COVER PAGE

TITLE: Artificial Intelligence Assisted Breast Ultrasound in Breast Cancer Screening

NCT number: 20210712

Date: 2020/08/17

Study Protocol and Statistical Analysis Plan

Study population

The program of "Two Cancer (breast and cervical cancer) Screening" has been launched in China since 2009 to provide free screening tests for millions of rural women aged 35-59 years (Ministry of Health of China and All-China Women's Federation). This program has been expanded to all eligible women in both rural and urban areas as a basic public health service since 2019. Based on this program, we conducted a prospective cluster randomized controlled screening trial in Shanghai, China.

Eligible women were aged 35–70 years without a history of breast cancer, including in-situ cancer or other cancers in the previous five years, without serious cardiopulmonary insufficiency, liver or kidney insufficiency or other systemic diseases, and with life expectancy of more than five years.

Definition of a cluster

We define one district as a cluster. In 2021, Shanghai has jurisdictions over 16 municipal districts. The division of districts was obtained from field measured data and reference to relevant geographical maps, using human-computer interaction to carry out the vectorization of administrative division maps. Of above divisions, seven are urban areas and nine are suburban. The urban area covered an area of 20-40 square kilometers with a resident population of 600,000 to 1 000,000. The suburban district covered an area of 300-1200 square kilometers with a population of 500,000 to 3,200,000. At least 20,000 women aged 35 to 70 were included in each district.

Randomization method

Randomization was conducted by cluster, where districts rather than individuals were chosen as units of randomization. In order to avoid transfer between control and intervention cluster, urban levels were used as our strata. We then randomly selected one district out of seven urban areas as intervention cluster and another one out of nine suburban areas as control cluster.

Participants and intervention

Two women and children's care centers from randomly selected two districts were the main study locations. Women in the intervention arm received AI-assisted ultrasound conducted by trained female primary health workers, while women in the control arm received routine ultrasound conducted by doctors from the Maternal and Child Health Hospital. Participants in both arms were eligible for further diagnostic evaluation and treatment at Fudan university Shanghai cancer center and women in both groups were offered access to obtain rapid treatment at the hospital. Recruitment was started in January 2021 and completed in December 2022. All participants were then followed up till the end of 2023. The database was sealed in January 2024 for analysis.

Screening procedure

Our screening procedure is based on the national Breast Cancer Screening Process Technical Guidelines issued in 2015 by the National Health Commission of People's Republic of China as follows:

(1) Clinical examination and initial breast ultrasound screening: all the participants were enrolled and received clinical examination and initial breast ultrasound or AI-assisted ultrasound. The initial screening test results were classified based on BI-RADS; (2) MG re-screening: participants with BI-RADS grade 0 or 3 were then suggested to receive MG. MG results were classified based on BI-RADS; (3) Histopathological examination (hereinafter referred to as biopsy): participants with BI-RADS grade 4 to 5 either in initial ultrasound test or MG test were suggested for further biopsy examination. For patients with MG rescreening of grade 0 or grade 3, short-term follow-up (three to six months) or further biopsy examination were suggested according to experts' evaluation; (4) Follow-up: participants with BI-RADS level 1 to 2 either in ultrasound or MG were under strict follow-up through visits and phone calls by trained medical social workers.

Final screening results (positive or negative) were based on a two-step approach. Initial screening results were determined by breast ultrasound and defined as negative with BI-RADS grade 1 or 2, positive with BI-RADS grade

4 or 5, and indeterminate with BI-RADS grade 0 or 3. Participants with indeterminate initial results were informed of the suspicious lesion detection and referred to MG assessment. The result in this second step was classified as negative or positive according to MG BI-RADS grade.

Participants receiving final negative results were exempt from additional diagnostic procedures. By contrast, those receiving final positive results were referred to Fudan University Shanghai Cancer Center for further diagnostic work-up to exclude or diagnose breast cancer. To assess interval cancers, we obtained data from all diagnosed breast cancers, plus an additional one year of follow-up from the Shanghai municipal cancer registry, one of the largest cancer registries globally, and an associate member of the International Association of Cancer Registries (IARC). For participants diagnosed with breast cancer outside of screening (ie, diagnosed with interval cancer), we collected all medical records. Two experienced radiologists reviewed the screening MG and ultrasound images from the study and the clinical MG and ultrasound images used for diagnosis of breast cancer, and reached a consensus on whether or not the breast cancer could retrospectively be identified in the screening.

Outcomes

The primary endpoint of this study was to assess, and compare, the difference in the incidence of early breast cancer between two groups using different breast cancer screening methods (AI-assisted ultrasound screening/routine ultrasound screening). Early cancers should meet any of the following criteria at the time of diagnosis: less than 20 milimeters in diameter, no metastatic lymph nodes in the axilla, no distant metastasis, non-invasive cancers, stage 0, stage I, and stage IIA IIB (T2N1M0) breast cancer according to AJCC 8th staging.

The secondary aim was to determine the sensitivity, specificity, positive predictive value, and negative predictive value of the screening methods (AI-assisted ultrasound screening/routine ultrasound screening). Screening-detected cancers were defined as breast cancers positive for screening tests. We defined interval cancers as: breast cancers diagnosed after a negative screening test; breast cancers diagnosed after an indeterminate screening

test, but without any follow-up mammography or diagnostic examinations. A true-positive screening result was a positive result in a participant who was diagnosed with breast cancer by further diagnostic work-up. A false-positive screening result was a positive result in the absence of breast cancer. A true-negative screening result was a negative result in the absence of breast cancer and a false-negative test result was a negative result followed by diagnosis of interval cancer.

Sample size calculation

Sample size was primarily calculated by expected detection and incidence data from breast cancer. According to the positive detection rate of Asian women with suspicious breast cancer lesion (5/1000), 16,000 participants were required. Based on the reported data by Shanghai Center for Disease Control and Prevention in 2017, the age-specific incidence rate of breast cancer among women aged 35-69 in urban Shanghai was 125.75/100,000, so that 20 breast cancer cases were expected to be screened. Our previous breast cancer screening study showed that the early stage rate of screen-detected breast cancer reached 83%, predicting about 17 breast cancer at early stage would be diagnosed in this study. All this would meet the evaluation needs of endpoint indicators.

Statistical analysis

Numerical data were shown as means and standard deviation (SD) and categorical variables in percentages. Sensitivity was determined via dividing the numbers of true-positive screening by the numbers of true-positive and false-positive screening. Specificity was determined via dividing the numbers of true-negative screening by the numbers of true-negative and false-negative screening. To calculate 95% CIs, we did bootstrapping based on 5000 samples. The significance of the differences of incidences was assessed using Poisson regression. For categorical variables we used Fisher's exact test or likelihood-based χ^2 test.

All analyses were done with IBM SPSS (version 20).

This study was performed according to the guidelines of the Helsinki Declaration and was approved by the Medical Ethics Committee of Fudan University Shanghai Cancer Center (No. 2008223-22). Informed consent was obtained from all individual participants included in the study.