PROTOCOL

Early Detection using Information Technology in Health (EDITH)

ISRCTN:	81384017
Co-sponsors:	University of Warwick and Cambridge University Hospitals NHS Foundation Trust
Funding body:	NIHR HTA Programme
Ethics approval date:	13 November 2025 by Cambridge East Research Ethics Committee
Version number:	Version: 4.0
Date:	19-May-26
Stage:	Final

This protocol has regard for current HRA guidance and content

Protocol modifications:

Modification no.	Date of modification	Date of approval
SM01	19-May-26	22-Jun-26

Funded by NIHR | National Institute for Health and Care Research

CONTACT DETAILS

Co-sponsor:	Associate Director Research Culture, Governance and Compliance University House University of Warwick Coventry CV4 8UW Tel: 024 7657 5733 Email: sponsorship@warwick.ac.uk	**Co-sponsor**	Cambridge University Hospitals NHS Foundation Trust Box 277, Addenbrooke's Hospital Cambridge CB2 0QQ Tel: 012 2321 7418 Email: cuh.research@nhs.net
Chief Investigator:	Prof Sian Taylor-Phillips Warwick Medical School University of Warwick Coventry CV4 7AL Tel: 024 7657 5882 Email: s.taylor-phillips@warwick.ac.uk	**Co-Chief Investigator** **Radiologist:**	Prof Fiona Gilbert Department of Radiology University of Cambridge Cambridge CB2 0QQ Tel: 012 2374 6439 Email: fjg28@medschl.cam.ac.uk
Trial Manager:	Amy Hopkins and Louisa Berridge Warwick Medical School University of Warwick Coventry CV4 7AL Tel: 024 7615 1912 Email: Edith@warwick.ac.uk	**Co-investigator** **Scottish radiologist lead:**	Dr Gerald Lip Clinical Director, Northeast of Scotland Breast Screening Service NHS Grampian Aberdeen Tel: 079 1733 6669 Email: gerald.lip@nhs.scot
Co-investigator **English radiologist lead:**	Dr Nisha Sharma Leeds Teaching Hospital NHS Trust Leeds LS9 7TF Tel: 011 326 3798 email: nisha.sharma2@nhs.net	**Co-investigator** **CTU lead:**	Prof Janet Dunn Warwick Medical School University of Warwick Coventry CV4 7AL Tel: 024 7657 5847 Email: J.A.Dunn@warwick.ac.uk
Co-investigator **AI and data analytics:**	Prof Hutan Ashrafian Imperial College London London SW7 2AZ Tel: 020 3312 7651 Email: h.ashrafian@imperial.ac.uk	**Co-investigator** **Information technology and monitoring:**	Prof Mark Halling-Brown Royal Surrey County Hospital Guildford GU2 7XX Tel: 079 7613 6248 Email: mhalling-brown@nhs.net
Co-investigator **Test evaluation methodology and statistician:**	Prof Sue Mallett University College London London WC1E 6BT Email: sue.mallett@ucl.ac.uk	**Co-investigator** **Health economics lead:**	Prof Katherine Payne The University of Manchester Manchester M139PL Tel: 016 1306 7906 Email: Katherine.Payne@manchester.ac.uk
Co-investigator **Health data science:**	Prof Yan Chen University of Nottingham Nottingham NG5 1PB Email: Yan.Chen@Nottingham.ac.uk	**Co-investigator** **Health data science Scotland:**	Prof Lesley Anderson University of Aberdeen Aberdeen AB25 2ZD Tel: 012 2443 7115 Email: lesley.anderson@abdn.ac.uk

Co-investigator Behavioural Scientist Qualitative lead:	Prof Jo Waller Queen Mary University of London London E1 4NS Email: jo.waller@qmul.ac.uk	**Co-investigator Radiologist:**	Dr Louise Wilkinson National Specialist Advisor for Breast Screening Vaccination and Screening Directorate NHSE Email: Louise.wilkinson11@nhs.net
Co-investigator Trial delivery:	Dr Karoline Freeman Warwick Medical School University of Warwick Coventry CV4 7AL Tel: 024 7657 4026 Email: k.freeman@warwick.ac.uk	**Co-investigator Test evaluation statistician:**	Dr Siew Wan Hee Warwick Medical School University of Warwick Coventry CV4 7AL Email: S.W.Hee@warwick.ac.uk
Co-investigator Imaging data and density:	Prof Susan Astley Theodossiadis The University of Manchester Manchester M139PL Tel: 016 1275 5162 Email: sue.astley@manchester.ac.uk	**Co-investigator Trial delivery Scotland:**	Dr Clarisse de Vries University of Glasgow Glasgow Email: Clarisse.deVries@glasgow.ac.uk
Co-investigator EDI champion:	Dr Gulnaz Iqbal Warwick Medical School University of Warwick Coventry CV4 7AL Tel: 024 7615 0178 Email: G.Iqbal@warwick.ac.uk	**Co-investigator PPI lead:**	Ms Debbie Keatley Email: debbiekeatley01@gmail.com
Co-investigator Statistician:	Dr Andrea Marshall Warwick Medical School University of Warwick Coventry CV4 7AL Email: Andrea.Marshall@warwick.ac.uk	**Co-investigator Health Economist:**	Dr Stuart Wright The University of Manchester Manchester M139PL Email: stuart.j.wright@manchester.ac.uk
Co-investigator Benchmarking and tuning AI tools:	Dr Nicholas Payne Department of Radiology University of Cambridge Cambridge CB2 0QQ Email: np520@medschl.cam.ac.uk		
For general queries and supply of trial materials please contact the coordinating centre: Warwick Clinical Trials Unit (WCTU) University of Warwick Gibbet Hill Road Coventry CV4 7AL Tel: 024 7615 1912 Email: edith@warwick.ac.uk			

TABLE OF CONTENTS PAGE

TRIAL SUMMARY 6

LIST OF ABBREVIATIONS/GLOSSARY 9

1. BACKGROUND 10

1.1 Current screening practice and limitations 10

1.2 Existing knowledge 10

1.3 Need for a trial 11

2. TRIAL DESIGN 12

3. AIMS AND OBJECTIVES 13

4. OUTCOME MEASURES 13

5. PARTICIPANT SELECTION AND ELIGIBILITY 14

5.1 Inclusion criteria 14

5.2 Exclusion criteria 14

5.3 Active dissent 14

6. RANDOMISATION AND INTERVENTION 17

6.1 Randomisation 17

6.2 Trial intervention 17

6.3 Pre-trial phase 18

6.4 Site staff training 19

6.5 Compliance/contamination 19

6.6 Blinding 20

6.7 Co-enrolment into other trials 20

6.8 End of trial 20

7. METHODS AND ASSESSMENTS 20

7.1 Overview of delivery of intervention and data flow 20

7.2 Qualitative assessments – nested study 21

7.3 Test accuracy of AI systems – nested study 22

8. ADVERSE EVENT MANAGEMENT 22

9. STATISTICAL CONSIDERATIONS 23

9.1 Power and sample size 23

9.2 Statistical analysis 23

10. HEALTH ECONOMIC EVALUATION 25

10.1 Cost of using AI to read mammograms 25

10.2 Decision-analytic model 25

10.3 Barriers affecting delivery of AI in NHSBSP for breast cancer 26

11. DATA MANAGEMENT 26
11.1 Data collection and management 26
11.2 Database 26
11.3 Data storage 27
11.4 Data access and quality assurance 27
11.5 Data shared with third parties 27
11.6 Archiving 27
12. TRIAL TIMETABLE AND MILESTONES 28
13. STUDY RESPONSIBILITIES, OVERSIGHT AND MANAGEMENT 28
13.1 Sponsor and governance arrangements 28
13.2 Ethical approval and considerations 29
13.3 Trial Registration 29
13.4 Notification of serious breaches to GCP and/or trial protocol 30
13.5 Indemnity 30
13.6 Administration 30
13.7 Trial Management Group (TMG) 30
13.8 Trial Steering Committee (TSC) 30
13.9 Data Monitoring Committee (DMC) 31
13.10 Clinical Safety Monitoring Committee (CSMC) 31
13.11 Essential documentation 31
13.12 Financial support 31
13.13 Safeguarding researchers and research participants 31
14. MONITORING, AUDIT AND INSPECTION 31
15. PATIENT AND PUBLIC INVOLVEMENT AND ENGAGEMENT (PPIE) 31
16. DISSEMINATION AND PUBLICATION 32
17. REFERENCES 33
18. APPENDIX 36
18.1 Appendix 1 Data flow map 36

LIST OF TABLES PAGE

Table 1 Progression criteria 28

LIST OF FIGURES PAGE

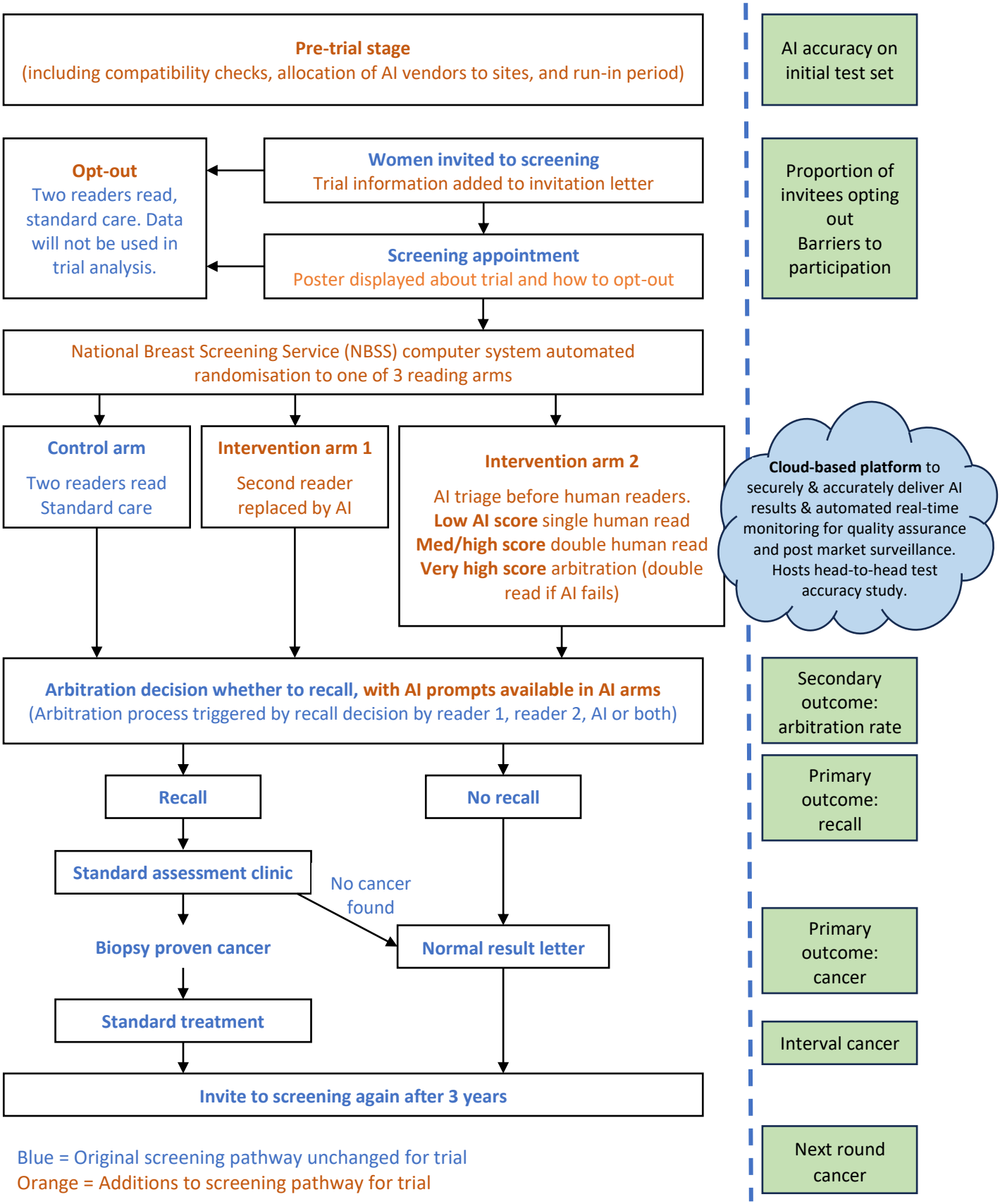
Figure 1 Trial flow diagram 8

TRIAL SUMMARY

Title	Early Detection using Information Technology in Health (EDITH)
Short title	EDITH
Clinical phase	Phase III
Trial design	Multi-centre, pragmatic, cluster, randomised controlled trial allocated 1:1:1 across control and intervention arms. • Control arm: current double human readers, • Intervention 1: replacing second human reader with AI, • Intervention 2: using AI for triage. Each site will undertake all three arms. Randomisation will be at the clinic day level across approximately 30 breast screening centres. A nested accuracy study will compare AI systems' accuracy by ethnicity, age, socioeconomic status and breast density. A health economic evaluation will be undertaken to evaluate the cost-effectiveness of implementing different AI systems and AI pathways. A qualitative sub-study will monitor reasons for women opting out.
Trial participants	Women invited to routine breast screening with mammography
Planned sample size	660,000 women
Follow-up duration	4 years from when women's screening episode is closed
Planned trial period	From 1st August 2025 – 31st July 2029
Objectives	• To integrate AI into the NHSBSP and measure the difference in clinical and cost-effectiveness between three study arms: replacing the second reader with AI (intervention 1); using AI for triage (intervention 2); and the current double reading pathway (control); • To understand differences in accuracy by AI vendor, mammographic acquisition system and by population subgroup (ethnicity, breast density, socioeconomic status) in a nested accuracy study; • To develop a comprehensive post-market surveillance methodology for national rollout; • To monitor reasons for women opting out; • To calculate the cost of embedding AI into the national screening programme.

Outcome measures	Primary 1. Cancer detection rate (CDR) (=cancers detected/women screened) 2. Recall rate (=women recalled/women screened) Secondary 3. Arbitration rate (=women requiring additional read /women screened) 4. Interval cancer rate (=interval cancers within 3 years/women screened) and next screen CDR 5. Characteristics of cancers detected (size, type and grade) 6. Outcomes 1-5 by population group (age, ethnicity, socioeconomic status (SES)) to explore inequalities and by breast density 7. AI failure rates and reasons 8. Atypia detection (incidental findings) 9. Workload (e.g. reading volume, arbitration volume, recall to assessment volume) 10. Reader behaviour and performance (by e.g. AI vendor, AI threshold, batch length, time on task, reader experience and reading volume) 11. Opt-out rates and reasons for opting out as a measure of acceptability in screening-eligible women (where provided by women) 12. Cost to the NHS of the three pathways 13. Cost-effectiveness by technology and population subgroup, and impact on NHS capacity 14. Test accuracy of AI from a nested test accuracy study overall and by population subgroup (e.g. age, ethnicity, breast density, socioeconomic status)
Inclusion	Women attending NHSBSP population breast screening at participating centres.
Exclusion	Women being screened in the moderate and high-risk screening service. Women being screened in the very high-risk screening service. Women who have actively dissented from participation in the study. Women who are not within the manufacturer defined intended purpose of the AI tool.
Project timetable	Recruitment will take place over 25 months. A 9-month internal pilot from the start of recruitment will assess site set up and recruitment. Primary analysis will be at the end of the four-year trial before interval cancers have been captured. The final analysis will be when all women have undergone a second round of screening or more than 3 years follow up to capture interval cancer rates.

Figure 1 Trial flow diagram



LIST OF ABBREVIATIONS/GLOSSARY

Abbreviation	Explanation
AE	Adverse Event
AI	Artificial Intelligence
CAG	Confidentiality Advisory Group
CDR	Cancer detection rate
CI	Chief Investigator
CONSORT	Consolidated Standards of Reporting Trials
COPE	Committee on Publication Ethics
CRF	Case Report Form
CSMC	Clinical Safety Monitoring Committee
CTU	Clinical Trials Unit
DICOM	Digital Imaging and Communications in Medicine
DMC	Data Monitoring Committee
FAQ	Frequently asked questions
GCP	Good Clinical Practice
IRAS	Integrated Research Application System
ISRCTN	International Standard Randomised Controlled Trial Number
NBSS	National Breast Screening System
NDOO	National Data Opt-Outs
NHSBSP	National Health Service Breast Screening Programme
PACS	Picture Archiving and Communication System
PI	Principal Investigator
PPIE	Patient & Public Involvement and Engagement
QALYs	Quality-Adjusted Life Years
QMUL	Queen Mary University London
RCT	Randomised Controlled Trial
RDO	Radiographic Diagnostic Officers
REC	Research Ethics Committee
R&D	Research and Development
SAE	Serious Adverse Event
SOP	Standard Operating Procedure
TSC	Trial Steering Committee
WCTU	Warwick Clinical Trials Unit

1. BACKGROUND

1.1 Current screening practice and limitations

Over six million women attend breast screening every three years in the UK, and each of the mammograms is assessed by two readers (radiologists, breast clinicians and trained radiographers). Double reading (two independent human readers) is the UK and European standard, as the second reader identifies an additional 6-10% of cancers and reduces false positive recalls (unnecessary further tests when the woman does not have cancer).[1]

There is a shortage of trained expert readers, with a 10%-11% vacancy rate in the consultant radiologist workforce. Unlike the rest of Europe, the UK uses trained radiographers (technologists rather than clinicians) as expert readers to meet the demand, and they undertake almost 50% of the work. However, increasing workloads are projected, potentially delaying results and treatment.[2]

Despite the high detection rate of the NHSBSP (31.4% of all diagnosed breast cancers in England are picked up through screening[3]), interval breast cancers remain a problem. These are cancers detected symptomatically between screening rounds, and have a worse prognosis than screen detected cancers. It has also been demonstrated that up to 25% of cancers presenting symptomatically in the interval between screens had a detectable abnormality which was missed at screening.[4-6]

Using AI to replace some of the work of human readers has the potential to address both these challenges, relieving some of the radiology burden while enhancing cancer detection rates and outcomes. It is expected that using AI will result in earlier detection of some cancers and improve outcomes. However, implementing the wrong AI system, pathway or test threshold could worsen these issues, for example by decreasing cancer detection rates so more cancers are missed, or increasing recall to assessment rates thereby flooding breast clinics with healthy women who will have unnecessary diagnostic tests and blocking workflow.

1.2 Existing knowledge

Multiple retrospective studies assessing the diagnostic accuracy of breast AI algorithms have been conducted. Freeman et al. (2021)[7] conducted a systematic review of these studies and found that the evidence is highly inconsistent, of poor quality, and does not clarify where AI algorithms best fit in the current breast screening pathway. Three studies compared AI systems with the clinical decisions of the original radiologist, including 79, 910 women, and found that 34 out of the 36 AI systems evaluated were less accurate than a single radiologist and all were less accurate than consensus of two or more radiologists (double reading). Five smaller studies (1086 women, 520 cancers), with a high risk of bias and low generalisability to the clinical context, found that all five of the evaluated AI algorithms evaluated either as standalone to replace a reader or as a reader aid, were more accurate than a single radiologist reading a test set in controlled conditions. A more recent review[8] conducted a meta-analysis of studies comparing AI to human readers, finding that the accuracy of AI may match human reader performance, although studies using internally validated test sets were included in this analysis. One larger UK retrospective study (78,849 women) compared the performance of three AI algorithms as a triage for single human reading. This study found that using AI to triage could mean that only 35.0% to 55.6% of mammograms would require a single human read, with only up to 0.1% of screen-detected cancers missed. Furthermore, their modelling suggested if an additional 1% of cases marked as high-suspicion by the AI had been recalled, 4.6-8.2% of interval and 5.2-6.1% of subsequent-round cancers may have been identified in the prior screening round. This needs to be tested prospectively.[8]

Since then, a limited number of prospective studies have been published or are ongoing. A randomised controlled trial (MASAI) has begun in Sweden at four sites comparing AI-supported screening using a single AI algorithm (Transpara version 1.70, intervention group) to standard double reading without AI (control group).[9,

[10] The trial showed an increased cancer detection rate of 1.4/1000 (29% more cancers) compared to double human reading, however this is accompanied by an 8% increase in recall rate.[10] ScreenTrustCAD, another study conducted in Sweden, compared 1) initial reading performed by one radiologist plus AI (Insight MMG AI system version 1.1.6) to double reading with two radiologists, and 2) single reading by AI compared to double reading by two radiologists.[11] The study found that the cancer detection rates for both AI roles was non-inferior to double reading at a non-inferiority margin of 15% relative reduction. In a Danish time-series analysis before and after the implementation of AI, cancer detection rates improved (7/1000 to 8.2/1000), while the false positive rate fell (2.4% to 1.6%).[12] However the Danish study only allowed their experienced readers to participate in this trial raising questions on how AI will affect the performance of less experienced readers.

In the UK, three prospective studies are being conducted to evaluate one AI system (Mia, Kheiron Medical Technologies) for different workflows in the screening pathway. All three of these studies are small, single centre and only evaluate parts of the screening pathway. ARIES was a retrospective study with Hologic mammograms across diverse UK regions: Southeast, West and North. GEMINI was a prospective non-randomised, non-interventional study where participants still received standard double reading, to assess use of Mia in dual roles – as an additional reader to estimate workload savings and improved early cancer detection. Data on the downstream impact of AI and how it would perform as part of a real-time NBSS workflow could not be collected. The aim of the third study, LIBRA, was slightly different in that it was a prospective feasibility study to evaluate the impact of adding Mia to standard double-reading on arbitration rates (i.e. as a third reader). A fourth small prospective non-interventional study, AIMS, which is evaluating a different AI system (developed by Google Health) is ongoing at 2 hospital sites in London where it runs in silent mode (not impacting any care pathway) to assess technical feasibility and identify workflow adaptations that may be required.

1.3 Need for a trial

Population breast screening using mammography is non-invasive, widely accepted, and inexpensive compared to other imaging modalities. However, it is resource intensive, as each mammogram requires at least two readers. This is becoming increasingly problematic in the UK due to the radiology workforce crisis. Artificial Intelligence (AI) for breast screening could replace some of the work of human readers, improve earlier cancer detection and outcomes, and create a more robust infrastructure for delivering screening at lower cost, freeing up expert staff for more complex tasks. There are a number of different potential roles for AI in breast screening, and it is currently unclear from the evidence how AI should be implemented to best support the NHS breast screening programme.

Retrospective studies do not provide any insight into the acceptance of AI by human readers and the impact of AI on reader behaviour and consequently recall rates with AI in a real-time NHSBSP system workflow. The three small prospective interventional studies assessing AI for breast screening in the UK, do not address all the issues. The LIBRA study encountered serious workload and workflow issues with too many false positives. Most UK studies are non-interventional and test one AI at once, or at a small number of screening sites which limits the evidence on generalisability, real time savings in routine care, and the downstream impacts of AI. There are also concerns about the impact of reducing the need for a second human reader as this may increase the number of cases needing further discussion (increasing arbitration rates), which would increase the workload for those under the greatest capacity pressures currently. There are a number of commercially available AI tools which differ and have different performances with different mammography acquisition systems. Not all AI tools and AI settings are the same and an evaluation of the various possible combinations with different acquisition systems is needed. Furthermore, it is unclear whether the demographics of the women screened, including breast density and ethnicity, or their amount of breast tissue have an impact on AI performance. In a study of women screened with 3D mammography in the US, AI performance was impacted by women's demographics, with black women and women with dense breasts most

disadvantaged.[13] There is also limited information on the impact of AI on interval cancer rates, cancers which are diagnosed when women develop symptoms between screening cycles. Consideration of interval cancers is crucial to understanding the potential health impact of AI and to reassure those providing screening programmes about the safety of substituting AI for human readers.[14] The quality of prospective studies, in particular the MASAI trial, from other countries is much better but the generalisability of these results to the UK setting is of concern. For example, Sweden has very different standards for breast screening (including different screening intervals [two-yearly], age ranges and qualification requirements for readers) and much lower recall rates. The studied AI tool (Screenpoint Transpara 1.7.0) was previously trained on 200,000 women's images with over 10,000 annotated cancers from 10 countries. The MASAI trial tested AI on a population which differs in demographics to the UK screening population, and with mammograms generated by a single mammography imaging system (GE Healthcare). In the UK, there is a wide variety of imaging systems,[15] which produce mammograms with varying image characteristics. AI in the other Swedish trial and the Danish study read Hologic mammograms only. AI performance can substantially worsen in response to minor changes in the mammography images. One study showed that the AI recall rate tripled following a software change in the acquisition system which did not negatively impact human readers' performance.[16] Subtle degradation of AI performance over time can also occur.[17] AI performance does not always generalise to different settings, requiring recalibration or retraining of the algorithm with new data.[18] Finally, the best way to monitor AI performance for clinical safety and flag either a sudden or more gradual negative change to its performance is unclear.

To address this, we will undertake the first randomised multi-vendor mammography and multi-AI algorithm trial in the world to assess whether AI can be used safely and sustainably in UK clinical practice to benefit all women attending screening. We will compare two AI use scenarios with standard human double reading to evaluate the impact of adding AI to the NHSBSP. We will assess which AI tool is best suited to which mammography machine.

2. TRIAL DESIGN

Multi-centre, pragmatic, cluster, randomised controlled trial (RCT) allocated 1:1:1 ratio across control (current double reading pathway) and intervention arms 1) replacing the second reader with AI, and 2) using AI for triage. Cluster randomisation will be at the clinic day level including 660,000 women across approximately 30 screening centres with opt-out consent. Each centre will undertake all three trial arms with one compatible, approved AI tool randomly allocated to be used at the centre.

A nested test accuracy study comparing AI systems from participating vendors will be undertaken towards the end of the study to establish which AI tool performs best with each of the mammography vendors.

A qualitative study exploring women's reasons for opt-out and acceptability of AI in breast screening will be undertaken in a small number of women.

A health economic evaluation will be undertaken to evaluate the cost-effectiveness of implementing different AI systems and AI pathways and includes (a) a micro costing study to calculate the resource use and cost of using AI to read mammograms; and (b) a qualitative study to identify barriers and facilitators to the introduction of AI which will inform the health economic model.

Target population and setting:
Women attending NHSBSP at participating NHSBSP centres in England will be studied, and the feasibility of Scotland, Wales and Northern Ireland joining will be explored.

Follow-up:
Women will be followed up until their screening episode is closed and then for a further four years through NHSBSP and cancer registries to ascertain if they develop a cancer in the future. The initial follow-up records whether or not the mammogram was normal, if an abnormality was detected by AI or by either of the readers, the arbitration outcome, the recall decision and the results from further assessment, if relevant, with histopathological confirmation of cancers where possible.

Interval cancers (3/1000) mostly occur 24-36 months after screening. It is important to demonstrate a reduction in interval cancers to prove that any additional cancers detected by the technology are not over-diagnoses (cancers that would not have been found in a woman's lifetime had the screening test not been undertaken). Robust interval cancer and next screening round detected cancer data requires four additional years' follow-up (3-year screening round and 1-year updating routine data sources). Analysis of interval cancers and next screening round detected cancers is subject to receiving further funding. Extended follow-up beyond these four additional years would enable analysis of longer-term outcomes and evaluation of different AI systems. Extended follow-up would be subject to receiving further funding.

3. AIMS AND OBJECTIVES

Aim: To assess whether AI can be used safely, sustainably, and equitably in the UK National Breast Screening Programme.

Objectives:

- To integrate AI into the NHSBSP to measure the difference in clinical and cost-effectiveness between three study arms: replacing the second reader with AI (intervention 1); using AI for triage (intervention 2); and the current double reading pathway (control)
- To understand differences in accuracy by AI vendor, mammographic acquisition system and by population subgroup (e.g. age, ethnicity, breast density, socioeconomic status) in a nested test accuracy study
- To develop a comprehensive post-market surveillance methodology for national rollout
- To monitor barriers to participation in women opting out
- To calculate the cost of the current pathway using two readers compared with replacing the second reader with AI or using AI for triage from the perspective of the NHS

4. OUTCOME MEASURES

Primary outcomes:

1. Cancer detection rate (CDR) (=cancers detected/women screened)
2. Recall rate (=women recalled/women screened)

Secondary outcomes:

3. Arbitration rate (=women requiring additional read/women screened)
4. Interval cancer rate (=interval cancers within 3 years/women screened) and next screen CDR
5. Characteristics of cancers detected (size, type and grade)

6. Outcomes 1-5 by population group (age, ethnicity, socioeconomic status (SES)) to explore inequalities and by breast density
7. AI failure rates and reasons
8. Atypia detection (incidental findings)
9. Workload (e.g. reading volume, arbitration volume, recall to assessment volume)
10. Reader behaviour and performance (by e.g. AI vendor, AI threshold, batch length, time on task, reader experience and reading volume)
11. Opt-out rates and reasons for opting out as a measure of acceptability in screening-eligible women (where provided by women)
12. Cost to the NHS of the three pathways
13. Cost-effectiveness by technology and population subgroup, and impact on NHS capacity
14. Test accuracy of AI from a nested test accuracy study overall and by population subgroup (e.g. age, ethnicity, breast density, socioeconomic status)

5. PARTICIPANT SELECTION AND ELIGIBILITY

All women attending NHSBSP at participating NHSBSP centres in England (and potentially Scotland, Wales and Northern Ireland) will be eligible and will be contacted via the routine breast screening invitation letter (or NHS app notification / SMS message) and informed about the trial via a link to the trial website.

5.1 Inclusion criteria

- Women attending NHSBSP breast screening at participating centres during the trial recruitment period, typically screening target age 50-70 years but pragmatically include all screened women even if outside target age range (for example older women are allowed to self-refer every three years)
- Women whose mammograms cannot be read by AI due to test failure (for instance due to not having standard 2D digital four-view mammograms acquired) will be included in the main trial intention to treat analysis comparing the trial arms. They will receive standard care (2 human readers) regardless of trial arm. This analysis will deliver a comparison of the AI pathway as it would actually be implemented for all women with current standard practice

5.2 Exclusion criteria

- Women being screened in the moderate and high-risk screening service (separate to NHSBSP)
- Women being screened following the NHSBSP very high-risk protocol (separate protocol to population risk women: annual mammography/MRI)
- Women who have opted out from participation in the study
- Women who are not within the manufacturer defined intended purpose (which forms the basis of the AI tool's regulatory status) of the AI tool in use at their participating centre. Specific exclusions vary between AI tools, but this may include women with breast implants or a pacemaker.

5.3 Active dissent

The EDITH trial will use an active dissent model for women to opt out of having their mammograms read by AI. This will support recruitment of large numbers, including women from all ethnic groups and all social economic status, allowing the study to be inclusive. The mammograms of women who opted out will not be read by AI but will revert to the standard of readings by two human readers.

In the absence of active consent, we also need to consider national data opt-outs (NDOO). Under guidance from the Confidentiality Advisory Group and Health Research Authority, women who have enrolled in the NDOO will be included in the aspects of trial delivery directly related to their clinical care; they will be randomised and may have their images read by AI. However, the woman's screening data will not be collected for the purposes of the research as part of the intervention or control arm. All women, including those enrolled in the NDOO, can register a study-specific opt out, which provides them with the choice to opt out of having their images read by AI.

This approach is permissible because randomisation is at the clinic day level therefore, at participating centres for the duration of the trial, the trial pathway becomes the standard care for all women invited to screening. There would be a risk, by completely excluding women who have enrolled in the NDOO, that women may be deprived access to enhanced care by being denied the option to have their images read by AI. They would be separated into own reading batches which would reveal the woman's NDOO choice to medical care team. It is known that radiologists reading behaviour is linked to batch length and so presents further risks to care.

We have obtained advice from the NHS Breast Screening Research, Innovation and Development Advisory Committee (RIDAC), and will obtain approvals from the Health Research Authority (HRA), NHS Research Ethics Committee and CAG prior to study start. Public facing material containing information of the trial (trial website, poster, invitation letter) and opt out information are co-developed by the EDITH trial PPIE representatives and PPIE researcher.

Justification for active dissent:

- Research has been published showing that some AI systems are as accurate as one human reader, and we will only include AI systems with both published data demonstrating accuracy and that pass our trial benchmark.
- All examinations will be read by at least one expert human reader.
- The PPIE group are supportive and advocate 'opt-out' rather than 'opt-in with individual consent' so that the trial is available to a wider group of women.
- An opt-out consent model has been chosen because gaining full consent is prohibitively expensive, and all clinical decisions are still made by NHS experts.
- By waiving active consent for the study, the sample obtained is more likely to be representative of a real-world screening population and a better evaluation of the behavioural impact and benefits that AI may have in real-world deployments.
- In all intervention arms human readers work with AI software and the responsibility of decision-making is with human readers. In intervention arm 1 (AI is second reader), mammograms are read by one human reader and by AI. In intervention arm 2 (AI triage to human reading) mammograms that received a low likelihood score from AI read will be read by one human reader. Arbitration practice will follow standard of care with decisions made by human readers.
- Women attending screening are generally not aware of the processes that follow the taking of the mammograms, particularly how many readers will assess the mammograms. An unpublished survey of ~3,400 women aged 45-70 years has found that only 6% of women knew that mammograms were usually looked at by at least two readers [personal communication Prof Jo Waller]. Focus groups with PPIE confirms that women are happy for the use of AI in breast screening as long as their mammograms have been read by at least one human reader and the use of AI does not delay results being sent to women [personal communication Prof Janet Dunn].

- There are precedents of using active dissent models in breast cancer screening trials (LIBRA and GEMINI) and cervical cancer screening trials.[19]

Opt-out prior to mammogram being read:

There will be two opportunities for women to opt out of having their mammograms read by AI and their data being used for the research:

1. Prior to the screening appointment by telephone: Women invited to screening will receive the invitation letter signposting them to trial information including how to opt out by contacting their screening unit. The text added to the invitation letter may be: "Our screening centre is testing how Artificial Intelligence (AI) can help trained experts to look for cancer in mammograms. For more information on the study, if you wish to opt out, or if you have registered with the National Data Opt-Out please visit https://warwick.ac.uk/EDITH." or similar subject to local adaptation by clinical team and patient representatives. With the introduction of the "ping and book" system (message on the NHS App or SMS message) we will request a link to the trial web site with information about the trial and how to opt out.
2. During the screening appointment in person: Posters with trial information and information on how to opt out will be displayed within breast screening clinics. Women will be able to ask staff any questions they may have at the time of the screening appointment. Women will be able to opt out of having their mammograms read by AI by speaking to a member of staff present on the mobile van or screening centre.

Opt-out decisions will be recorded on the NHSBSP systems by screening centre staff and women will receive double human reading as standard of care. The woman's screening data will not be collected for the research as part of the intervention or control arm.

A woman's decision to opt-out without giving reason will be respected but reasons will be collected if participants wish to disclose this in the qualitative sub-study (see below).

Opt-out after mammogram has been read:

Once a woman's mammogram has been read by the AI tool it is not possible to undo this action. The result will be recorded in the woman's medical record (NBSS) and it is not permitted to delete information from a medical record. However, after their mammograms have been read women can still opt out of having their data used in the analysis. Women who contact the screening programme and wish to leave the trial after their mammograms have been read will have their decision recorded on the NHSBSP systems. The NHSBSP systems will be queried once at the end of the trial for a list of withdrawals and their data will be excluded from analysis and deleted from the trial dataset.

Opt-out after analysis and publication:

After the trial data has been analysed and the results published it will not be possible to delete a woman's data because trial data is required to be retained to ensure the research is reliable, reproducible and compliant with legal and ethical rules.

Invitation to take part in a qualitative study to explore reasons for opt-out:

Women who decide to opt out of having their mammograms read by AI, will be offered the opportunity to provide reasons for opting out a) on an online proforma and b) by taking part in a qualitative interview which

aims to explore reasons of opt out and attitudes towards AI. Staff recording the decision to opt out (over telephone or in person) will invite women to complete an anonymous online proforma collecting reasons for opting out and to take part in a qualitative interview study by providing interested women with the link to the survey and a trial email address to contact the study team for further information and consent forms. The information provided will describe the option of an interview if they would like to talk about their decision to opt out. Any women who are interested in taking part in the qualitative study can contact the study team directly to receive further information and a consent form for participation.

6. RANDOMISATION AND INTERVENTION

6.1 Randomisation

Clusters (clinic days) will be randomised at site level. Randomisation of clinic days takes place automatically within the NHSBSP systems after clinic days are closed.

All women within the cluster (anticipated median 36 women) will be included in the trial unless they opted out or meet study exclusion criteria.

6.2 Trial intervention

Current pathway is double human reading where two experts (radiologists, radiography advanced practitioners or other qualified experts) independently examine each woman's mammograms and decide whether to recall her for further tests. All women recalled by one reader (and at some centres consensus recalls as well) have their mammograms reviewed by a third reader or group of readers by arbitration to make a final recall decision. As part of the trial, centres will be asked, if possible, to arbitrate all recalls (discordant and consensus recalls) in all trial arms including the control arm. This is designed to reduce unnecessary recalls to assessment and optimise centre capacity. This may be adapted to fit local centre workflows and requests within the pragmatic trial design. Recall after arbitration follows the standard pathway.

The control arm consists of standard double human reading with arbitration.

Two intervention pathways were chosen based on published trial and test accuracy results, studies of human behaviour, and a consultation meeting at the Royal College of Radiology in September 2023 with 95 experts, including people from 25 screening centres and four PPI representatives.

In intervention pathway 1, the second reader is replaced by AI and, if technically feasible, the first reader makes their decision blinded to AI results and AI image annotations.

In intervention pathway 2, AI triages mammograms into:

- "low likelihood of cancer", to be read by a single human reader, or
- "medium to high likelihood of cancer", to be read by two human readers, or
- "very high likelihood of cancer" to be referred to arbitration, in order to operationalise this within NHSBSP systems the mammograms may be required to be read by one or two human readers before referral to arbitration.

Women with a very high AI score (top 1%) are recommended to be referred to arbitration regardless of human decision. If technically feasible this referral of women with a very high AI score to arbitration will be automated, but if not then readers will be recommended to refer to arbitration and communicate their own clinical opinion alongside that of the AI result to the arbitration reader(s).

In intervention arm 2 the readers have access to the AI result. In intervention arm 2 the readers have access to the AI result and AI image annotations for women with a risk score above a certain threshold. The trial aims to take a similar approach to MASAI trial, where AI image annotations were available to readers for women with a risk score above 70%. But this specific threshold will be adapted in collaboration with each AI vendor to align with their device's intended purpose.

Arbitration readers have access to AI results and images annotated by AI in both intervention arms 1 and 2.

Where supported by the AI tool, one or more prior mammograms will be used by the AI as part of intervention pathways 1 and 2.

6.3 Pre-trial phase

AI independent benchmarking

AI vendors will be eligible for the trial if they meet the benchmark performance, have lesion location software and can triage cases into likelihood of cancer categories, demonstrate robust processing and be integrated into the NHS PACS (Picture Archiving and Communication System - a medical imaging technology which provides storage and access to images for transmission, review and archiving patient information and images) that meets study timings. Prior to deployment in the study, approved UK curated screening datasets will be used to benchmark AI tool performance. Images will be previously unseen by AI vendors and will be made up of images from a range of mammography vendors and include screen detected, interval and next round cancers. AI performance will be analysed against these validation datasets in a retrospective manner to ensure each proposed AI system meets the required quality standards for use in the trial.

AI implementation

Where possible within the confines of a pragmatic trial and technical limitations, the AI tool will be randomly matched to a site with compatible mammographic equipment. For AI tools without preset operating points for recall by the AI company, we will set thresholds for each acquisition software using mammograms from a recent software version as a pragmatic approach. The AI tool will be tuned to ensure an appropriate recall rate/false positive rate is achieved for each mammographic acquisition vendor. This threshold will be taken forward to the next stage. For AI tools with fixed, validated operating points for recall that received appropriate market approval, the vendor specific recall threshold will be taken forward to the next stage.

Workflow testing at each site of the prospective AI reading will start in background shadow mode to check workflow and identify safety issues. No action will be taken by the readers based on the AI results during shadow mode. Once all approvals have been received and end-to-end testing has been completed, sites will begin live recruitment. During the live trial, ongoing real-time monitoring will detect major changes in AI recall rates considering background variation/noise. If AI performance varies outside recommended limits, the Clinical Safety Monitoring Committee (CSMC) will consider pausing the trial at that site and either allow the AI vendor to re-threshold or switch the centre to a different AI vendor. Preplanned safety measures include: at 1 month after live trial implementation an end-to-end check will be undertaken to ensure all women are accounted for; at 7,000 women after live trial implementation, site specific data will be reviewed and if necessary used to fine tune the AI threshold to the desired recall rate.

Interim two-arm phase

An objective of the trial is to integrate AI into the NHSBSP to enable measurement of the difference in clinical and cost-effectiveness between three study arms. AI tools and AI settings differ and in order to be inclusive of all AI systems, that meet the required quality standards for use in the trial, there are technical limitations to

be worked through that vary between AI manufacturers. While these technical constraints are worked through with each AI vendor, implementation of AI at site may begin with randomisation between two study arms initially, either:

- (1) current double reading pathway (control) and (2) double reading pathway where readers have access to AI results/annotations (intervention), or
- (1) current double reading pathway (control) and (2) using AI for triage. With the triage arm as described in section 6.2.

The duration of this interim two-arm phase will vary between AI vendors and participating sites, due to the specific nature of the technical constraints to be addressed. This pragmatic approach will enable readers at participating sites to become familiar with mammography AI tools.

Updates to AI technology

Updates to AI technology will be allowed during the trial to pragmatically reflect practice. Updated versions must pass the test set performance standards and be CE/UKCA marked (where required by the Medical Device Regulations) before implementation. New AI vendors can join the trial in advance of platform procurement if they meet the requirements set out above. New vendors can participate in the nested test accuracy study at the end of the study with less stringent criteria provided that funding is available, and it is technically possible.

6.4 Site staff training

Screening centre staff

Screening centre staff will be informed about the trial and trained to answer questions about the AI trial and how to document opt outs. They will be provided with standard responses to questions about the trial and women who wish to opt out. It is likely women will ask about the trial at two points. Firstly, on receiving their invitation letter asking them to phone to make a screening appointment, office staff may be asked questions. Secondly, during the screening mammogram appointment, radiographers, radiographic diagnostic officers (RDO), radiologists and breast clinicians may be asked about the trial. -Staff will receive training to ensure they are fully informed about the trial and will be given a Frequently Asked Questions (FAQ) sheet and a trial contact to discuss any further questions. Information and training will be delivered at the site initiation visit. Information about the trial is already being disseminated at national breast screening meetings.

Mammogram readers

It is anticipated that most readers will be unfamiliar with mammography AI tools. There will be online or face-to-face training provided to explain about the trial and how to use the AI tool. We will curate and make available a training set to readers at participating trial site. The aim is for readers to understand how AI decision support affects their decision-making process and get an idea of how the AI algorithms score challenging cases. The training will include guidance on handling very high AI scores and understanding false prompts.

6.5 Compliance/contamination

Adherence to the reading protocols with and without AI will be monitored and where possible controlled centrally. In the triage arm, IT systems enforce AI as first reader and controls the timing of availability of prompts, with programmed single or double human reading and arbitration.

In the AI as second reader arm, the NHSBSP systems are programmed that AI is the second reader and AI prompts will not be available to study readers but will be made available for arbitration.

In both arms, when AI is available to readers, its use will not be enforced. Whether readers use it or not is part of this pragmatic trial.

The use of the AI prompts will be documented by NHSBSP systems and examined at a case level as part of the assessment of the impact of AI on staff if the software change required to record this is feasible.

To ensure balance between trial arms and to ensure that the trial does not affect routine monitoring of reader quality we will work with the Director of Screening at each site to ensure readers do not disproportionately examine cases from one arm of the trial, using training and monitoring.

6.6 Blinding

Participating women attending screening are blinded to allocation. Blinding of readers is not appropriate in this pragmatic trial as the impact of incorporating AI on reader decisions is of interest. Non-blinding will not affect the integrity of the study.

6.7 Co-enrolment into other trials

EDITH is a large pragmatic trial and co-enrolment in other trials is expected. The EDITH team will work with other triallists to enable centres to take part in multiple trials, while ensuring there is not an impact on primary and secondary (interval cancer) outcomes, being particularly cautious on impacts that differ between trial arms.

Co-enrolment will be permitted on a trial-by-trial basis in consultation with the EDITH trial team. With the exception of trials of treatment, for which there is no requirement to seek agreement from the EDITH trial to co-enrol a participant in another trial.

6.8 End of trial

The end of the trial is defined as the date when all women enrolled into the trial will have received a recall/no recall decision and results from further assessment have been recorded if relevant. Interval cancer data and next screening round cancer data for enrolled women will be taken directly from NHSBSP after an additional four years' follow-up.

The trial will be stopped prematurely if:

- Mandated by the Ethics Committee
- Following recommendations from the Data Monitoring Committee (DMC)
- Funding for the trial ceases

The Research Ethics Committee will be notified in writing within 90 days when the trial has been concluded or within 15 days if terminated early.

7. METHODS AND ASSESSMENTS

7.1 Overview of delivery of intervention and data flow

The intervention in this trial is AI assistance for mammography readers in reading screening mammograms (see section 6.2). We will assess AI as: arm 1) a standalone tool to replace a second reader and arm 2) a triage tool to reduce volume of reading. The detailed data flow map (Appendix 1) illustrates the data flow. In brief: Women who have not opted out will have their mammograms read in line with the cluster randomisation.

Mammograms from women allocated to the control arm, women who opted out and women who are ineligible will receive standard human double reading with arbitration. Furthermore, in cases where AI failed, mammograms will be read by two human readers as in standard care.

Data from NHSBSP systems, PACS and the AI tools will be pseudonymised as it arrives at the continuous monitoring system (CMS) which processes data and sends pseudonymised, processed data for analysis to Warwick Clinical Trials Unit (WCTU) and for safety monitoring by the oversight committees. Data flow will be orchestrated by a platform for secure and accurate delivery of data between NHSBSP systems, AI tools and the CMS.

7.2 Qualitative assessments – nested study

The trial includes a qualitative sub-study to understand reasons for opt-out and broader attitudes to the use of AI in breast screening. A second qualitative sub-study of adverse psychological responses to participation in the trial is currently being developed.

Understanding reasons for opt-out and broader attitudes to the use of AI in breast screening

Aim: To explore women's reasons for opting out and barriers to participation in AI-assisted screening and concerns to inform future communication materials.

Good uptake of breast screening is key to the success of the programme. Demographic differences in participation can exacerbate social inequalities in health. It is, therefore, essential to consider the impact of changes to existing screening programmes (like the introduction of AI-assisted screening) on patterns of participation. Although non-participation in screening is different from opting out of research, understanding reasons for opting out of the EDITH trial may shed light on potential concerns among screening eligible women about AI, allowing us to develop appropriate information and support for future implementation.

All women who opt out of AI assisted reading, will be asked if they are willing to share their reasons for opting out using a short online proforma. Women opting out by phone will be directed to the EDITH trial website where they can find information and a link to a fully anonymous online survey hosted by Queen Mary University London (QMUL) by the call handler. Those opting out in-person at their appointment will be handed paper-based information including a survey URL link, a QR code linking to the survey and an email address for the QMUL study team. In both cases, completion will be entirely optional.

In addition to being asked for brief information about their reason for opting out, women will be given information about opting *in* to a semi-structured qualitative interview (circa n=30) to explore barriers to AI assisted mammogram assessment. Women will be provided with an email address/telephone number with which to get in touch with the researchers at QMUL. In response, potential interview participants will be contacted by phone or email by the QMUL team and will be sent a separate participant information sheet and consent form for the interview sub-study. Based on our previous experience as part of the LIBRA study, we anticipate very low rates of consent to interview so we are unlikely to be able to carry out purposive sampling, but where possible, we will try to ensure we include women from a range of socioeconomic and ethnic backgrounds and a range of geographical areas. Women will be asked to complete a short demographic and screening history questionnaire if they are interested in participating, to allow purposive sampling where possible.

Interviews will be carried out by phone or on Teams and will be audio recorded. They will be semi-structured and guided by a topic guide developed based on previous literature (e.g. Gatting et al).[20] They will cover women's previous breast screening history and experience, their attitudes to AI in the healthcare context, and their attitudes to taking part in medical research. The interviewer will be trained in qualitative methods and

will probe women's understanding of the EDITH trial and reasons for opting out in a non-judgmental way, respecting their right not to take part. Interviews are expected to last 30-60 minutes and participants will be given a shopping voucher to acknowledge their time and input.

Interviews will be transcribed verbatim by a transcription agency and will be analysed using Framework Analysis to identify key themes.[21]

Exploring potential adverse psychological responses to participation in the trial

Breast screening participation can have negative psychological consequences, especially related to false positive results and associated investigations.[22] Little is known about how the introduction of AI-assisted screening might affect responses to abnormal screening results. Given that false positive results may reduce attendance at the next round of screening, understanding the possible impact of AI on psychological outcomes is important.

A study of adverse psychological responses to participation in the trial is currently being developed with the PPIE group. Once developed this will be submitted for the relevant approvals (e.g. REC, HRA) as a protocol amendment before any activity commences on this sub-study.

7.3 Test accuracy of AI systems – nested study

Aim: To compare AI system's accuracy by ethnicity, age, socioeconomic status and breast density

The performance of approved AI systems will be directly compared using the same cases from a national contemporaneous dataset of mammograms and clinical outcomes from trial sites. Towards the end of the trial, AI systems will re-read mammograms of a subset of the 660,000 women using their latest software version. AI decisions of recall / no recall will be compared to clinical outcomes in a retrospective study. Test accuracy performance metrics including sensitivity and specificity will be determined and reported by AI system, mammography vendor, and patient characteristics. Age, ethnicity and socioeconomic status will be extracted from NHSBSP systems as available. If ethnicity data is insufficient (e.g. less than 60% coverage) or unavailable, alternative external resources will be explored. Breast density will be provided by AI systems with the capability on processed images unless further funding is secured to use unprocessed images. A pragmatic approach to density estimation will be employed, taking into account system-generated outputs, and standardised across mammogram systems using available software.

8. ADVERSE EVENT MANAGEMENT

An Adverse Event (AE) is defined as any untoward medical occurrence in a randomised trial participant and which does not necessarily have a causal relationship with their involvement in the trial. A Serious Adverse Event (SAE) is an AE that fulfils one or more of the following criteria:

- Results in death
- Is immediately life-threatening
- Requires hospitalisation or prolongation of existing hospitalisation
- Results in persistent or significant disability or incapacity
- Is a congenital abnormality or birth defect
- Is an important medical condition

The intervention in this trial is AI assistance for mammography readers in reading screening mammograms. In the intervention arms human readers work with AI software and the responsibility of decision-making is with human readers. All clinical interventions received by trial participants are identical to those given in normal screening practice, for which there is extensive safety data already available. All Adverse Events (AEs) are therefore expected and, for the purposes of this trial, further adverse event data is not collected as it is not required for trial analysis. There is no requirement for expedited of any Serious Adverse Events (SAEs) in the trial.

9. STATISTICAL CONSIDERATIONS

9.1 Power and sample size

The primary comparison of the control arm against each AI pathway requires confirmation of superiority in cancer detection and non-inferiority of recall rate. Thus, as this three-arm study has independent hypotheses comparing intervention 1 with control and intervention 2 with control and requires both co-primary outcomes to be significant for a single hypothesis, multiple testing is not required.[23] With an expected cancer detection rate of 8.8/1000 for the current pathway,[24] 193,619 women in each arm are required to detect a difference of 1/1000 screened (similar to Swedish trial[9]) with a 5% two-sided significance level and 90% power. In addition, with an expected recall rate of 3.8% in the UK[24] for the control arm, a minimum of 192,055 women in each trial arm are required to demonstrate that the recall rate from using AI in the screening pathway compared to current practice is not increased by more than an absolute difference of 0.2% (agreed by PPIE representatives and researchers) with a 2.5% one-sided significance level and 90% power. This would equate to a recall rate in the AI screening pathway being no more than 4%. To allow for clustering at the clinic day level, with an average cluster size 36 and an intercluster correlation of 0.002, the design effect is 1.07. Thus, randomising a total sample size of 660,000 women (220,000 to each of the 3 trial arms) will account for clustering and allow for 5% loss to follow up.

The MASAI Swedish trial[9] required 100,000 women to demonstrate non-inferiority to within a 20% non-inferiority margin with 80% power for interval cancers and 80,000 women for a cancer detection rate of no worse than 3 per 1000 from the current rate of 5 per 1000 in a two arm study with 80% power. Thus, considerably more women are required in our study to have the required higher power of 90% to reliably detect the smaller observed difference of 1 per 1000 and, because there are three arms, to efficiently be able to answer two questions instead of one question in a two-arm study.

9.2 Statistical analysis

Primary and secondary analyses

The primary comparisons are intervention 1 compared to control, and intervention 2 compared to control, designed to answer whether each intervention is more clinically effective than current practice. All other comparisons are secondary. Cancer detection and recall rates will be analysed separately using hierarchical models. To account for clustering, hierarchical models will be considered for woman screened (level 1) nested in a clinic day (level 2) within a centre (level 3). Random effects will be considered for centre and clinic day, and fixed effects for women's characteristics (age, ethnicity, socioeconomic status, breast density and previous attendance) in a similar manner to our previous trial in 46 hospitals and 1.1 million women.[25] The intervention pathway effect will be estimated from the model along with 95% confidence intervals (CI). Similar analyses will be undertaken to assess the arbitration rate and interval cancers. A comparison of the results from the first reader in intervention arm 1 without using AI and in intervention arm 2 with the aid of AI will be used to assess the impact of AI on their decision making taking into consideration reader experience and reading volume if available. An intention to treat analysis will be performed. Descriptive statistics using means

and standard deviations, medians and interquartile ranges or frequency and percentages depending on the distribution of the data will be used to describe the characteristics of any cancer detected (type, size, grade), incidental findings, and workforce utilisation (numbers read, arbitrated and recalled for additional tests or procedures, and time taken for reading).

Nested test accuracy study

This is a retrospective sub-study of mammograms from the main study taking place towards the end of the trial. This study has no impact on patient care pathway, as it does not involve any further contact with participants. As there is no comparison of AI systems in the main trial, this sub-study allows comparison and investigation of how the AI systems perform.

Outcomes and analysis

1. Compare the different AI systems using accuracy (diagnostic outcomes: sensitivity, specificity)

Comparison using multi-level multivariable regression model (random effects: level 1 woman, level 2 image, level 3 AI system. fixed variables: AI system). Descriptive estimates will be reported by each AI system (sensitivity, specificity) with 95% CI.

2. Investigate and present accuracy results visually according to single characteristics (participants, hardware and software) and participant subgroups of interest

Descriptive presentation of accuracy results with 95% CI will use visual presentation with Forest plots (sensitivity, specificity) and tabulation by AI system and according to characteristics (participants, hardware and software). For descriptive diagnostic outcomes: sensitivity, specificity, hypothetical population of participants in terms of true positives, false negatives, false positives, true negative.

3. Investigate and present descriptive summary of AI test failures

Descriptive presentation of AI test failures by AI, characteristics (participants, hardware and software) with subgroup analysis by type of failure (unable to complete AI reading, inconclusive readings). Categorical results will be expressed as: percentage, for continuous and count data; mode, median, IQR values, min, max.

Detailed description of analysis methods and selection of women's images will be in the Statistical Analysis Plan.

Selection of women's images

In this retrospective sub-study, we will include all women's images identified with a diagnosis of breast cancer and a subset of women's images without a diagnosis of breast cancer as part of the trial. Five women without a diagnosis of breast cancer will be randomly selected from the same clinic day, matched for each woman with a diagnosis of breast cancer. The number of women diagnosed with breast cancer is estimated as 8.8 of 1000 women recruited, so approximately 5,800 and 29,000 women's images will be selected with and without a diagnosis of breast cancer, respectively.

Source of covariate data

Covariate data will be sourced from the main trial. No additional data from participants will be collected.

10. HEALTH ECONOMIC EVALUATION

10.1 Cost of using AI to read mammograms

Aim: To calculate cost of the current pathway using two readers compared with replacing the second reader with AI or using AI for triage from the perspective of the NHS.

Methods: A micro-costing study will be used to identify the resource use and total cost (2024 price year associated with each AI-driven intervention as delivered in the trial (standalone tool; triage tool).[26] To identify the relevant resource use, AI-driven intervention pathways will be created using structured interviews with 8 experts. The sampling frame for these experts will include radiologists, administrators in the NHSBSP in the UK, NHS IT specialists, and representatives from the AI providers. A cost will then be produced for the current mammography reading pathway and for each of AI-driven intervention (standalone; triage) for each of the proposed five AI-systems.

10.2 Decision-analytic model

Aim: To evaluate the cost-effectiveness of implementing different AI systems and AI pathways into the UK National Breast Cancer Screening Programme

A decision-analytic model (NHS perspective; lifetime horizon; 2025 price-year) will be conceptualised in line with published recommendations for conducting economic evaluations of AI interventions.[27, 28] Strategies will be developed to represent each AI-driven intervention (standalone; triage) using each of the five available AI-systems compared with both the current screening programme and no screening. The decision-analytic model will adapt an existing validated model called MANC-RISKSCREEN which is an openly available, validated, discrete event simulation programmed in R.[29] The decision-analytic model will simulate a population of woman (age 50-years and over) living in the UK following the no screening or screening pathways, with or without the use of AI potentially delivered in two different strategies by five providers. The costs of each AI strategy will be estimated in the micro-costing study detailed above. Each AI strategy will also produce a health impact through its effect on the sensitivity and specificity of screening in detecting cancers. These parameters will be estimated as part of the clinical trial and used to adjust the probabilities used in the screening event contained in the model. The sensitivity of screening will be updated to reflect the relative sensitivity of each AI tool and strategy, taking into account the impact of breast density on this sensitivity. Differences in specificity across AI interventions and providers will be reflected in differing recall rates in the model. Higher recall rates result in additional numbers of follow-up appointments and biopsies for women without cancer leading to higher costs for the health system. In addition to the key parameters of AI cost, sensitivity, and specificity, all the input parameters in the MANC-RISK-SCREEN model will be updated to reflect the best currently available data. Sources for data include: published systematic reviews; de novo systematic reviews; trial data; the costing analysis. The code and data for AI-decision-analytic model will be made open access (using GitHub,[30]). The analysis will estimate mean (deterministic) values for healthcare costs, life-years and quality-adjusted life years (QALYs) for each AI-driven intervention (standalone; triage) and AI-system; current NHSBSP; no screening. A fully incremental analysis will be completed. The difference (incremental) healthcare costs and patient outcomes (life-years; QALYs) between each strategy (standalone; triage; five providers; compared with no screening current NHSBSP) will be calculated. Uncertainty in the data driving the observed calculated expected costs, life-years and QALYs will be measured using probabilistic sensitivity analysis together with one-way sensitivity analyses. Additional analysis will explore the cost-effectiveness of the AI strategies in pre-specified sub-groups of the population (such as by ethnic group; women with dense breasts).

10.3 Barriers affecting delivery of AI in NHSBSP for breast cancer

Aim: To identify barriers and facilitators to the introduction of AI in a National Breast Cancer Screening Programme

Following the work of Wright et al, an analysis will be conducted to take account of imperfect screening implementation with data sourced from the trial.[31, 32] Key healthcare system constraints linked to using AI-driven interventions in NHSBSP, as well as constraints in delivering the current NHSBSP, will be identified using published sources supported by qualitative interviews (using a sample of ~twelve experts in radiology, breast cancer, pathology).[33] Relevant individuals will be identified through known contacts and relevant publications of breast screening trials and snowball sampling. Structured online (e.g. MS Teams/Zoom) one-to-one interviews (~one-hour) will describe key healthcare system constraints potentially affecting the delivery of the AI-driven NHSBSP. The interviews will be transcribed, and data will be analysed using qualitative content analysis to identify themes and patterns.[33] Data analysis will characterise a list of key constraints subsequently structured by categories derived from the updated Consolidated Framework for Implementation Research (CFIR).[34] In addition to exploring the range of different barriers and facilitators to implementing AI-assisted breast cancer screening, the experts will also be asked which barriers or facilitators they believe will have the biggest impact on the health of patients and the costs of the programme. These will be consolidated into a shortlist of constraints to incorporate into the MANC-RISK-SCREEN model. The impact of these factors on the incremental cost-effectiveness ratio and population monetary net benefit of each AI strategy and manufacturer will be quantified. The impact on resource use, such as health professional reading time, in the screening programme will also be estimated and reported.

11. DATA MANAGEMENT

11.1 Data collection and management

Personal data collected during the trial will be handled and stored in accordance with the 2018 Data Protection Act and the UK GDPR. The legal basis for access to confidential medical records will be via section 251 approval from CAG.

NHSBSP will hold data on participants who attend for screening which will be used for the trial. Data will be collected directly from NBSS. A platform company will host the AI tools in a secure and trusted environment. Identifiable data and mammogram images from NHSBSP systems and PACs will flow to the platform and back again. This data will be identifiable because it is for direct care (similar to data being stored in the PACS system on the cloud). The platform will host a CMS de-identification tool which will automatically de-identify data points required by the externally hosted CMS and trial database. The CMS send pseudonymised data for analysis to the trial database at WCTU and safety monitoring by trial oversight committees. This centralised automated process will ensure consistency in the de-identification process and allow only pseudonymised data to be utilised in the CMS and trial database. Mammogram images with pseudonymised trial ID will be collected and stored in a secure cloud-based object storage or trusted research environment at Royal Surrey NHS Foundation Trust and University of Cambridge.

11.2 Database

The trial database will be developed by the Programming Team at WCTU and all specifications (i.e. data linkage, database variables, validation checks, screens) will be agreed between the programmer and appropriate trial staff.

11.3 Data storage

All essential documentation and trial records will be stored by WCTU in conformance with the applicable regulatory requirements and access to stored information will be restricted to authorised personnel.

Pseudonymised trial data will be stored in a trial database at WCTU, which will be developed by the WCTU Programming Team in accordance with regulatory requirements and best practice guidelines.

An encrypted trial lookup will be held in a secure cloud environment managed by the Scientific Computing team at the Royal Surrey NHS Foundation Trust, where identifiers will be stored in an encrypted form, against a pseudonymised trial ID to enable for future linkage (analysis of interval cancers/next round cancers).

Images with pseudonymised trial ID will be stored in a secure cloud-based object storage or trusted research environment managed by the Royal Surrey NHS Foundation Trust and University of Cambridge, and the AI and density measures for the nested test accuracy study applied there. AI results and density results will be sent to the University of Warwick to analyse the nested test accuracy study.

During the trial, NHS England will be developing an NHS research platform on which images from sites participating in the second half of the study will be stored together with their metadata.

Data related to the qualitative interviews of women who opted out will be stored at Queen Mary University London.

Data related to the qualitative interviews for the micro-costing study and the study identifying barriers to implementation of AI in NHSBSP will be stored at the University of Manchester.

11.4 Data access and quality assurance

Data will be entered on the trial database via NHSBSP data linkage. Periodically data will undergo checks to ensure consistency with remote data linkage and any additional data collected on trial Case Report Forms (CRFs).

11.5 Data shared with third parties

Images from the participants (screening mammograms), meta-data and outcomes will be rooted through the platform provider. Data will be held by the platform provider for a limited time period for operational reasons (maximum 30 days) and then permanently deleted. The images, meta-data and outcomes will not be held by the AI companies at any time.

At the end of the trial third parties may request access to the trial data and images. The trial will form a Data Access Committee (which will include the Chief Investigator(s) and TMG members) to consider requests. All requests must have the approval of the Data Access Committee. As data controller, University of Warwick is responsible for ensuring all sharing of data complies with University of Warwick policies, principles of Good Clinical Practice (GCP), the UK GDPR and the common law duty of confidentiality where they are applicable. Data will not be shared if the conditions of these are not met. If any images or meta data are released to a third party, full anonymisation will take place.

11.6 Archiving

Trial documentation, data and images will be archived for up to 20 years after completion of the trial. This is in accordance with Warwick SOPs which states a minimum retention period of five years. The justification for retaining for 20 years is because follow-up data collection (directly from NHSBSP) for interval cancer data and

next screening round cancer will continue for an additional four years after the trial end date, and beyond for longer-term outcomes.

Trial sites will be responsible for archiving of their Investigator Site File and any other trial records held by the site for 10 years after the completion of the trial.

12. TRIAL TIMETABLE AND MILESTONES

EDITH aims to recruit 660,000 women from circa 30 screening centres across the UK over 25 months. It is projected that sites will recruit on average 20,000 women per year and thus recruit an average of 1,667 women per month per site. A 9-month internal pilot from the start of recruitment on a commercial platform has been incorporated to assess site set up and recruitment.

Table 1 Progression criteria

Progression criteria	**Red**	**Amber**	**Green**
Trial recruitment (% threshold)	<60%	≥60%	≥100%
Number of sites opened	<4	4-8	≥9
Total number of participants recruited	<19,000	~24,000	≥40,000
Recruitment rate/ site/ month	<1000	≥1000	≥1667
Action to be taken	Urgent recovery plan, agree with TSC and HTA; consider stopping trial	Continue, consider adding additional sites	Continue, no action

Year 1: Software platform pilot development and testing if relevant, changes to NHSBSP systems, ethics approval, national information governance template, site setup, AI independent benchmarking; AI implementation silently in background for safety, workflow and failsafe checks

Year 2: Commence recruitment (recruitment projections for 30 sites with an average recruitment of 1,667 women per site per month) and prospective reading with AI

Year 3: Complete site setup

Year 4: Complete recruitment, extract and clean data, analyse and report

Year 8: Interval cancer analysis (further funding will be required)

13. STUDY RESPONSIBILITIES, OVERSIGHT AND MANAGEMENT

13.1 Sponsor and governance arrangements

University of Warwick and Cambridge University Hospitals NHS Foundation Trust will act as co-sponsors for this study. The trial will be conducted in full conformance with the principles of the Declaration of Helsinki and to Good Clinical Practice (GCP) guidelines. It will also comply with all applicable Warwick Standard Operating Procedures (SOPs).

Co-sponsor's responsibilities include, but are not limited to:

- Study set up and regulatory approval
- Reporting to the independent steering committee according to the study monitoring plan
- Ensuring a risk assessment has been carried out and adequate monitoring is in place.

13.2 Ethical approval and considerations

Ethical approvals

All required ethical approvals for the trial will be sought using the Integrated Research Application System. The trial will be conducted in accordance with all relevant regulations. The trial will seek REC, HRA and CAG section 251 approval.

Before enrolling patients into the trial, each trial site must ensure that the local conduct of the trial has the agreement of the relevant NHS Trust Research & Development (R&D) department. Sites will not be permitted to enrol patients into the trial until written confirmation of R&D agreement is received by the EDITH team at Warwick Clinical Trials Unit.

Ethical considerations

Active dissent / opt out: see section 5.3 for details.

Potential bias of AI tools by ethnicity: AI tools are developed on existing mammography datasets which may or may not have the same ethnicity representation as in the UK. In order to ensure that the AI tools perform equally well across all age groups, breast densities, mammography types, ethnicities and geographic areas this data will be recorded, and AI performance will be compared at the end of the trial against these variables.

Processing of identifiable data by platform company: At present platform companies process identifiable images within the constraints of information governance and data controller regulations. In this study we are exploring utilising identifiable images through the trial data flows via the platform company. Within the trial Digital Imaging and Communications in Medicine (DICOM) images and clinical data points from the National Breast Screening System (NBSS) from each site will be transferred to a third-party platform company. The platform company will host the AI tools in a trusted environment and the results returned to the NHSBSP and the local PACS system. The platform will host a de-identification tool which will automatically de-identify data points required by the externally hosted Continuous Monitoring System (CMS). Pseudonymised data in the CMS will be transferred to the trial database in Warwick. Although we could insist that images are deidentified our preferred option is to transfer identifiable images as we believe that it is safer to maintain the ID and remove additional steps of de- and re-identification. Retaining images and data that are processed by the AI tools in an identifiable format will avoid the large-scale and complex process of automatically pseudonymising DICOM and HL7 data at all participating sites and the associated risks of errors preventing the merging of the patients' records. We will seek section 251 support from CAG to allow an automated pseudonymisation layer to sit on the platform cloud to process the personal information for the purposes of pseudonymisation before ingestion into the CMS and trial database. This centralised automated process will ensure consistency in the de-identification process and allow only pseudonymised data to be utilised in the CMS and trial database.

13.3 Trial Registration

The trials will be registered with the International Standard Randomised Controlled Trial Number (ISRCTN) Register.

13.4 Notification of serious breaches to GCP and/or trial protocol

A "serious breach" is a breach which is likely to effect to a significant degree –

a) the safety or physical or mental integrity of the subjects of the trial; or
b) the scientific value of the trial

The sponsor will be notified immediately of any case where the above definition applies during the trial conduct phase.

13.5 Indemnity

NHS indemnity covers NHS staff, medical academic staff with honorary contracts, and those conducting the trial. NHS bodies carry this risk themselves or spread it through the Clinical Negligence Scheme for Trusts, which provides unlimited cover for this risk. The University of Warwick provides indemnity for any harm caused to participants by the design of the research protocol.

13.6 Administration

The trial co-ordination will be based at Warwick Medical School, University of Warwick.

13.7 Trial Management Group (TMG)

The Core Trial Management Group (CTMG), consisting of the project staff and co-investigators involved in the day-to-day running of the trial, will meet regularly throughout the project. Significant issues arising from management meetings will be referred to the other oversight committees as appropriate.

The coinvestigators, including PPI members, will sit on a Wider Trial Management group (WTMG) to provide oversight of the trial; this will allow joint applicants in different institutions to communicate and monitor progress of the project.

In accordance with Committee on Publication Ethics (COPE) guidance, any potential conflicts of interest will be declared on all publications.

13.8 Trial Steering Committee (TSC)

The trial will be guided by a group of respected and experienced personnel and trialists as well as at least one 'lay' representative. The TSC will have an independent Chairperson. Meetings will be held at regular intervals determined by need but not less than once a year.

The Steering Committee, in the development of this protocol and throughout the trial will take responsibility for:

- Major decisions such as a need to change the protocol for any reason
- Monitoring and supervising the progress of the trial
- Reviewing relevant information from other sources
- Considering recommendations from the DMC
- Informing and advising on all aspects of the trial

The full remit and responsibilities of the TSC will be documented in the Committee Charter which will be signed by all members.

13.9 Data Monitoring Committee (DMC)

An independent data and safety monitoring committee will meet at least annually and review confidential reports containing recruitment, protocol compliance, safety data and interim assessments of outcomes. The DMC will advise the TSC as to whether there is evidence or reason why the trial should be amended or terminated.

DMC meetings will also be attended by the Chief Investigator(s) and Trial Manager (for non-confidential parts of the meeting) and the Trial Statistician.

The full remit and responsibilities of the DMC will be documented in the Committee Charter which will be signed by all members.

13.10 Clinical Safety Monitoring Committee (CSMC)

The clinical safety monitoring committee will comprise a senior representative from the screening programmes from all involved nations to review site level safety data from the study at four-monthly intervals. They will flag any concerns directly to the triallists to either pause a site or change AI tool.

AI tool performance will be monitored on an on-going basis to ensure that the level of recall, plus monitoring the delivery of the screening programme (e.g. time for results, time to assessment), and specificity remain within agreed thresholds and that any changes to the data source data (e.g. change in modality, software version or protocol etc) do not alter or affect the AI tool's performance.

13.11 Essential documentation

A Trial Master File will be set up according to Warwick SOPs and held securely at Warwick Clinical Trials Unit who will provide Investigator Site Files to all recruiting centres involved in the trial.

13.12 Financial support

The trial has been funded by a grant from NIHR HTA Programme (NIHR162840).

13.13 Safeguarding researchers and research participants

The study will adhere to the University of Warwick's safeguarding policy whose principles comply with the NIHR's safeguarding guidance.

14. MONITORING, AUDIT AND INSPECTION

A Trial Monitoring Plan will be developed and agreed by the TMG and TSC based on the trial risk assessment which may include on site monitoring. The NHSBSP system will undergo required monitoring by the CSMC to ensure all sites are adhering to the national standards.

15. PATIENT AND PUBLIC INVOLVEMENT AND ENGAGEMENT (PPIE)

The funding bid was extensively supported from inception by a team of PPIE co-applicants with lived experience of cancer, a population perspective and relatives with cancer. They have actively contributed to all team discussions, including representation on each of the study subgroup teams, and drafting this document. We have supported expenses and remuneration for their time in line with INVOLVE and NIHR guidance. They

will be members of the wider trial management group, and their input will be critical to effective dissemination and implementation.

Consent was also discussed with the PPIE team, favouring opt-out rather than individual participant consent so that a wider group of participants are included in the trial and to avoid delays to the screening service that would result from the time needed to obtain individual consent. This needs to be managed alongside information about the trial made available to the wider public, translated into different languages in line with the current screening invitation. All patient facing materials will be reviewed and have input from the PPI team before release which will include a website on the Warwick CTU webpages with links to breast cancer charities and the screening service. The established PPIE group will be expanded to better reflect the diversity of the screening population, and a PPIE researcher will contribute to the research going forward and support and guide the trial.

16. DISSEMINATION AND PUBLICATION

The results of the trial will be reported first to trial collaborators. The main report will be drafted by the trial co-ordinating team, and shared with the Trial Steering Committee and Data Monitoring Committee before submission for publication.

The success of the trial implementation depends on the collaboration of clinicians, radiologists, site staff and researchers from across the UK. Equal credit will be given to those who have wholeheartedly collaborated in the trial.

The trial will be reported in accordance with the Consolidated Standards of Reporting Trials (CONSORT) guidelines (www.consort-statement.org).

PPIE is fundamental to this study. Our PPIE group will ensure that we maintain a trial website with access to publications alongside a plain English summary of results and other communication materials (such as infographics/stories/quotes) co-designed with our PPIE team. The website will be the same as that accessed when women are making an informed choice about whether or not to opt out, and women will be informed at that point that they can check back in the same place for results of the trial.

Peer reviewed publication alongside Plain English Summary and links to further materials on our website will be shared with UKNSC, NHS England, Integrated cancer boards and cancer alliances. The PPIE team will present at patient related conferences such as the Multinational Association of Supportive Care in Cancer (MASCC). We will present at national and international radiology and cancer conferences to engage breast screening clinicians.

Rapid dissemination will be critical, due to workforce crisis, rapidly changing AI technology and wider pressures to implement AI. In addition to publication and conference presentation we will work closely with the NHSBSP and the UKNSC who will advise the four UK Governments on adoption of AI.

17. REFERENCES

1. Perry N, Broeders M, de Wolf C, Törnberg S, Holland R, von Karsa L. European guidelines for quality assurance in breast cancer screening and diagnosis. Fourth edition—summary document. *Annals of Oncology*. 2008;19(4):614-22. doi: https://doi.org/10.1093/annonc/mdm481.

2. Mottram R, Knerr WL, Gallacher D, Fraser H, Al-Khudairy L, Ayorinde A, Williamson S, Nduka C, Uthman O, A, Johnson S, Tsertsvadze A, Stinton C, Taylor-Phillips S, Clarke A. Factors associated with attendance at screening for breast cancer: a systematic review and meta-analysis. *BMJ Open*. 2021;11(11):e046660. doi: 10.1136/bmjopen-2020-046660.

3. Service NDR. Routes to diagnosis. 2019 [cited; Available from: https://nhsd-ndrs.shinyapps.io/routes_to_diagnosis/.

4. Hovda T, Hoff SR, Larsen M, Romundstad L, Sahlberg KK, Hofvind S. True and missed interval cancer in organized mammographic screening: a retrospective review study of diagnostic and prior screening mammograms. *Academic Radiology*. 2022;29:S180-S91.

5. Houssami N, Hunter K. The epidemiology, radiology and biological characteristics of interval breast cancers in population mammography screening. *npj Breast Cancer*. 2017;3(1):12. doi: 10.1038/s41523-017-0014-x.

6. England PH. Interval cancers explained in the NHS Breast Screening Programme. 2020.

7. Freeman K, Geppert J, Stinton C, Todkill D, Johnson S, Clarke A, Taylor-Phillips S. Use of artificial intelligence for image analysis in breast cancer screening programmes: systematic review of test accuracy. *BMJ*. 2021;374:n1872. doi: 10.1136/bmj.n1872.

8. Hickman SE, Woitek R, Le EPV, Im YR, Mouritsen Luxhøj C, Aviles-Rivero AI, Baxter GC, MacKay JW, Gilbert FJ. Machine Learning for Workflow Applications in Screening Mammography: Systematic Review and Meta-Analysis. *Radiology*. 2021;302(1):88-104. doi: 10.1148/radiol.2021210391.

9. Lång K, Josefsson V, Larsson A-M, Larsson S, Högberg C, Sartor H, Hofvind S, Andersson I, Rosso A. Artificial intelligence-supported screen reading versus standard double reading in the Mammography Screening with Artificial Intelligence trial (MASAI): a clinical safety analysis of a randomised, controlled, non-inferiority, single-blinded, screening accuracy study. *The Lancet Oncology*. 2023;24(8):936-44. doi: 10.1016/S1470-2045(23)00298-X.

10. Hernström V, Josefsson V, Sartor H, Schmidt D, Larsson A-M, Hofvind S, Andersson I, Rosso A, Hagberg O, Lång K. Screening performance and characteristics of breast cancer detected in the Mammography Screening with Artificial Intelligence trial (MASAI): a randomised, controlled, parallel-group, non-inferiority, single-blinded, screening accuracy study. *The Lancet Digital Health*. 2025;7(3):e175-e83. doi: 10.1016/S2589-7500(24)00267-X.

11. Dembrower K, Crippa A, Colón E, Eklund M, Strand F. Artificial intelligence for breast cancer detection in screening mammography in Sweden: a prospective, population-based, paired-reader, non-inferiority study. *The Lancet Digital Health*. 2023;5(10):e703-e11. doi: 10.1016/S2589-7500(23)00153-X.

12. Lauritzen AD, Lillholm M, Lynge E, Nielsen M, Karssemeijer N, Vejborg I, Moy L. Early Indicators of the Impact of Using AI in Mammography Screening for Breast Cancer. *Radiology*. 2024;311(3):e232479. doi: 10.1148/radiol.232479.

13. Nguyen DL, Ren Y, Jones TM, Thomas SM, Lo JY, Grimm LJ, Goh V, Gamagami E. Patient Characteristics Impact Performance of AI Algorithm in Interpreting Negative Screening Digital Breast Tomosynthesis Studies. *Radiology*. 2024;311(2):e232286. doi: 10.1148/radiol.232286.

14. Houssami N, Marinovich ML. AI for mammography screening: enter evidence from prospective trials. *The Lancet Digital Health*. 2023;5(10):e641-e2.

15. Mackenzie A, Loveland J, van Engen R. Survey of image processing settings used for mammography systems in the United Kingdom: how variable is it?: SPIE; 2024.

16. de Vries CF, Colosimo SJ, Staff RT, Dymiter JA, Yearsley J, Dinneen D, Boyle M, Harrison DJ, Anderson LA, Lip G, Black C, Murray AD, Wilde K, Blackwood JD, Butterly C, Zurowski J, Eilbeck J, McSkimming C. Impact of Different Mammography Systems on Artificial Intelligence Performance in Breast Cancer Screening. *Radiology: Artificial Intelligence*. 2023;5(3):e220146. doi: 10.1148/ryai.220146.

17. Sahiner B, Chen W, Samala RK, Petrick N. Data drift in medical machine learning: implications and potential remedies. *British Journal of Radiology*. 2023;96(1150). doi: 10.1259/bjr.20220878.

18. Condon JJJ, Trinh V, Hall KA, Reintals M, Holmes AS, Oakden-Rayner L, Palmer LJ. Impact of Transfer Learning Using Local Data on Performance of a Deep Learning Model for Screening Mammography. *Radiology: Artificial Intelligence*. 2024;6(4):e230383. doi: 10.1148/ryai.230383.

19. Lim AWW, Deats K, Gambell J, Lawrence A, Lei J, Lyons M, North B, Parmar D, Patel H, Waller J, Warwick J, Sasieni PD, Lawrence A, Stephens C, Bhuiya A, Bojang F, Nestor C, Turabi N, Norman H, Sanger K, Quaye M, Gilani F, Ladva M, Lim A, Sasieni P, Waller J, Lyons M, Gambell J, Deats K, Wright A-M, Masson A, Pearmain P, Stubbs R, Ruwende J, Patel H, Gandy M, Roberts P, Fisher P, Lydon-Burgan A, Silver L, Ruane K, Winfield N, Dunn M, Platt G, Taylor M, Shepherd A, Coppini E, Cowie A, Cook C, Horner C, Fairbairn E, Aracena J, Douglas W, McLaughlin L, Siegal G. Opportunistic offering of self-sampling to non-attenders within the English cervical screening programme: a pragmatic, multicentre, implementation feasibility trial with randomly allocated cluster intervention start dates (YouScreen). *eClinicalMedicine*. 2024;73. doi: 10.1016/j.eclinm.2024.102672.

20. Lauren G, Syeda A, Priscilla M, Rumana N, Angie AK, Jo W. Acceptability of artificial intelligence in breast screening: focus groups with the screening-eligible population in England. *BMJ Public Health*. 2024;2(2):e000892. doi: 10.1136/bmjph-2024-000892.

21. Gale NK, Heath G, Cameron E, Rashid S, Redwood S. Using the framework method for the analysis of qualitative data in multi-disciplinary health research. *BMC Medical Research Methodology*. 2013;13(1):117. doi: 10.1186/1471-2288-13-117.

22. Bond M, Pavey T, Welch K, Cooper C, Garside R, Dean S, Hyde C. Systematic review of the psychological consequences of false-positive screening mammograms. *Health Technol Assess*. 2013;17(13):1-170, v-vi. doi: 10.3310/hta17130.

23. Howard DR, Brown JM, Todd S, Gregory WM. Recommendations on multiple testing adjustment in multi-arm trials with a shared control group. *Statistical Methods in Medical Research*. 2016;27(5):1513-30. doi: 10.1177/0962280216664759.

24. NHS Digital. Breast Screening Programme, England 2018-19 [NS]. 2020.

25. Taylor-Phillips S, Wallis MG, Jenkinson D, Adekanmbi V, Parsons H, Dunn J, Stallard N, Szczepura A, Gates S, Kearins O, Duncan A, Hudson S, Clarke A. Effect of Using the Same vs Different Order for Second Readings of Screening Mammograms on Rates of Breast Cancer Detection: A Randomized Clinical Trial. *Jama*. 2016;315(18):1956-65. doi: 10.1001/jama.2016.5257.

26. Frick KD. Microcosting quantity data collection methods. *Medical care*. 2009;47(7_Supplement_1):S76-S81.

27. Roberts M, Russell LB, Paltiel AD, Chambers M, McEwan P, Krahn M. Conceptualizing a model: a report of the ISPOR-SMDM modeling good research practices task force–2. *Medical Decision Making*. 2012;32(5):678-89.

28. Elvidge J, Hawksworth C, Avşar TS, Zemplenyi A, Chalkidou A, Petrou S, Petykó Z, Srivastava D, Chandra G, Delaye J. Consolidated Health Economic Evaluation Reporting Standards for Interventions that use Artificial Intelligence (CHEERS-AI). *Value in Health*. 2024.

29. Wright SJ, Gray E, Rogers G, Donten A, Payne K. A structured process for the validation of a decision-analytic model: application to a cost-effectiveness model for risk-stratified national breast screening. *Applied health economics and health policy*. 2024:1-16.

30. GitHub. 2024 [cited; Available from: https://github.com/.

31. Wright SJ, Paulden M, Payne K. Implementing interventions with varying marginal cost-effectiveness: an application in precision medicine. *Medical Decision Making*. 2020;40(7):924-38.

32. Wright SJ, Newman WG, Payne K. Quantifying the impact of capacity constraints in economic evaluations: an application in precision medicine. *Medical Decision Making*. 2022;42(4):538-53.

33. Hsieh H-F, Shannon SE. Three approaches to qualitative content analysis. *Qualitative health research*. 2005;15(9):1277-88.

34. Damschroder LJ, Reardon CM, Opra Widerquist MA, Lowery J. Conceptualizing outcomes for use with the Consolidated Framework for Implementation Research (CFIR): the CFIR Outcomes Addendum. *Implementation science*. 2022;17(1):7.

18. APPENDIX

18.1 Appendix 1 Data flow map

