SPECIAL ARTICLE

Combined performance of physical examination, mammography, and ultrasonography for breast cancer screening among Chinese women: a follow-up study

Y. Huang MD,* M. Kang MB,[†] H. Li MB,[‡] J.Y. Li PhD,*
J.Y. Zhang PhD,[§] L.H. Liu MB,[†] X.T. Liu MB,[†] Y. Zhao MS,*
Q. Wang MD,* C.C. Li MPH,* and H. Lee MPH*

ABSTRACT

Objective

We aimed to determine which combination of physical examination (PE), mammography (MAM), and ultrasonography (US) would optimize breast cancer detection in China.

Methods

We conducted a trial of screening with PE, MAM, and US among Chinese women 25 years of age and older. All initial screenings using the three modalities were completed within 30 days of each other, and subjects were followed approximately 1 year later. The performances of the three screening methods used alone, in parallel, or in series were compared. Data were analyzed using exact confidence intervals (CIS) and the McNemar test.

Results

Between March 2009 and July 2011, 3028 eligible women completed all study examinations. At a mean follow-up of 1.3 years, 33 breast cancers were identified in the study population. Mammography detected 28 cancers; US, 24 cancers; and PE, 22 cancers. During the follow-up period, 2 false-negative cases occurred clinically. The highest sensitivity for breast cancer screening (93.9%) was achieved by paralleling MAM with US, but came at the cost of a higher recall rate (12.15%). Using US alone at the first stage, followed by MAM when indicated, offered high specificity (99.4%) and the lowest recall rate (1.82%), which were not reached at the expense of sensitivity (84.8%). Used in series, US and MAM achieved a sensitivity similar to that for the same modalities used in parallel (McNemar *p* > 0.05).

Conclusions

Taking limited health resources into consideration, the strategy of screening with US alone at the first stage, followed by MAM when indicated, may optimize breast cancer detection in most regions of China.

KEY WORDS

Breast cancer, mass screening, mammography, ultrasonography, follow-up study, Chinese women

1. INTRODUCTION

Breast cancer is by far the most common female cancer (21.6/100,000) and one of the primary causes of death (5.7/100,000) among Chinese women, with an estimated 169,452 new cases and 44,908 deaths in 2008 [1]. The total annual number of new cases in regions with limited health resources continues to increase [1,2].

The causes of breast cancer remain unclear. Early detection is believed to be the best means of reducing mortality and improving postoperative quality of life for patients with breast cancer [3]. Mammography (MAM) is the only evidence-based screening technology available for that purpose, as verified by randomized clinical trials in Western countries [4]. Although organizations have released conflicting recommendations on the risks and benefits of MAM in clinical and societal contexts [5], many Western countries have successfully implemented mass MAM screening for breast cancer [6]. However, evidence to justify population-based breast cancer screening by MAM for women in China is currently lacking. To date, no randomized controlled trials of MAM screening have been conducted among Chinese women, and evidence that such screening would reduce breast cancer mortality in China is insufficient.

One disadvantage of MAM is its limitations in detecting small lumps in dense breast tissue, because the dense parenchyma in women before menopause can obscure tumour shadows [7,8]. Dense parenchyma is common in Chinese women [9]. Additionally, mass MAM screening in China would likely not be cost effective. Wong and colleagues [10] developed a Markov model to simulate MAM screening, breast cancer

diagnosis, and treatment in a hypothetical cohort of Chinese women and reported that, compared with no screening, at least US$61,600 would be spent for each quality-adjusted life year, which is significantly higher than the standard of US$50,000 per quality-adjusted life year accepted by the World Health Organization [11]. As a result, mass MAM screening has not been recommended in China, especially in regions with limited health resources.

Apart from MAM, physical examination (PE) and ultrasonography (US) are the techniques most commonly used to diagnose breast cancer. So far, no studies have proved a survival benefit for screening by PE, whether performed by physicians or patients, in Asian countries [12–14]. Ultrasonography has been used widely in China because it is safe, cheap, and convenient, and it detects lumps more sensitively in dense breast tissue [15]. However, although US has been suggested both in the West and in Asia as an adjunct tool to elevate the cancer detection rate of MAM or PE alone [7,8,16,17], its utility as a population-based breast cancer screening method is doubtful because of its high rates of both false positives and false negatives [18]. To avoid such problems, combined screening modalities, including parallel and series approaches, are currently used in Asian countries. In developed regions (Japan, for example), screening MAM paralleled with US is recommended for breast cancer control [16]; in developing regions (for instance, China), mass screenings begin with PE, which is followed with imaging techniques (MAM or US) if suspicious symptoms are found [19].

However, the screening performances of various combination approaches (parallel and series) of PE, MAM, and US for breast cancer have not been compared in a Chinese population. Therefore, to determine the combination that optimizes breast cancer detection in China and other Asian countries with limited health resources, we conducted a blind comparison trial with a 1-year follow-up.

2. METHODS

2.1 Study Subjects

Study subjects consisted of an organized screening population (ORSP—healthy physical examinees recruited from the Qingyang community in Chengdu by local government) and an opportunistic screening population (OPSP—outpatients recruited from the Chengdu Women's and Children's Central Hospital). Women who were 25 years of age or older, who had resided in local communities for more than 3 years, and who were willing to participate in the screening were eligible for the study. Pregnant or lactating women were asked to defer their participation. Women with existing untreated malignancies, known metastatic disease, or psychiatric conditions that precluded fully informed consent were excluded from the study.

2.2 Study Protocol and Screening Procedure

The study protocol was reviewed and approved by the Institutional Review Board of Sichuan University, and written informed consent was obtained from all study participants. The trial consisted of two rounds of breast cancer screening. At the primary screening, all eligible women underwent PE, MAM, and US within a period of 30 days. One year later, a second screening with PE was performed to detect false-negative cases.

2.3 Mammography

All MAM examinations were performed using a Giotto MAM unit (Giotto 2000: I.M.S. Internazionale Medico Scientifica, Bologna, Italy) and a dedicated cassette (Fujifilm, Tokyo, Japan). Conventional 4-view film or screen mammograms (mediolateral–oblique and craniocaudal views) were obtained and read independently by 2 radiologists who had more than 10 years of clinical experience in breast imaging. Magnified spot views were added if necessary. The reader was blinded to the results of other screening examinations. Diagnoses by MAM were coded according to the Breast Imaging Reporting and Data System (BI-RADS) categories: 1, negative; 2, benign findings; 3, probably benign findings, short-term follow-up recommended; 4, suspicious abnormality, biopsy recommended; and 5, highly suggestive of malignancy.

The mammographic density of the breast tissue was evaluated according to the BI-RADS density categories: 1, <25% dense (almost entirely fat); 2, 25%–50% dense (scattered fibroglandular tissue); 3, 51%–75% dense (heterogeneous); and 4, >75% dense (extremely dense breast). For analyses, category 1 was categorized as "fatty breasts," and categories 2–4 were defined as dense breasts [7].

2.4 Ultrasonography

An experienced physician who was blinded to the results of the other screening examinations used a color Doppler US system with a 12-MHz transducer (Siemens G60S: Siemens Medical Solutions, Malvern, PA, U.S.A.) to systematically examine the entire breast and regional lymphatic areas. The examiner had more than 10 years of clinical experience in breast imaging. The US images were also interpreted using the BI-RADS categories.

2.5 Physical Examination

Clinical PE of the breasts and regional lymphatic areas (including the lateral and medial borders and axilla) was performed by a surgeon who had more than 20 years of clinical experience in breast diseases. The examiner was blinded to the results of the MAM and US examinations. The PE findings were classified into

5 categories: 1, negative; 2, benign findings; 3, probably benign findings; 4, suspicious abnormality; and 5, highly suggestive of malignancy [17].

2.6 Breast Biopsies

Women in diagnostic categories 3–5 after the first screening were recalled to undergo the next examination. The diagnoses of women who were rated category 3 by any screening technique were compared with the results of the other two examinations to determine whether short-term follow-up (repeat examination after 6 months) or biopsy was indicated. Women who were rated category 4 or 5 by any of the screening techniques were recommended to undergo a biopsy. Fine-needle aspiration cytology, core-needle biopsy, or surgical biopsy was selected according to the patient's condition. Not all solid masses were subjected to biopsy.

2.7 Follow-Up and Validation of Examination Diagnoses

All normal and benign results were followed by a telephone interview and a second round of screening with PE. For each subject, the final diagnosis was reached by either histology (mostly for positive outcomes) or follow-up (mostly for negative outcomes). If no evidence of breast cancer was found at the primary screening, but breast cancer was identified clinically (for example, it became palpable) during the 1-year follow-up or at the second screening, we reviewed the follow-up data to determine whether it constituted a false negative case or a new tumour.

2.8 Clinical Management

All breast cancers detected in the present study were transferred to West China Hospital or Sichuan Cancer Hospital and treated according to the criteria of the National Comprehensive Cancer Network.

2.9 Screening Modalities

We designed 11 different screening modalities (single, in parallel, and in series) and tested them statistically using data from the 3 initial examinations and the follow-up. For the single and parallel modalities, we regarded women who were rated category 4 or 5 by any screening method as positive cases and recommended that they undergo biopsy; for series modalities, we regarded women who were rated category 3 or higher at the first stage of screening as suspicious cases, and we assumed that they would undergo the next stage of screening. Diagnostic categories 1–3 for each screening modality were defined as negative results; categories 4 and 5 were defined as positive results.

2.10 Statistical Analysis

Data were input into EpiData 3.1 (EpiData Association, Odense, Denmark) with dual-entry verification, consistency, and logic error checking. Statistical analyses were performed using the SPSS software application (version 18.0: SPSS, Chicago, IL, U.S.A.), with $p < 0.05$ as the threshold of significance. Descriptive analyses used analysis of variance and the chi-square test to estimate the statistical differences in the risk factors for breast cancer of the women of various demographics undergoing the screening trial. The tumour sizes, lymph node statuses, and TNM stages of breast cancers detected in screening populations were compared by chi-square test. Using the results of biopsies and 1-year follow-ups as the "gold standard," the number of true positives, true negatives, false positives, and false negatives for each screening method or each combination (parallel or series) was determined. The cancer detection rate and the sensitivity, specificity, positive predictive value, negative predictive value, and accuracy of each screening modality for breast cancer were also calculated. McNemar tests were performed to compare the statistical differences in sensitivity and specificity between screening modalities for subgroups of women identified by age, body mass index, and breast density.

3. RESULTS

3.1 General Characteristics of Study Subjects

Between March 2009 and July 2011, 3028 eligible women (ORSP and OPSP) underwent 2 rounds of breast cancer screening, and all were subsequently followed for more than 1 year (mean observation period: 1.3 years). Compared with subjects from the OPSP, those from the ORSP were younger and more likely to have a lower body mass index and dense breasts. The proportions of postmenopausal women and of women with histories of benign breast disease were higher in the OPSP group than in the ORSP group (Table I).

Among the 3028 eligible women, 33 with breast cancer (1.1%) were indentified, including 18 (54.5%) with early-stage disease (TNM 0, I, and II). Another 840 women (27.7%) were found to have benign breast disease. Compared with the women having benign or normal results, the women with cancer were significantly older and had a higher mean body mass index. Compared with women in the benign and normal groups, those in the group found to have cancer were more likely to have fatty breasts (BI-RADS density category 1) and to be postmenopausal. Women found to have benign breast disease were particularly liable to relapse (Table I).

TABLE I   Risk factors for breast cancer in women enrolled in the screening trial

Characteristic	Screening population (n=3028)			Screening results (n=3028)			
	Organized (n=2471)	Opportunistic (n=557)	Comparative statistic	Cancer (n=33)	Benign (n=840)	Normal (n=2155)	Comparative statistic
Age (years)							
Mean[a]	44.25±6.10	49.11±11.33	198.360[b,c]	49.82±10.54	44.61±7.67	45.28±7.47	8.726[b,c]
Range	25–68	27–80		33–72	25–75	27–80	
<40 (n=587)	468	119	132.428[c,d]	4	189	394	25.367[c,d]
40–50 (n=1841)	1605	236		13	504	1324	
>50 (n=600)	398	202		16	147	437	
Body mass index (kg/m²)							
Mean[a]	21.99±2.48	22.27±2.49	5.731[b,c]	22.86±1.90	21.92±2.22	22.08±2.59	3.148[b,c]
Range	16.41–37.78	16.18–31.25		18.37–26.04	16.41–30.82	16.18–37.78	
Breast density by BI-RADS [n (%)]							
Almost entirely fat	355 (14.4)	111 (19.9)	15.381[c,d]	9 (27.3)	83 (9.9)	363 (16.8)	78.499[c,d]
Scattered fibroglandular tissue	931 (37.7)	173 (31.1)		12 (36.4)	251 (29.9)	852 (39.5)	
Heterogeneously dense	706 (28.6)	170 (30.5)		8 (24.2)	285 (33.9)	583 (27.1)	
Extremely dense	479 (19.4)	103 (18.5)		4 (12.1)	221 (26.3)	357 (16.6)	
Menstrual status [n (%)]							
Premenopausal	2084 (84.3)	370 (66.4)	94.916[c,d]	19 (57.6)	688 (81.9)	1747 (81.1)	12.236[c,d]
Postmenopausal	387 (15.7)	187 (33.6)		14 (42.4)	152 (18.1)	408 (18.9)	
History of hormone use [n (%)]							
Yes	263 (10.6)	53 (9.5)	0.619[d]	5 (15.2)	90 (10.7)	221 (10.3)	1.207[e]
No	2208 (89.4)	504 (90.5)		28 (84.8)	750 (89.3)	1934 (89.7)	
Family history of breast cancer [n (%)]							
Yes	63 (2.5)	17 (3.1)	0.446[d]	2 (6.1)	22 (2.6)	56 (2.6)	1.984[e]
No	2408 (97.5)	540 (96.9)		31 (93.9)	818 (97.4)	2099 (97.4)	
History of benign breast disease [n (%)]							
Yes	146 (5.9)	61 (11.0)	18.150[c,d]	2 (6.1)	104 (12.4)	101 (4.7)	50.661[c,e]
No	2325 (94.1)	496 (89.0)		31 (93.9)	736 (87.6)	2054 (95.3)	
History of breast cancer [n (%)]							
Yes	10 (0.4)	3 (0.5)	0.191[e]	1 (3.0)	2 (0.2)	10 (0.4)	4.798[e]
No	2461 (99.6)	554 (99.5)		32 (97.0)	838 (99.8)	2145 (99.6)	

a   With standard deviation.
b   By analysis of variance.
c $p < 0.05$—That is, statistically significant difference between the subgroups.
d   By chi-square test.
e   By Fisher exact test.

BI-RADS = Breast Imaging Reporting and Data System (categories: 1, negative; 2, benign findings; 3, probably benign findings; 4, suspicious abnormality, biopsy recommended; and 5, highly suggestive of malignancy).

3.2 Breast Cancers Detected by the Various Screening Methods

Compared with breast cancers detected in the OPSP, cancers detected in the ORSP were mostly identified in dense breasts (BI-RADS density categories 2–4) and were significantly smaller and at earlier TNM stages (Table II). Of the 33 breast cancers, 22 were recognized by PE, 28 by MAM, and 24 by US. The MAM modality recognized more breast cancers with small size (<2 cm), negative lymph nodes, and early TNM stage (0, I, or II). One cancer at TNM stage I was detected only by MAM, and not by US or PE. Compared with MAM, US showed a poor capacity to detect cancers with negative lymph nodes and at early TNM stages. However, US detected 3 cancers in dense breasts that went undetected by MAM and PE. The PE modality detected no cases that were missed by MAM, and it missed more cancers within dense breasts. All tumours identified by PE were at least 1 cm in diameter (Table II).

Of the 33 breast cancers, 31 (93.9%) were detected by primary screening. The other 2 were identified clinically during the follow-up period and were considered false negatives, because those women had undergone 3 examinations during the first round of screening, each with a negative result. At the second round of screening in October 2010, 1 woman with a pathology diagnosis of carcinoma *in situ* (TNM 0) was found; we regarded that finding as a new case and excluded it from the statistical analysis.

TABLE II   Diagnostic yields for 33 breast cancers by screening population and screening method

Variable	Cases[a] (n=33)	Detections by screening population[a]			Detections by screening method[b]		
		Organized (n=14)	Opportunistic (n=19)	Fisher exact test[a]	PE (n=22)	MAM (n=28)	US (n=24)
Tumour size, centimetres [n (%)]							
<1	3 (9.1)	3 (21.4)	0 (0.0)	6.745[c]	0 (0.0)	1 (33.3)	2 (66.7)
1–2	13 (39.4)	7 (50.0)	6 (31.6)		10 (76.9)	13 (100.0)	11 (84.6)
>2	17 (51.5)	4 (28.6)	13 (68.4)		12 (70.6)	14 (82.4)	11 (64.7)
Nodal status [n (%)]							
Positive	23 (69.7)	9 (64.3)	14 (73.7)	0.337	14 (60.9)	19 (82.6)	20 (87.0)
Negative	10 (30.3)	5 (35.7)	5 (26.3)		8 (80.0)	9 (90.0)	4 (40.0)
TNM stage							
0 or I	3 (9.1)	3 (21.4)	0 (0.0)	7.392[c]	1 (33.3)	2 (66.7)	0 (0.0)
II	15 (45.5)	8 (57.1)	7 (36.8)		11 (73.3)	12 (80.0)	9 (60.0)
III	15 (45.5)	3 (21.4)	12 (63.2)		10 (66.7)	14 (93.3)	15 (100.0)
Pathologic diagnosis [n (%)]							
Carcinoma *in situ*	1 (3.0)	1 (7.1)	0 (0.0)	2.617	0 (0.0)	0 (0.0)	0 (0.0)
Invasive ductal carcinoma	26 (78.8)	11 (78.6)	15 (78.9)		16 (61.5)	22 (84.6)	21 (80.8)
Other invasive carcinoma	2 (6.1)	0 (0.0)	2 (10.5)		2 (100.0)	2 (100.0)	1 (50.0)
Other	4 (12.1)	2 (14.3)	2 (10.5)		4 (100.0)	4 (100.0)	2 (50.0)
Breast density, BI-RADS [n (%)]							
Almost entirely fat	9 (27.3)	1 (7.1)	8 (42.1)	8.870[c]	6 (66.7)	9 (100.0)	9 (100.0)
Scattered fibroglandular tissue	12 (36.4)	5 (35.7)	7 (36.8)		8 (66.7)	10 (83.3)	8 (66.7)
Heterogeneously dense	8 (24.2)	4 (28.6)	4 (21.1)		5 (62.5)	6 (75.0)	4 (50.0)
Extremely dense	4 (12.1)	4 (28.6)	0 (0.0)		3 (75.0)	3 (75.0)	3 (75.0)

[a]   Proportion of specified screening population within each subgroup.
[b]   Detection rate for the specified screening method within each subgroup.
[c] $p < 0.05$ (Indicates a statistically significant difference between the various screening populations for that subgroup).
PE = physical examination; MAM = mammography; US = ultrasonography; BI-RADS = Breast Imaging Reporting and Data System (categories: 1, negative; 2, benign findings; 3, probably benign findings, short-term follow-up recommended; 4, suspicious abnormality, biopsy recommended; and 5, highly suggestive of malignancy).

3.3 Performance Characteristics of Screening Modalities

Our study analyzed 11 screening modalities, including single modalities, parallel modalities, and series modalities (Table III). Among the single modalities, MAM identified the fewest suspicious cases (*n* = 105) and detected the most cancers (*n* = 28). More category 3 results were diagnosed by PE and US, which could not discern benign from malignant lumps. The sensitivity of MAM for breast malignancy was 84.8%, which was higher than that for US (72.7%) or PE (66.7%). However, the specificity of MAM (98.4%) was lower than that for US (99.4%) or PE (99.2%). Compared with the single modalities, the 3 parallel modality combinations were more sensitive and detected more cancers, but their specificities were lower and more suspicious cases were recalled. The 4 series modality combinations had the highest specificities and lowest recall numbers, but their sensitivities and cancer detection rates were lower.

Modality 6 (MAM in parallel with US) identified the most cancer cases (31 of 33) and had the highest sensitivity (93.9%), but at the cost of a higher recall rate (12.15%). Modality 9 (US alone at the start, and then MAM when indicated) offered high specificity (99.4%) and the lowest recall rate (1.82%); it was also comparatively sensitive (84.8%). The sensitivity (84.4%) and cancer detection rate (0.92%) of modality 9 were higher than those of modality 3 (US alone) and equivalent to those of modality 2 (MAM alone). Modality 11, a combination of PE, MAM, and US, did not improve screening performance compared with modality 9 (Table III).

We observed no significant difference in sensitivity between modality 6 and modality 9 for any subgroup of women identified by age, body mass index, or breast density. Modality 9 had a higher specificity than modality 6 in most subgroups (Table IV).

4. DISCUSSION

Although China has a relatively low incidence of breast cancer (21.6/100,000), the incidence has recently climbed rapidly, as has the proportion of young women diagnosed with breast cancer[1,2]. The purpose of screening is to detect cancer at an early stage (while the tumour is small and lymph nodes are not yet involved), thereby improving postoperative quality of life, reducing mortality, and lightening the financial burden of prolonged cancer treatment. However, to date, no guidelines have been set for early breast cancer detection in China.

The screening strategies of Western countries may not be suitable for China because of the many differences between Asian and Caucasian women with respect to the physiology of mammary glands (breast size and density) and the clinical characteristics of breast cancers (peak age incidence)[9]. The strategies of wealthier countries, including developed Asian countries, may also be unsuitable for China because of differences in health care resources. Thus, we conducted a prospective blind trial to determine the screening modality or combination of modalities that optimizes breast cancer detection in China.

Our study identified 33 cancers, with 28 being detected by MAM, 24 by US, and 22 by PE. Our results show that MAM is more sensitive than US or PE among Chinese women, although no method found all of the cancers, which is consistent with previously published studies in Asian women[20,21]. More early-stage cancers with microcalcifications, small size, and no lymph node involvement were identified by MAM, which is consistent with other evidence that has made MAM the method of choice for breast cancer screening in Western countries[3]. However, 3 cancers in dense breasts were detected by US but not by MAM in our trial. Physical examination did not identify cancers missed by MAM. Ultrasonography has proven effective in detecting clinically and mammographically occult cancers in the dense parenchyma of women before menopause[7–8,16,17], a common circumstance among Chinese women with breast cancer. And US is a useful tool for differentiating cystic lesions from solid tumours in the breast, potentially lowering the rate of unnecessary biopsies for benign lesions by up to 25%[22]. However, compared with MAM, the sensitivity of US depends more on the operator's experience and the equipment used[21,22].

On the other hand, the recall rate with MAM was lower than those with US and PE. The poor capacity of US or PE to differentiate benign from malignant tumours led to higher recall and biopsy rates, which brought about overdiagnosis and overtreatment. False-positive outcomes may affect the well-being and behavior of women and may also lead to greater expense[23]. Thus, the use of US or PE alone as a screening method for the general population is not practical at present because of inadequate sensitivity and unsatisfactory false-positive rates.

We investigated whether a combination of screening methods might be superior to any single modality. We analyzed 8 combinations of screening modalities, including parallel and series approaches, to determine which combination achieves both high sensitivity and high specificity among Chinese women. Compared with the single modalities, the parallel combinations increased sensitivity and decreased specificity, and the series modes decreased sensitivity and increased specificity. Of the combined modalities, the highest sensitivity and highest cancer detection rate were achieved by MAM in parallel with US (modality 6), a result that accords with results from previous studies in Asia[16,17,24]. It is worthwhile to note that US alone at the first stage, followed by MAM when indicated, (modality 9) yields a similar cancer detection rate and a lower recall rate. In modality 9, women with category 3 US results ("probably benign") entered the next stage of screening by MAM to determine whether a lump was malignant, reducing the false-positive and false-negative diagnoses.

TABLE III Solo and combined performance of the various screening modalities at 1-year follow-up in 3028 women with confirmed 33 breast cancer cases

Screening modality or modalities	Recalls [n (%)][a]	Cancers detected [n (%)]	Positives True (n)	Positives False (n)	Negatives True (n)	Negatives False (n)	Sensitivity[b] (95% CI)	Specificity[c] (95% CI)	Predictive value Positive[d] (95% CI)	Predictive value Negative[e] (95% CI)	Accuracy[f] (95% CI)
Modality 1 PE alone	257 (8.49)	22 (0.73)	22	23	2972	11	0.667 (0.482 to 0.820)	0.992 (0.988 to 0.995)	0.489 (0.337 to 0.642)	0.996 (0.993 to 0.998)	0.989 (0.984 to 0.992)
Modality 2 MAM alone	105 (3.47)	28 (0.92)	28	48	2947	5	0.848 (0.681 to 0.949)	0.984 (0.979 to 0.988)	0.368 (0.261 to 0.487)	0.998 (0.996 to 0.999)	0.982 (0.977 to 0.987)
Modality 3 US alone	318 (10.50)	24 (0.79)	24	19	2976	9	0.727 (0.545 to 0.867)	0.994 (0.990 to 0.996)	0.558 (0.399 to 0.709)	0.997 (0.994 to 0.999)	0.991 (0.987 to 0.994)
Modality 4 MAM paralleled with PE	315 (10.40)	28 (0.92)	28	60	2935	5	0.848 (0.681 to 0.949)	0.980 (0.974 to 0.985)	0.318 (0.223 to 0.426)	0.998 (0.996 to 0.999)	0.979 (0.973 to 0.983)
Modality 5 US paralleled with PE	420 (13.87)	30 (0.99)	30	36	2959	3	0.909 (0.757 to 0.981)	0.988 (0.983 to 0.992)	0.455 (0.331 to 0.582)	0.999 (0.997 to 1.000)	0.987 (0.982 to 0.991)
Modality 6 MAM paralleled with US	368 (12.15)	31 (1.02)	31	59	2936	2	0.939 (0.798 to 0.993)	0.980 (0.975 to 0.985)	0.344 (0.247 to 0.452)	1.000 (0.998 to 1.000)	0.980 (0.974 to 0.985)
Modality 7 PE followed by MAM when indicated	47 (1.55)	23 (0.76)	23	17	2978	10	0.697 (0.513 to 0.844)	0.994 (0.991 to 0.997)	0.575 (0.409 to 0.730)	0.997 (0.994 to 0.999)	0.991 (0.987 to 0.994)
Modality 8 PE followed by US when indicated	155 (5.12)	20 (0.66)	20	10	2985	13	0.606 (0.421 to 0.771)	0.997 (0.994 to 0.998)	0.667 (0.472 to 0.827)	0.996 (0.993 to 0.998)	0.992 (0.989 to 0.995)
Modality 9 US followed by MAM when indicated	55 (1.82)	28 (0.92)	28	19	2976	5	0.848 (0.681 to 0.949)	0.994 (0.990 to 0.996)	0.596 (0.443 to 0.736)	0.998 (0.996 to 0.999)	0.992 (0.988 to 0.995)
Modality 10 MAM followed by US when indicated	55 (1.82)	22 (0.73)	22	9	2986	11	0.667 (0.482 to 0.820)	0.997 (0.994 to 0.999)	0.710 (0.520 to 0.858)	0.996 (0.993 to 0.998)	0.993 (0.990 to 0.996)
Modality 11 PE paralleled with US, followed by MAM when indicated	62 (2.05)	28 (0.92)	28	23	2972	5	0.848 (0.681 to 0.949)	0.992 (0.988 to 0.995)	0.549 (0.403 to 0.689)	0.998 (0.996 to 0.999)	0.991 (0.987 to 0.994)

a   Women who were rated BI-RADS category 3 or higher by the given screening modality were recalled for further examination (either short-term follow-up and re-examination, or biopsy).
b   Calculated as true positives divided by the sum of true positives and false negatives.
c   Calculated as true negatives divided by the sum of true negatives and false positives.
d   Calculated as true positives divided by the sum of true positives and false positives.
e   Calculated as true negatives divided by the sum of true negatives and false negatives.
F   Calculated as the sum of the true positives and true negatives divided by all detections.
CI = confidence interval; PE = physical examination; MAM = mammography; US = ultrasonography.

TABLE IV   Sensitivity and specificity of mammography (MAM) paralleled with ultrasonography (US) compared with US followed by MAM in various subgroups among 3028 women with confirmed 33 breast cancer cases

Subgroup	Members (n)	Sensitivity			Specificity		
		Modality 6	Modality 9	p *Value*[a]	Modality 6	Modality 9	p *Value*[a]
All eligible women	3028	0.939 (31/33)	0.848 (28/33)	0.375	0.980 (2936/2995)	0.994 (2976/2995)	<0.001
Age (years)							
<50	2343	1.000 (17/17)	0.824 (14/17)	—[b]	0.984 (2288/2326)	0.996 (2316/2326)	<0.001
≥50	685	0.875 (14/16)	0.875 (14/16)	1.000	0.969 (648/669)	0.987 (660/669)	0.017
Body mass index (kg/m^2)							
<25	2714	0.929 (26/28)	0.821 (23/28)	0.250	0.981 (2635/2686)	0.997 (2679/2686)	<0.001
≥25	314	1.000 (5/5)	1.000 (5/5)	—[b]	0.981 (303/309)	0.994 (307/309)	0.289
Breast density by BI-RADS							
Fatty breast	467	1.000 (9/9)	1.000 (9/9)	—[b]	0.978 (448/458)	0.993 (455/458)	0.065
Dense breast	2561	0.917 (22/24)	0.792 (19/24)	0.250	0.981 (2490/2537)	0.994 (2521/2537)	<0.001

[a]   By McNemar test, *p* < 0.05 for modality 6 compared with modality 9 (statistically significant difference).
[b]   Not computed because of a small sample number.
Modality 6 = MAM paralleled with US; Modality 9 = US alone at the first stage, followed by MAM when indicated.

The screening accuracies (sensitivity and specificity) of both modality 6 (MAM and US in parallel) and modality 9 (US followed by MAM when indicated) were acceptable. However, accuracy is not the only relevant factor in making a modality practical and beneficial for nationwide mass cancer screening. Other factors that need to be considered include the modality's likely effects on breast cancer mortality, financial cost, population selection, technical disparities, appropriate intervals between screenings, and optimal start age for screening. Reducing breast cancer mortality is the most important index of cancer screening, but it takes several decades to show true effectiveness.

With respect to screening cost, MAM in China costs about 240 yuan (approximately US$36.00), while US costs about 90 yuan (approximately US$13.50). Thus, modality 6 would be the most expensive of the screening modalities tested and could likely be implemented only in a few developed regions in China. The screening cost of modality 9 would be lower than that of MAM alone (modality 2), and we expect that it would result in better screening compliance because US requires less exposure to radiation than MAM and is available at more hospitals [25]. However, economic evaluations are needed to further investigate whether modality 9 would be suitable as a national breast cancer screening strategy for a general population or for high-risk populations in developing or underdeveloped regions in China and other Asian countries.

Our trial included both ORSP and OPSP. More early-stage cancers of small size and with an absence of lymph node involvement were detected in the ORSP, but the overall breast cancer detection rate was higher in the OPSP. Further evaluations are needed to determine which screening mode is more cost effective.

The strengths of the present study include the use of blinding to avoid examiner bias, and the fact that screening performances were calculated on the basis of either histology or follow-up results, which revealed the true performances of the various screening modalities. However, the study also has several limitations. First, it was conducted at a single site, which might have resulted in selection bias. Large-sample and multicentre clinical studies, or a meta-analysis, are needed to verify the general applicability of our findings. Second, the proportion of early-stage cancers detected in our study was small, which may have lessened the power of the study to compare the performances of various screening modalities. Third, some subgroup analyses were based on small sample sizes, which limited their statistical power.

5.   CONCLUSIONS

Screening with US alone at the first stage, followed by MAM when indicated, may optimize breast cancer detection in regions of China with limited health resources. This finding may also be applicable to other

developing East Asian countries. However, analyses of cost effectiveness and survival benefits are needed to more clearly address whether such a strategy would ultimately reduce mortality in a cost-effective way.

6. ACKNOWLEDGMENTS

We thank Shawna Williams for editing this text.

7. CONFLICT OF INTEREST DISCLOSURES

The authors have no financial conflicts of interest to report.

8. REFERENCES

1. Ferlay J, Shin HR, Bray F, Forman D, Mathers C, Parkin DM. GLOBOCAN 2008. Cancer incidence, mortality and prevalence worldwide in 2008. Lyon, France: International Agency for Research on Cancer; 2010.
2. PR China, Ministry of Health, Disease Prevention and Control Bureau, National Central Cancer Registry, National Office for Cancer Prevention and Control. *Chinese Cancer Registry Annual Report* [Chinese]. Beijing, PR China: Military Medical Science Press; 2010.
3. Humphrey LL, Helfand M, Chan BK, Woolf SH. Breast cancer screening: a summary of the evidence for the U.S. Preventive Services Task Force. *Ann Intern Me*d 2002;137:347–60.
4. Strax P. Control of breast cancer through mass screening: from research to action. *Cancer* 1989;63:1881–7.
5. Hirsch BR, Lyman GH. Breast cancer screening with mammography. *Curr Oncol Rep* 2011;13:63–70.
6. Shapiro S, Coleman EA, Broeders M, *et al.* Breast cancer screening programmes in 22 countries: current policies, administration and guidelines. *Int J Epidemiol* 1998;27:735–42.
7. Kolb TM, Lichy J, Newhouse JH. Comparison of the performance of screening mammography, physical examination, and breast US and evaluation of factors that influence them: an analysis of 27 825 patient evaluations. *Radiology* 2002;225:165–75.
8. Buchberger W, DeKoekkoek–Doll P, Springer P, Obrist P, Dünser M. Incidental findings on sonography of the breast: clinical significance and diagnostic workup. *AJR Am J Roentgenol* 1999;173:921–7.
9. Li J, Zhang BN, Fan JH, *et al.* A nation-wide multicenter 10-year (1999–2008) retrospective clinical epidemiological study of breast cancer in China. *BMC Cancer* 2011;11:364.
10. Wong IO, Kuntz KM, Cowling BJ, Lam CL, Leung GM. Cost effectiveness of mammography screening for Chinese women. *Cancer* 2007;110:885–95. [Erratum in: *Cancer* 2010;116:540]
11. Owens DK. Interpretation of cost-effectiveness analyses. *J Gen Intern Med* 1998;13:716–17.
12. Ota J, Horino T, Taguchi T, *et al.* Mass screening for breast cancer: comparison of the clinical stages and prognosis of breast cancer detected by mass screening and in out-patient clinics. *Jpn J Cancer Res* 1989;80:1028–34.
13. Morimoto T, Komaki K, Mori T, *et al.* The quality of mass screening for breast cancer by physical examination. *Surg Today* 1993;23:200–4.
14. Gao DL, Hu YW, Wang WW, *et al.* Evaluation on the effect of intervention regarding breast self-examination for decreasing breast cancer mortality (Chinese). *Zhonghua Liu Xing Bing Xue Za Zhi* 2006;27:985–90.
15. Lin BY. *Breast Cancer* [Chinese]. Beijing, PR China: China Medical Science and Technology Press; 2007: 110,131,134–7.
16. Uchida K, Yamashita A, Kawase K, Kamiya K. Screening ultrasonography revealed 15% of mammographically occult breast cancers. *Breast Cancer* 2008;15:165–8.
17. Honjo S, Ando J, Tsukioka T, *et al.* Relative and combined performance of mammography and ultrasonography for breast cancer screening in the general population: a pilot study in Tochigi prefecture, Japan. *Jpn J Clin Oncol* 2007;37:715–20.
18. Teh W, Wilson AR. The role of ultrasound in breast cancer screening. A consensus statement by the European Group for Breast Cancer Screening. *Eur J Cancer* 1998;34:449–50.
19. Huang Y, Zhou K, Li H, *et al.* Knowledge, attitudes, and behavior regarding breast cancer screening among women from different socio-economic regions in southwest China: a cross-sectional study. *Asian Pac J Cancer Prev* 2011;12:203–9.
20. Kang M, Pang Y, Li JY, Liu LH, Liu XT. Accuracy evaluation of mammography in the breast cancer screening in Asian women: a community-based follow-up study and meta analysis [Chinese]. *Zhonghua Zhong Liu Za Zhi* 2010;32:212–16.
21. Huang Y, Pang Y, Wang Q, Li JY. Evaluation on the accuracy of high-frequency ultrasound being used in the breast cancer screening program in women from Asian countries: a systematic review [Chinese]. *Zhonghua Liu Xing Bing Xue Za Zhi* 2010;31:1296–9.
22. Yang WT, Mok CO, King W, Tang A, Metreweli C. Role of high frequency ultrasonography in the evaluation of palpable breast masses in Chinese women: alternative to mammography? *J Ultrasound Med* 1996;15:637–44.
23. Brewer NT, Salz T, Lillie SE. Systematic review: the long-term effects of false-positive mammograms. *Ann Intern Med* 2007;146:502–10.
24. Ji YJ, Chang C, Ding JH, Zeng W, Chen M, Liu GY. The value of breast ultrasound in a mammography-based Chinese breast screening project [Chinese]. *J Surg Concepts Pract* 2010;15:486–9.
25. Tsai HW, Twu NF, Ko CC, *et al.* Compliance with screening mammography and breast sonography of young Asian women. *Eur J Obstet Gynecol Reprod Biol* 2011;157:89–93.

Correspondence to: Jiayuan Li, Department of Epidemiology, West China School of Public Health, Sichuan University, 16 South Road RenMin, Chengdu, Sichuan Province 610041 PR China.
E-mail: lijiayuan73@163.com

* Department of Epidemiology, West China School of Public Health, Sichuan University, Chengdu, Sichuan, PR China.
† Department of Radiology, Chengdu Women's and Children's Central Hospital, Chengdu, Sichuan, PR China.
‡ Department of Breast, Second People's Hospital of Sichuan, Chengdu, Sichuan, PR China.
§ Department of Biostatistics, West China School of Public Health, Sichuan University, Chengdu, Sichuan, PR China.