ScreenTrust MRI – supplemental documentation

OFFICIAL TITLE: Image Analysis With Artificial Intelligence to Increase Precision in Breast Cancer Screening - the ScreenTrust MRI Substudy: a Prospective Trial of AI to Select Women for Supplemental Screening MRI

UNIQUE PROTOCOL ID: KSRAD001

DOCUMENT DATE: March 8, 2023

NCT NUMBER: NCT04832594

CONTENTS:

Revision history

Version	Date	Reason for change	Approved by
1.0	March 31, 2020	N/A	Fredrik Strand, Martin Eklund
1.1	March 8, 2023	Early stop of invitations Analysis of non-invitations	Fredrik Strand, Martin Eklund

Comprehensive Study Protocol

for

Image analysis with artificial intelligence to increase precision in breast cancer screening - the ScreenTrust MRI substudy: a prospective trial of AI to select women for supplemental screening MRI

Contents

World Health Organization Trial Registration Dataset

Primary Registry
The trial was registered with clinicaltrials.gov (NCT NCT04832594) with an initial release date of March 31, 2021.

Trial Identifying Number
KSRAD001

Date of registration in Primay Registry
TBD

Secondary Identifying Numbers
- Ethical Review Authority, Sweden, EPM 2020-00487
- Karolinska University Hospital, Sweden, K 2020-0807

Source(s) of Monetary or Material Support
- Karolinska University Hospital, Stockholm, academic hospital (PI salary)
- Region Stockholm, regional authority (funding from programs: Medtechlabs, Clinical postdoctoral researcher)
- Karolinska Institute, Stockholm, university (in-kind)

- Royal Institute of Technology, Stockholm, university
- Swedish breast cancer association, Sweden, patient organization (funding)
- Lunit Inc., South Korea, commercial company (software use free of charge, no funding)

Primary Sponsor
- Karolinska University Hospital, Stockholm

Secondary Sponsor
- Royal Institute of Technology, Stockholm

Contact for Public Queries
- screentruststudy@gmail.com, +46 517 700 00

Contact for Scientific Queries
- Principal Investigator: Fredrik Strand, MD, fredrik.strand@sll.se, +46 517 700 00, Bröstradiologi, NB1:03, Gävlegatan 55, 171 64 Solna, Sweden

Public Title
- ScreenTrust MRI – artificial intelligence to select women for supplemental MRI in breast cancer screening

Scientific Title
- Image analysis with artificial intelligence to increase precision in breast cancer screening, the ScreenTrust MRI substudy: a prospective trial of AI to select women for supplemental MRI in breast cancer screening

Countries of Recruitment
- Sweden

Health Condition or Problem Studied
- Breast cancer

Intervention
- Intervention Name: Artificial intelligence-based framework to select for MRI (AI MRI)
- Intervention Description: An AI-based framework has been developed by researchers at Karolinska Institute (led by Dr. Fredrik Strand) and Royal Institute of Technology (led by Dr: Kevin Smith). The specific AI-implementation (AI tool) in this study is a result of AI predictions from three equally weighted component AI models analyzing mammograms: (i) masking predictor, (ii) risk predictor and (iii) cancer signs predictor (by one commercial CAD model and one in-house academic CAD model); the age of the woman is also taken into account by multiplying the score with (110-age)/70. The purpose of the age factor is to attain a relatively similar proportion of MRI exams in the lower and higher age groups. The aim of the AI tool is to identify women with the highest probability of having a delay in cancer detection, i.e., having had a false negative screening mammogram. The specific AI tool and its settings will remain the same during the study. For each examination, the AI tool will produce an AI Joint Score and an AI Masking Score. The AI Masking Score cut-off point was defined by the median of examinations collected during the initial period of March 1 to March 24, 2021. The cut-off point of the AI Joint Score was defined by the 92nd percentile of the initial population. Women meeting these criteria will be invited to the study, and randomized to MRI or no-MRI (standard-of-care).

Key Inclusion and Exclusion Criteria

All women attending screening mammography at Karolinska University Hospital will be included in the study. The ethical review authority has waived the need to obtain individual written informed consent to be included in this initial stage of the study. However, for women that will be invited to

undergo MRI, a written informed consent will be collected.

- Inclusion criterion

    - women for whom a standard four-view mammography examination was acquired at regular screening (right MLO, right CC, left MLO, left CC; exam code 66200)

- Exclusion criteria:

    - women attending screening as part of a surveillance program (i.e, referred due to high risk of breast cancer, or a personal history of breast cancer) - defined by having an exam code different from regular screening

    - women having breast implant(s)

    - women recalled for diagnostic work-up in the current regular screening process

    - women whose exams the AI tool cannot process (e.g., corrupt file structures, irregular file paths, missing dicom tags describing window center, window width, image laterality, patient age, pixel array)

Study Type

Type of Study: Randomized clinical trial

Study design: All women with negative screening mammograms less than one month before the selection day will be analyzed for MRI eligibility. We will exclude women who participate in special surveillance programs, who have breast implants, who had prior breast cancer, who have breast implant, who are breast feeding or have any MRI contraindication requiring radiologist assessment. Then AI scores will be calculated as described above (Intervention). The workflow for the study until randomization is described in **Figure 1**. The number of women randomized to the MRI-group and the non-MRI group will be adapted to available MRI capacity over time.

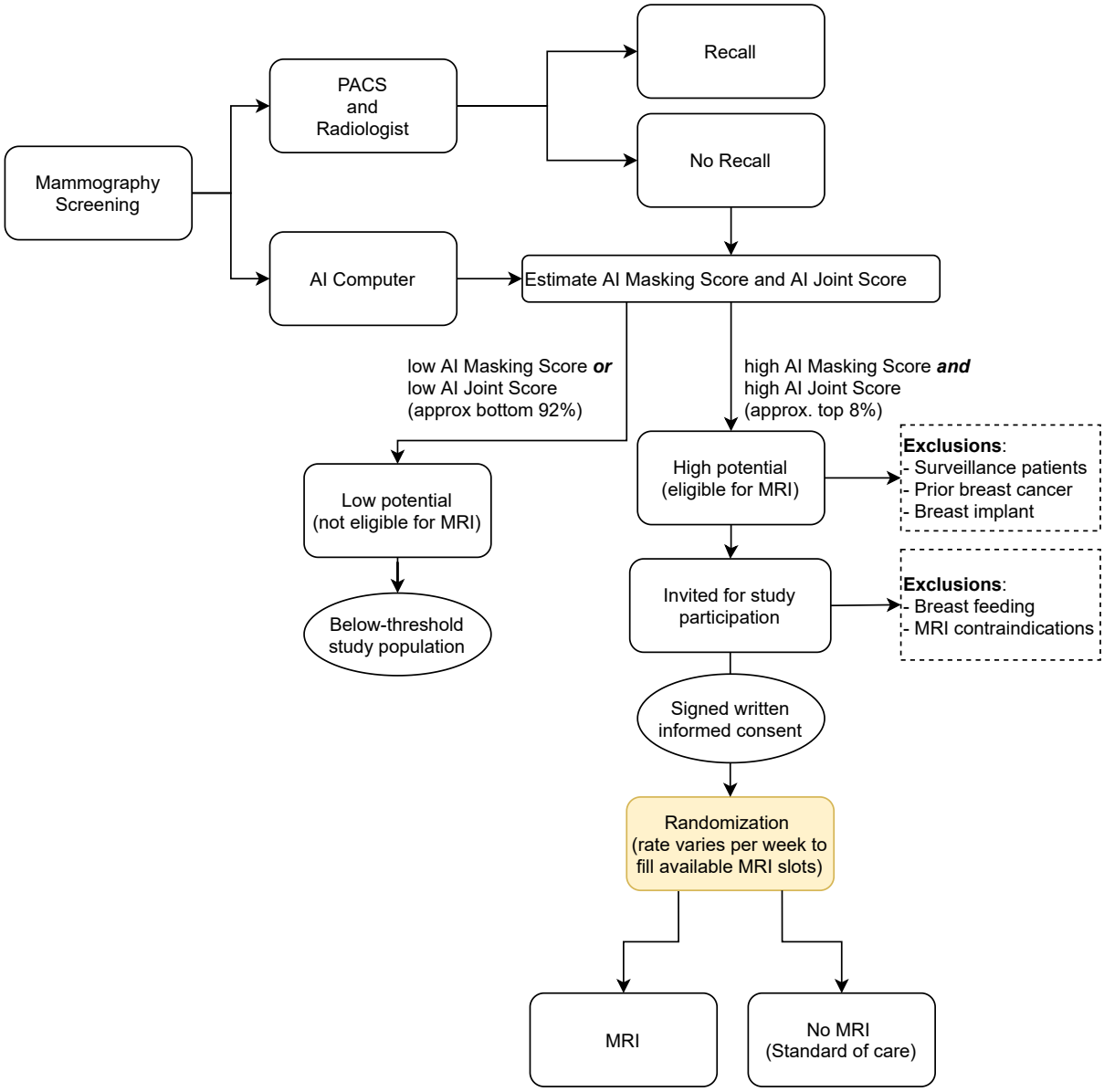

Figure 1. *Study flow chart until Randomization.*

Regular contraindications for MRI will apply. A Signa Premier 3T MRI scanner from GE Healthcare will be used. The MRI protocol will contain a T2-weighted Dixon sequence and a T1-weighted dynamic contrast enhanced series, and will remain the same through the course of the study. All MRI exams will be assessed by two radiologists, where the second reader will have access to the assessment of the first reader. In case of disagreement, a consensus discussion between two radiologists will be held. The MRI exams will be assessed according to BI-RADS, and follow-up will depend on the BI-RADS category (see **Figure 2**).

Follow-up of breast cancer status and tumor characteristics will continue until, and including, the next planned screening examination (within 27 months of the first included screening examination).

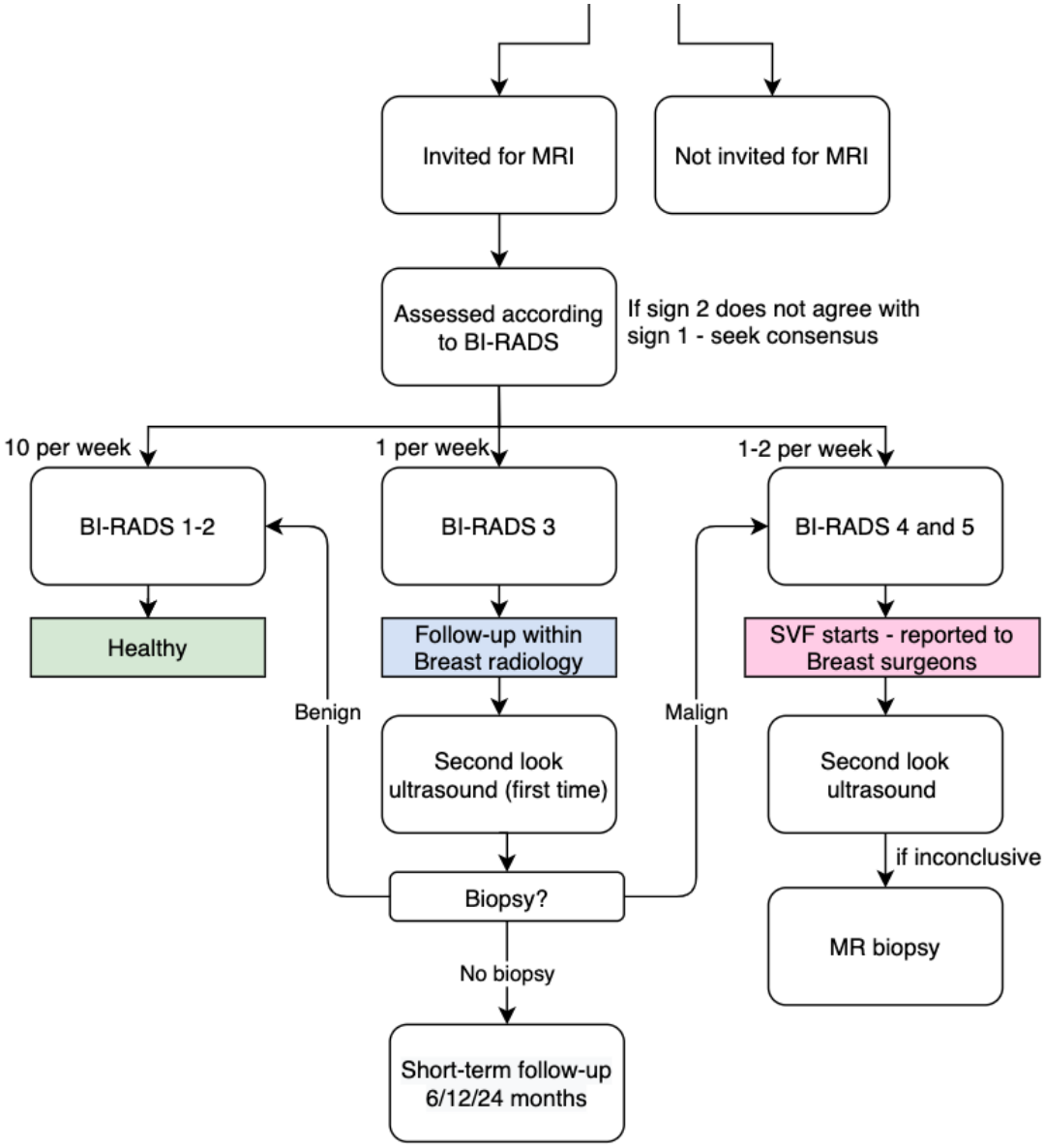

Figure 2. *Study flowchart for women undergoing MRI. Number of weekly exams for each category are prior estimates based on the DENSE trial.*

Date of First Enrollment

Study start is estimated to take place in April 1, 2021.

Sample Size

The study shall enroll study persons until 1000 women have undergone MRI scanning. With an 95% attendance rate after invitation to MRI, we estimate that around 1050 women have been invited for MRI (and thus included in the MRI group). Given the variable randomization ratio, described above, we estimate that a slightly larger number of women have been included in the non-MRI group.

AMENDMENT IN MARCH 2023: Due to a decision to cease to invite women after May 2023, the sample size of 1000 performed MRI exams will not be attainable. An assessment of MRIs revealed a

higher cancer detection rate than we had anticipated, which should results in a bigger effect size than we had assumed in our sample size calculations for the study. Thus, we believe that the statistical power in the trial will be similar to what we had originally anticipated, even with the smaller number of participants. See also "completion date" below as well as the section about sample size calculations in the SAP.

Recruitment Status

Active

Primary Outcome(s)

The primary outcome is a composite endpoint representing **early detection failure** – a breast cancer diagnosed with any of the below characteristics:

- Interval cancer

- Cancer with invasive component larger than 15 mm

- Cancer with lymph node metastasis

The follow-up time starts once the initial mammography and screening MRI are fully processed and any cancer detected at the initial examination has been diagnosed. Thus, the cancers that will be considered for primary outcome are the women that are healthy after the initial examination. A positive primary outcome will be based on interval cancers and the cancers that are screen-detected during a 27-month follow-up time after initial screening and fulfil the above criteria.

Secondary Outcome(s)

To the extent that a secondary outcome measure does not overlap with the primary outcome, it may be included in interim reporting.

1. Women invited, women declined and women accepted to participate; including reason for non-participation (contraindication, patient choice, no response).

2. Distribution of AI scores for: i) Joint model, ii) AI risk score, iii) AI masking score, iv) AI CAD score(s).

3. a) Breast cancer diagnosed by screening MRI.
b) All breast cancers diagnosed during the study time including the initial screening, including mode of detection (MRI, mammographic detection, clinical detection including time since screening) for all women for whom AI scores were calculated (above and below threshold)

For all cancers in the outcome measures, the following tumor characteristics will be reported:

a) Invasiveness (in situ or invasive; micro-invasive counted as in situ)

b) Histology (ductal, lobular, mucinous, tubular, other)

c) Lymph node status (0, 1-3, or more than 4 lymph node metastases)

d) Tumor size (in mm) – in situ and invasive component separately measured by pathologist

e) Ki-67 percent. With a 20% binary cut-off, and an ex 14% cut-off level

d) Molecular subtype (Luminal A-like defined as estrogen and/or progesterone receptor positive and HER2 negative and KI-67 <14%, Luminal B-like defined as estrogen and/or progesterone receptor positive and HER2 positive/negative and KI-67 >=14% , HER2-positive defined as estrogen and progesterone receptor negative and HER2 positive, and

Triple negative defined as estrogen and progesterone receptor and HER2 negative)

4. Radiological process measures: distribution of BI-RADS scores (amount of fibroglandular tissue, background parenchymal enhancement, and any lesion); number and outcome of second look ultrasound, biopsies, short-term MRI follow-up. We will report operational characteristics of the screening MRI (cancer detection rate, recall rate, positive predictive value for 2nd look ultrasound decision and for biopsy decision).

5. Questionnaire for women undergoing MRI: self-examination habit (yes, sometimes, often); previous MRI (yes, no); prior breast cancer (yes, no, which breast and year); prior non-malignant breast disease (yes, no, which year); first-degree relative with breast cancer (yes, no, at what earliest age); first-degree relative with ovarian cancer (yes, no, earliest age); Age of menarche, Parity, Age of first pregnancy, Number of children born, Time of breast-feeding, date of latest menstruation; reason for no menstruation, use of hormonal medications; use of Cozaar or Losartan; Current length and weight. They may also fill in a questionnaire concerning their experience of undergoing a screening MRI.

Explorative: The influence on cancer incidence and detection of the following potential predictors or modifiers will be explored (further detailed in the Statistical Analysis Plan): Radiologist assessments of screening mammogram at inclusion, Age, Density, AI scores: Overall pipeline (the basis for cut-off point), and various cut-off points for each of the three AI components (risk, masking and CAD).

Ethics Review

Approved on April 28, 2020, with registration id EPM 2020-00487 by the Ethical Review Authority of Sweden (email: registrator@etikprovning.se, phone: +46 10 475 08 00).

Completion date

Patient invitations to the study will cease in May 2023, even if we do not reach the original target of 1,000 performed MRI exams. The reason is that there is an increasing risk that the same women that were included in April 2021 will return for their next round of regular screening. Having a second round for some women and not for others would be inconsistent and we have decided to avoid this situation. See also "Sample size" above.

Summary Results

N/A

IPD sharing statement

We plan to share individual participant-level data to the extent that the data can be considered anonymous by the responsible research body. A transfer agreement for academic research purposes will be required.

SPIRIT-AI additional items

Protocol Version

1.0, March 10, 2020.

Funding

See WHO above.

Roles and responsibilities

Protocol contributors

Fredrik Strand, MD PhD, Karolinska University Hospital: Principal Investigator, Radiology team leader

Martin Eklund, PhD, Karolinska Institutet: Biostatistician

Kevin Smith, PhD, SciLife Lab / Royal School of Engineering, Stockholm: Computer science team leader

Hossein Azizpour, PhD, Division of Robotics, Perception and Learning, Royal School of Engineering

Trial sponsor:

See WHO above.

Role of study sponsor and funders

The primary sponsor of the study, Karolinska University Hospital, represented by the PI is responsible for collection and management of the study; for study design, analysis and interpretation of data, writing of the report, and the decision to submit the report for publication (the PI will have ultimate authority to decide over these activities). The funding has been provided by regional authorities and the breast cancer patient organization, which will have no influence over study management.

Composition, roles and responsibilities of steering committee

The steering committee consists of (all from Karolinska University Hospital, except for Dr. Smith from SciLifeLab/KTH and Dr. Eklund from Karolinska Institute): Fredrik Strand, PI; Kevin Smith, computer science team leader; Martin Eklund, biostatistics; Athanasios Zouzos, head of breast radiology; Irma Fredriksson, breast surgeon; Hanna Fredholm, breast surgeon; Theodoros Foukakis, breast oncologist.

Introduction

Background and rationale

Breast cancer is the most common cancer for women. Though the patients have a relatively good probability of survival, around 1500 women die each year in Sweden from the disease. Mammographic screening has been shown to lower mortality by around 30 (1). However, in the screening programs large resources are consumed and around 30 percent of cancers go undetected and the women find them by noticing a lump in the breast (2). MRI has consistently been shown to have a higher sensitivity than mammography, but often a lower specificity. MRI is also several times more expensive than mammography, and there is limited capacity for additional examinations in the current installations. Thus, even if MRI would increase early cancer detection, it would be offered only to a limited group of women. In a Dutch study, MRI was offered to women at average risk but extremely high mammographic density. The

researchers found that the additional cancer detection rate was 16.5 per 1000 MRI screenings (compared to around 5 per 1000 mammography screenings in the general population) (3). For women undergoing MRI they found a decrease from 5.0 interval cancers / 1000 women to 0.8 interval cancers / 1000 women; as well as a decrease in the lymph node positivity from 45% to 15%. Since 2018 our group, together with collaborators, has developed AI-based tools that in retrospective analysis seems to offer higher precision for short-term breast cancer risk prediction than using mammographic density (4). In this prospective clinical study, we will examine how using an AI pipeline to select women who are invited for MRI will affect the number of screen-detected cancers and the number of cancers with poor prognosis shown by lymph node positivity and/or interval cancer detection.

1. Weedon-Fekjær H, Romundstad PR, Vatten LJ. Modern mammography screening and breast cancer mortality: population study2014 2014-06-17 22:30:49.
2. Törnberg S, Kemetli L, Ascunce N, Hofvind S, Anttila A, Sèradour B, et al. A pooled analysis of interval cancer rates in six European countries. European journal of cancer prevention. 2010;19(2):87-93.
3. Bakker MF, de Lange SV, Pijnappel RM, Mann RM, Peeters PHM, Monninkhof EM, et al. Supplemental MRI Screening for Women with Extremely Dense Breast Tissue. The New England journal of medicine. 2019;381(22):2091-102.
4. Dembrower K, Liu Y, Azizpour H, Eklund M, Smith K, Lindholm P, et al. Comparison of a Deep Learning Risk Score and Standard Mammographic Density Score for Breast Cancer Risk Prediction. Radiology. 2019:190872.

Objectives

The overall aim of the project is to examine how an AI pipeline to select women for supplemental MRI at breast cancer screening affects cancer detection.

Trial design
See WHO above, "Study Type"

Study setting
The study will be conducted in the Karolinska University hospital which has a defined geographical uptake area for breast cancer screening. The breast imaging department has around seven dedicated breast radiologists. The x-ray equipment for screening mammography is from Hologic Inc, and the MRI scanner is a 3T Signa Premier from GE Healthcare Inc.

Eligibility criteria, Interventions, Outcomes
See WHO above

Participant timeline
Enrollment will start as soon as all IT systems have been integrated and tested, and after a pilot week where patient inclusion, administrative processes and technical integration is tested in practice. For each participant there will be a follow-up period of 27 months for collecting information on both screen-detected cancer and interval cancer (i.e., cancer diagnosed after the current screening and before the next planned one). Enrollment will be paused during public holidays and during the summer when MRI staffing is low.

Sample size
See WHO above

Recruitment

Recruitment will continue until the target of 1,000 women having undergone MRI examination is reached

Sequence generation, allocation concealment mechanism, implementation
Women that are selected into the targeted group by the AI pipeline, will be invited to participate in the study, and then randomized to MRI or no MRI (standard of care). The randomization will be carried out in a computerized process over which the operator has no influence.

Blinding
The initial mammography examination will be assessed by radiologists as standard of care at the Karolinska University Hospital and will be finalized prior to any selection of study persons for this study. The radiologists are therefore intrinsically blinded to whether the woman may be invited for MRI afterwards or not.

Data collection methods
Screening images including DICOM data will be transferred from the mammography equipment to a secure AI computer. On the AI computer, the images will be processed by the three component networks in the defined AI pipeline. The output will be reported in a CSV file. The CSV file will be transferred to a clinical work station where the three scores will be combined based on a predefined model and checked for being above the cut-off point for the targeted group. On the same computer, randomization of women in the targeted group will be carried out, and the patient workflow will be continuously recorded including keeping track of which women were invited, how they responded, how they were randomized and whether an MRI exam has been carried out. Data from the radiology system and the AI pipeline will be stored. Data on all cancer diagnoses will be collected through linking to the regional cancer registry and hospital records. All data will be linked based on the unique national personal identity number of each study person.

Data management
Data will remain stored at the study hospital during the course of the study. Regular backups will be performed. When required for predefined interim or final analysis, data will be extracted for research purposes under the responsibility of the PI. This data will be stored following the usual practice in the breast imaging research group at the Karolinska University Hospital, including pseudonymization before any statistical analysis. Each data parameter will undergo type and range checks for validity. Further details on data management can be found in the Data Policy document of the breast imaging research group.

Statistical methods
See separate Statistical Analysis Plan (SAP) document.

Data monitoring
There is no data monitoring committee. This has not been deemed necessary since all continuously collected data will be automatically recorded in the radiology system.

Final analysis of each outcome will take place once all relevant data has already been collected. The outcome that can be analyzed first, will be the number of cancers detected by MRI which can be carried out as soon as the last included study person has undergone MRI and that examination has been fully assessed. Analysis will be conducted under the supervision of the PI. Interim analysis will be carried out to assess any secondary outcome to the extent that it does not overlap with the primary outcome.

Harms
MRI examination confers relatively little risk. In the large Dutch DENSE study referred to above, they performed 4783 screening MRIs and encountered only 8 complications of which 5 were

categorized as serious (3 vasovagal reactions and 2 allergic reactions) and 3 less serious (2 extravasation of contrast media and 1 should dislocation).

Auditing
No auditing is planned.

Research ethics approval
See WHO above.

Protocol Amendments
Any protocol amendment will be decided by the Steering Committee. If the amendment is deemed to require additional permission by the Ethical Review Authority this will be sought before communication. Then, the amendment will be communicated to the Steering Committee and included in the updated study protocol.

Consent or ascent
Women that decide to participate in the randomization to MRI or no MRI will have signed a written informed consent. The Ethics Review Authority has waived the need for individual written informed consent from women who only undergo the regular screening mammography.

Confidentiality
All personal information will be handled according to GDPR and other applicable laws. Data will be pseudonymized before statistical analysis is performed.

Declaration of interests
The study is mainly funded by the regional authority, Region Stockholm, responsible for public health care in the area. The breast cancer patient association has also contributed funding. For activities outside this study, the principal investigator receives occasional fees for public presentations from Lunit Inc., a South Korean manufacturer of AI CAD systems for cancer detection.

Access to data
The final trial dataset will be available for the research team of the principal investigator. Pseudonymized data can be made available for external research audit. Anonymous data may be shared with academic researchers.

Ancillary and post-trial care
Not applicable. Patients are always covered by the national Swedish patient insurance.

Dissemination policy
Investigators plan to communicate findings primarily through original research papers and through participation in professional meetings. In addition, the investigators will communicate with the general public through media and through presentations at patient association gatherings. For research papers, the inclusion of co-authors will follow ICMJE recommendations. We do not intend to use professional writers outside the investigator team. While access may be granted to academic researchers, public access to complete participant-level data will not necessarily be granted. The statistical code may be shared publicly.

Appendices
None.

Approvals

ScreenTrust MRI Statisical Analysis Plan

Authors	Fredrik Strand, PI
Approved by	Martin Eklund, study statistician

STATISTICAL ANALYSIS PLAN

Revision history

Version	Date	Reason for change
1.0	March 31, 2020	
1.1	March 8 , 2023	Early stop of invitations Analysis of non-invitations

1 Preface

This Statistical Analysis Plan (SAP) describes the planned analyses for
Image analysis with artificial intelligence to increase precision in breast cancer screening, the ScreenTrust MRI substudy, a prospective trial of AI triaging to invite women for supplemental screening MRI ("ScreenTrust MRI")

The trial was registered with clinicaltrials.gov (NCT NCT04832594) with an initial release date of March 31, 2021.

The planned analyses identified in this SAP will be included in future manuscripts. Exploratory analyses not necessarily identified in this SAP may be performed to support planned analyses. Any post-hoc exploratory or unplanned analyses not specified in this SAP before database lock will be identified as such in manuscripts for publication, and added as amendments to this SAP.
This SAP was written by statistician and investigator who were blinded to any assessments already performed by AI or human radiologists, and to outcomes.

2 Background

Breast cancer is the most common cancer for women. Though the patients have a relatively good probability of survival, around 1500 women die each year in Sweden from the disease. Mammographic screening has been shown to lower mortality by around 30 (1). However, in the screening programs large resources are consumed and around 30 percent of cancers go undetected and the women find them by noticing a lump in the breast (2).

MRI has consistently been shown to have a higher sensitivity than mammography, but often a lower specificity. MRI is also several times more expensive than mammography, and there is limited capacity for additional examinations in the current installations. Thus, even if MRI would increase early cancer detection, it would be offered only to a limited group of women. In a Dutch study, MRI was offered to women at average risk but extremely high mammographic density. The researchers found that the additional cancer detection rate was 16.5 per 1000 MRI screenings (compared to around 5 per 1000 mammography screenings in the general population) (3). Since 2018 our group, together with collaborators, has developed AI-based tools that in retrospective analysis seems to offer higher precision than using mammographic density (4).

In this prospective clinical study, we will examine how using an AI pipeline to select women who are invited for MRI will affect the diagnosed breast cancers.

3 Design

This is a prospective clinical trial aiming to determine the ability of an AI pipeline to identify women who would benefit from supplemental MRI in terms of decreasing the number of cancers having a significantly delayed detection (defined in the section "4.1 Primary endpoint" below).

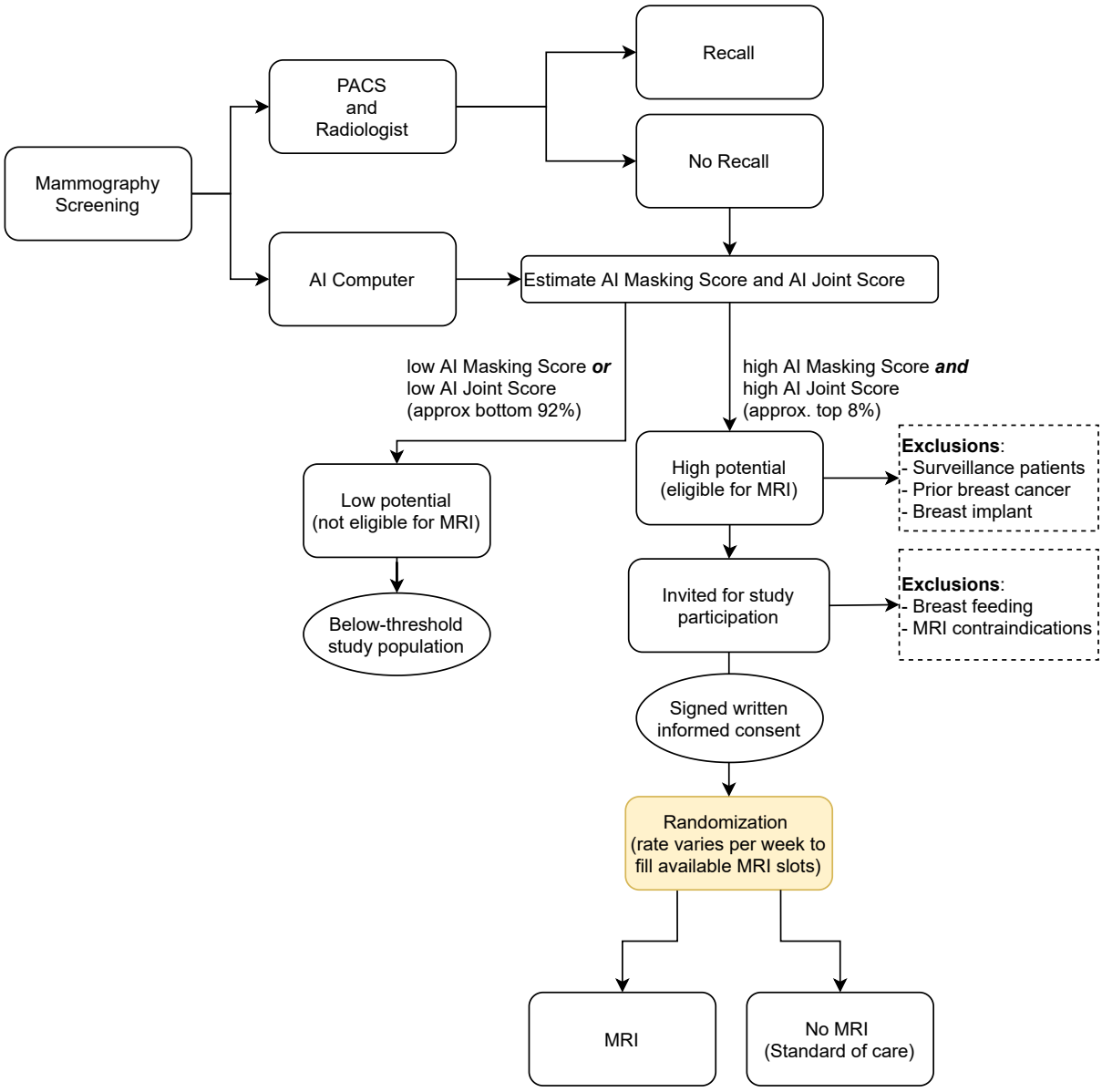

Figure 1. *Study flow chart until Randomization.*

All women attending mammography screening at Karolinska University Hospital will have their mammograms analyzed by AI (Figure 1). The specific AI-implementation (AI tool) in this study is a result of AI predictions from three equally weighted component AI models analyzing mammograms: (i) masking predictor, (ii) risk predictor and (iii) cancer signs predictor (by one commercial CAD model and one in-house academic CAD model); the age of the woman is also taken into account by multiplying the score with (110-age)/70. The purpose of the age factor is to attain a relatively similar proportion of MRI exams in the

lower and higher age groups. The aim of the AI tool is to identify women with the highest probability of having a delay in cancer detection, i.e., having had a false negative screening mammogram.

The specific AI tool and its settings will remain the same during the study. For each examination, the AI tool will produce an AI Joint Score. The AI Joint Score calculation is specified in section 5.3. Women who decide to participate, will be randomized to MRI or no-MRI (standard-of-care).

A Signa Premier 3T MRI scanner from GE Healthcare will be used. The MRI protocol will contain a T2-weighted Dixon sequence and a T1-weighted dynamic contrast enhanced series, and will remain the same through the course of the study. All MRI exams will be assessed by two radiologists, where the second reader will have access to the assessment of the first reader. In case of disagreement, a consensus discussion between two radiologists will be held. The MRI exams will be assessed according to BI-RADS, and follow-up will depend on the BI-RADS category (Figure 2). Women with BI-RADS 1-2 will have no further diagnostics and will be sent a 'healthy letter'. Women with BI-RADS 3 to 5 will be recalled for 2nd look ultrasound. Women with BI-RADS 4-5 will be included in the regular process for established cancer suspicion and be discussed in a multidisciplinary team conference. For women with BI-RADS 3, the follow-up will be handled within the breast radiology unit.

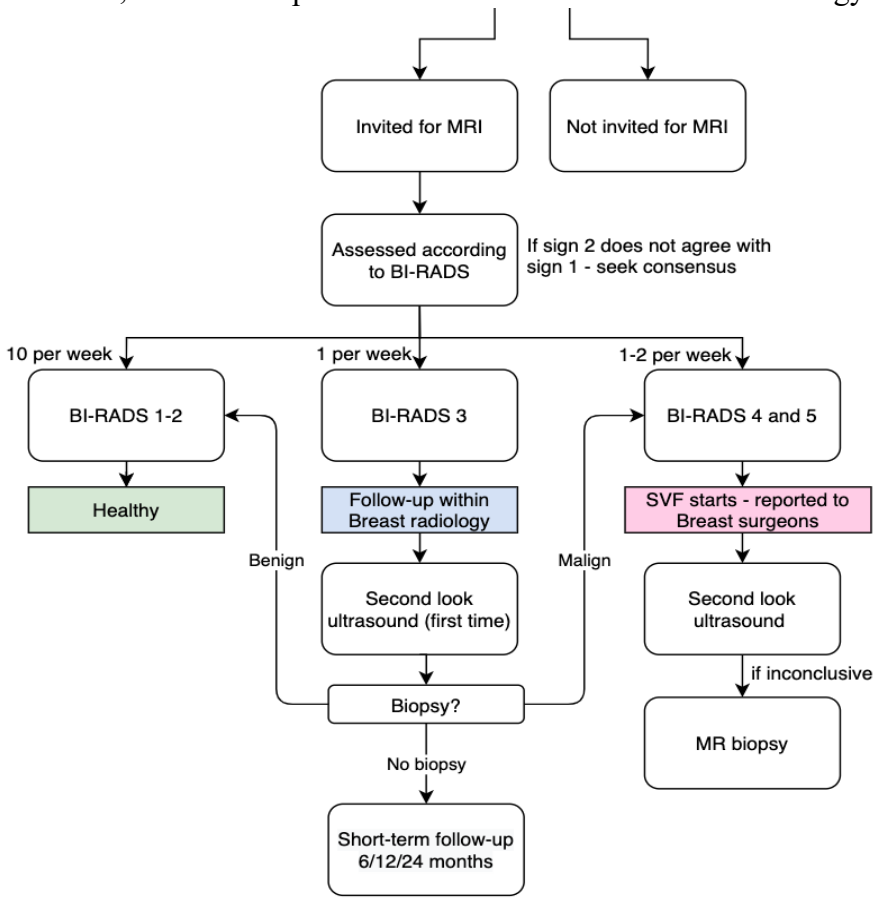

Figure 2. Workflow for patients invited for MRI.

3.1 Study population

The study will include healthy women attending regular mammographic screening program at Karolinska University hospital in Stockholm. All consecutive women will be included; the ethical review authority waived the need for individual informed consent at this stage.

- Inclusion criterion
    - Women in the population-based screening program for whom a standard four-view mammography examination was acquired (right MLO, right CC, left MLO, left CC)
- Exclusion criteria
    - Women in surveillance program referred from the hereditary cancer unit
    - Breast implants
    - Women in surveillance program due to prior breast cancer
    - Breast feeding
    - MRI contraindication requiring radiologist assessment

3.2 Study period

The study started on April 1, 2021. Invitations will not be sent after May 2023. The inclusion may be paused during general holidays, due to staffing difficulties, specifically the summer holiday which encompasses around 2 months from second half of June to second half of August.

AMENDMENT IN MARCH 2023 (replacing any conflicting statements compared to the above):

The last participant invitations will be sent out in May 2023. The reason is that some women included in April 2021 will return for their next screening, and we wish to avoid the inconsistency that some women would have two screening rounds included while others would have only one round. This will likely lead to that we have performed less than the 1,000 MRI exams that we aimed for. An assessment of MRIs revealed a higher cancer detection rate than we had anticipated, which should results in a bigger effect size than we had assumed in our sample size calculations for the study. Thus, we believe that the statistical power in the trial will be similar to what we had originally anticipated, even with the smaller number of participants. See also the section about sample size calculations (Section 5.7).

4 Objectives and Endpoints

The purpose of the study is to examine AI-based triaging of a small proportion of women with the objective to reduce the number of significantly delayed cancer diagnoses through the use of supplemental screening MRI.

4.1 Primary endpoint

Our study has a primary composite endpoint (CEP) representing **significantly delayed cancer detection** – having any of the below characteristics:

- Interval Cancer

- Cancer with lymph node metastasis

- Cancer with invasive component larger than 15 mm

4.2 Secondary endpoints

To the extent that a secondary outcome measure does not overlap with the primary outcome, it may be included in interim reporting.

4.2.1. Number and age of women eligible and not eligible for MRI, women invited, women declined and women accepted to participate; including reason for non-participation (contraindication, patient choice, no response).
4.2.2. Distribution of AI scores for: i) Joint model, ii) AI risk score, iii) AI masking score, iv) AI CAD score(s).
4.2.3. a) Breast cancer diagnosed by screening MRI.
b) All breast cancers diagnosed during the study time including the initial screening, including mode of detection (MRI, mammographic detection, clinical detection including time since screening) for all women for whom AI scores were calculated (above and below threshold)
For all cancers in the outcome measures, the following tumor characteristics will be reported:

a) Invasiveness (in situ or invasive; micro-invasive counted as in situ)

b) Histology (ductal, lobular, mucinous, tubular, other)

c) Lymph node status (0, 1-3, or more than 4 lymph node metastases)

d) Tumor size (in mm) – in situ and invasive component separately measured by pathologist

e) Ki-67 percent. With a 20% binary cut-off, and an ex 14% cut-off level

d) Molecular subtype (Luminal A-like defined as estrogen and/or progesterone receptor positive and HER2 negative and KI-67 <14%, Luminal B-like defined as estrogen and/or progesterone receptor positive and HER2 positive/negative and KI-67 >=14% , HER2-positive defiined as estrogen and progesterone receptor negative and HER2 positive, and Triple negative defined as estrogen and progesterone receptor and HER2 negative)

4.2.4. Radiological process measures: distribution of BI-RADS scores (amount of fibroglandular tissue, background parenchymal enhancement, and any lesion); number and outcome of second look ultrasound, biopsies, short-term MRI follow-up. We will report operational characteristics of the screening MRI (cancer detection rate, recall rate, positive predictive value for 2nd look ultrasound decision and for biopsy decision).

4.2.5. Questionnaire for women undergoing MRI: self-examination habit (yes, sometimes, often); previous MRI (yes, no); prior breast cancer (yes, no, which breast and year); prior non-malignant breast disease (yes, no, which year); first-degree relative with breast cancer (yes, no, at what earliest age); first-degree relative with ovarian cancer (yes, no, earliest age); Age of menarche, Parity, Age of first pregnancy, Number of children born, Time of breast-feeding, date of latest menstruation; reason for no menstruation, use of hormonal medications; use of Cozaar or Losartan; Current length and weight. They may also fill in a questionnaire concerning their experience of undergoing a screening MRI.

4.3 Exploratory objectives and endpoints

The first exploratory objective is to examine the MRI cancer detection rate if having the cut-off point at each percentile above the actual cut-off point (e.g., at the 95th, 96th, 97th, 98th, 99th percentile).
The second exploratory objective is to determine the overall cancer detection rate for each cut-off points at each percentile of the AI score (1st ,2nd ,3rd,…,98th ,99th).
The third exploratory objective is to examine the effect of each of the three AI networks in the pipeline by one-at-a-time setting the AI score equal to the population average, and for each of these three set-ups re-analyze the overall cancer detection rate according to the above second exploratory objective.
The fourth exploratory objective is to examine the influence on cancer incidence and detection of the following potential predictors or modifiers will be explored (further detailed in the Statistical Analysis Plan): Radiologist assessments of screening mammogram at inclusion, Age, Density, AI scores: Overall pipeline (the basis for cut-off point), and various cut-off points for each of the three AI components (risk, masking and CAD).
The fifth exploratory objective is to conduct a reader study to understand if the mammograms of the women selected by the AI pipeline and in which MRI detected cancer, could represent a screening radiologist mistake. This will be examined by performing a blinded review of the mammograms of women with MRI-detected cancer mixed with mammograms of other women and a random selection of mammograms in the invited group who did not have MRI-detected cancer.
The sixth exploratory objective is to conduct a hypothesis-generating reader review of the MRI examination for women who had MRI but nevertheless were diagnosed with interval cancer or screen-detected cancer at the following screening
The seventh exploratory objective is to conduct a hypothesis-generating reader review of the women who were assigned a BI-RADS 3 category at MRI, in order to understand who the number of women in this category could be minimized

5 Statistical methods

The methods and statistical analysis used in the study is described here. Analysis of the primary endpoint will take place only after the follow-up period has passed for all study persons; there will be no public interim analysis. However, there may be a statistical analysis during the second year of the study to ascertain that the study is likely to reach the predefined statistical power. For secondary endpoints there may be interim analyses, to the extent that these do not overlap with the analysis of the primary endpoint.
All statistics will be performed using Stata version ≥16 or R version ≥ 4.0.

5.1 Populations

The study population consists of all consecutive women attending regular screening mammography at Karolinska University Hospital in Stockholm having a normal assessment (i.e., not being recalled for further assessment by the consensus discussion). The prior data for each woman consists of: mammographic images, date of examination, age at examination, assessment by first radiologist, assessment by second radiologist, decision by consensus discussion (if any), AI scores from the three component models and a summary score, pathology-verified cancer diagnosis (if any) including key cancer characteristics (listed in section 4.2).

5.2 Demographics and baseline data

All data will be presented using descriptive statistics. Continuous variables will be summarized using number of women, mean, standard deviation, median, interquartile range (IQR), minimum and maximum. Categorical variables will be categorized as described in the end-points and summarized using the number and percentage of cases in each category.

5.3 Setting the operating point of the AI system

The AI tool will be calibrated based on a reference population of all consecutive screening mammograms collected during March 1 to March 17, 2021. From the distribution of the output of each AI model, the reference mean and reference standard deviation will be calculated and used to define the parameters for standardization of the scores from each component network (standardized value = (score – reference mean) / reference standard deviation). The joint score will be calculated by equal weighting of AI scores from the three components (risk prediction, masking potential, cancer detector). For the cancer detector, there will be two models used, each contributing equally. All model parameters will be fixed before study start and remain the same throughout the study. Finally, the joint score will be weighted to reflect remaining life years by multiplication with: (110 years – current age)/70. The purpose of the age factor is to achieve an approximately even distribution of MRI across the youngest and oldest age groups. The AI Masking Score cut-off point was defined by the median of examinations collected during the initial period of March 1 to March 24, 2021. The AI Joint Score cut-off point was defined to have 8% of the initial population in the eligible group to be invited. The reason behind selecting approximately 8% is that it roughly corresponds to the proportion of women having the highest BI-RADS density score of extremely dense breast. Another reason is that the selected population should contain a relatively high concentration of our composite endpoint of cancer of significantly delayed detection (i.e., around 14 per 1000 women).

5.4 Primary analyses: Superiority

Superiority analyses in terms of the absolute incidence of the primary CEP (significantly delayed cancer detection) further defined in the primary end-point (4.1). The study has been powered for the primary analysis of difference in CEP between the MRI and the no-MRI (standard of care) arms of the study.

Analyses will compare the difference CEP between study arms. This can be done on an absolute or a relative scale. The primary analysis will be performed on an absolute scale and supportive analyses may use a relative scale.

Absolute scale. The absolute difference in CEP is defined as the CEP the standard arm minus the CEP in the experimental arm ($\Delta CEP = CEP_{Std} - CEP_{Exp}$) or vice versa, as appropriate. It is estimated by plugging into the formula the observed proportions. An approximate 100(1-α)% two-sided Wald confidence interval for ΔCEP is calculated as

$$\widehat{\Delta CEP} \pm z_{\alpha/2}\sqrt{\frac{\widehat{CEP}_{Exp}*(1-\widehat{CEP}_{Exp})}{n_{Exp}} + \frac{\widehat{CEP}_{Std}*(1-\widehat{CEP}_{Std})}{n_{Std}}}.$$

Relative scale. The relative difference in CEP is defined as the CEP in the experimental arm divided by the CEP in the standard arm ($rCEP = CEP_{Exp}/CEP_{Std}$) or vice versa, as appropriate.

It is estimated by plugging into the formula the observed proportions. An approximate 100(1-α)% two-sided Wald confidence interval for rCEP is calculated as

$$exp\left(log(\widehat{rCEP}) \pm z_{\alpha/2}\sqrt{\frac{1}{\widehat{CEP}_{Exp}*n_{Exp}} - \frac{1}{n_{Exp}} + \frac{1}{\widehat{CEP}_{Std}*n_{Std}} - \frac{1}{n_{Std}}}\right)$$

5.5 Secondary analyses

5.5.1. We will calculate summary statistics for the age and number of women not eligible for MRI, eligible and invited for MRI, women declined MRI and women undergoing MRI; including reason for non-participation (contraindication, patient choice, no response): for the entire study period, and how it changes over time.

5.5.2. We will calculate summary statistics and graphically show the distribution of AI scores for: i) Joint model, ii) AI risk score, iii) AI masking score, iv) AI CAD score(s). We will subgroup these analysis by age group at the time of initial screening: 40-49, 50-59, 60-69 and 70-74 years. We will subgroup these analysis by participation category (eligible vs. not eligible for MRI, invited vs. not invited for MRI, eligible women undergoing vs. not undergoing MRI, declined MRI (by reason)).

5.5.3. a) For women invited to MRI, we will calculate the number and proportion of breast cancer diagnosed by screening MRI, and perform logistic regression analysis to examine associations with AI scores (joint and component scores), age of the woman, by responses in the study persion questionnaire (see section 4.2.5), by BI-RADS scores of amount of fibgroglandular tissue and background parenchymal enhancement. Similar analysis will be carried out for all cancers diagnosed during the study time. We will repeat the analyses for women actually undergoing MRI.

b) We will perform similar analyses as above for all women undergoing mammography only, but by necessity exclude parameters that are not collected for those women.

For all cancers diagnosed, the tumor characteristics listed in section 4.2.3 will be reported, overall and divided by women eligible or not eligible for MRI, as well as for women invited vs. not invited to MRI, eligible women undergoing vs. not undergoing MRI.

5.5.4. We will calculate summary statistics for the radiological process measures described in 4.2.4.

5.5.5. We will calculate summary statistics for each of the questionnaire responses described in 4.2.5.

5.7 Sample size

The sample size calculation relates to the primary analysis after women have undergone the baseline mammogram. We have estimated that the study would have a at least 90% power in detection of a statistically significant ($p<0.05$) difference in the absolute incidence of the primary endpoint of significantly delayed cancer (defined in 4.1), under the following assumptions:

- 50% of the women invited will agree to participate (not relevant)
- 84% of the significantly delayed cancers are averted by undergoing MRI (based on DENSE trial)

- 80% of the women will participate in the subsequent screening (around two years after the initial screening)
- 28.6 (approximately twice as many cancers detected as we assumed before study start) per 1000 women having a significantly delayed detection in the population selected by the AI tool
- The randomization rate will be around 1:1 for MRI vs. no-MRI (standard of care)

5.8 Adjustment for multiplicity

We will not perform any correction for multiple comparisons. Each analysis will be presented with unadjusted 95% confidence interval.

5.9 Handling of missing data

Endpoints

Our primary approach is of intention-to-treat-type (defined by which arm they are randomized to, MRI or not MRI); for any examination where the I process failed, study inclusion is not possible. In addition, we will perform a per-protocol analysis (defined by whether they underwent MRI or not).

Patient characteristics

Since data collection is performed through the electronic medical record, missing data is expected to be minimal. For analyses where patient characteristics are used (e.g., analyzing the interaction between a patient characteristics and study outcomes), we will exclude patients with missing data. As a sensitivity analysis, we will impute the missing data using the mean or median, whichever is appropriate.

5.10 Analysis of non-invitations of eligible women

Interim analysis performed in March, 2023, showed that 423 women had an AI Joint Score above the threshold but had not been invited to randomization due to technical errors. These women were erroneously excluded from randomization and therefore never had the chance to be selected for MRI. We have no reason to believe that the technical error was related to the endpoints of this study, but we nevertheless intend to perform analysis to estimate whether this exclusion corresponds to a random sample by comparing the age and mammographic density distribution between the invited eligible women and the non-invited eligible women. If we determine that it was not a random sample, we will, in sensitivity analysis, employ randomization and imputation of the clinical outcome for women randomized to MRI.