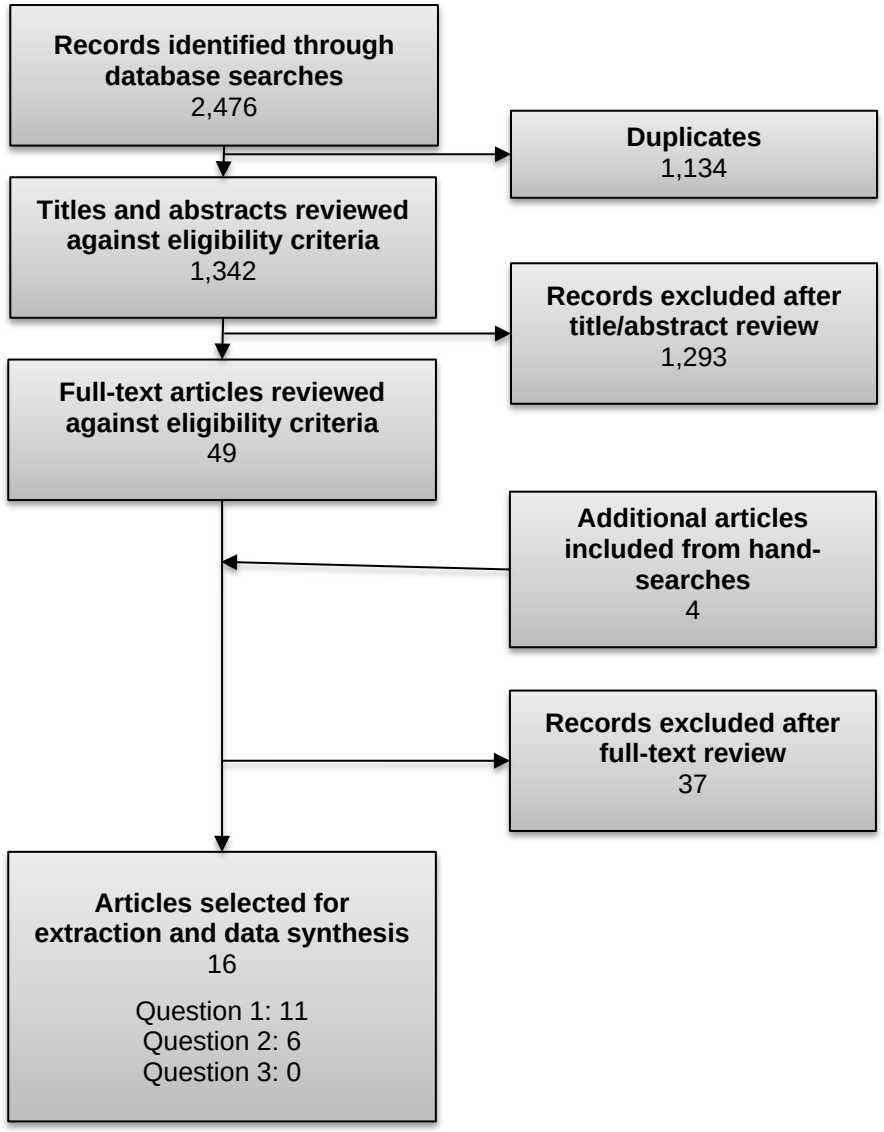
After the exclusion of duplicates, 1,342 references remained.

Appendix 2 — Included and excluded studies

PRISMA flowchart

Figure 1 summarises the volume of publications included and excluded at each stage of the review. Fifty-three publications were ultimately judged to be relevant to one or more review questions and were considered for extraction. Publications that were included or excluded after the review of full-text articles are detailed below.

Figure 1. Summary of publications included and excluded at each stage of the review



Williams *et al.* (2020)[7] was included in both question 1 and question 2.

Publications included after review of full-text articles

The 16 publications included after review of full texts are summarised in Table 7 below.

Studies were prioritised for extraction and data synthesis. It was planned *a priori* that the following approach would be taken to prioritise studies for extraction:

1. Systematic reviews and meta-analyses would be considered the highest quality of evidence if any were found
2. Studies from the UK screening programmes would be prioritised but, in the absence of such studies, those from other countries could be reported
3. For question 1, studies in randomly assigned or consecutively enrolled populations would be prioritised. Case-control studies could be considered if no other types of studies were available
4. For question 1, only studies separately reporting outcomes for breast or cervical cancer would be included
5. For questions 2 and 3, studies relating to the use of digital pathology in breast and cervical screening would be prioritised. However, in the absence of such studies or if the volume of evidence was low, studies looking at the use of digital pathology more broadly could also be reported

In addition, the following criteria for this evidence summary were applied after completion of the evidence map assessing the overall volume and type of evidence available:

6. Prioritise for inclusion, extraction and quality assessment all studies in which a pathologist's interpretation of whole slide images is compared to a pathologist's interpretation of glass slides for primary diagnosis of disease
7. Prioritise for inclusion, extraction and quality assessment all studies comparing a pathologist interpretation of whole slide images with a pathologist interpretation of glass slides for grading of cancer
8. Deprioritise studies that evaluate some form of automated image analysis of whole slide images rather than a pathologist's interpretation of whole slide images
9. Deprioritise cervical cancer studies based on liquid-based cytology

Publications not selected for extraction and data synthesis are clearly detailed in Table 8 below.

Table 7. Summary of publications included after review of full-text articles, and the question(s) each publication was identified as being relevant to

Study	The test	The screening programme	Comments
Al-Janabi *et al.* 2016[24]	X		Grading of breast cancer (non-UK study)
Babwale *et al.* 2021[19]	X		Primary diagnosis of breast cancer (UK study)

Borowsky *et al.* 2020[20]	X		Primary diagnosis of breast cancer (non-UK study)
Davidson *et al.* 2019[25]	X		Grading of breast cancer (non-UK study)
Elmore *et al.* 2017[21]	X		Primary diagnosis of breast cancer (non-UK study)
Mukhopadhyay *et al.* 2018[22]	X		Primary diagnosis of breast cancer (non-UK study)
Ordi *et al.* 2015[28]	X		Primary diagnosis of cervical cancer (non-UK study)
Rakha *et al.* 2018[26]	X		Grading of breast cancer (UK study)
Wilbur *et al.* 2015[27]	X		Grading of breast cancer (non-UK study)
Williams *et al.* 2018[23]	X		Primary diagnosis of breast cancer (UK study)
Williams *et al.* 2020[7]	X	X	Primary diagnosis of breast cancer and acceptability (UK study)
Browning *et al.* 2020[39]		X	Acceptability (UK study)
Dessauvagie *et al.* 2018[30]		X	Acceptability (UK study)
Turnquist *et al.* 2019[41]		X	Acceptability (UK study)
Williams *et al.* 2019[42]		X	Acceptability (UK study)
Williams *et al.* 2018[40]		X	Acceptability (UK study)

Publications excluded after review of full-text articles

Of the 53 publications included after the review of titles and abstracts, 37 were ultimately judged not to be relevant to this evidence summary. These publications, along with reasons for exclusion, are listed in Table 8.

Table 8. Publications excluded after review of full-text articles

	Reference	Reason for exclusion
1	Azam AS, Miligy IM, Kimani PK, Maqbool H, Hewitt K, Rajpoot NM, *et al.* Diagnostic concordance and discordance in digital pathology: a systematic review and meta-analysis. Journal of Clinical Pathology. 2020;15:15.	Study includes a mix of samples with no separate reporting for breast or cervical cancer outcomes
2	Baidoshvili A, Stathonikos N, Freling G, Bart J, t Hart N, van der Laak J, *et al.* Validation of a whole-slide image-based teleconsultation network. Histopathology. 2018;73(5):777-83.	Study includes a mix of samples with no separate reporting for breast or cervical cancer outcomes
3	Bauer TW, Slaw RJ, McKenney JK, Patil DT. Validation of whole slide imaging for frozen section diagnosis in surgical pathology. Journal of Pathology Informatics. 2015;6:49.	Study includes a mix of samples with no separate reporting for breast or cervical cancer outcomes
4	Bongaerts O, Clevers C, Debets M, Paffen D, Senden L, Rijks K, *et al.* Conventional microscopical versus digital whole-slide imaging-based diagnosis of thin-layer cervical specimens: A validation study. Journal of Pathology Informatics. 2018;9:29.	Study about liquid-based cytology
5	Bongaerts O, van Diest PJ, Pieters M, Nap M. Working toward consensus among professionals in the identification of classical cervical cytomorphological characteristics in whole slide images. Journal of Pathology Informatics. 2015;6:52.	Study about liquid-based cytology
6	Brunye TT, Mercan E, Weaver DL, Elmore JG. Accuracy is in the eyes of the pathologist: The visual interpretive process and diagnostic accuracy with digital whole slide images. Journal of Biomedical Informatics. 2017;66:171-9.	Study does not compare whole slide imaging and light microscopy
7	Cheng CL, Azhar R, Sng SH, Chua YQ, Hwang JS, Chin JP, *et al.* Enabling digital pathology in the diagnostic setting: navigating through the implementation journey in an academic medical centre. Journal of Clinical Pathology. 2016;69(9):784-92.	Study includes a mix of samples with no separate reporting for breast or cervical cancer outcomes
8	Cima L, Brunelli M, Parwani A, Girolami I, Ciangherotti A, Riva G, *et al.* Validation of remote digital frozen sections for cancer and transplant intraoperative services. Journal of Pathology Informatics. 2018;9:34.	Study includes a mix of samples with no separate reporting for breast or cervical cancer outcomes
9	Chow ZL, Thike AA, Li HH, Nasir NDM, Yeong JPS, Tan PH. Counting mitoses with digital pathology in breast phyllodes tumors. Archives of Pathology & Laboratory Medicine. 2020;09:09.	Study does not compare whole slide imaging and light microscopy

10	Dessauvagie BF, Lee AHS, Meehan K, Nijhawan A, Tan PH, Thomas J, *et al.* Interobserver variation in the diagnosis of fibroepithelial lesions of the breast: a multicentre audit by digital pathology. Journal of Clinical Pathology. 2018;71(8):672-9.	Study does not compare whole slide imaging and light microscopy
11	Elmore JG, Shucard H, Lee AC, Wang PC, Kerr KF, Carney PA, *et al.* Pathology trainees' experience and attitudes on use of digital whole slide images. Academic Pathology. 2020;7:2374289520951922.	Non-UK study on acceptability. UK studies available
12	Evans AJ, Brown RW, Bui MM, Chlipala EA, Lacchetti C, Milner DA, *et al.* Validating whole slide imaging systems for diagnostic purposes in pathology: Guideline update from the College of American Pathologists in collaboration with the American Society for Clinical Pathology and the Association for Pathology Informatics. Archives of Pathology & Laboratory Medicine. 2021;18:18	Guideline update. Mentioned in the evidence summary introduction
13	Ginter PS, Idress R, D'Alfonso TM, Fineberg S, Jaffer S, Sattar AK, *et al.* Histologic grading of breast carcinoma: a multi-institution study of interobserver variation using virtual microscopy. Modern Pathology. 2021;34(4):701-9	Study does not compare whole slide imaging and light microscopy
14	Girolami I, Pantanowitz L, Marletta S, Brunelli M, Mescoli C, Parisi A, *et al.* Diagnostic concordance between whole slide imaging and conventional light microscopy in cytopathology: A systematic review. Cancer Cytopathology. 2020;128(1):17-28.	Gynaecological studies included in this systematic review were about liquid-based cytology
15	Hanna MG, Monaco SE, Cuda J, Xing J, Ahmed I, Pantanowitz L. Comparison of glass slides and various digital-slide modalities for cytopathology screening and interpretation. Cancer Cytopathology. 2017;125(9):701-9.	Study about liquid-based cytology
16	Hanna MG, Reuter VE, Hameed MR, Tan LK, Chiang S, Sigel C, *et al.* Whole slide imaging equivalency and efficiency study: experience at a large academic center. Modern pathology. 2019.	Study includes a mix of samples with no separate reporting for breast or cervical cancer outcomes
17	Hanna MG, Reuter VE, Ardon O, Kim D, Sirintrapun SJ, Schüffler PJ et al. Validation of a digital pathology system including remote review during the COVID-19 pandemic. Modern Pathology. 2020, 33(11):2115-2127.	Non-UK study on acceptability. UK studies available
18	Hoda RS, Brogi E, D'Alfonso TM, Grabenstetter A, Giri D, Hanna MG, *et al.* Interobserver variation of PD-L1 SP142 immunohistochemistry interpretation in breast carcinoma: A study of 79 cases using whole slide imaging. Archives of Pathology & Laboratory Medicine. 2021;08:08.	Study does not compare whole slide imaging and light microscopy
19	Jhun I, Levy D, Lim H, Herrera Q, Dobo E, Burns D, *et al.* Implementation of collodion bag protocol to improve whole-slide imaging of scant gynecologic curettage specimens. Journal of Pathology Informatics. 2021;12:2.	Study does not compare whole slide imaging and light microscopy
20	Jones NC, Nazarian RM, Duncan LM, Kamionek M, Lauwers GY, Tambouret RH, *et al.* Interinstitutional whole slide imaging teleconsultation service development: assessment using internal training and clinical consultation cases. Archives of Pathology & Laboratory Medicine. 2015;139(5):627-35.	Study includes a mix of samples with no separate reporting for breast or cervical cancer outcomes
21	Lundstrom C, Waltersson M, Persson A, Treanor D. Summary of third Nordic symposium on digital pathology. Journal of Pathology Informatics. 2016;7:12.	Non-UK study on acceptability. UK studies available

22	McIntire PJ, Zhong E, Patel A, Khani F, D'Alfonso TM, Chen Z, *et al.* Hotspot enumeration of CD8+ tumor-infiltrating lymphocytes using digital image analysis in triple-negative breast cancer yields consistent results. Human Pathology. 2019;85:27-32.	Study does not compare whole slide imaging and light microscopy
23	Mercan E, Shapiro LG, Brunye TT, Weaver DL, Elmore JG. Characterizing diagnostic search patterns in digital breast pathology: Scanners and drillers. Journal of Digital Imaging. 2018;31(1):32-41.	Study does not compare whole slide imaging and light microscopy
24	Mills AM, Gradecki SE, Horton BJ, Blackwell R, Moskaluk CA, Mandell JW, *et al.* Diagnostic efficiency in digital pathology: A comparison of optical versus digital assessment in 510 surgical pathology cases. American Journal of Surgical Pathology. 2018;42(1):53-9.	Study includes a mix of samples with no separate reporting for breast or cervical cancer outcomes
25	Nagarkar DB, Mercan E, Weaver DL, Brunye TT, Carney PA, Rendi MH, *et al.* Region of interest identification and diagnostic agreement in breast pathology. Modern Pathology. 2016;29(9):1004-11.	Study does not compare whole slide imaging and light microscopy
26	Samuelson MI, Chen SJ, Boukhar SA, Schnieders EM, Walhof ML, Bellizzi AM, *et al.* Rapid validation of whole-slide imaging for primary histopathology diagnosis. American Journal of Clinical Pathology. 2021;155(5):638-48	Study includes a mix of samples with no separate reporting for breast or cervical cancer outcomes
27	Snead DRJ, Tsang YW, Meskiri A, Kimani PK, Crossman R, Rajpoot NM, *et al.* Validation of digital pathology imaging for primary histopathological diagnosis. Histopathology. 2016;68(7):1063-72.	Study includes a mix of samples with no separate reporting for breast or cervical cancer outcomes
28	Stathonikos N, Nguyen TQ, Spoto CP, Verdaasdonk MAM & van Diest PJ. Being fully digital: perspective of a Dutch academic pathology laboratory. Histopathology. 2019, 75(5):621-635	Non-UK study on acceptability. UK studies available
29	Tabata K, Mori I, Sasaki T, Itoh T, Shiraishi T, Yoshimi N, *et al.* Whole-slide imaging at primary pathological diagnosis: Validation of whole-slide imaging-based primary pathological diagnosis at twelve Japanese academic institutes. Pathology International. 2017;67(11):547-54.	Study includes a mix of samples with no separate reporting for breast or cervical cancer outcomes
30	Tawfik O, Davis M, Dillon S, Tawfik L, Diaz FJ, Amin K, *et al.* Whole-slide imaging of pap cellblock preparations is a potentially valid screening method. Acta Cytologica. 2015;59(2):187-200.	Study about liquid-based cytology
31	Unternaehrer J, Grobholz R, Janowczyk A, Zlobec I. Current opinion, status and future development of digital pathology in Switzerland. Journal of Clinical Pathology. 2020;73(6):341-6.	Non-UK study on acceptability. UK studies available
32	van Seijen M, Jozwiak K, Pinder SE, Hall A, Krishnamurthy S, Thomas JS, *et al.* Variability in grading of ductal carcinoma in situ among an international group of pathologists. The Journal of Pathology Clinical Research. 2021;7(3):233-42.	Study does not compare whole slide imaging and light microscopy
33	Villa I, Mathieu MC, Bosq J, Auperin A, Pomerol JF, Lacroix-Triki M, *et al.* Daily biopsy diagnosis in surgical pathology: Concordance between light microscopy and whole-slide imaging in real-life conditions. American Journal of Clinical Pathology. 2018;149(4):344-51.	Study includes a mix of samples with no separate reporting for

		breast or cervical cancer outcomes
34	Vodovnik A. Diagnostic time in digital pathology: A comparative study on 400 cases. Journal of Pathology Informatics. 2016;7:4.	Study includes a mix of samples with no separate reporting for breast or cervical cancer outcomes
35	Vodovnik A, Aghdam MRF. Complete routine remote digital pathology services. Journal of Pathology Informatics. 2018;9:36.	Study includes a mix of samples with no separate reporting for breast or cervical cancer outcomes
36	Wack K, Drogowski L, Treloar M, Evans A, Ho J, Parwani A, *et al.* A multisite validation of whole slide imaging for primary diagnosis using standardized data collection and analysis. Journal of Pathology Informatics. 2016;7:49	Study includes a mix of samples with no separate reporting for breast or cervical cancer outcomes
37	Williams BJ, DaCosta P, Goacher E, Treanor D. A systematic analysis of discordant diagnoses in digital pathology compared with light microscopy. Archives of Pathology & Laboratory Medicine. 2017;141(12):1712-8.	Study includes a mix of samples with no separate reporting for breast or cervical cancer outcomes

Appendix 3 — Summary and appraisal of individual studies

Data extraction and appraisal for quality and risk of bias

Question 1: What is the evidence on the accuracy of digital pathology (whole slide imaging) in breast and cervical cancer screening?

Breast cancer studies

Table 9: Al-Janabi *et al.* 2016[24]

Publication	Al-Janabi S, Horstman A, van Slooten HJ, Kuijpers C, Lai AFC, van Diest PJ, *et al.* Validity of whole slide images for scoring HER2 chromogenic in situ hybridisation in breast cancer. Journal of Clinical Pathology. 2016;69(11):992-7.
Study details	Validation study (grading)
Study objectives	To test the validity of whole slide imaging in assessing human epidermal growth factor receptor (HER2)[§§§] status in breast cancer specimens
Study setting	The Netherlands, single centre
Inclusions	HER2 chromogenic *in situ* hybridisation slides [****]
Exclusions	Areas with necrosis or overlapping nuclei
Population	96 selected cases from the archives of one centre in the Netherlands. Consisting of invasive ductal carcinoma (n=83), invasive lobular carcinoma (n=11) and mucinous carcinoma (n=2)
Test	Pathologist review of digital whole slide images (2 pathologists, one scoring 67 cases and one 29 cases)
Comparator / reference standard	Pathologist review of glass slides of the same cases on a light microscope using either scores retrieved from pathology reporting systems or from re-scoring after a minimum 6-week washout period (2 pathologists, one scoring 67 cases and one 29 cases)
Outcomes	**Concordance between digital and light microscopy (HER2 score)** • agreement: 73.1% (95%CI 63.9 to 82.3) • kappa score: 0.588 Confidence intervals not reported HER2 scores were underestimated in 59 cases using digital whole slide images and higher in 11 cases Samples with an average of <4 spots/nucleus were considered non-amplified (normal); samples with between 4 and 6 spots/nucleus were considered equivocal; samples with ≥6 spots/nucleus were considered amplified

[§§§] HER2 is used to assess which patients will be responsive to particular therapies
[****] Chromogenic *in situ* hybridisation is a morphological test that allows the evaluation of HER2 gene by assessing small nuclear signals within tumour cells using a glass slide and bright field microscopy

	Most of the discordant cases were in the equivocal category Test metrics for scoring the normal and amplified categories: • sensitivity: 95.45% • specificity: 100% • positive predictive value: 100% • negative predictive value: 97.14% Confidence intervals not reported In 3 cases, HER2 status could not be established **Time spent in scoring (n=30 cases):** • digital whole slide image: 81.7 seconds • light microscopy of glass slide: 86.9 seconds

Quality appraisal using Quality Assessment of Diagnostic Accuracy Studies (QUADAS-2) tool

Question	Assessment (Y, N, unclear)	Risk of Bias (low, high, unclear)	Supporting info
Domain I: Patient selection			
Consecutive or random sample of population enrolled?	N	High	Cases described as selected
Case-control design avoided?	N/A	N/A	Assessing grade in cancer cases
Inappropriate exclusions avoided?	Y	Low	
Domain II: Index Test			
Index test results interpreted without knowledge of reference standard results?	Y	Low	
Threshold pre-specified?	Y	Low	
Domain III: Reference standard			
Reference standard likely to correctly classify condition?	Y	Low	Archived or re-reviewed cases using light microscopy
Reference standard results interpreted without knowledge of index test results?	Y	Low	Archived or re-reviewed cases using light microscopy
Domain IV: Test strategy flow and timing			
Appropriate interval between index test and reference standard?	Y	Low	Archived or re-reviewed cases using light microscopy after ≥6-week washout period
Did all participants receive same reference standard?	Y	Low	
All patients included in analysis?	Y	Low	All cases on which a decision could be made included in the analysis
Domain V: Applicability			

Applicable to UK screening population of interest?	Unclear	Unclear	Data from breast cancer cases. Unclear if any cases were detected by screening
Applicable to UK screening test of interest?	Y	Low	
Target condition measured by reference test applicable to UK screening condition of interest?	Y	Low	
Other comments	Leica SCN400 or Philips scanning hardware was used with x40 magnification. Cases were viewed by the same pathologists. No statement was made about their experience of using digital pathology or any training received. No power calculation was reported relating to the number of cases included in the study. The number of slides available for each case was not specified. It is not clear if any of the cases were detected through screening. The study authors concluded that whole slide images scanned on a single focal plane are not suitable for assessing HER2 chromogenic in situ hybridisation. This was because of a tendency to underestimate the number of HER2 spots which lead to missing clinically relevant HER2 amplification.		

Table 10: Babawale *et al.* 2021[19]

Publication	Babawale M, Gunavardhan A, Walker J, Corfield T, Huey P, Savage A, *et al.* Verification and validation of digital pathology (whole slide imaging) for primary histopathological diagnosis: All Wales experience. Journal of Pathology Informatics. 2021;12:4.
Study details	Validation study
Study objectives	To validate and verify digital pathology for routine diagnostic histopathological services
Study setting	Wales, single centre
Inclusions	Samples representing all tissue types received in Glan Clwyd Hospital, Wales from April to December 2016
Exclusions	None stated
Population	47 breast cases from a study of 3,001 cases of different tissue types. No further details on type of breast cases reported Pathologists could request further information, including extra levels, special stains or immunohistochemistry
Test	Pathologist review of digital whole slide images following participation in a pilot study (pathologists from Glan Clwyd Hospital) or access to tests sets (60 cases) (pathologists from 7 Cellular Pathology departments across Wales) (number of pathologists for breast cases not clear)
Comparator / reference standard	Comparison with authorised reports on glass slides Discrepant cases were assessed by review of digital and glass slides by a third pathologist with the sub-specialty interest. More difficult cases were reviewed under multi-header microscope by a panel of expert pathologists
Outcomes	**Concordance** • breast cases: 95.7%

	Confidence intervals not reported for breast cases **Discordant cases** • breast cases: 4.3% **Clinically significant discordance** • breast cases: 0% Concordance was described as agreement in diagnosis between digital and glass slide reports. Discordant cases were divided into cases of no clinical significance and no impact on patient management or discordance of clinical significance. Examples of clinically significant discordant cases included discordance between benign and malignant, grading of dysplasia or missing findings that change patient management Other outcomes not reported separately for breast cases

Quality appraisal using Quality Assessment of Diagnostic Accuracy Studies (QUADAS-2) tool

Question	Assessment (Y, N, unclear)	Risk of Bias (low, high, unclear)	Supporting info
Domain I: Patient selection			
Consecutive or random sample of population enrolled?	Unclear	Unclear	Not clear if all samples received during the study period included
Case-control design avoided?	Y	Low	
Inappropriate exclusions avoided?	Y	Low	No exclusions stated
Domain II: Index Test			
Index test results interpreted without knowledge of reference standard results?	Y	Low	Whole slide images reviewed first
Threshold pre-specified?	Y	Low	
Domain III: Reference standard			
Reference standard likely to correctly classify condition?	Y	Low	Authorised reports on glass slides
Reference standard results interpreted without knowledge of index test results?	Y	Low	Authorised reports on glass slides
Domain IV: Test strategy flow and timing			
Appropriate interval between index test and reference standard?	Y	Low	Authorised reports on glass slides
Did all participants receive same reference standard?	Y	Low	Authorised reports on glass slides
All patients included in analysis?	Y	Low	
Domain V: Applicability			

Applicable to UK screening population of interest?	Unclear	Unclear	Data from breast biopsies. Unclear if any cases were detected by screening
Applicable to UK screening test of interest?	Y	Low	
Target condition measured by reference test applicable to UK screening condition of interest?	Y	Low	
Other comments	The Leica Aperio AT2 scanning hardware was used with x20 magnification. None of the pathologists had any previous experience in assessing digital whole slide images. However, all either took part in a pilot study or had access to test sets prior to taking part in this study. The power calculation reported by the study authors related to the whole case set for all tissue types. The breast tissue cases formed 1.6% of the overall study cases. It is not clear if any of the cases were detected through screening. The study authors concluded that there was no clinically significant discordance between digital whole slide images and glass slides for breast cases and that the scanner used produces adequate quality images for routine histopathologic diagnosis.		

Table 11: Borowsky *et al.* 2020[20]

Publication	Borowsky AD, Glassy EF, Wallace WD, Kallichanda NS, Behling CA, Miller DV, *et al.* Digital whole slide imaging compared with light microscopy for primary diagnosis in surgical pathology. Archives of Pathology & Laboratory Medicine. 2020;144(10):1245-53.
Study details	Double-blind randomised non-inferiority study
Study objectives	To compare pathologists' primary diagnoses derived from whole slide imaging versus the standard light microscope
Study setting	Wales, single centre
Inclusions	Consecutive routine care cases, enriched for more difficult diagnostic categories, from archived slides and recut sections of tissue blocks previously used for patient care at 5 US study sites
Exclusions	Cases where special stains or immunohistochemical studies were required for diagnosis, but were not available
Population	304 breast cases from a study of 2,045 consecutive routine care cases. Consisting of benign/ atypical core needle biopsy (n=57), benign/ atypical lumpectomy (n=44), in situ carcinoma core needle biopsy (n=51), in situ carcinoma lumpectomy (n=53), invasive carcinoma core needle biopsy (n=53) and invasive carcinoma lumpectomy (n=46) In some cases, all slides for a case were included. A representative set of slides could be selected for a case if these were considered sufficient for determination of the reference diagnosis Pathologists were randomised to the case order and assignment of digital slide or glass slide for first review

Test	Pathologist review of digital whole slide images with a minimum washout period of 31 days between modality reads (9 reviewing pathologists at 5 sites) (number of pathologists for breast cases not clear)
Comparator / reference standard	Pathologist review of glass slide by light microscopy with a minimum washout period of 31 days between modality reads (9 reviewing pathologists at 5 sites) (number of pathologists for breast cases not clear) The reference standard was the case's original diagnosis made using light microscopy. Comparisons between the pathologist review and original diagnosis made by separate adjudicating pathologists who were blinded to other adjudications and modality
Outcomes	**Major discrepancy rate for breast cases compared to reference standard** • digital whole slide images: 4.29% • glass slide by light microscopy: 3.53% • difference in major discrepancy rate: 0.76% Confidence intervals not reported A major discrepancy was defined as different diagnoses associated with different patient management **Unbalanced diagnoses** • rate of digital whole slide image discordance with glass slide by light microscopy concordance: 15/306[††††] (4.9%) • rate of digital whole slide image concordance with glass slide by light microscopy discordance: 13/302 (4.3%) An unbalanced diagnosis was defined as a case diagnosis in which one modality diagnosis was a major discrepancy and the other modality was concordant with the reference standard. Other outcomes not reported separately for breast cases

Quality appraisal using Quality Assessment of Diagnostic Accuracy Studies (QUADAS-2) tool

Question	Assessment (Y, N, unclear)	Risk of Bias (low, high, unclear)	Supporting info
Domain I: Patient selection			
Consecutive or random sample of population enrolled?	Y	Low	Consecutive sample of archived cases
Case-control design avoided?	Y	Low	
Inappropriate exclusions avoided?	Y	Low	Cases were excluded in the absence of information required for diagnosis
Domain II: Index Test			
Index test results interpreted without	Y	Low	Pathologists were blinded to the diagnosis made by

[††††] Figures as reported in the paper. The total number of cases was reported as 304. However, the denominator for this outcome was reported as 306

knowledge of reference standard results?			others or by their own interpretation of their first review by digital or glass slide
Threshold pre-specified?	Y	Low	
Domain III: Reference standard			
Reference standard likely to correctly classify condition?	Y	Low	Reference standard was the original diagnosis
Reference standard results interpreted without knowledge of index test results?	Y	Low	Reference standard was the original diagnosis
Domain IV: Test strategy flow and timing			
Appropriate interval between index test and reference standard?	Y	Low	Reference standard was the original diagnosis
Did all participants receive same reference standard?	Y	Low	Reference standard was the original diagnosis
All patients included in analysis?	Y	Low	
Domain V: Applicability			
Applicable to UK screening population of interest?	Unclear	Unclear	Data from routine care cases. Unclear if any cases were detected by screening
Applicable to UK screening test of interest?	Y	Low	
Target condition measured by reference test applicable to UK screening condition of interest?	Y	Low	
Other comments	The Aperio AT2 DX system scanning hardware was used with x20 magnification. Pathologists could request a x40 scan for digital slides. It is not clear how much experience or training the reading pathologists had in assessing digital whole slide images. All slides for a case were included unless a representative set of slides were considered sufficient for determination of the reference diagnosis. No power calculation was reported relating to the number of cases included in the study. It is not clear if any of the cases were detected through screening. The authors overall conclusion was that digital shole slide images are noninferior to glass slides for primary diagnosis in anatomic pathology.		

Table 12: Davidson *et al.* 2019[25]

Publication	Davidson TM, Rendi MH, Frederick PD, Onega T, Allison KH, Mercan E, *et al.* Breast cancer prognostic factors in the digital era: Comparison of Nottingham grade using whole slide images and glass slides. Journal of Pathology Informatics. 2019;10:11.

Study details	Subgroup analysis from a randomised prospective validation study (Elmore *et al.* 2017)
Study objectives	To assess reproducibility and accuracy of Nottingham grade scores for invasive breast cancer cases using digital whole slide images compared to glass slides
Study setting	US, registry data
Inclusions	Breast excisional and core biopsy specimens from pathology registries in 2 US states
Exclusions	None stated
Population	22 cases of invasive breast cancer divided into 4 biopsy test sets, each of 5 to 6 cases. A single slide best representing the reference diagnosis was selected for each case Pathologists from 8 US states were randomised to a biopsy set. They were then randomised twice to either digital whole slide images or glass slides
Test	Phase 1: Pathologists randomly assigned to review one of 4 test sets using a digital whole slide (digital) interpretive format (93 pathologists) Phase 2: Pathologists randomly assigned to review the same cases using a digital interpretive format after a wash-out period of at least 9 months (86 pathologists)
Comparator / reference standard	Phase 1: Pathologists randomly assigned to review one of 4 test sets using a glass slide light microscopy (glass) interpretive format (115 pathologists) Phase 2: Pathologists randomly assigned to review the same cases using a glass interpretive format after a wash-out period of at least 9 months (86 pathologists) Reference standard was a consensus diagnosis from 3 experienced breast pathologists

Outcomes	**Concordance in grading (Nottingham grade) for pathologists who interpreted the same cases in phase 1 and phase 2 (intraobserver)**
	• different format in each phase (one digital, one glass) (82 pathologists)
	  • agreement: 63% (95%CI 59 to 68)
	  • kappa score: 0.38 (95%CI 0.30 to 0.46)
	• digital whole slide image format in both phases (41 pathologists)
	  • agreement: 68% (95%CI 61 to 75)
	  • kappa score: 0.48 (95%CI 0.37 to 0.58)
	• glass slide format in both phases (49 pathologists)
	  • agreement: 73% (95%CI 68 to 78)
	  • kappa score: 0.57 (95%CI 0.48 to 0.66)
	
	Comparison of agreement in Nottingham grading for digital slide format in both phases (68%) vs glass slide format in both phases (73%): p=0.22
	
	Comparison of agreement in Nottingham grading when the format changed (63%) vs glass slides in both phases (73%): p=0.004
	
	Concordance in grading of components for pathologists who interpreted the same cases in phase 1 and phase 2 (intraobserver)
	• different format in each phase (one digital, one glass) (82 pathologists)

- agreement tubular score: 72% (95%CI 68 to 76)
- agreement nuclear pleomorphism score: 63% (95%CI 58 to 68)
- agreement mitotic score: 75% (95%CI 71 to 79)
- kappa tubular score: 0.50 (95%CI 0.43 to 0.57)
- kappa nuclear pleomorphism score: 0.32 (95%CI 0.24 to 0.40)
- kappa mitotic score: 0.40 (95%CI 0.32 to 0.48)

- digital whole slide image format in both phases (41 pathologists)
  - agreement tubular score: 72% (95%CI 65 to 78)
  - agreement nuclear pleomorphism score: 69% (95%CI 62 to 75)
  - agreement mitotic score: 72% (95%CI 65 to 79)
  - kappa tubular score: 0.48 (95%CI 0.38 to 0.59)
  - kappa nuclear pleomorphism score: 0.41 (95%CI 0.30 to 0.53)
  - kappa mitotic score: 0.37 (95%CI 0.26 to 0.49)
- glass slide format in both phases (49 pathologists)
  - agreement tubular score: 84% (95%CI 79 to 88)
  - agreement nuclear pleomorphism score: 68% (95%CI 62 to 73)
  - agreement mitotic score: 79% (95%CI 73 to 84)
  - kappa tubular score: 0.73 (95%CI 0.66 to 0.81)
  - kappa nuclear pleomorphism score: 0.44 (95%CI 0.34 to 0.54)
  - kappa mitotic score: 0.52 (95%CI 0.42 to 0.62)

Concordance in grading (Nottingham grade) between different pathologists interpreting the same cases, phase 1 data (interobserver)

- digital whole slide image format (93 pathologists)
  - agreement: 60% (95%CI 57 to 62)
  - kappa score: 0.32 (95%CI 0.31 to 0.34)
- glass slide format (115 pathologists)
  - agreement: 68% (95%CI 66 to 70)
  - kappa score: 0.48 (95%CI 0.47 to 0.49)

Comparison of kappa score in Nottingham grading for digital slide format in both phases (0.32) vs glass slide format in both phases (0.48): $p<0.001$

Concordance in grading (Nottingham grade) between different pathologists interpreting the same cases, phase 2 data (interobserver)

- digital whole slide image format (86 pathologists)
  - agreement: 62% (95%CI 60 to 64)
  - kappa statistic: 0.36 (95%CI 0.34 to 0.37)
- glass slide format (86 pathologists)
  - agreement: 69% (95%CI 67 to 71)
  - kappa statistic: 0.49 (95%CI 0.48 to 0.51)

Phase 2 results were stated to be consistent with phase 1 results

13/22 cases were assigned one of all 3 of the different Nottingham grade categories (low to high) by different pathologists using whole slide imaging

8/22 cases were assigned one of all 3 of the different Nottingham grade categories (low to high) by different pathologists using glass slides

Diagnostic concordance in grading of components between different pathologists interpreting the same cases, phase 1 data (interobserver)

- digital whole slide image format (93 pathologists)
  - agreement tubular score: 67% (95%CI 65 to 70)
  - agreement nuclear pleomorphism score: 58% (95%CI 56 to 61)
  - agreement mitotic score: 70% (95%CI 67 to 73)
  - kappa tubular score: 0.40 (95%CI 0.39 to 0.42)
  - kappa nuclear pleomorphism score: 0.22 (95%CI 0.20 to 0.23)
  - kappa mitotic score: 0.25 (95%CI 0.23 to 0.27)
- glass slide format (115 pathologists)
  - agreement tubular score: 71% (95%CI 69 to 73)
  - agreement nuclear pleomorphism score: 58% (95%CI 56 to 59)
  - agreement mitotic score: 74% (95%CI 72 to 77)
  - kappa tubular score: 0.51 (95%CI 0.50 to 0.52)
  - kappa nuclear pleomorphism score: 0.22 (95%CI 0.21 to 0.24)
  - kappa mitotic score: 0.42 (95%CI 0.40 to 0.43)

Diagnostic concordance in grading of components between different pathologists interpreting the same cases, phase 2 data (interobserver)

- digital whole slide image format (86 pathologists)
  - agreement tubular score: 65% (95%CI 62 to 67)
  - agreement nuclear pleomorphism score: 56% (95%CI 53 to 58)
  - agreement mitotic score: 68% (95%CI 65 to 70)
  - kappa tubular score: 0.37 (95%CI 0.35 to 0.38)
  - kappa nuclear pleomorphism score: 0.17 (95%CI 0.15 to 0.19)
  - kappa mitotic score: 0.23 (95%CI 0.21 to 0.25)
- glass slide format (86 pathologists)
  - agreement tubular score: 71% (95%CI 68 to 73)
  - agreement nuclear pleomorphism score: 56% (95%CI 54 to 59)
  - agreement mitotic score: 76% (95%CI 73 to 78)
  - kappa tubular score: 0.47 (95%CI 0.45 to 0.48)
  - kappa nuclear pleomorphism score: 0.25 (95%CI 0.23 to 0.27)
  - kappa mitotic score: 0.45 (95%CI 0.44 to 0.47)

Quality appraisal using Quality Assessment of Diagnostic Accuracy Studies (QUADAS-2) tool

Question	Assessment (Y, N, unclear)	Risk of Bias (low, high, unclear)	Supporting info
Domain I: Patient selection			

Consecutive or random sample of population enrolled?	N	High	Subgroup of invasive cancer cases
Case-control design avoided?	N/A	N/A	Assessing grade in subgroup of invasive cancer cases
Inappropriate exclusions avoided?	Y	Low	
Domain II: Index Test			
Index test results interpreted without knowledge of reference standard results?	Y	Low	
Threshold pre-specified?	Y	Low	
Domain III: Reference standard			
Reference standard likely to correctly classify condition?	Y	Low	Consensus diagnosis
Reference standard results interpreted without knowledge of index test results?	Y	Low	
Domain IV: Test strategy flow and timing			
Appropriate interval between index test and reference standard?	Y	Low	Washout period ≥9 months
Did all participants receive same reference standard?	Y	Low	Consensus diagnosis
All patients included in analysis?	Y	Low	
Domain V: Applicability			
Applicable to UK screening population of interest?	Unclear	Unclear	Data from breast biopsies. Unclear if any cases were detected by screening
Applicable to UK screening test of interest?	Y	Low	
Target condition measured by reference test applicable to UK screening condition of interest?	Y	Low	
Other comments	iScan Coreo Au® scanning hardware was used with x40 magnification. Pathologists were instructed to review the biopsy cases as they would in their routine clinical practice. Pathologists were not provided with written instructions or training sets. The authors stated that there was no intent to standardise diagnostic criteria. Between 45% and 56% of the pathologists participating in the different phases of the study had any experience of using digital whole slide images in their professional work. Results were reported for pathologists who interpreted the same cases in both phases of the study, however they were not informed that the		

	cases in phase 2 were the same re-ordered cases that they had already seen in phase 1. No power calculation was reported relating to the number of cases included in the study. A single representative slide was selected for each case. It was not stated whether any specimens were screen detected cases.

Table 13: Elmore *et al.* 2017[21]

Publication	Elmore JG, Longton GM, Pepe MS, Carney PA, Nelson HD, Allison KH, *et al.* A Randomized study comparing digital imaging to traditional glass slide microscopy for breast biopsy and cancer diagnosis. Journal of Pathology Informatics. 2017, 8:12.
Study details	Randomised prospective validation study
Study objectives	To evaluate the results of pathologists randomly assigned to interpret breast biopsy specimens in either traditional glass slide or digital whole slide imaging format
Study setting	US, registry data
Inclusions	Breast biopsy cases randomly selected from pathology registries with oversampling for cases with atypia and ductal carcinoma in situ and from women aged 40 to 49 years and with dense breasts. Cases were from core needle (n=138) and excisional (n=102) biopsies and consisted of 30% benign without atypia, 30% atypia, 30% ductal carcinoma *in situ* and 10% invasive carcinoma
Exclusions	None stated
Population	240 breast biopsy cases, divided into 4 test sets of 60 cases, interpreted by pathologists from 8 US states who had completed residency training, had interpreted breast specimens for ≥1 year, and intended to continue interpreting breast specimens for ≥1 year 208 pathologists from 8 US states were randomised to a biopsy set. They were then randomised twice to either digital whole slide images or glass slides 48% of the pathologists reported using the digital format in their professional work
Test	Phase 1: Pathologists randomly assigned to review one of 4 test sets using a digital whole slide (digital) interpretive format (93 pathologists) Phase 2: Pathologists randomly assigned to review the same cases using a digital interpretive format after a wash-out period of at least 9 months (86 pathologists). Pathologists were not aware that the phase 2 cases were the same as the phase 1 cases
Comparator / reference standard	Phase 1: Pathologists randomly assigned to review one of 4 test sets using a glass slide light microscopy (glass) interpretive format (115 pathologists) Phase 2: Pathologists randomly assigned to review the same cases using a glass interpretive format after a wash-out period of at least 9 months (86 pathologists). Pathologists were not aware that the phase 2 cases were the same as the phase 1 cases Reference standard determined by 3 experienced breast pathologists by consensus agreement for each case in glass format using standardised diagnostic categories

Outcomes	**Case agreement rates, comparing phase 1 digital and glass interpretation against the consensus reference standard**
	Reported by diagnostic category:
	• benign without atypia: digital 82% (95%CI 79 to 85) vs glass 87% (95%CI 85 to 89), p<0.01
	• atypia: digital 43% (95%CI 39 to 47) vs glass 48% (95%CI 44 to 52), p=0.08
	• ductal carcinoma *in situ*: digital 79% (95%CI 77 to 82) vs glass 84% (95%CI 82 to 86), p<0.01
	• invasive carcinoma: digital 93% (95%CI 90 to 95) vs glass 96% (95%CI 94 to 97), p=0.04
	
	Concordance in diagnosis for phase 1 and phase 2 interpretations by pathologists:
	172 pathologists completed both phases
	• intraobserver agreement with a different format in each phase (for example, one digital, one glass): 77% (95%CI 75 to 78) (82 pathologists)
	• intraobserver agreement with digital slide format in both phases: 73% (95%CI 71 to 76) (41 pathologists)
	• intraobserver agreement with glass slide format in both phases: 79% (95%CI 77 to 81) (49 pathologists)
	
	Comparison of agreement for digital slide format in both phases (73%) vs glass slide format in both phases (79%): p<0.001
	
	Comparison of agreement for different format in each phase (77%) vs glass slide format in both phases (79%): p=0.08
	
	Intraobserver agreement was high for invasive carcinoma regardless of format (between 93% and 97%)
	
	Intraobserver agreement was low for cases in categories such as atypia regardless of format (between 56% and 62%)
	
	Predictive values of initial digital or glass interpretation compared to confirmation by a reference panel (calculated by combining phase 1 data with the prevalence of diagnostic outcomes in US women aged 50-59 years who received breast biopsies after screening)
	Likelihood that initial interpretation is confirmed:
	• benign without atypia
	  • digital: 95.7% (95%CI 95.0 to 96.4)
	  • glass: 97.1% (95%CI 96.7 to 97.4)
	• atypia
	  • digital: 27.8% (95%CI 23.9 to 32.5)
	  • glass: 37.8% (95%CI 33.6 to 42.7)
	• ductal carcinoma *in situ*
	  • digital: 57.1% (95%CI 50.6 to 64.8)
	  • glass: 69.6% (95%CI 64.4 to 75.3)
	• invasive carcinoma
	  • digital: 97.2% (95%CI 95.6 to 98.6)
	  • glass: 97.7% (95%CI 96.5 to 98.7)

	The authors reported that the estimated predictive values were statically significantly lower for digital compared to glass for atypia (p=0.002) and ductal carcinoma in situ (p=0.007). No p value was reported for benign without atypia and invasive carcinoma **Confidence** • confidence in interpretive format: digital 78.6% vs glass 81.7%, p=0.22 • percentage of interpretations marked as borderline: digital 24.6% vs glass 26.1%, p=0.35 • images rated as challenging cases: digital 38.5% vs glass 30.0%, p=0.003 • pathologists desiring a second opinion: digital 42.5% vs glass 35.5%, p=0.03 **Interpretation time** Percentage of pathologists spending 20 hours participating in the study (the maximum allowed): digital 76% vs glass 51%, p = 0.01 No absolute values for time to diagnosis reported

Quality appraisal using Quality Assessment of Diagnostic Accuracy Studies (QUADAS-2) tool

Question	Assessment (Y, N, unclear)	Risk of Bias (low, high, unclear)	Supporting info
Domain I: Patient selection			
Consecutive or random sample of population enrolled?	Y	Low	Cases randomly selected
Case-control design avoided?	Y	Low	
Inappropriate exclusions avoided?	Y	Low	No exclusions stated
Domain II: Index Test			
Index test results interpreted without knowledge of reference standard results?	Y	Low	All cases viewed without knowledge of diagnosis
Threshold pre-specified?	Y	Low	
Domain III: Reference standard			
Reference standard likely to correctly classify condition?	Y	Low	Consensus agreement in glass format using standardised diagnostic categories
Reference standard results interpreted without knowledge of index test results?	Y	Low	
Domain IV: Test strategy flow and timing			
Appropriate interval between index test and reference standard?	Y	Low	Analysis of images by separate pathologists using registry cases

Did all participants receive same reference standard?	Y	Low	
All patients included in analysis?	Y	Low	
Domain V: Applicability			
Applicable to UK screening population of interest?	Unclear	Unclear	Data from breast biopsies. Unclear if any cases were detected by screening
Applicable to UK screening test of interest?	Y	Low	
Target condition measured by reference test applicable to UK screening condition of interest?	Y	Low	
Other comments	iScan Coreo Au® scanning hardware was used with x40 magnification. In phase 1 pathologists were randomised to a test set and interpretive format with stratification by clinical expertise. In phase 2, pathologists were randomised to interpretive format with stratification by phase 1 interpretative format and clinical expertise. It is not clear how many slides were provided for each case. It was not stated whether any specimens were screen detected cases. Pathologists interpreted the same cases in both phases of the study, however they were not informed that the cases in phase 2 were the same re-ordered cases that they had already seen in phase 1. It is not clear how much training the reading pathologists had in assessing digital whole slide images. Between 45% and 50% of the pathologists participating in the study had any experience of using digital whole slide images in their professional work. The authors stated that they had sufficient statistical power for case agreement rates from Phase 1 data. However, no details of a power calculation were reported. The study authors concluded that digital format interpretations were similar to glass slide interpretations for benign and invasive breast cancer cases. However, they also concluded that cases in the middle of the spectrum may be more problematic in digital format		

Table 14: Mukhopadhyay *et al.* 2018[22]

Publication	Mukhopadhyay S, Feldman MD, Abels E, Ashfaq R, Beltaifa S, Cacciabeve NG, *et al.* Whole slide imaging versus microscopy for primary diagnosis in surgical pathology: A multicenter blinded randomized noninferiority study of 1992 cases (pivotal study). American Journal of Surgical Pathology. 2018;42(1):39-52.
Study details	Blinded randomised non-inferiority validation study
Study objectives	To demonstrate that whole slide imaging is non-inferior to light microscopy for primary diagnosis in surgical pathology
Study setting	US, 4 centres

Inclusions	Consecutive surgical pathology cases from 4 US institutions (formalin-fixed paraffin-embedded biopsies and resections) including haematoxylin and eosin, immunohistochemistry and special stains. Cases reflected routine clinical practice with enrichment for more difficult malignant cases. The interval between accession of cases and inclusion in the study was ≥1 year
Exclusions	Slides for a case not available at the site; control slides for immunohistochemistry or special stains not available; slide selected did not match any subtype of the organ for which the case was selected; clinical information available to the sign-out pathologist in the pathology requisition form could not be obtained; selected slides contained indelible markings; more than one case selected for a patient; case consisted of frozen section slides only, or case consisted of gross specimens only
Population	299 breast cases from a study of 1,991 consecutive routine care surgical pathology cases with between one and 16 slides per case. Cases consisted of benign/ atypical core needle biopsy (n=50), benign/ atypical lumpectomy (n=50), in situ carcinoma core needle biopsy (n=49), in situ carcinoma lumpectomy (n=50), invasive carcinoma core needle biopsy (n=50) and invasive carcinoma lumpectomy (n=50) Pathologists were randomised to the assignment of digital whole slide image or glass slide for first review and the order of batches of 20 cases to review Each pathologist followed standard training, including self-familiarisation with the digital whole slide image viewer
Test	Pathologist review of digital whole slide images with a minimum washout period of 4 weeks between modality reads (16 pathologists at 4 sites) (4 pathologists viewed each case by both modalities)
Comparator / reference standard	Pathologist review of glass slide images using light microscopy with a minimum washout period of 4 weeks between modality reads (16 pathologists at 4 sites) (4 pathologists viewed each case by both modalities) The reference standard was the original diagnosis made during routine patient care
Outcomes	**Major discordance rate for breast cases compared to reference standard** • digital whole slide images: 4.2% • glass slide by light microscopy: 4.3% • difference in major discordance rate: 0.2%‡‡‡‡ Confidence intervals not reported Major discordance was defined as a difference in diagnosis associated with a difference in patient management. Other outcomes not reported separately for breast cases

Quality appraisal using Quality Assessment of Diagnostic Accuracy Studies (QUADAS-2) tool

Question	Assessment (Y, N, unclear)	Risk of Bias (low, high, unclear)	Supporting info
Domain I: Patient selection			

‡‡‡‡ Figures for major discordance rate and difference as stated in paper. Differences may be due to rounding

Consecutive or random sample of population enrolled?	Y	Low	Consecutive samples
Case-control design avoided?	Y	Low	
Inappropriate exclusions avoided?	Y	Low	Exclusions related to lack of available information
Domain II: Index Test			
Index test results interpreted without knowledge of reference standard results?	Y	Low	Reference standard was the original diagnosis
Threshold pre-specified?	Y	Low	
Domain III: Reference standard			
Reference standard likely to correctly classify condition?	Y	Low	Reference standard was the original diagnosis
Reference standard results interpreted without knowledge of index test results?	Y	Low	Reference standard was the original diagnosis
Domain IV: Test strategy flow and timing			
Appropriate interval between index test and reference standard?	Y	Low	Reference standard was the original diagnosis
Did all participants receive same reference standard?	Y	Low	Reference standard was the original diagnosis
All patients included in analysis?	Y	Low	
Domain V: Applicability			
Applicable to UK screening population of interest?	Unclear	Unclear	Data from routine care cases. Unclear if any cases were detected by screening
Applicable to UK screening test of interest?	Y	Low	
Target condition measured by reference test applicable to UK screening condition of interest?	Y	Low	
Other comments	The Philips IntelliSite Pathology Solutions system was used with x40 magnification. The pathologists followed standard training, including self-familiarisation with the digital whole slide image viewer. Breast cases represented 15% of the study sample. No power calculation was reported relating to the number of cases included in the study. Cases were routine care surgical pathology cases. Between one and 16 slides were available per case. Ten cases had 10 or more slides. It is not clear if any of the cases were detected through screening.		

	The study authors concluded that digital whole slide imaging is noninferior to light microscopy for primary diagnosis across a wide variety of organ systems.

Table 15: Rakha *et al.* 2018[26]

Publication	Rakha EA, Aleskandarani M, Toss MS, Green AR, Ball G, Ellis IO, *et al.* Breast cancer histologic grading using digital microscopy: Concordance and outcome association. Journal of Clinical Pathology. 2018, 71(8):680-6.
Study details	Validation study (grading)
Study objectives	To investigate the agreement between breast cancer grading using traditional light microscopy and digital whole slide imaging with consideration of reproducibility and impact on outcome prediction
Study setting	UK, one centre
Inclusions	Patients with early-stage invasive primary operable breast cancer who presented to Nottingham City Hospital between 1999 to 2006
Exclusions	None stated
Population	1,675 cases of early-stage invasive primary operable breast cancer. Consisting of ductal no special type (n=1,258), lobular (n=102), tubular/invasive cribriform (n=60), pure mucinous (n=22), invasive micropapillary (n=13) and other types including medullary-like (n=220). A single representative slide was selected for each case. It is not clear if the pathologist had training or experience in digital pathology
Test	Pathologist review of digital whole slide images on 2 separate occasions after a 3-month wash-out time using College of American Pathologists' criteria for cancer grading which are reported to be essentially the same as the original Nottingham criteria (one pathologist)
Comparator / reference standard	Data retrieved from patient notes on breast cancer grading using the Nottingham grading system during routine pathology reporting utilising all tumour glass slides by light microscope
Outcomes	**Concordance between digital and light microscopy grading:** • exact agreement of grade: 68% • kappa score: 0.51 (95%CI 0.47 to 0.54) • Cramner's V: 0.58 **Concordance between digital and light microscopy for high grade (grade 3) vs not high (grades 1 and 2):** • kappa statistic: 0.51 • Cramner V: 0.66 Confidence intervals not reported **Concordance between digital and light microscopy grading for components:** • exact agreement of tubule formation: 76.6% • exact agreement of pleomorphism: 60.1% • exact agreement of mitotic counts: 69.4% • kappa score tubules: 0.48 (95%CI 0.44 to 0.52) • kappa score pleomorphism: 0.27 (95%CI 0.24 to 0.31) • kappa score mitosis: 0.46 (95%CI 0.43 to 0.50) • Cramner's V tubules: 0.53 • Cramner's V pleomorphism: 0.41 • Cramner's V mitosis: 0.51 **Discordance**

- discordance between digital and light microscopy grade: 32.3%
- high vs low/intermediate discordance of grade: 17%
- grade assignments attributable to high vs low-grade discrepancy: 1.5%

Discordance was described as largely being between adjacent levels of grade

Intraobserver agreement for the 2 separate digital whole slide image readings

- kappa score: 0.65 (95%CI 0.60 to 0.68)
- Cramner V: 0.65

Quality appraisal using Quality Assessment of Diagnostic Accuracy Studies (QUADAS-2) tool

Question	Assessment (Y, N, unclear)	Risk of Bias (low, high, unclear)	Supporting info
Domain I: Patient selection			
Consecutive or random sample of population enrolled?	Y	Low	
Case-control design avoided?	N/A	N/A	Assessing grade in invasive cancer cases
Inappropriate exclusions avoided?	Y	Low	No exclusions stated
Domain II: Index Test			
Index test results interpreted without knowledge of reference standard results?	Y	Low	Whole slide images reviewed first and diagnosis recorded
Threshold pre-specified?	Y	Low	
Domain III: Reference standard			
Reference standard likely to correctly classify condition?	Y	Low	Archived light microscopy reports
Reference standard results interpreted without knowledge of index test results?	Y	Low	Archived light microscopy reports
Domain IV: Test strategy flow and timing			
Appropriate interval between index test and reference standard?	Y	Low	Archived light microscopy reports
Did all participants receive same reference standard?	Y	Low	Archived light microscopy reports
All patients included in analysis?	Y	Low	
Domain V: Applicability			
Applicable to UK screening population of interest?	Unclear	Unclear	Data from breast biopsies. Unclear if any cases were detected by screening

Applicable to UK screening test of interest?	Y	Low
Target condition measured by reference test applicable to UK screening condition of interest?	Y	Low
Other comments	The 3DHistech Panoramic 250 Flash II scanning hardware was used with x20 magnification. Slides were reviewed by a single pathologist. It is not clear if the pathologist had training or experience in digital pathology. A single representative slide was selected for each case. No power calculation was reported relating to the number of cases included in the study. It is not clear if any of the cases were detected through screening. The study authors concluded that digital whole slide imaging is a reliable and reproducible method for assessing breast cancer histological grade.	

Table 16: Wilbur *et al.* 2015[27] (study 1)

Publication	Wilbur DC, Brachtel EF, Gilbertson JR, Jones NC, Vallone JG, Krishnamurthy S. Whole slide imaging for human epidermal growth factor receptor 2 immunohistochemistry interpretation: Accuracy, precision, and reproducibility studies for digital manual and paired glass slide manual interpretation. Journal of Pathology Informatics. 2015;6:22.
Study details	2 validation studies using data from one tissue bank. This table includes details for study 1 (grading)
Study objectives	To assess whole slide imaging in interpreting human epidermal growth factor (HER2) immunohistochemistry in breast cancer specimens
Study setting	Denmark, tissue bank
Inclusions	Known breast cancer patients selected from a Danish tissue bank. Slide sets were constructed to include all score categories with an equal distribution of categories to reduce bias toward any particular result type. Pathologists were blinded to the slide set construction criteria
Exclusions	None stated
Population	195 breast cancer cases
Test	Pathologist review of digital whole slide images (3 pathologists)
Comparator / reference standard	Pathologist review of glass slides by light microscopy (3 pathologists) Glass slides were read before the digital whole slide images with a minimum washout period of 7 days (range 7 to 51). The order in which paired cases were read was randomised for each pathologist. Pathologists were blinded to any prior results
Outcomes	180 cases had scores for both formats and all 3 pathologists Dichotomous categories were created of negative results (HER2 score of 0 and 1+) and positive results (HER2 score of 2+ and 3+) **Agreement between digital whole slide image and light microscopy** • overall agreement: 87.2% (95%CI 84.1 to 89.8) • positive percentage agreement: 94.9% (95%CI 91.3 to 97.1) • negative percentage agreement: 81.3% (95%CI 76.6 to 85.3)

	Number of outliers • digital whole slide images: 15.2% • glass slides: 14.3% The binary agreement rates between individual pathologists ranged from 81% to 92% using different interpretation formats, from 86% to 92% using the same digital format and from 88% to 94% using the same light microscopy format. The study authors reported that there was no statistically significant difference in the agreement rates between the digital and light microscopy formats (p not stated) Results for comparisons within formats (for example digital vs digital or glass vs glass) not extracted

Quality appraisal using Quality Assessment of Diagnostic Accuracy Studies (QUADAS-2) tool

Question	Assessment (Y, N, unclear)	Risk of Bias (low, high, unclear)	Supporting info
Domain I: Patient selection			
Consecutive or random sample of population enrolled?	N	High	Cases described as selected
Case-control design avoided?	N/A	N/A	Assessing grade in cancer cases
Inappropriate exclusions avoided?	Y	Low	None stated
Domain II: Index Test			
Index test results interpreted without knowledge of reference standard results?	Y	Low	Pathologists were blinded to any prior results
Threshold pre-specified?	Y	Low	
Domain III: Reference standard			
Reference standard likely to correctly classify condition?	Y	Low	Light microscopy
Reference standard results interpreted without knowledge of index test results?	Y	Low	Pathologists were blinded to any prior results
Domain IV: Test strategy flow and timing			
Appropriate interval between index test and reference standard?	Unclear	Unclear	The minimum washout period was 7 days, but the range was 7 to 51 days
Did all participants receive same reference standard?	Y	Low	Light microscopy
All patients included in analysis?	Y	Low	
Domain V: Applicability			
Applicable to UK screening population of interest?	Unclear	Unclear	Data from breast cancer cases. Unclear if any cases were detected by screening

Applicable to UK screening test of interest?	Y	Low
Target condition measured by reference test applicable to UK screening condition of interest?	Y	Low
Other comments	Philips Digital Pathology Solution™ scanning hardware was used. Magnification not stated. All pathologists received training with standard sets. The data were from breast cancer cases. No power calculation was reported relating to the number of cases included in the study. The number of slides available for each case was not specified. It is not clear if any of the cases were detected through screening.	

Table 17: Wilbur *et al.* 2015[27] (study 2)

Publication	Wilbur DC, Brachtel EF, Gilbertson JR, Jones NC, Vallone JG, Krishnamurthy S. Whole slide imaging for human epidermal growth factor receptor 2 immunohistochemistry interpretation: Accuracy, precision, and reproducibility studies for digital manual and paired glass slide manual interpretation. Journal of Pathology Informatics. 2015;6:22.
Study details	2 validation studies using data from one tissue bank. This table includes details for study 2 (grading)
Study objectives	To assess whole slide imaging in interpreting human epidermal growth factor (HER2) immunohistochemistry in breast cancer specimens
Study setting	Denmark, tissue bank
Inclusions	Known breast cancer patients selected from a Danish tissue bank. Slide sets were constructed to include all score categories with an equal distribution of categories. Pathologists were blinded to the slide set construction criteria
Exclusions	None stated
Population	200 breast cancer cases (73 cases were also used in study 1, these were re-randomised and re-labeled for study 2)
Test	Pathologist review of digital whole slide images (3 pathologists)
Comparator / reference standard	Pathologist review of glass slides by light microscopy (3 pathologists) The order in which digital whole slide images and glass slides were read was randomised with a minimum washout period of 7 days (range 7 to 51). The order in which paired cases were read was randomised for each pathologist. Pathologists were blinded to any prior results
Outcomes	184 cases had scores for both formats and all 3 pathologists Dichotomous categories were created of negative results (HER2 score of 0 and 1+) and positive results (HER2 score of 2+ and 3+) **Agreement between digital whole slide image and light microscopy** • overall agreement: 88.8% (95%CI 85.7 to 91.7) • positive percentage agreement: 95.7% (95%CI 93.1 to 98.0) • negative percentage agreement: 82.8% (95%CI 77.9 to 87.3) **Number of outliers**

- digital whole slide images: 13.4%
- glass slides: 12.0%

The binary agreement rates between individual pathologists ranged from 83% to 92% using different interpretation formats, from 78% to 91% using the same digital format and from 86% to 91% using the same light microscopy format. No statement was made about the significance of any differences

Results for comparisons within formats (for example digital vs digital or glass vs glass) not extracted

Quality appraisal using Quality Assessment of Diagnostic Accuracy Studies (QUADAS-2) tool

Question	Assessment (Y, N, unclear)	Risk of Bias (low, high, unclear)	Supporting info
Domain I: Patient selection			
Consecutive or random sample of population enrolled?	N	High	Cases described as selected
Case-control design avoided?	N/A	N/A	Assessing grade in cancer cases
Inappropriate exclusions avoided?	Y	Low	None stated
Domain II: Index Test			
Index test results interpreted without knowledge of reference standard results?	Y	Low	Pathologists were blinded to any prior results
Threshold pre-specified?	Y	Low	
Domain III: Reference standard			
Reference standard likely to correctly classify condition?			
Reference standard results interpreted without knowledge of index test results?	Y	Low	Pathologists were blinded to any prior results
Domain IV: Test strategy flow and timing			
Appropriate interval between index test and reference standard?	Unclear	Unclear	The minimum washout period was 7 days, but the range was 7 to 51 days
Did all participants receive same reference standard?	Y	Low	Light microscopy
All patients included in analysis?	Y	Low	
Domain V: Applicability			
Applicable to UK screening population of interest?	Unclear	Unclear	Data from breast cancer cases. Unclear if any cases were detected by screening

Applicable to UK screening test of interest?	Y	Low
Target condition measured by reference test applicable to UK screening condition of interest?	Y	Low
Other comments	Philips Digital Pathology Solution™ scanning hardware was used. Magnification not stated. All pathologists received training with standard sets. The data were from breast cancer cases. No power calculation was reported relating to the number of cases included in the study. The number of slides available for each case was not specified. It is not clear if any of the cases were detected through screening.	

Table 18: Williams *et al.* 2018[23]

Publication	Williams BJ, Hanby A, Millican-Slater R, Nijhawan A, Verghese E, Treanor D. Digital pathology for the primary diagnosis of breast histopathological specimens: an innovative validation and concordance study on digital pathology validation and training. Histopathology. 2018, 72(4):662-71.
Study details	Validation study
Study objectives	To train and individually validate a group of breast pathologists in specialty-specific digital primary diagnosis by using a novel protocol endorsed by the Royal College of Pathologists' new guideline for digital pathology
Study setting	UK, one centre
Inclusions	All breast histopathology slides from each participant's breast pathology workload were prospectively scanned prior to laboratory send out from August 2016, including immunostains and special stains
Exclusions	None stated
Population	Total breast pathology workload of 3 Consultant Breast Histopathologists (694 complete cases) at St James University Hospital, Leeds Consisting of B1 normal tissue (n=85), B2 benign lesion (n=308), B3 lesion of uncertain malignant potential (n=51), B4 suspicious lesion (n=5), B5a malignant *in situ* (n=43), B5b malignant invasive (n=145), LB1 no lymphoid tissue (n=1), LB2 benign lymphoid tissue (n=22), LB5 malignant, metastatic carcinoma or other (n=5) and other (n=29)
Test	Pathologist review of digital whole slide images after completing training in the use of the digital microscopy system (3 pathologists)
Comparator / reference standard	Pathologist immediate glass slide review by light microscopy for reconciliation before final reporting (3 pathologists) All discordances were discussed at validation meetings with review of both digital and glass slides by participants and the validator
Outcomes	**Complete concordance:** • all combined data: 96.2% • pathologist 1: 95.0% • pathologist 2: 96.2% • pathologist 3: 97.4% Confidence intervals not reported **Any observable difference:** • all combined data: 3.8%

- pathologist 1: 5.0%
- pathologist 2: 3.8%
- pathologist 3: 2.6%

Confidence intervals not reported

Complete clinical concordance:

- all combined data: 98.8%
- pathologist 1: 99.3%
- pathologist 2: 99.1%
- pathologist 3: 98.5%

Confidence intervals not reported

Clinically significant discordances:

- all combined data: 1.2%
- pathologist 1: 0.7%
- pathologist 2: 0.9%
- pathologist 3: 1.5%

Mean (range) diagnostic confidence (0=not at all confident, 7=very confident):

Digital slides

- pathologist 1: 6.70 (4-7)
- pathologist 2: 6.90 (4-7)
- pathologist 3: 6.79 (0-7)

Glass slides

- pathologist 1: 6.80 (4-7)
- pathologist 2: 6.90 (4-7)
- pathologist 3: 6.99 (6-7)

No definitions were provided for 'complete concordance' or 'complete clinical concordance'. Clinically significant discordances concerned the mitotic count component of invasive tumour grading, identification of weddellite calcification, identification of isolated tumour cells, assessment of fibroepithelial lesion for cellularity and identification of focal epithelial atypia. The authors stated that the 2 most significant discordances both concerned the diagnosis of ductal carcinoma in situ.

Quality appraisal using Quality Assessment of Diagnostic Accuracy Studies (QUADAS-2) tool

Question	Assessment (Y, N, unclear)	Risk of Bias (low, high, unclear)	Supporting info
Domain I: Patient selection			
Consecutive or random sample of population enrolled?	Y	Low	All workload prospectively scanned
Case-control design avoided?	Y	Low	
Inappropriate exclusions avoided?	Y	Low	No exclusions stated
Domain II: Index Test			
Index test results interpreted without	Y	Low	All cases viewed digitally in the first instance

knowledge of reference standard results?			
Threshold pre-specified?	Y	Low	
Domain III: Reference standard			
Reference standard likely to correctly classify condition?	Y	Low	Glass slides reviewed
Reference standard results interpreted without knowledge of index test results?	N	High	Corresponding glass slides viewed immediately after whole slide images
Domain IV: Test strategy flow and timing			
Appropriate interval between index test and reference standard?	N	High	No washout period between viewing digital glass slides
Did all participants receive same reference standard?	Y	Low	
All patients included in analysis?	Unclear	Unclear	States a technical failure rate of 1.2%. Not clear if these cases were included in the analysis
Domain V: Applicability			
Applicable to UK screening population of interest?	Unclear	Unclear	Data from breast biopsies. Unclear if any cases were detected by screening
Applicable to UK screening test of interest?	Y	Low	
Target condition measured by reference test applicable to UK screening condition of interest?	Y	Low	
Other comments	The Leica Aperio AT2 and CS2 scanning hardware was used with x40 magnification for standard slides and x20 magnification for large slides. Pathologists completed training in the use of a digital microscopy system and were exposed to a training set of 20 challenging cases. Although the number of slides was not stated, all complete breast histopathology slides were scanned digitally prior to sending them out to laboratories. No power calculation was reported relating to the number of cases included in the study. It is not clear if any of the cases were detected through screening. The difference in the terminology used for the reporting of results was not defined. However, the figure for complete clinical concordance (98.2%) is the converse of the figure for clinically significant discordances (1.2%) and the figure for complete concordance (96.2%) is the converse of the figure for any observable difference (3.8%). The study authors concluded that individual training and validation allows pathologists to develop competence and confidence in their digital diagnostic skills.		

Table 19: Williams *et al.* 2020[7] (study 1)

Publication	Williams B, Hanby A, Millican-Slater R, Verghese E, Nijhawan A, Wilson I, *et al.* Digital pathology for primary diagnosis of screen-detected breast lesions - experimental data, validation and experience from four centres. Histopathology. 2020, 76(7):968-75.
Study details	3 validation studies using experimental data from 4 centres. This table includes details for study 1
Study objectives	To establish if digital slides are diagnostically equivalent to the glass slides they represent
Study setting	UK, one centre; Lithuania, one centre
Inclusions	Complete breast pathology cases, including immunohistochemistry and special stains where applicable selected from departmental archives
Exclusions	None stated
Population	475 complete breast pathology cases selected from departmental archives at University Hospitals Coventry (250 cases detected through the NHS Breast Cancer Screening Programme) and the Centre for Pathology, Vilnius, Lithuania (225 cases) No breakdown of the type of cases reported. Number of slides not specified
Test	Pathologist review of digital whole slide images (2 pathologists in Coventry, one pathologist in Vilnius)
Comparator / reference standard	Pathologist review of archived light microscopy reports (2 pathologists in Coventry, one pathologist in Vilnius) In cases of disagreement, both glass and digital slides were reviewed by an expert consensus panel
Outcomes	**Clinical concordance:** • all combined data: 98.7% • Coventry: 99.6% • Vilnius: 96.0% Confidence intervals not reported **Clinically significant discordances:** • all combined data: 10 • Coventry: 1 • Vilnius: 9 A clinically significant discordance was defined as any material difference in the diagnosis, regardless of whether or not this would have affected patient prognosis or treatment. The majority of discordances were differences in invasive tumour grading attributable to mitotic count-scoring.

Quality appraisal using Quality Assessment of Diagnostic Accuracy Studies (QUADAS-2) tool

Question	Assessment (Y, N, unclear)	Risk of Bias (low, high, unclear)	Supporting info
Domain I: Patient selection			
Consecutive or random sample of population enrolled?	Unclear	Unclear	States that cases were selected from archives. No further detail provided
Case-control design avoided?	Y	Low	

Inappropriate exclusions avoided?	Y	Low	No exclusions stated
Domain II: Index Test			
Index test results interpreted without knowledge of reference standard results?	Y	Low	Whole slide images reviewed first and diagnosis recorded
Threshold pre-specified?	Y	Low	
Domain III: Reference standard			
Reference standard likely to correctly classify condition?	Y	Low	Archived light microscopy reports
Reference standard results interpreted without knowledge of index test results?	Y	Low	Archived light microscopy reports
Domain IV: Test strategy flow and timing			
Appropriate interval between index test and reference standard?	Y	Low	Archived light microscopy reports
Did all participants receive same reference standard?	Y	Low	
All patients included in analysis?	Y	Low	All selected cases included in analysis
Domain V: Applicability			
Applicable to UK screening population of interest?	Y	Low	UK data from NHS Breast Screening programme
Applicable to UK screening test of interest?	Y	Low	
Target condition measured by reference test applicable to UK screening condition of interest?	Y	Low	
Other comments	In Coventry, Omnyx scanning hardware was used with x40 magnification. In Vilnius, Leica Aperio Scanscope was used with x20 magnification. It is not clear if the pathologists received any training in digital pathology prior to participation. The UK cases were from the NHS Breast Screening Programme. The detection route for the cases from Vilnius was not reported. It is not clear if the selection of cases was consecutive and complete. Although the number of slides was not stated, the cases were described as complete and the authors stated that pathologists viewed cases as recorded their diagnosis as they would in their routine practice. No power calculation was reported relating to the number of cases included in the study.		

	The authors concluded that digital pathology is safe for the primary diagnosis of NHS Breast Screening Programme breast histology specimens and does not increase the risk of misclassification.

Table 20: Williams *et al.* 2020[7] (study 2)

Publication	Williams B, Hanby A, Millican-Slater R, Verghese E, Nijhawan A, Wilson I, *et al.* Digital pathology for primary diagnosis of screen-detected breast lesions - experimental data, validation and experience from four centres. Histopathology. 2020, 76(7):968-75.
Study details	3 validation studies using experimental data from 4 centres. This table includes details for study 2
Study objectives	To train and validate individual pathologists for the primary digital diagnosis of breast pathology using a direct comparison method endorsed by the Royal College of Pathologists and evaluate clinical concordance rates throughout the validation process
Study setting	UK, 2 centres
Inclusions	'Live' breast histopathology work of all participating pathologists, scanned prospectively
Exclusions	None stated
Population	1,077 live complete breast histopathology cases detected through the NHS Breast Cancer Screening Programme from the case load of 5 Consultant Breast Histopathologists at Leeds Teaching Hospitals NHS Trust (896 cases) and United Lincolnshire Hospitals NHS Trust (181 cases) No breakdown of the type of cases reported. Number of slides not specified
Test	Pathologist review of digital whole slide images after completing training in the use of the digital microscopy system (4 pathologists in Leeds, one pathologist in Lincolnshire)
Comparator / reference standard	Pathologist immediate glass slide review by light microscopy for reconciliation before final reporting (4 pathologists in Leeds, one pathologist in Lincolnshire)
Outcomes	**Clinical concordance rate:** • all combined data: 99.1% • Leeds: 99.0% • Lincolnshire: 99.4% Confidence intervals not reported **Clinically significant discordances:** • all combined data: 10 • Leeds: 9 • Lincolnshire: 1 A clinically significant discordance was defined as any material difference in the diagnosis, regardless of whether or not this would have affected patient prognosis or treatment. The majority of discordances were differences in invasive tumour grading attributable to differences in mitotic count-scoring and the detection of small diagnostic objects such as isolated tumour cells in a sentinel lymph node

Quality appraisal using Quality Assessment of Diagnostic Accuracy Studies (QUADAS-2) tool

Question	Assessment (Y, N, unclear)	Risk of Bias (low, high, unclear)	Supporting info
Domain I: Patient selection			

Consecutive or random sample of population enrolled?	Y	Low	All 'live' work prospectively scanned
Case-control design avoided?	Y	Low	
Inappropriate exclusions avoided?	Y	Low	No exclusions stated
Domain II: Index Test			
Index test results interpreted without knowledge of reference standard results?	Y	Low	All cases viewed digitally in the first instance
Threshold pre-specified?	Y	Low	
Domain III: Reference standard			
Reference standard likely to correctly classify condition?	Y	Low	Glass slides viewed
Reference standard results interpreted without knowledge of index test results?	N	High	Corresponding glass slides viewed immediately after whole slide images
Domain IV: Test strategy flow and timing			
Appropriate interval between index test and reference standard?	N	High	No washout period between viewing digital glass slides
Did all participants receive same reference standard?	Y	Low	
All patients included in analysis?	Y	Low	
Domain V: Applicability			
Applicable to UK screening population of interest?	Y	Low	Data from NHS Breast Screening programme
Applicable to UK screening test of interest?	Y	Low	
Target condition measured by reference test applicable to UK screening condition of interest?	Y	Low	
Other comments	In Lincolnshire, Omnyx VL 120 hardware was used with x40 magnification. In Leeds, Leica Aperio AT2 and CS2 scanning hardware was used with x40 magnification. Pathologists received training in digital pathology and had access to training sets of cases. The cases were from the NHS Breast Screening Programme. Although the number of slides was not stated, the cases were described as the 'live' breast histopathology work of all participating consultant. No power calculation was reported relating to the number of cases included in the study.		

	The authors concluded that digital pathology is safe for the primary diagnosis of NHS Breast Screening Programme breast histology specimens and does not increase the risk of misclassification.

Table 21: Williams *et al.* 2020[7] (study 3)

Publication	Williams B, Hanby A, Millican-Slater R, Verghese E, Nijhawan A, Wilson I, *et al.* Digital pathology for primary diagnosis of screen-detected breast lesions - experimental data, validation and experience from four centres. Histopathology. 2020, 76(7):968-75.
Study details	3 validation studies using experimental data from 4 centres. This table includes details for study 3
Study objectives	To focus on the ability to categorise borderline lesions on the ductal atypia spectrum
Study setting	UK, one centre
Inclusions	Anonymised breast biopsy screening specimens of diagnostically challenging B2, B3 and B5a cases, selected from the departmental archives
Exclusions	None stated
Population	50 diagnostically challenging breast biopsy specimens from the NHS Breast Cancer Screening Programme selected from the archive of the Department of Histopathology at St James' University Hospital, Leeds reviewed by 3 Consultant Breast Histopathologists. A single representative slide was selected for each case. Each pathologist viewed each case on 4 separate occasions
Test	Pathologist review of digital whole slide images on 2 separate occasions with a washout period of 2 weeks between slide reads of the same case (3 pathologists)
Comparator / reference standard	Pathologist glass slide review by light microscopy on 2 separate occasions with a washout period of 2 weeks between slide reads of the same case (3 pathologists)
Outcomes	**Intraobserver agreement:** • digital vs glass: 87% (kappa value 0.80 (95%CI 0.70 to 0.90)) • 2 digital reads (digital vs digital): 87% (kappa value 0.80 (95%CI 0.72 to 0.87)) • 2 glass reads (glass vs glass): 85% (kappa value 0.78 (95%CI 0.57 to 0.81)) Agreement for breast lesion classification (not further defined)

Quality appraisal using Quality Assessment of Diagnostic Accuracy Studies (QUADAS-2) tool

Question	Assessment (Y, N, unclear)	Risk of Bias (low, high, unclear)	Supporting info
Domain I: Patient selection			
Consecutive or random sample of population enrolled?	N	High	States that cases were selected from archives
Case-control design avoided?	Y	Low	
Inappropriate exclusions avoided?	Y	Low	No exclusions stated
Domain II: Index Test			
Index test results interpreted without knowledge of reference standard results?	Y	Low	Each slide interpreted as they would in normal clinical practice

Threshold pre-specified?	Y	Low	
Domain III: Reference standard			
Reference standard likely to correctly classify condition?	Y	Low	
Reference standard results interpreted without knowledge of index test results?	Y	Low	Each slide interpreted as they would in normal clinical practice
Domain IV: Test strategy flow and timing			
Appropriate interval between index test and reference standard?	Y	Low	2-week washout period between slide reads of the same case
Did all participants receive same reference standard?	Y	Low	
All patients included in analysis?	Y	Low	
Domain V: Applicability			
Applicable to UK screening population of interest?	Y	Low	Data from NHS Breast Screening programme
Applicable to UK screening test of interest?	Y	Low	
Target condition measured by reference test applicable to UK screening condition of interest?	Y	Low	
Other comments	Leica Aperio AT2 and CS2 scanning hardware was used with x40 magnification. It is not clear if the pathologists received any training in digital pathology prior to participation. Cases were from the NHS Breast Screening Programme. A single representative slide was selected for each case. No power calculation was reported relating to the number of cases included in the study. The authors concluded that digital pathology is safe for the primary diagnosis of NHS Breast Screening Programme breast histology specimens and does not increase the risk of misclassification.		

Cervical cancer studies

Table 22: Ordi *et al.* 2015[28]

Publication	Ordi J, Castillo P, Saco A, Del Pino M, Ordi O, Rodriguez-Carunchio L, *et al.* Validation of whole slide imaging in the primary diagnosis of gynaecological pathology in a University Hospital. Journal of Clinical Pathology. 2015, 68(1):33-9.
Study details	Validation study
Study objectives	To determine the accuracy of interpretation of whole slide imaging compared with conventional light microscopy in the diagnosis of routine gynaecological biopsies
Study setting	Spain, one centre
Inclusions	All gynaecological specimens consecutively received over a 2-month period (July to August 2013) at the Department of Pathology of the Hospital Clinic of Barcelona
Exclusions	None stated
Population	157 biopsies or excisions of the uterine cervix from patients referred to colposcopy because of an abnormal Pap smear from a study of 452 gynaecological samples. Consisting or normal/reactive, benign tumours, low-grade premalignant lesions, high-grade premalignant lesions and malignant tumours (number for lesion type not reported for the 157 cases of patients referred to colposcopy)
Test	Digital whole slide images reviewed by one pathologist who had previously had a 1-week training course on the use of whole slide imaging
Comparator / reference standard	Glass side review by light microscopy by one pathologist Cases with discrepant results were reviewed by both of the 2 study pathologists and revisions made using light microscopy
Outcomes	**Complete concordance between digital and light microscopy** • 86.6% (95%CI 80.3 to 91.5) • kappa score: 0.832 (95%CI 0.757 to 0.906) Major discrepancies: 5.1% Minor discrepancies: 8.3% Major discrepancies were defined as differences with clinical and/or prognostic implications for the patients. Minor discrepancies were defined as mild differences which would not have any clinical or prognostic implications The authors stated that discrepancies were mostly associated with different interpretations of difficult or borderline cases or with the presence of small lesions overlooked in the evaluation

Quality appraisal using Quality Assessment of Diagnostic Accuracy Studies (QUADAS-2) tool

Question	Assessment (Y, N, unclear)	Risk of Bias (low, high, unclear)	Supporting info
Domain I: Patient selection			
Consecutive or random sample of population enrolled?	Y	Low	All specimens received in a 2-month period
Case-control design avoided?	Y	Low	

Inappropriate exclusions avoided?	Y	Low	No exclusions stated
Domain II: Index Test			
Index test results interpreted without knowledge of reference standard results?	Y	Low	Index test and reference standard reviewed by different, blinded, pathologists
Threshold pre-specified?	Y	Low	
Domain III: Reference standard			
Reference standard likely to correctly classify condition?	Y	Low	Glass slide review
Reference standard results interpreted without knowledge of index test results?	Y	Low	Index test and reference standard reviewed by different, blinded, pathologists
Domain IV: Test strategy flow and timing			
Appropriate interval between index test and reference standard?	Y	Low	Index test and reference standard reviewed by different pathologists
Did all participants receive same reference standard?	Y	Low	
All patients included in analysis?	Y	Low	
Domain V: Applicability			
Applicable to UK screening population of interest?	Y	Low	Data from patients referred to colposcopy due to an abnormal Pap smear (Spanish data)
Applicable to UK screening test of interest?	Y	Low	
Target condition measured by reference test applicable to UK screening condition of interest?	Y	Low	
Other comments	The Ventana iScan HT scanning hardware was used with x200[§§§§] magnification. The cases were viewed by different pathologists using either whole slide images or glass slides. The design of the study means it is not possible to determine if the differences in agreement observed are due to the different viewing format or due to differences between the 2 pathologists. The pathologist reading the digital whole slide images had received training.		

[§§§§] A magnification of 200x is stated in the publication

	The power calculation reported by the study authors related to the whole case set for all gynaecological specimens. The cervical tissue cases formed 35% of the overall study cases. The number of slides read for each case was not clear.

Question 2: Is the use of digital pathology in breast and cervical cancer screening clinically, socially and ethically acceptable to health professionals and the public?

Table 23: Browning *et al.* 2020[39]

Publication	Browning L, Fryer E, Roskell D, White K, Colling R, Rittscher Jens *et al*. Role of digital pathology in diagnostic histopathology in the response to COVID-19: results from a survey of experience in a UK tertiary referral hospital. Journal of Clinical Pathology Published Online First: 02 July 2020.
Study details	Online survey
Study objectives	To gather opinions with regard to the use of digital pathology within the clinical setting and to assess how this may have been impacted by the COVID-19 pandemic
Study setting	UK, one centre
Inclusions	Pathologists within the Cellular Pathology Department at the Oxford University Hospitals NHS Foundation Trust
Exclusions	None stated
Population	18/34 pathologists completed the survey (53% response rate)
Intervention	Digital pathology system
Comparator	N/A
Outcomes	**Response to the implementation of digital pathology:** • agreed that digital pathology is a positive step for their speciality team: 14/18 (78%) • agreed that they would likely continue to report digitally beyond the COVID-19 pandemic: 16/18 (89%) No responders stated that they would not report digitally in the future as a result of their digital pathology experience to date **Impact of access to digital pathology during the COVID-19 pandemic** • agreed that digital pathology had facilitated maintenance of their diagnostic practice while remote working: 9/18 (50%) • agreed that digital pathology had eased workforce crises: 6/18 (33%) • agreed that digital pathology had reduced potential impact on turnaround times: 7/18 (39%) **Use of digital pathology** • prior to the COVID-19 pandemic: 9/18 (50%) • during the COVID-19 pandemic: 14/18 (78%) • to report all clinical cases: 3/18 (17%) • to report some clinical cases: 6/18 (33%) • for a quick review of a case to determine if levels/ immunohistochemistry/ special stains were needed: 15/18 (83%) • for second opinions (within the Trust): 12/18 (67%) • to demonstrate digital images in a multi-disciplinary team meeting: 10/18 (56%) • to prepare/ review a case prior to a multi-disciplinary team meeting: 14/18 (78%) The term 'reporting' was not specific and encompassed digital viewing of whole slide images within the diagnostic process. It was not limited to the authorisation of a case based on review of whole slide images **Challenges of using digital pathology (reported as free text comments)** • set-up for remote working, including internet speed and workstations • personal investment needed to upgrade internet access and equipment • occasional out of focus slides

	The study authors commented that none of the challenges appear to have impacted on the uptake of digital pathology

Quality appraisal using the CASP critical appraisal checklist for qualitative research	
Question	**Assessment (Y, N, unclear)**
Was there a clear statement of the aims of the research?	Y
Is a qualitative methodology appropriate?	Y
Was the research design appropriate to address the aims of the research?	Y
Was the recruitment strategy appropriate to the aims of the research?	unclear
Was the data collected in a way that addressed the research issue?	Y
Has the relationship between the research and participants been adequately considered?	unclear
Have ethical issues been taken into consideration?	unclear
Was the data analysis sufficiently rigorous?	Y
Is there a clear statement of findings?	Y
Other comments	The authors stated that the survey was circulated to 34 pathologists within the hospital. It is not clear if it was circulated to all pathologists within the hospital. Approximately half of the 34 pathologists responded to the survey. It is not clear if the views of the responders are representative of those who did not participate. No reasons for non-participation were stated. The relationship between the researchers and the participants was not clear and it is not clear if the survey was completed anonymously. It is not clear if this would have affected the responses provided. The analysis reported was limited to descriptive statements, however an indication was given of the percentage of participants who gave a particular response. The statements of findings were appropriate for the study's aim. The study authors concluded that in their institution, digital pathology has demonstrated current and future potential to increase resilience in diagnostic practice and has highlighted some of the challenges that need to be considered.

Table 24: Dessauvagie *et al.* 2018[30]

Publication	Dessauvagie BF, Lee AHS, Meehan K, Nijhawan A, Tan PH, Thomas J, *et al.* Interobserver variation in the diagnosis of fibroepithelial lesions of the breast: a multicentre audit by digital pathology. Journal of Clinical Pathology. 2018, 71(8):672-9.
Study details	Multicentre diagnostic audit and feedback survey
Study objectives	To conduct a multi-centre review of core needle biopsies and excisions of fibroepithelial lesions of the breast in order to assess the degree of interobserver variability. This included a survey of experience, comfort and confidence in using digital pathology
Study setting	Study coordinated from the UK and conducted in 3 unspecified countries, 4 centres
Inclusions	Pathologists who took part in the audit study
Exclusions	None stated
Population	8 pathologists from 4 tertiary pathology institutions in 3 unspecified countries. Pathologists ranged in experience from recently qualified pathologists (less than 10 years specialist experience) to specialised breast pathologists (more than 10 years specialist experience with a dominant practice in breast pathology)
Intervention	Digital whole slide images
Comparator	N/A

Outcomes	The survey results found that: • 7/8 pathologists (88%) reported feeling comfortable using the digital platform for the audit and were satisfied with the quality of the digital images • 4/8 (50%) reported use of digital pathology in routine diagnostic work • 8/8 (100%) reported use of digital pathology in some aspect of their clinical practice • 6/8 (75%) felt comfortable using digital pathology for diagnosis in routine practice, 1/8 (13%) was neutral and 1/8 (13%) was uncomfortable Limitations included slower visual scanning of slides at low power and worse resolution compared with conventional microscopy. These were worse with poor viewing screen quality. Participants also reported that mitoses were harder to find, more time consuming to identify and, where found, were difficult to quantify by microscope fields Results relating to the concordance between pathologists are not eligible for inclusion in this evidence summary as no comparison to light microscopy was reported
Quality appraisal	This study was not formally appraised with a checklist as the study was designed as an audit of interobserver variability which was outside the scope of this evidence summary. The outcomes extracted were from a feedback survey about working digitally for the participating pathologists. All 8 of the participating pathologists provided feedback and descriptive results were provided. Limitations of working digitally were reported, although it is not clear how many of the participating pathologists experienced the issues raised. The Aperio AT2 slide scanner was used with x40 magnification. The study authors did not draw any specific conclusions about the acceptability of digital pathology.

Table 25: Turnquist *et al.* 2019[41]

Publication	Turnquist C, Roberts-Gant S, Hemsworth H, White K, Browning L, Rees G, *et al*. On the edge of a digital pathology transformation: Views from a cellular pathology laboratory focus group. Journal of Pathology Informatics. 2019;10:37.
Study details	Focus group
Study objectives	To ascertain the benefits and challenges of transitioning to digital pathology from pathologists and biomedical scientists in a department about to transition from diagnostic reporting via traditional microscopy to digital pathology
Study setting	UK, one centre
Inclusions	Participants were selected by strategic sampling. The selected pathologists and biomedical scientists were those most involved in the digital pathology pilot
Exclusions	None stated
Population	Staff in a cellular pathology department in a large NHS teaching hospital in Oxford, UK. Participants in the focus group included a laboratory manager, a quality manager, an IT manager, 2 urological pathologists and one haematopathologist. Comments were separately sought from a breast pathologist who was not present at the focus group
Intervention	Focus group discussion using open questions
Comparator	N/A
Outcomes	The focus group was analysed using content analysis Benefits of transitioning to digital pathology included: • **collaboration:** improved due to an environment of sharing and openness and an increased referral rate. The process of validation could enhance collaboration and discussion with colleagues

- **training and teaching:** may benefit from access to an archive of digital slides and trainees may have access to a greater range of cases, rare cases and small samples that would not otherwise be available. May enhance student engagement
- **cost savings:** by reducing production of glass slide teaching sets, reducing purchase and maintenance of microscopes, reducing turn-around times and reducing error logs that require investigation
- **research:** may benefit as the reconstitution of old cohorts will not be necessary, there may be decreased requirement for glass slide storage and there may be a reduction in samples that are damaged or lost or need to be re-cut or re-stained. May also assist with artificial intelligence research by providing infrastructure to build cohorts
- **growth of speciality:** may attract more trainees. New technologies may bring in those with an engineering background with the possibility of building algorithms. The field may become more teamwork focused
- **improved multi-disciplinary team meetings:** may reduce the time it takes to show slides and allow them to be displayed in meetings, for example, showing inconclusive cases and what the uncertainty is or visually communicating details about surgical margins. May foster greater communication between clinicians and pathologists
- **patient-centred care:** patients may have access to their own images fostering greater patient involvement. Pathologists may interact more with patients, increase patient-centred care and enhance communication between doctor and patient

Barriers in the implementation of digital pathology included:

- **standardisation:** difficulties associated with standardisation across departments of NHS Trusts. For example, in variability of reporting on different microscopes, variations in haematoxylin and eosin staining and discrepancies in protocols
- **validation:** some concerns over less details in guidelines for validation in some areas, including determining when it might be appropriate to sign out a case on digital versus glass and what amount of validation is required in terms of stages 1 and 2. Also concerns about how to conduct robust validation and whether this should be self-reflective or externally administered, for example, by a departmental governance committee. The participants favoured an open culture to improve self-validation systems and avoid a ‘testing’ process
- **national implementation:** concerns about the transition from local pilots implementing digital pathology to a national programme in terms of the procedure and who will oversee it
- **storage and backups:** uncertainty about whether glass slides will serve as back-ups to use if the digital archive fails and the implications for equipment requirements and cost savings if microscopes are still required. Suggestions that most departments will have a hybrid of digital and glass slides, as some samples, such as micrometastasis, require viewing on glass slides
- **training:** concerns about variation in training due to differences in digital pathology implementation. Concerns about the standardisation of a trainee curriculum
- **technical:** concerns about logical implementation in terms of which platforms would be used, what is stored and for how long, who is overseeing the management of the archive and how it will be funded
- **cost-effectiveness:** emphasis on the need to establish measurements for cost-effectiveness
- **workload:** concerns that digital pathology may increase the demand for referrals and second opinions for pathologists at tertiary centres

	• **privacy/ legality:** concerns that legal issues such as privacy data and consent laws may interfere with the development of digital pathology and the use of artificial intelligence for algorithms. Need for guidance and risk assessment specific to digital pathology

Quality appraisal using the CASP critical appraisal checklist for qualitative research

Question	Assessment (Y, N, unclear)
Was there a clear statement of the aims of the research?	Y
Is a qualitative methodology appropriate?	Y
Was the research design appropriate to address the aims of the research?	Y
Was the recruitment strategy appropriate to the aims of the research?	unclear
Was the data collected in a way that addressed the research issue?	Y
Has the relationship between the research and participants been adequately considered?	unclear
Have ethical issues been taken into consideration?	unclear
Was the data analysis sufficiently rigorous?	N
Is there a clear statement of findings?	Y
Other comments	Seven pathologists and staff participated. It is not clear what proportion of such staff at the Trust this represents. The authors stated that the participants were selected as those most involved in the digital pathology pilot. It is not clear if the views of the participants are representative of those who were not invited to participate. The relationship between the researchers and the participants was not clear. Content analysis was performed with descriptive themes. It was not clear how many of the participants held the views or concerns expressed. The statements of findings were appropriate for the study's aim. The study authors concluded that many benefits of digital pathology were identified, but that key barriers need to be addressed for digital pathology to be fully implemented on a trust and national level.

Table 26: Williams *et al.* 2020[7]

Publication	Williams B, Hanby A, Millican-Slater R, Verghese E, Nijhawan A, Wilson I, *et al.* Digital pathology for primary diagnosis of screen-detected breast lesions - experimental data, validation and experience from four centres. Histopathology. 2020, 76(7):968-75.
Study details	3 validation studies using experimental data from 4 centres
Study objectives	To provide a comprehensive assessment of digital primary diagnosis of screen-detected breast lesions
Study setting	UK, 3 centres; Lithuania, one centre
Inclusions	Pathologists participating in a validation study
Exclusions	None stated
Population	Study 1: 2 pathologists from University Hospitals Coventry and one pathologist from the Centre for Pathology, Vilnius, Lithuania Study 2: 4 pathologists from Leeds Teaching Hospitals NHS Trust and one pathologist from United Lincolnshire Hospitals NHS Trust Study 3: 3 pathologists from Leeds Teaching Hospitals NHS Trust
Intervention	Pathologist review of digital whole slide images
Comparator	Pathologist review of glass slides by light microscopy (studies 2 and 3) or review of archived light microscopy reports (study 1)
Outcomes	Pathologists at the 4 clinical sites identified benefits of reporting their work digitally: • loss of glass slide transport and loss of transfer delays • rapid and convenient availability of images for sharing and second opinion • rapid access to previous biopsies for comparison with resection/repeat biopsies

	• perceived increased efficiency in the diagnosis of large volume biopsies/ multi-slide and multi-level cases • occupational health benefits, for example one pathologist would have been unable to complete her breast screening workload on the light microscope on one day due to a neck injury, but was able to complete her work digitally • enhanced opportunities to demonstrate pathology in multi-disciplinary team discussions and clinicopathological and departmental pathology review meetings • useful for teaching a larger cohort of trainees and facilitating the inclusion of trainees from distant sites • feasibility of applying artificial intelligence-based tools in the routine setting of breast pathology reporting
Quality appraisal	The studies reported in this paper were appraised using the QUADAS-2 checklist in relation to the diagnostic accuracy outcomes reported, reflecting the primary aims of these validation studies (see Appendix Tables 19 to 21). Feedback relating to working digitally was reported in the discussion section of the paper. This was reported to be from pathologists at all 4 clinical sites, although it is not clear how many pathologists contributed feedback, held a particular view or how this information was collected. No statements were reported about any potential disadvantages of digital working, or whether this aspect was explored. The cases used in this study were mostly taken from the NHS Breast Cancer Screening Programme. The study authors did not draw any specific conclusions about the acceptability of digital pathology.

Table 27: Williams *et al.* 2019[42]

Publication	Williams BJ, Jayewardene D, Treanor D. Digital immunohistochemistry implementation, training and validation: experience and technical notes from a large clinical laboratory. Journal of Clinical Pathology. 2019, 72(5):373-8.
Study details	Validation study and lessons learned from the experience of implementing digital pathology
Study objectives	To consider the value proposition of digitisation of clinical immunohistochemistry services and to develop an approach to digital immunohistochemistry implementation and validation in a large clinical laboratory
Study setting	UK, one centre
Inclusions	Pathologists from Leeds Teaching Hospitals NHS Trust
Exclusions	None stated
Population	24 Consultant Pathologists from Leeds Teaching Hospitals NHS Trust who had completed digital immunohistochemistry training
Intervention	Pathologist review of digital whole slide images
Comparator	Pathologist review of glass side by light microscopy
Outcomes	**Mean satisfaction scores with digital immunohistochemistry slides** (1=not at all satisfied, 7 = very satisfied): 5.91 (range 2-7) No satisfaction ratings were reported for glass slides **Mean confidence scores (1-7 scale)** • digital immunohistochemistry slide assessment: 6.1 (range 2-7) • glass slide assessment: 6.9 (range 6-7) No statistical comparison for digital and glass slide assessment reported

	Digital cases scoring low for confidence contained particular immunohistochemistry stains which were reported as being difficult to assess digitally. The study authors reported that scanning selected immunostained slides at x40 rather than x20 improved the ability of pathologists to make a confident diagnosis **Comments from pathologists with high satisfaction (score of 6 or 7):** • digital as quick and as easy as glass slide • easier to spot areas of concern at low power on digital slides than on glass • positive results spotted more quickly on digital slide • easier to assess a multi-slide case digitally • easy to use and interpret • quicker looking at digital images • digital immunohistochemistry seems more crisp **Comments from pathologists with low satisfaction or confidence (score of < 6):** • headache from screening large volumes of tissue for rare positive cells • took longer to scroll through all the tissue at high power digitally than on light microscope • need higher magnification scanning for some stains • *H pylori* blurry and difficult to spot Concordance outcomes from this study were not eligible for inclusion as they were not presented by speciality
Quality appraisal	This study was not formally appraised with a checklist as the study was designed as a validation study. The validation study aspect was outside the scope of this evidence summary. The outcomes extracted were from a feedback survey about working digitally for the participating pathologists. 24 pathologists took part in the study. However, it is not clear how many pathologists contributed feedback, held a particular view or how this information was collected. It is not clear if all pathologists in the hospital department took part in this study and how representative the views expressed were. Leica Aperio AT2 and CS2 slide scanners were used, primarily with x20 magnification. The study authors did not draw any specific conclusions about the acceptability of digital pathology.

Table 28: Williams *et al.* 2018[40]

Publication	Williams BJ, Lee J, Oien KA, Treanor D. Digital pathology access and usage in the UK: results from a national survey on behalf of the National Cancer Research Institute's CM-Path initiative. Journal of Clinical Pathology. 2018, 71(5):463-466.
Study details	National survey
Study objectives	To canvass the UK pathology community to ascertain current levels of digital pathology usage in clinical and academic histopathology departments, and prevalent attitudes to digital pathology
Study setting	UK, 41 centres
Inclusions	Pathology departments who were Cellular Molecular Pathology members. Academic and clinical pathology departments without a Cellular Molecular Pathology member
Exclusions	None stated
Population	41 NHS (n=34) and academic (n=6) pathology departments or institutes from across the UK responded to the survey from February to July 2017. Of the 34 NHS clinical departments, 10 were in district general hospitals and 24 tertiary referral centres

	Responses were sought at a departmental or institutional level. Department heads were asked to complete the survey themselves or forward it to the most relevant individual in their department
Intervention	Digital pathology
Comparator	N/A
Outcomes	**Access** Access to a digital pathology scanner: 23/39 (60.0%) • NHS-owned scanner: 8/23 (34.8%) • university-owned scanner: 10/23 (43.5%) • other ownership: 5/23 (21.7% Access to a digital pathology workstation: 24/40 (60.0%) Access to a digital slide archive/ library: 18/39 (46.2%) **Current usage** • currently produce digital slides: 14/34 (41.2%) (annual total slides range 50 to 30,000) • currently use digital slides for primary diagnosis: 31% (n not stated)[*****] • currently use digital slides for secondary diagnosis: 36% (n not stated) The authors stated that the most popular current applications of digital pathology were undergraduate and postgraduate teaching, research and quality assurance **Image analysis usage** Currently use image analysis on digital slides: 16/39 (41.0%) Examples of current use of image analysis included immunoscoring, tumour environment assessment, basic measurements, tumour cell proportions and tumour segmentation **Attitudes to digital pathology adoption and usage** • investigation and use of digital pathology a high or essential priority at their institution: 24/41 (58.5%) Digital pathology would improve (% agree or strongly agree): • safety: 37% • collaboration: 97% • turnaround time: 56% • staff time: 56% • cost: 20% • efficiency: 70% **Predicted usage** Prediction of digital slide use in one years' time (% always or often) • undergraduate teaching: 33% • postgraduate teaching: 56% • research tool (such as immunoscoring of tissues): 46% • subject of research: 43% • quality assurance: 26% • primary clinical diagnosis: 28% • secondary clinical diagnosis: 25%

[*****] Figures taken from the paper text. Percentage presented graphically differ from the percentages presented in the text

- multi-disciplinary team meetings: 27%

Barriers

Current barriers to wider digital pathology use at institution (% agree or strongly agree)

- safety concerns: 15%
- not perceived as useful by department: 33%
- apathy amongst staff: 13%
- lack of training: 31%
- time cost: 49%
- financial cost: 83%

Factors identified that could enable institutions to increase their digital pathology use included funding (96%) for initial hardware, software and staff outlay, training for pathologists (73%), guidance from the Royal College of Pathologists (78%) and further evidence regarding accuracy (51%). Other enabling factors listed as free text included: relevant UK data proving cost savings, PHE approval for screening specimens, clear and strong stance from NHS England, improved internet connections, algorithms which improve reporting standards, a change in attitude from managers, information technology infrastructure and personnel support

Quality appraisal using the CASP critical appraisal checklist for qualitative research

Question	**Assessment (Y, N, unclear)**
Was there a clear statement of the aims of the research?	Y
Is a qualitative methodology appropriate?	Y
Was the research design appropriate to address the aims of the research?	Y
Was the recruitment strategy appropriate to the aims of the research?	unclear
Was the data collected in a way that addressed the research issue?	Y
Has the relationship between the research and participants been adequately considered?	unclear
Have ethical issues been taken into consideration?	unclear
Was the data analysis sufficiently rigorous?	Y
Is there a clear statement of findings?	Y

Other comments

41 departments responded. No response rate was reported. An individual was asked to respond on behalf of a department. It is not clear how representative the views expressed are of the wider population of pathologists.

The survey was developed by members of the technology and workstream of the National Cancer Research Institute's Cellular Molecular Pathology initiative. The relationship between the researchers and the participants was not clear and it is not clear if the survey was completed anonymously. It is not clear if this would have affected the responses provided.

The analysis reported was limited to descriptive statements, however an indication was given of the percentage of participants who gave a particular response. The statements of findings were appropriate for the study's aim.

The study authors concluded that interest in digital pathology adoption is high in the UK with usage likely to increase in the coming years.

Question 3: Are there any health economic/ cost-effectiveness analyses and/or models on the use of digital pathology in breast and cervical cancer screening compared to the use of light microscopy? If so, what do they show?

No studies were identified on the cost-effectiveness of digital pathology in breast and cervical cancer screening, or on the cost-effectiveness of digital pathology more broadly, compared to the use of light microscopy.

Appendix 4 – UK NSC reporting checklist for evidence summaries

All items on the UK NSC Reporting Checklist for Evidence Summaries have been addressed in this report. A summary of the checklist, along with the page or pages where each item can be found in this report, is presented below.

Table 29. UK NSC reporting checklist for evidence summaries

	Section	Item	Page no.
1.	TITLE AND SUMMARIES		
1.1	Title sheet	Identify the review as a UK NSC evidence summary.	Title page
1.2	Plain English summary	Plain English description of the executive summary.	5
1.3	Executive summary	Structured overview of the whole report. To include: the purpose/aim of the review; background; previous recommendations; findings and gaps in the evidence; recommendations on the screening that can or cannot be made on the basis of the review.	6
2.	INTRODUCTION AND APPROACH		
2.1	Background and objectives	Background – Current policy context and rationale for the current review – for example, reference to details of previous reviews, basis for current recommendation, recommendations made, gaps identified, drivers for new reviews Objectives – What are the questions the current evidence summary intends to answer? – statement of the key questions for the current evidence summary, criteria they address, and number of studies included per question, description of the overall results of the literature search. Method – briefly outline the rapid review methods used.	11
2.2	Eligibility for inclusion in the review	State all criteria for inclusion and exclusion of studies to the review clearly (PICO, dates, language, study type, publication type, publication status etc.) To be decided *a priori*.	17
2.3	Appraisal for quality/risk of bias tool	Details of tool/checklist used to assess quality, e.g. QUADAS 2, CASP, SIGN, AMSTAR.	21
3.	SEARCH STRATEGY AND STUDY SELECTION (FOR EACH KEY QUESTION)		

3.1	Databases/ sources searched	Give details of all databases searched (including platform/interface and coverage dates) and date of final search.	21
3.2	Search strategy and results	Present the full search strategy for at least one database (usually a version of Medline), including limits and search filters if used. Provide details of the total number of (results from each database searched), number of duplicates removed, and the final number of unique records to consider for inclusion.	Appendix 1
3.3	Study selection	State the process for selecting studies – inclusion and exclusion criteria, number of studies screened by title/abstract and full text, number of reviewers, any cross checking carried out.	17
4.	STUDY LEVEL REPORTING OF RESULTS (FOR EACH KEY QUESTION)		
4.1	Study level reporting, results and risk of bias assessment	For each study, produce a table that includes the full citation and a summary of the data relevant to the question (for example, study size, PICO, follow-up period, outcomes reported, statistical analyses etc.). Provide a simple summary of key measures, effect estimates and confidence intervals for each study where available. For each study, present the results of any assessment of quality/risk of bias.	Appendix 3
5.	QUESTION LEVEL SYNTHESIS		
5.1	Description of the evidence	For each question, give numbers of studies screened, assessed for eligibility, and included in the review, with summary reasons for exclusion.	23, 44
5.2	Combining and presenting the findings	Provide a balanced discussion of the body of evidence which avoids over reliance on one study or set of studies. Consideration of four components should inform the reviewer's judgement on whether the criterion is 'met', 'not met' or 'uncertain': quantity; quality; applicability and consistency.	24, 45
5.3	Summary of findings	Provide a description of the evidence reviewed and included for each question, with reference to their eligibility for inclusion. Summarise the main findings including the quality/risk of bias issues for each question. Have the criteria addressed been 'met', 'not met' or 'uncertain'?	42, 51, 52
6.	REVIEW SUMMARY		
6.1	Conclusions and	Do findings indicate whether screening should be recommended?	53

	implications for policy	Is further work warranted? Are there gaps in the evidence highlighted by the review?	
6.2	Limitations	Discuss limitations of the available evidence and of the review methodology if relevant.	54

References

[1] Goacher E, Randell R, Williams B, Treanor D. The diagnostic concordance of whole slide Imaging and light microscopy: A systematic review. Archives of Pathology & Laboratory Medicine. 2017, 141(1):151-61.

[2] The Royal College of Pathologists. Digital Pathology. Available from https://www.rcpath.org/profession/digital-pathology.html (Accessed June 2021).

[3] Cross S, Furness P, Igali L, Snead D, Treanor D. Best practice recommendations for implementing digital pathology. The Royal College of Pathologists. January 2018. Available from https://www.rcpath.org/uploads/assets/f465d1b3-797b-4297-b7fedc00b4d77e51/Best-practice-recommendations-for-implementing-digital-pathology.pdf (Accessed June 2021).

[4] Treanor D, Williams B. The Leeds guide to digital pathology. July 2018. Available from http://www.virtualpathology.leeds.ac.uk/Research/clinical/docs/2018/pdfs/18778_Leeds%20Guide%20to%20Digital%20Pathology_Brochure_A4_final_hi.pdf (Accessed June 2021).

[5] Richards M. Diagnostics: Recovery and renewal, 2020. Available from https://www.england.nhs.uk/wp-content/uploads/2020/10/BM2025Pu-item-5-diagnostics-recovery-and-renewal.pdf (Accessed June 2021).

[6] NHS England. National Pathology Programme. Digital First: Clinical transformation through pathology innovation, 2014. Available from https://www.england.nhs.uk/wp-content/uploads/2014/02/pathol-dig-first.pdf (Accessed June 2021).

[7] Williams B, Hanby A, Millican-Slater R, Verghese E, Nijhawan A, Wilson I, *et al*. Digital pathology for primary diagnosis of screen-detected breast lesions - experimental data, validation and experience from four centres. Histopathology. 2020, 76(7):968-75.

[8] Cancer screening: Call for evidence. Response from the Royal College of Pathologists, 2019. Available from https://www.rcpath.org/uploads/assets/ac167630-1814-4ac5-95559d001cfd8595/RCPath-response-to-Cancer-Screening-Call-for-Evidence.pdf (June 2021).

[9] The Royal College of Pathologists. Letter from the Royal College of Pathologists to Public Health England. 2020.

[10] Baxter G, Jones V, Milnes V, Oduko J, Philips V, Sellars S, Vegnuti Z. Routine quality control tests for full-field digital mammography systems: Equipment report 1303: fourth edition. Public Health England, October 2013. Available from Breast screening: routine quality control tests for FFDM - GOV.UK (www.gov.uk) (Accessed August 2021).

[11] National Institute for Health Research. Research Award: Multi-centred validation of digital whole slide imaging for routine diagnosis. Available from https://fundingawards.nihr.ac.uk/award/17/84/07 (Accessed June 2021).

[12] Pantanowitz L, Sinard JH, Henricks WH, Fatheree LA, Carter AB, Contis L *et al*. Validating whole slide imaging for diagnostic purposes in pathology: Guideline from the College of American Pathologists Pathology and Laboratory Quality Center. Archives of Pathology & Laboratory Medicine. 2013, 137(12):1710-22.

[13] Evans AJ, Brown RW, Bui MM, Chlipala EA, Lacchetti C, Milner DA, *et al*. Validating whole slide imaging systems for diagnostic purposes in pathology: Guideline update from the College of American Pathologists in collaboration with the American Society for Clinical Pathology and the Association for Pathology Informatics. Archives of Pathology & Laboratory Medicine. 2021;18:18.

[14] Evans AJ, Salama ME, Henricks WH, Pantanowitz L. Implementation of whole slide imaging for clinical purposes: Issues to consider from the perspective of early adopters. Archives of Pathology & Laboratory Medicine. 2017, 141(7):944-59.

[15] The Royal College of Pathologists. Digital Pathology Strategy, 2019. Available from: https://www.rcpath.org/uploads/assets/2248bb71-b773-4693-945bffda593f2f2f/cf251e84-f7d0-415d-bb67217219203066/Digital-Pathology-Strategy.pdf (Accessed June 2021).

[16] Aslam M, Barrett P, Bryson G, Cross S, Snead D, Treanor D, Verrill C, Williams B. Guidance for remote reporting of digital pathology slides during periods of exceptional service pressure. The Digital Pathology Committee of the Royal College of Pathologists. March 2020. Available from https://www.rcpath.org/uploads/assets/626ead77-d7dd-42e1-949988e43dc84c97/RCPath-guidance-for-remote-digital-pathology.pdf (Accessed June 2021).

[17] Williams BJ, Brettle D, Aslam M, Barrett P, Bryson G, Cross S, Snead D, Verrill C, Clarke E, Wright A, Treanor D. Guidance for remote reporting of digital pathology slides during periods of exceptional service pressure: an emergency response from the UK Royal College of Pathologists. Journal of Pathology Informatics 2020, 11:12.

[18] Williams BJ, Knowles C, Treanor D. Maintaining quality diagnosis with digital pathology: a practical guide to ISO 15189 accreditation. Journal of Clinical Pathology 2019, 72: 663-668.

[19] Babawale M, Gunavardhan A, Walker J, Corfield T, Huey P, Savage A, *et al.* Verification and validation of digital pathology (whole slide imaging) for primary histopathological diagnosis: All Wales experience. Journal of Pathology Informatics. 2021;12:4.

[20] Borowsky AD, Glassy EF, Wallace WD, Kallichanda NS, Behling CA, Miller DV, *et al*. Digital whole slide imaging compared with light microscopy for primary diagnosis in surgical pathology. Archives of Pathology & Laboratory Medicine. 2020, 144(10):1245-53.

[21] Elmore JG, Longton GM, Pepe MS, Carney PA, Nelson HD, Allison KH, *et al*. A randomized study comparing digital imaging to traditional glass slide microscopy for breast biopsy and cancer diagnosis. Journal of Pathology Informatics. 2017, 8:12.

[22] Mukhopadhyay S, Feldman MD, Abels E, Ashfaq R, Beltaifa S, Cacciabeve NG, *et al*. Whole slide imaging versus microscopy for primary diagnosis in surgical pathology: A multicenter blinded randomized noninferiority study of 1992 cases (pivotal study). American Journal of Surgical Pathology. 2018, 42(1):39-52.

[23] Williams BJ, Hanby A, Millican-Slater R, Nijhawan A, Verghese E, Treanor D. Digital pathology for the primary diagnosis of breast histopathological specimens: an innovative validation and concordance study on digital pathology validation and training. Histopathology. 2018, 72(4):662-71.

[24] Al-Janabi S, Horstman A, van Slooten HJ, Kuijpers C, Lai AFC, van Diest PJ, *et al*. Validity of whole slide images for scoring HER2 chromogenic in situ hybridisation in breast cancer. Journal of Clinical Pathology. 2016, 69(11):992-7.

[25] Davidson TM, Rendi MH, Frederick PD, Onega T, Allison KH, Mercan E, *et al*. Breast cancer prognostic factors in the digital era: Comparison of Nottingham Grade using whole slide images and glass slides. Journal of Pathology Informatics. 2019, 10:11.

[26] Rakha EA, Aleskandarani M, Toss MS, Green AR, Ball G, Ellis IO, *et al*. Breast cancer histologic grading using digital microscopy: Concordance and outcome association. Journal of Clinical Pathology. 2018, 71(8):680-6.

[27] Wilbur DC, Brachtel EF, Gilbertson JR, Jones NC, Vallone JG, Krishnamurthy S. Whole slide imaging for human epidermal growth factor receptor 2 immunohistochemistry interpretation: Accuracy, precision, and reproducibility studies for digital manual and paired glass slide manual interpretation. Journal of Pathology Informatics. 2015;6:22.

[28] Ordi J, Castillo P, Saco A, Del Pino M, Ordi O, Rodriguez-Carunchio L, *et al.* Validation of whole slide imaging in the primary diagnosis of gynaecological pathology in a University Hospital. Journal of Clinical Pathology. 2015, 68(1):33-9.

[29] McHugh ML. Interrater reliability: the kappa statistic. Biochem. Med. 2012. 22(3):276-282.

[30] Dessauvagie BF, Lee AHS, Meehan K, Nijhawan A, Tan PH, Thomas J, *et al*. Interobserver variation in the diagnosis of fibroepithelial lesions of the breast: a multicentre audit by digital pathology. Journal of Clinical Pathology. 2018, 71(8):672-9.

[31] van Seijen M, Jozwiak K, Pinder SE, Hall A, Krishnamurthy S, Thomas JS, *et al*. Variability in grading of ductal carcinoma in situ among an international group of pathologists. The Journal of Pathology Clinical Research. 2021, 7(3):233-42.

[32] Wu W, Li B, Mercan E, Mehta S, Bartlett J, Weaver DL, *et al*. MLCD: A unified software package for cancer diagnosis. JCO Clinical Cancer Informatics. 2020, 4:290-8.

[33] Mercan E, Mehta S, Bartlett J, Shapiro LG, Weaver DL, Elmore JG. Assessment of machine learning of breast pathology structures for automated differentiation of breast cancer and high-risk proliferative lesions. JAMA Network Open. 2019, 2(8):e198777.

[34] Saikia AR, Bora K, Mahanta LB, Das AK. Comparative assessment of CNN architectures for classification of breast FNAC images. Tissue & Cell. 2019, 57:8-14.

[35] Gecer B, Aksoy S, Mercan E, Shapiro LG, Weaver DL, Elmore JG. Detection and classification of cancer in whole slide breast histopathology images using deep convolutional networks. Pattern Recognition. 2018, 84:345-56.

[36] Yuan C, Yao Y, Cheng B, Cheng Y, Li Y, Li Y, *et al*. The application of deep learning based diagnostic system to cervical squamous intraepithelial lesions recognition in colposcopy images. Scientific Reports. 2020, 10(1):11639.

[37] Sornapudi S, Stanley RJ, Stoecker WV, Long R, Xue Z, Zuna R, *et al*. DeepCIN: attention-based cervical histology image classification with sequential feature modeling for pathologist-level accuracy. Journal of Pathology Informatics. 2020, 11:40.

[38] Huang P, Tan X, Chen C, Lv X, Li Y. AF-SENet: Classification of cancer in cervical tissue pathological images based on fusing deep convolution features. Sensors. 2020, 21(1):27.

[39] Browning L, Fryer E, Roskell D, White K, Colling R, Rittscher Jens *et al*. Role of digital pathology in diagnostic histopathology in the response to COVID-19: results from a survey of experience in a UK tertiary referral hospital. Journal of Clinical Pathology Published Online First: 02 July 2020.

[40] Williams BJ, Lee J, Oien KA, Treanor D. Digital pathology access and usage in the UK: results from a national survey on behalf of the National Cancer Research Institute's CM-Path initiative. Journal of Clinical Pathology. 2018, 71(5):463-466.

[41] Turnquist C, Roberts-Gant S, Hemsworth H, White K, Browning L, Rees G, *et al*. On the edge of a digital pathology transformation: Views from a cellular pathology laboratory focus group. Journal of Pathology Informatics. 2019;10:37.

[42] Williams BJ, Jayewardene D, Treanor D. Digital immunohistochemistry implementation, training and validation: experience and technical notes from a large clinical laboratory. Journal of Clinical Pathology. 2019, 72(5):373-8.

[43] Elmore JG, Shucard H, Lee AC, Wang PC, Kerr KF, Carney PA, *et al*. Pathology trainees' experience and attitudes on use of digital whole slide images. Academic Pathology. 2020, volume 7.

[44] Hanna MG, Reuter VE, Ardon O, Kim D, Sirintrapun SJ, Schüffler PJ *et al*. Validation of a digital pathology system including remote review during the COVID-19 pandemic. Modern Pathology. 2020, 33(11):2115-2127.

[45] Hanna MG, Reuter VE, Hameed MR, Tan LK, Chiang S, Sigel C, *et al*. Whole slide imaging equivalency and efficiency study: experience at a large academic center. Modern pathology. 2019, 32(7):916-928.

[46] Lundström C, Waltersson M, Treanor D *et al*. Summary of third Nordic symposium on digital pathology. Journal of Pathology Informatics. 2016, 7:12.

[47] Stathonikos N, Nguyen TQ, Spoto CP, Verdaasdonk MAM & van Diest PJ. Being fully digital: perspective of a Dutch academic pathology laboratory. Histopathology. 2019, 75(5):621-635.

[48] Unternaehrer J, Grobholz R, Janowczyk A, Zlobec I; Swiss Digital Pathology Consortium (SDiPath). Current opinion, status and future development of digital pathology in Switzerland. Journal of Clinical Pathology. 2020, 73(6):341-346.