# Artificial intelligence (AI) technologies to assist histopathology for breast cancer diagnosis Draft Protocol

| | |
|---|---|
| Topic Number: | ID6732 |
| Produced by: | Bristol Technology Assessment Group |
| Version Number: | 1.0 |
| Date: | 7 May 2026 |


Correspondence to:

Penny Whiting
Professor of Clinical Epidemiology
Bristol TAG Population Health Sciences
Bristol Medical School, University of Bristol
Bristol, BS8 2PS
Penny.whiting@bristol.ac.uk



This project was commissioned by the National Institute for Health and Care Research (NIHR) Evidence Synthesis Programme as NIHR178075.

## Team members:

| NAME | ROLE ON PROJECT | AFFILIATION |
|---|---|---|
| Penny Whiting | Clinical Lead | Bristol TAG |
| Howard Thom | Health Economics Lead | Bristol TAG |
| Katie Kerr | Reviewer | Bristol TAG |
| Marie Westwood | Reviewer | Bristol TAG |
| Catalina Lopez-Manzano | Reviewer | Bristol TAG |
| Chris Cooper | Information Specialist | Bristol TAG |
| Hayley Jones | Statistical/Methodological Expertise | Bristol TAG |
| Joe Carroll | Health Economist | Bristol TAG |
| Mary Ward | Health Economist | Bristol TAG |
| Joanna Thorn | Health Economist | Bristol TAG |
| Colin Alderson | Clinical Advisor | York and Scarborough Teaching Hospitals NHS Trust |
| Roisin Bradley | Clinical Advisor | York and Scarborough Teaching Hospitals NHS Trust |
| Clare McGenity | Clinical Advisor | University of Leeds |
| TBC | Patient Representative | |
| TBC | Patient Representative | |

# 1 Plain English Summary

## What is the breast cancer?

Breast cancer happens when some cells in the breast grow in an unusual way and form a lump (tumour). It is one of the most common types of cancer. It can affect both women and men, but it is much more common in women. It can sometimes be hard to notice early on because it may not cause symptoms at first. If symptoms do appear, they may include a lump in the breast or under the arm, changes in the size or shape of the breast, changes to the skin (such as dimpling), or fluid coming from the nipple.

Breast cancer is usually diagnosed using a combination of tests. This often starts with imaging, such as a mammogram (a special X-ray of the breast) or an ultrasound (which uses sound waves to create a picture), to look for any unusual areas. If something abnormal is seen, the next step is usually a biopsy. A biopsy means taking a small sample of tissue from the breast using a thin needle. The tissue sample is then looked at in a laboratory under a microscope. This is called histopathology. Increasingly, these slides are scanned to create high-quality digital images, known as digital pathology. These images can be viewed on a computer instead of under a microscope. Specialists examine the cells closely to see if cancer is present. If it is, they can also tell what type it is and how it may behave. These tests help make a clear diagnosis and guide treatment.

## What is artificial intelligence (AI)?

Artificial intelligence (AI) is a type of computer technology that is trained to recognise patterns in data. In this case, it can be used to look at images of tissue samples and help identify areas that may show signs of cancer. It does not replace doctors, but it can support their work.

## What are we trying to find out?

We want to find out if AI can help doctors diagnose breast cancer when they examine breast tissue samples as the first step in deciding if somebody has breast cancer. We also want to find out it is an effective use of NHS resources. To do this, we will look at research studies which answer:

- How does using AI to help look at breast tissue samples affect patient outcomes?
- How accurate is AI at helping to diagnose breast cancer?
- How does AI affect the way pathology services work, including speed of reporting, workload, service capacity, and the experiences of patients and clinicians?
- What research has already looked at the costs of using AI to help diagnose breast cancer?
- Is using AI to help diagnose breast cancer good value for money for the NHS?
- What information is still missing or unclear in the research that would help decision-making?

# 2 Decision Problem

This assessment will address the decision problem set out in the NICE draft scope for the evaluation of AI technologies to assist histopathology for breast cancer diagnosis.[1] The overall aim of this appraisal is to determine whether artificial intelligence (AI) technologies to assist histopathology for breast cancer diagnosis are clinically and -cost effective to the NHS.

Table 1 summarises the decision problem to be addressed in this assessment. Further detail on each element can be found in the published scope for the assessment. No deviations from the published scope are planned.

**Table 1 Summary table of the decision problem**

| Component | Description | EAG comment |
|---|---|---|
| **Intervention(s)** | AI technologies that assist histopathological detection of breast cancer in WSI of core or vacuum assisted biopsies for initial diagnosis, including:<br>• Aiforia Breast Cancer Suite (Aiforia Technologies Plc)<br>• AIRABreast (AIRA Matrix Private Limited)<br>• Ibex Breast, Ibex Breast IHC (Ibex Medical Analytics)<br>• Mindpeak Breast Suite (Mindpeak GmbH)<br>• Cleo Breast (Primaa)<br>• Paige Breast Suite (Tempus)<br>• Insight (Visiopharm A/S) | As defined in the scope |
| **Population(s)** | Adults who have had core needle or vacuum assisted biopsy and are awaiting initial diagnosis for suspected breast cancer outside of the NHS BSP pathway.<br><br>Adults who have previously been treated for breast cancer and did not have any tumour present when discharged from that episode are included. | As defined in the scope |
| **Subgroups** | People who have had systemic treatment (like chemotherapy or endocrine therapy) or radiotherapy for a previous breast cancer episode that may have affected the region from which the biopsy was taken. | Subgroup analysis may be performed if data permit |
| **Comparators** | Histopathologist review of breast core needle or vacuum assisted biopsies for initial diagnosis of breast cancer without AI assistance.<br><br>Reference standard for test accuracy will be determined by the evidence. | As defined in the scope |
| **Outcomes** | **Clinical outcomes**<br>• time to disease-free status<br>• overall survival<br>• breast cancer-specific mortality<br>• progression free survival, disease-free survival, distant disease-free survival<br>• adverse effects including under or overtreatment<br><br>**Patient reported outcomes**<br>• health-related quality of life<br>• service user and carer acceptability and views | As defined in the scope |

| Component | Description | EAG comment |
|---|---|---|
| | **Other intermediate outcomes**<br>• diagnostic accuracy (sensitivity, specificity, positive predictive value and negative predictive value)<br>• case review time/ turnaround time (slide review time/number of cases reviewed per session, time to produce report for MDT)<br>• time to diagnosis (referral to diagnosis, biopsy to MDT)<br>• time to initiate treatment (referral to treatment, MDT to treatment)<br>• concordance between AI and pathologist review<br>• need for or use of additional tests (repeat biopsies, repeat IHC, genetic testing)<br>• need for second pathologist read<br>• proportion of slides not appropriate for AI review/ repeat slide scanning, and reason for it<br>• technical failure<br>• effect of WSI quality and acquisition methods on accuracy<br>• impact on clinical decision-making including staging and treatment selection<br><br>**Other outcomes**<br>• ease of use or user acceptability<br><br>**Resource use**<br>• cost of technology, considering:<br>  ○ procurement<br>  ○ implementation<br>  ○ ongoing running costs<br>  ○ IT set-up<br>  ○ updates<br>  ○ data storage<br>  ○ training<br>• costs of additional tests (for example, for IHCs or ISH/FISH)<br>• cost of second reads<br>• cost of repeat procedures (repeat biopsies, repeat slide scanning, repeat IHC)<br>• cost of managing cancer, related to missed cancers or overdiagnosis<br>• total volume of cases through per session.<br><br>**Cost-effectiveness outcomes**<br>• Total quality adjusted life years (QALYs)<br>• Total costs<br>• Incremental Cost-Effectiveness Ratio (ICER)<br>• Net monetary benefit | |
| **Setting** | NHS histopathology services that do breast cancer diagnosis, outside of the NHS BSP. | Digital pathology infrastructure prerequisite |

| Component | Description | EAG comment |
|---|---|---|
| **Economic analysis** | The NICE reference case stipulates that the cost effectiveness of treatments should be expressed in terms of incremental cost per quality-adjusted life year.<br><br>The NICE reference case stipulates that the time horizon for estimating clinical and cost effectiveness should be sufficiently long to reflect any differences in costs or outcomes between the technologies being compared.<br><br>Costs will be considered from an NHS and Personal Social Services perspective.<br><br>The availability of any commercial arrangements for the intervention, comparator and subsequent treatment technologies will be taken into account. | As per NICE reference case |

# 3 Aim and Objectives

The overall aim of this project is to determine whether artificial intelligence (AI) technologies to assist histopathology for initial diagnosis and biomarker analysis of breast cancer are clinically- and cost-effective to the NHS.

We have identified the following objectives to address this aim:

1. What is the impact of AI-assisted histopathology for initial diagnosis and biomarker analysis of breast cancer on patient outcomes?
2. What is the diagnostic test accuracy of AI-assisted histopathology for initial diagnosis and biomarker analysis of breast cancer?
3. What is the impact of AI-assisted histopathology on pathology workflow, reporting efficiency, service capacity for initial diagnosis and biomarker analysis of breast cancer and patient and clinician experience?
4. What previous economic evaluations have looked at AI technologies to assist histopathology for breast cancer diagnosis?
5. What is the cost-effectiveness of AI technologies to assist histopathology for initial diagnosis and biomarker analysis of breast cancer?
6. What are the main gaps and uncertainties in the current evidence base relevant to NICE decision making?

# 4 Evidence review methods

A systematic review will be conducted to summarise the evidence on the accuracy, technical performance and clinical effects of artificial intelligence (AI) technologies to assist histopathology for initial diagnosis and biomarker analysis of breast cancer, and previous economic models in this area. The systematic review will follow the principles outlined in the Centre for Reviews and Dissemination (CRD) guidance for undertaking reviews in health care, the Cochrane Handbook for Systematic Reviews of Diagnostic Test Accuracy and the NICE Health Technology Evaluations Manual. [2 3 4]

## 4.1 Inclusion and exclusion criteria

Studies that meet the criteria summarised in Table 2 will be eligible for inclusion:

**Table 2 Inclusion Criteria**

| | **Obj 1: Clinical Impact** | **Obj 2: Accuracy** | **Obj 3: Technical performance** | **Obj 4: Previous economic models** |
|---|---|---|---|---|
| Participants | Patients who have had core needle or vacuum assisted breast biopsy samples and are awaiting results.<br><br>If insufficient studies are available in this population, we will also consider studies based solely on interpretation of images of stored samples. | | | |
| Technology | Any of the following AI technologies that assist histopathological detection of breast cancer in WSI of core or vacuum assisted biopsies for initial diagnosis, including:<br>• Aiforia Breast Cancer Suite (Aiforia Technologies Plc)<br>• AIRABreast (AIRA Matrix Private Limited)<br>• Cleo Breast (Primaa)<br>• Ibex Breast, Ibex Breast IHC (Ibex Medical Analytics)<br>• Insight (Visiopharm A/S)<br>• Mindpeak Breast Suite (Mindpeak GmbH)<br>• Paige Breast Suite (Tempus)<br><br>Evaluations of the package as a whole or separately for the following modules will be eligible:<br>• Detection module<br>• HER-2 module<br>• ER and PR<br>• Grading either using mitotic classification or the Nottingham grading scales<br><br>Studies that evaluate the use of AI as an addition to human interpretation will be prioritised. Where insufficient studies of AI plus human interpretation are available, we will also consider studies that evaluate AI alone.<br><br>Studies that evaluate these technologies for risk stratification and prognosis, on surgical specimens or on lymph node biopsies | | | Any AI technology. If no economic evaluations of AI technologies are identified, we will broaden to economic models of histopathology. |

| | Obj 1: Clinical Impact | Obj 2: Accuracy | Obj 3: Technical performance | Obj 4: Previous economic models |
|---|---|---|---|---|
| | will be excluded. Studies of mixed samples will only be included if data are reported separately for initial diagnosis of breast cancer. | | | |
| Comparator | Standard histopathological diagnostic pathway using digital pathology (WSI) or standard light microscopy | | Standard histopathological diagnostic pathway using digital pathology (WSI) | |
| Reference standard | NA | Any reported reference standard based on histo-pathological review | NA | NA |
| Outcomes | • Time to disease-free status<br>• Overall survival<br>• Breast cancer-specific mortality<br>• Progression free survival, disease-free survival, distant disease-free survival<br>• Adverse effects including under or overtreatment<br>• Health-related quality of life | • Sufficient data to construct a 2x2 table of test performance<br>• Area under the receiver operating characteristic (ROC) curve (AUC).<br>• Effect of WSI quality and acquisition methods on accuracy<br>• | • Service user and carer acceptability and views<br>• Case review time/ turnaround time<br>• Time to diagnosis<br>• Time to initiate treatment<br>• Concordance between AI and pathologist review<br>• Need for or use of additional tests (repeat biopsies, repeat IHC, genetic testing)<br>• Need for second pathologist read<br>• Proportion of slides not appropriate for AI review/ repeat slide scanning, and reason for it<br>• Technical failure<br>• Impact on clinical decision-making including staging and treatment selection<br>• Ease of use<br>• Cost of technology, | • Direct costs (e.g., medical, treatment, testing, travel, and administrative)<br>• Indirect costs (e.g., productivity losses, absenteeism, presenteeism)<br>• Cost per QALY<br>• Life-years gained (LYG)<br>• Incremental costs<br>• Incremental QALYs/LYGs/DALYs<br>• Incremental cost-effectiveness ratio (ICER) |

| | Obj 1: Clinical Impact | Obj 2: Accuracy | Obj 3: Technical performance | Obj 4: Previous economic models |
|---|---|---|---|---|
| | | | • Costs of additional tests<br>• Cost of second reads<br>• Cost of repeat procedures<br>• Cost of managing cancer, related to missed cancers or overdiagnosis<br>• Total volume of cases through per session | |
| Study design | RCT or non-randomised studies of interventions (NRSI) | Diagnostic test accuracy (DTA) studies. One-gate studies will be preferred. Where insufficient one-gate studies are available for any technology multi-gate studies will also be eligible. | If sufficient studies for obj 1 and 2 are identified that report on eligible technical performance outcomes, the review will be restricted to these studies. Where insufficient data on any outcome are available for each technology, observational studies of other designs, including qualitative studies, will be considered. Case reports will be excluded. | Economic modelling studies including cost-effectiveness analyses (CEA), cost benefit analyses (CBA), cost-utility analyses (CUA), and cost-consequence analyses (CCA). |

## 4.2 Study identification

Studies will be identified using bibliographic and non-bibliographic search methods following guidance in the NICE technology appraisal manual and recent guidance on searching.[3]

### 4.2.1 Bibliographic searching

The following databases will be searched:

- MEDLINE (Ovid SP)
- EMBASE (Ovid SP)
- EconLIT (EBSCO)
- SCI-Expanded (Clarivate)
- Conference Proceedings citation index (CPCI-S) (Clarivate)

We will use a sensitive search strategy structured as follows:

1. terms for breast cancer or diagnostic context
2. terms for histopathology and digital pathology
3. a search filter for AI or AI interventional studies developed by NICE
4. 1 and 2 and 3
5. terms for the technologies/manufacturers of the technologies
6. 4 or 5

The search will not be limited by date, publication type, study design, or language. A draft search strategy is reported in Appendix 10.1.

### 4.2.2 Non-bibliographic search methods

Completed and ongoing trials will be identified through searches of the following trial registries:

- ClinicalTrials.gov via https://www.clinicaltrials.gov/
- WHO International Clinical Trials Registry Platform (ICTRP) via https://www.who.int/clinical-trials-registry-platform

Additional relevant studies will be identified by:

- Screening reference lists of any reviews (systematic or non-systematic) identified by our searches
- Reviewing the reference lists of all included studies
- Hand searching the websites of the manufacturer/or licence holders for each test
- Screening the AI Device Register (https://osf.io/gb84r/files/n78x5)
- Information submitted by test manufacturers and other stakeholders

### 4.2.3 Managing the searches

Search results will be exported to EndNote 20. They will then be imported into Nested Knowledge systematic review software (nested-knowledge.com) for deduplication and screening.

## 4.3 Review strategy

Prior to screening, we will implement several steps to categorise the unscreened records, as the search strategy is sensitive. Studies published prior to 2019 will be excluded, as the technologies of interest were first introduced after this – the earliest technology to become available was the Ibex Breast algorithm (formerly known as Galen Breast) which was launched in 2019. Studies excluded prior to 2019 will be categorised separately to allow for later exploration if needed.

Conference abstracts will only be eligible if they mention the specific technology evaluated in the abstract; full study reports will not be restricted in this way as they may only mention the technology in the full text of the report. The Nested Knowledge criteria-based screening (CBS) workflow, which includes the large language model (LLM) AI feature, 'Smart Screen',[5] will be used to identify conference abstracts which do not mention the technologies of interest, which will be bulk excluded prior to screening.

Two reviewers will independently screen the remaining titles and abstracts identified by the searches. Full copies of all reports considered potentially relevant will be obtained and two reviewers will independently assess these for inclusion. Study prioritisation may be undertaken where the evidence base is large, focusing on studies most relevant to the NICE decision problem as set out in the eligibility criteria. Any disagreements will be resolved by consensus or discussion with a third reviewer.

Multiple publications of the same studies will be linked and considered together. Data will be extracted using standardised data extraction forms developed in Nested Knowledge. Data extraction forms will be piloted on three papers by two human reviewers and adapted as necessary. Data will then be extracted for the remaining studies using the Nested Knowledge LLM AI extraction feature, Adaptive Smart Tags (AST).[6] The AI extracted data will be checked in detail by a human reviewer. Where substantial edits from the human reviewer are made to the AI suggested extractions, or where any calculations from extracted data are made, this will be checked by a second human reviewer. Any disagreements will be resolved by consensus or discussion with a third reviewer.

Data will be extracted on the following: study design (RCT, NRSI, DTA, economic model or other), objective that the study addresses, funding sources (public, industry, mixed), study location, indication for biopsy, sex, age, inclusion criteria, AI technology details, existing infrastructure, how AI interacts with the human, datasets used to inform AI algorithms, comparator or reference standard test(s), and outcomes specified in inclusion criteria (section 4.1).[7]

We will consider the PROGRESS-Plus population factors, where reported.[7] PROGRESS-Plus is an acronym that describes factors that contribute to health inequity. PROGRESS stands for: place of residence, race/ethnicity/culture/language, occupation, gender/sex, religion, education, socioeconomic status, and social capital. “Plus” stands for any additional factors considered important for the specific topic under review. We consider the main factors relevant to this appraisal to be sex and ethnicity, with men often diagnosed at a later stage. There are no additional factors beyond the standard PROGRESS factors relevant for this appraisal. We will extract whether each PROGRESS-Plus factor was reported at baseline (y/n), the baseline data concerning the factor as reported by the authors, and whether the study reports results data stratified by the factor. Where stratified data are reported, these will be extracted.

Dichotomous clinical impact data will be extracted as number of patients with events and/or number of events and total number of patients in each treatment arm. For categorical data, we will extract details on the categories assessed, the total number of patients in each treatment arm and the number of patients in each outcome category. For continuous data we will extract means/medians together with ranges, standard deviations (SD), standard errors (SE) and/or confidence intervals (CIs) for the outcome at baseline, follow-up and for change from baseline in each treatment group. For all types of clinical impact data, effect estimates together with 95% CIs and p-values for comparisons between groups, together with details on the methods of analysis, any variables controlled for in the analysis and the test statistic will be extracted.

Accuracy data will be extracted as 2x2 tables comparing the AI technology with the reference standard where available. If measure of accuracy (e.g. sensitivity, specificity, AUC for the ROC plot) are reported without providing the information needed to calculate 2x2 tables, then these data will be extracted, alongside 95% CIs where reported.

Where multiple sets of 2x2 data are reported in a single study, for example for different tests, target conditions, thresholds, or subgroups of interest, all data will be extracted.

For economic models we will extract data on study objectives, country/location, year of cost estimates, analytic viewpoint (perspective), time horizon, discount rate, population, histopathology strategies included in model, key assumptions, type of model (e.g., decision tree, Markov model, decision tree combined with Markov model, discrete event simulation, budget impact model), cycle length (if Markov), states/events in the model, sources of evidence on transition/event rates, sources of evidence on utilities and costs, results (e.g. incremental cost-effectiveness ratio [ICER], net monetary benefit), key uncertainties (e.g., results of scenario analyses, deterministic sensitivity analyses and probabilistic sensitivity analysis), conclusions and limitations (noted by the author).

## 4.4 Quality assessment strategy

The methodological quality of included RCTs will be assessed using the updated Cochrane Risk of Bias Tool (ROB 2.0).[8] NRSI will be assessed using the ROBINS-I tool.[8] DTA studies will be assessed for methodological quality using QUADAS-3.[9] An add on to QUADAS-3 for studies evaluating AI technologies, QUADAS-AI is under development.[10] We will pilot this tool as part of this assessment.[11] The reporting quality of economic evaluations will be assessed using the Consolidated Health Economic Evaluation Reporting (CHEERS-AI) checklist,[8] and the methodological quality of the analytic decision model will be assessed using the ECO-BIAS checklist.[12] Quality assessment will be undertaken by one reviewer and checked by a second reviewer. Any disagreements will be resolved by consensus or discussion with a third reviewer.

## 4.5 Synthesis methods

For each of the four systematic review objectives (obj 1 to 4), a narrative summary of all of the included studies will be presented. This will include a summary of the study characteristics (e.g. study designs, sample size, geographical location, year, baseline population characteristics, AI technology evaluated, how the AI is used in conjunction with human interpreters), outcomes reported and study quality. All syntheses will be stratified by technology evaluated.

If sufficient data are available for any reported outcome, meta-analysis will be carried out to generate summary estimates. For studies of effectiveness, random effects meta-analysis will be performed to allow for between-studies heterogeneity. A restricted maximum likelihood (REML) approach will be used to estimate the between-study heterogeneity parameter, tau. Heterogeneity and inconsistency across studies will be quantified using the tau and $I^2$ statistics.[13] Fixed effect meta-analyses will be performed as sensitivity analyses, or as the sole analyses if insufficient data are available to estimate tau. Where NRSI are synthesised, estimates that have been adjusted for potential confounders will be used where available.

For accuracy data, bivariate random effects meta-analysis of sensitivity and specificity will be performed, with binomial likelihoods.[14, 15] Analyses will be stratified according to AI technology and whether this is evaluated alone or in combination with human interpretation. Summary estimates of sensitivity and specificity together with 95% confidence intervals (CIs) will be calculated. Coupled forest plots of sensitivity and specificity will be used to display results from individual studies, to allow visual assessment of heterogeneity. Study-level and pooled results will also be plotted in Receiver Operating Characteristic (ROC) space, with 95% confidence ellipses around pooled estimates representing the joint uncertainty in sensitivity and specificity. We do not anticipate having sufficient studies for formal investigation of heterogeneity for any objectives. However, if sufficient data are available, we will consider population, and how the human interacts with the AI technologies, whether participants have previously been treated for breast cancer, presence of rare tumour morphologies, and borderline biomarker status as potential sources of heterogeneity.

A detailed description of any gaps in the evidence will be provided together with any methodological limitations of the existing studies.

### 4.6 The use of supportive AI tools

AI tools will be used in a limited and supportive capacity within this review. Within the Nested Knowledge platform, AI-enabled features will be used to assist with study processing. The 'Smart Screener' feature will be used to support the refinement of search results prior to formal screening.[5] In addition, the 'Adaptive Smart Tags' (AST) feature will be used to assist with initial data extraction.[6]

Separately, general-purpose AI tools, including Microsoft Copilot,[16] may be used to support comprehension of complex technical reports and for editorial refinement of text. These tools will not be used to generate, analyse, or interpret study data.

All AI-assisted outputs will be subject to human oversight. AI-supported refinement of search results will not replace independent dual-reviewer screening for the majority of abstracts. A 10% random sample of records excluded during search result refinement using Smart Screener will be independently reviewed by at least one reviewer. Data extracted using AST will be checked in detail by at least one reviewer, with a second reviewer involved where substantial amendments or calculations are required.

Potential risks associated with the use of AI tools include misinterpretation or oversimplification of technical material, inaccurate or incomplete data extraction, and inappropriate influence on the interpretation of findings. To mitigate these risks, all use of AI tools will be subject to full human oversight. Reviewers will retain responsibility for study selection, data extraction, and quality assessment; interpretation of the evidence; drafting and approval of all outputs; and the formulation of final decisions and conclusions.

# 5 Economic analysis methods

Subject to data availability, decision analytic modelling will be undertaken to determine the cost-effectiveness of artificial intelligence (AI) technologies to assist histopathology for initial diagnosis and biomarker analysis of breast cancer in adults who have had core needle or vacuum assisted biopsy and are awaiting initial diagnosis for suspected breast cancer outside of the NHS BSP pathway. This includes adults who have previously been treated for breast cancer and did not have any tumour present when discharged from that episode.

## 5.1 Diagnosis and treatment strategies

The analysis will consider the diagnostic performance, and long-term consequences of the different AI technologies that assist histopathology for breast cancer diagnosis. For technologies for which the diagnostic value is unclear, when feasible, assumptions will be made to provide some indication of the range of cost-effectiveness outcomes.

The comparator will be histopathologist review of breast core needle or vacuum assisted biopsies for initial diagnosis of breast cancer without AI assistance. Reference standard for test accuracy will be determined by the evidence.

## 5.2 Model structure

The modelling approach (e.g., decision tree followed by Markov model, discrete-event simulation) and model structure will be determined following the results of the systematic review of economic models described in Section 4 and discussion with clinical and PPIE advisors.

## 5.3 Model inputs

The systematic reviews described in Section 4 will provide estimates of diagnostic test accuracy, clinical impact, and technical performance that will inform the model.

Targeted reviews of the literature regarding published utility and cost studies will be performed.[17] The intention of this explorative review is to identify studies that can be used to support the development of a health economic model, and to estimate the model input parameters, that will aim to answer the research questions of this assessment, but not to perform a systematic review.

### 5.3.1 Health outcomes

We will include the impact of the different diagnostic strategies (use of AI technologies or standard care) on health-related quality of life and mortality in the model. Utility values will be based on literature sources identified through our review of previous cost-effectiveness models and other targeted literature searches, and quality-adjusted life-years (QALYs) will be derived for each strategy. We will also include the impact on carers in the model if appropriate evidence is available.

### 5.3.2 Costs

Resource utilisation will be estimated from an NHS and Personal Social Services (PSS) perspective. Unit cost data for the cost analyses will be drawn from routine sources (e.g. NHS

National Cost Collection,[18] Unit Costs of Health and Social Care,[19] British National Formulary (BNF)),[20] discussions with clinical advisers and with the manufacturers of the technologies, and via our review of previous cost-effectiveness models, supplemented with targeted literature searches. We will consider the following resources:

- Diagnostic technology costs, including licence fees and maintenance
- IT requirements, including staffing, integration with digital pathology systems and data storage costs
- Staff training costs and other healthcare staffing and process implications
- MDT costs associated with the management of cases
- Ongoing treatment and monitoring costs
- Costs of additional tests and second reads
- Costs associated with adverse events or technical failures
- End of life care associated with breast cancer mortality

As the infrastructure required for digital pathology (a pre-requisite for the use of AI technology) forms standard of care in this assessment, costs associated with digital pathology itself will not be considered.

## 5.4 Planned analyses

The economic analysis will be conducted in line with the NICE reference case.[3] Perspective will be that of the NHS and PSS. We will estimate total costs, total QALYs, incremental costs, incremental QALYs, and the ICER for each of the interventions. A full incremental analysis will be conducted if multiple interventions and comparators are modelled. Longer term costs and consequences will be discounted using the UK discount rates of 3.5% of both costs and effects.

Probabilistic sensitivity analyses will be performed using parameter distributions instead of fixed values. Deterministic sensitivity & scenario analyses will be performed, if necessary. Scenarios will include exploration of the impact of varying histopathologist experience, laboratory characteristics, presence of rare tumour morphologies, and borderline biomarker status, where data availability permits. Decision uncertainty regarding mutually exclusive alternatives will be reflected using cost-effectiveness planes and cost-effectiveness acceptability curves.

Subgroup analyses will be explored if feasible, potentially by sex and ethnicity. This will include consideration of people who have had systemic treatment (like chemotherapy or endocrine therapy) or radiotherapy for a previous breast cancer episode that may have affected the region from which the biopsy was taken.

# 6 Handling information from the companies and other stakeholder

All data submitted by the companies in evidence and information requests by NICE, or data submitted by other stakeholders will be considered by the EAG if received by 30/6/2026. Information arriving after this date will not be considered. If the data included in the information provided meets the inclusion criteria for the review, they will be extracted and quality assessed following the procedures outlined in this protocol. The EAG may seek clarification or additional information from companies and other stakeholders where necessary. All correspondence between the EAG and companies will happen through NICE.

Any 'commercial in confidence' data provided by manufacturers, and specified as such, will be highlighted in blue and underlined in the assessment report (followed by company name in parentheses). Any 'academic in confidence' data provided by manufacturers, and specified as such, will be highlighted in yellow and underlined in the assessment report. Any confidential data used in the cost-effectiveness models will also be highlighted. If confidential information is included in economic models then a version using dummy data or publicly available data in place of confidential data will be provided.

# 7 Competing interests of authors

None of the authors have any competing interests.

# 8 References


1. National Institute for Health and Care Excellence. *[Final scope] Artificial intelligence (AI) technologies to assist histopathology for breast cancer diagnosis*.2026 (Accessed May 2026).
2. Centre for Reviews & Dissemination. *CRD's guidance for undertaking reviews in health care. Centre for Reviews and Dissemination*. York: University of York; 2009.
3. The National Institute for Health Care Excellence. *NICE health technology evaluations: the manual*.2026. URL: https://www.nice.org.uk/process/pmg36/chapter/introduction-to-health-technology-evaluation (Accessed April 2026).
4. Deeks J, Bossuyt P, Leeflang M, Y T, Flemyng E. Cochrane Handbook for Systematic Reviews of Diagnostic Test Accuracy. 2022; https://doi.org/10.1002/9781119756194
5. NESTED Knowledge. *Smart Screener*.2026. URL: https://about.nested-knowledge.com/docs/smart-screener/ (Accessed April 2026).
6. NESTED Knowledge. *Adaptive Smart Tags*.2026. URL: https://about.nested-knowledge.com/docs/smart-tag-recommendations/ (Accessed April 2026).
7. O'Neill J, Tabish H, Welch V, Petticrew M, Pottie K, Clarke M*, et al.* Applying an equity lens to interventions: using PROGRESS ensures consideration of socially stratifying factors to illuminate inequities in health. *Journal of Clinical Epidemiology* 2014;**67**(1): 56-64
8. Elvidge J, Hawksworth C, Avşar TS, Zemplenyi A, Chalkidou A, Petrou S*, et al.* Consolidated Health Economic Evaluation Reporting Standards for Interventions That Use Artificial Intelligence (CHEERS-AI). *Value Health* 2024;**27**(9): 1196-1205
9. Adarkwah CC, van Gils PF, Hiligsmann M, Evers SM. Risk of bias in model-based economic evaluations: the ECOBIAS checklist. *Expert Review of Pharmacoeconomics & Outcomes Research* 2016;**16**(4): 513-523
10. Sounderajah V, Ashrafian H, Rose S, Shah NH, Ghassemi M, Golub R*, et al.* A quality assessment tool for artificial intelligence-centered diagnostic test accuracy studies: QUADAS-AI. *Nature Medicine* 2021;**27**(10): 1663-1665
11. Drummond M, Sculpher M, Torrance G, O'Brien B, Stoddart G. *[Drummond checklist] Methods for the economic evaluation of health care programs*. Oxford: Oxford University Press; 2005.
12. Adarkwah CC, van Gils PF, Hiligsmann M, Evers SMAA. Risk of bias in model-based economic evaluations: the ECOBIAS checklist. *Expert Review of Pharmacoeconomics & Outcomes Research* 2016;**16**(4): 513-523
13. Higgins J, Thompson SG, Deeks JJ, Altman DG. Measuring inconsistency in meta-analyses. *BMJ* 2003;**327**(7414): 557-560
14. Reitsma JB, Glas AS, Rutjes AWS, Scholten RJPM, Bossuyt PM, Zwinderman AH. Bivariate analysis of sensitivity and specificity produces informative summary measures in diagnostic reviews. *Journal of Clinical Epidemiology* 2005;**58**(10): 982-990
15. Chu H, Cole SR. Bivariate meta-analysis of sensitivity and specificity with sparse data: a generalized linear mixed model approach. *Journal of Clinical Epidemiology* 2006;**59**(12): 1331-1332
16. Microsoft. *Microsoft Copilot*..n.r. URL: https://copilot.microsoft.com/ (Accessed April 2026).
17. NHS England. *2024/25 National Cost Collection Data Publication*.2026. URL: https://www.england.nhs.uk/publication/2024-25-national-cost-collection-data-publication/ (Accessed April 2026).

18. NHS England. *National Cost Collection for the NHS*..n.r. URL: https://www.england.nhs.uk/costing-in-the-nhs/national-cost-collection/. (Accessed May 2026).
19. University of Kent. *Unit Costs of Health and Social Care 2024 Manual*.2020. URL: https://kar.kent.ac.uk/109563/ (Accessed April 2026).
20. National Institute for Health and Care Excellence. *British National Formulary (BNF)*.2026. URL: https://bnf.nice.org.uk/ (Accessed April 2026).
21. Campbell SM. *Retrieve Studies related to Social Media from the OVID EMBASE Database*.2022. URL: https://docs.google.com/document/d/1eWyO0jv9_6FYsxyC5LUYwFe9eH_3h83-tPNZ6wmos18/edit?tab=t.0#heading=h.qi55eeyvgzy9 (Accessed April 2026).
22. Ayiku L, Finnegan A, Hudson T, Walsh N, Adams R. Development and validation of the NICE artificial intelligence (AI) medical device intervention search filters for MEDLINE and Embase (Ovid). *International Journal of Technology Assessment in Health Care* 20251-27

# 9 Appendices

## 9.1 Literature searches

Search purpose: to identify studies reporting data on the clinical or cost effectiveness, accuracy or the technical performance, and experience and views, of the technologies specified in the scope.

Database: MEDLINE (MEDALL)
Host: Ovid
Data parameters: 1946 to April 14, 2026
Date of SAMPLE search: 20 April 2026

| # | Search strategy | Hits | Search narrative |
|---|---|---|---|
| 1 | exp Breast Neoplasms/ | 381636 | **Terms for breast cancer and diagnostic context:** Lines 1 and 2 define the disease domain (breast cancer). Line 3 broadens the search to include diagnostic and imaging-related terms associated with breast evaluation. The combination (Line 4) ensures comprehensive retrieval of studies relating to breast cancer and its diagnostic assessment. |
| 2 | (Breast* adj3 (carcin* or cancer* or neoplas* or tumor* or tumour* or metasta* or malig*)).ti,ab,kw,kf. | 474021 | |
| 3 | (breast adj3 (mammary or mammography or mammographers or tomosynth* or patholog or exam* or ultrasound or MRI or DBT or image or imaging)).ti,ab,kw,kf. | 33109 | |
| 4 | 1 or 2 or 3 | 547530 | |
| 5 | (histopatholog* or "histo patholog*" or histolog* or "tissue pathology" or "tissue diagnosis").ti,ab,kw,kf. | 10047565 | **Terms for histopathology and digital pathology:** Lines 5 and 6 capture histopathology and tissue-based diagnostic terminology. Line 7 introduces terms related to digital pathology and image representation, including whole-slide imaging and computational image handling. Their combination (Line 8) ensures retrieval of studies involving digital histopathological analysis, which underpins AI-assisted pathology applications. |
| 6 | Histology.ti,ab,kw,kf. | 175170 | |
| 7 | (((digital* or gigapixel*) adj2 imag*) or digital slide or (digital* adj2 scan*) or digital* pathology or virtual slide* or mounted section or whole slide image* or whole slide imaging or tiling or patch extract*).ti,ab,kw,kf. | 38737 | |
| 8 | 5 or 6 or 7 | 1038089 | |
| 9 | exp Artificial Intelligence/ or (AI or "artificial intelligence" or AIVI or "classification algorithm*" or "computer heuristic*" or "computer assisted" or "convolutional network*" or DALL-E or "decision support system*" or "decision tree" or DeepAI or "deep learning" or "data science" or "feature detection" or "generative pre-trained transformer" or "generative pretrained transformer" or Invideo or "language learning model*" or "large language model*" or "learning algorithm*" or "machine learning" or "machinelearn*" or "natural language process*" | 681574 | **Terms for Artificial Intelligence:** We have combined and adapted the filters by Campbell (2025),[21] by removing the terms for common LLMs and adding the term 'computer assisted', and the NICE AI intervention filter by Ayiku and colleagues,[22] for use in this appraisal. |

| # | Search strategy | Hits | Search narrative |
|---|---|---|---|
| | or "nearest neighbo*" or "neural network*" or "outlier detection" or "pattern recognition" or "outcome prediction" or "support vector machine*").ti,ab,kw,kf. | | |
| 10 | algorithm*.ti,kf. or (algorithm* adj2 (learn* or automate* or detect* or predict* or treatment* or therap* or radiolog* or AI or DL or data or dataset* or base* or classif*)).ab. | 195468 | |
| 11 | 9 or 10 791336 | 793246 | |
| 12 | 4 and 8 and 11 | 2338 | Line 12 completes the primary search. It combines terms for breast cancer or diagnostic context (Line 4) AND terms for histopathology or digital pathology (Line 8) AND terms for AI (Line 11).<br><br>The MEDLINE search is not limited by publication date or type, by language, or by study design. |
| 13 | 13 (Aiforia* or "Ibex Breast*" or "Galen Breast" or "Ibex medical" or "Ibex IHC" or "Ibex AI" or "Cleo Breast*" or Primaa).ti,ab,kf. [in scope] | 31 | Lines 13 and14 represent the names of the AI tools or their manufacturers. |
| 14 | (AIRABreast* or "AIRA Matrix*" or Mindpeak or "Paige* AND Breast" or "Paige* AND suite" or (tempus and (breast* or AI)) or (Visiopharm* or Oncotopix* or "HER2-CONNECT") or (insight and AI and breast*)).ti,ab,kf. [in scope, but awaiting confirmation] | 160 | |
| 15 | 13 or 14 | 196 | |
| 16 | 12 or 15 | 2516 | Line 16 combines the primary search with the search for AI tools under review. These separate searches are combined using the Boolean connector OR. This means that results from both search sets are pooled for study selection. |