A multi-centre randomised trial comparing ultrasound vs mammography for screening breast cancer in high-risk Chinese women

The Harvard community has made this article openly available. **Please share** how this access benefits you. Your story matters

<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Published Version</td>
<td>doi:10.1038/bjc.2015.33</td>
</tr>
<tr>
<td>Citable link</td>
<td><a href="http://nrs.harvard.edu/urn-3:HUL.InstRepos:15034941">http://nrs.harvard.edu/urn-3:HUL.InstRepos:15034941</a></td>
</tr>
<tr>
<td>Terms of Use</td>
<td>This article was downloaded from Harvard University's DASH repository, and is made available under the terms and conditions applicable to Other Posted Material, as set forth at <a href="http://nrs.harvard.edu/urn-3:HUL.InstRepos:dash.current.terms-of-use#LAA">http://nrs.harvard.edu/urn-3:HUL.InstRepos:dash.current.terms-of-use#LAA</a></td>
</tr>
</tbody>
</table>
A multi-centre randomised trial comparing ultrasound vs mammography for screening breast cancer in high-risk Chinese women

S Shen, Y Zhou, Y Xu, B Zhang, X Duan, R Huang, B Li, Y Shi, Z Shao, H Liao, J Jiang, N Shen, J Zhang, C Yu, H Jiang, S Han, J Ma and Q Sun*

Department of Breast Surgery, Peking Union Medical College Hospital, Beijing, China

Background: Chinese women tend to have small and dense breasts and ultrasound is a common method for breast cancer screening in China. However, its efficacy and cost comparing with mammography has not been evaluated in randomised trials.

Methods: At 14 breast centres across China during 2008–2010, 13 339 high-risk women aged 30–65 years were randomised to be screened by mammography alone, ultrasound alone, or by both methods at enrolment and 1-year follow-up.

Results: A total of 12 519 and 8692 women underwent the initial and second screenings, respectively. Among the 30 cancers (of which 15 were stage 0/I) detected, 5 (0.72/1000) were in the mammography group, 11 (1.51/1000) in the ultrasound group, and 14 (2.02/1000) in the combined group (P = 0.12). In the combined group, ultrasound detected all the 14 cancers, whereas mammography detected 8, making ultrasound more sensitive (100 vs 57.1%, P = 0.04) with a better diagnostic accuracy (0.999 vs 0.766, P = 0.01). There was no difference between mammography and ultrasound in specificity (100 vs 99.9%, P = 0.51) and positive predictive value (72.7 vs 70.0%; P = 0.87). To detect one cancer, the costs of ultrasound, mammography, and combined modality were $7876, $45 253, and $21 599, respectively.

Conclusions: Ultrasound is superior to mammography for breast cancer screening in high-risk Chinese women.
Breast cancer incidence has been dramatically increased during the past two decades and is now the most common cancer among women in China (Chen et al, 2013; Fan et al, 2014). To address this issue, the Ministry of Health of China and All-China Women’s Federation jointly launched a 3-year national ‘Two Cancer (breast and cervical cancer) Screening’ campaign between 2009 and 2011 to provide free screening for 1.46 million rural women, aged 35–59 years (Ministry of Health of China and All-China Women’s Federation, 2009). This pilot programme has entered its second phase (2012–2015) to provide a free breast cancer screening for six million rural women, aged 35–64 years.

Although mammography-based screening strategy has been widely adopted in developed countries for over 30 years, a few reasons make mammography-based screening less attractive in China. First, Chinese women tend to have small and dense breasts (Stomper et al, 1996; Zulfiqar et al, 2011), which is known to reduce the diagnostic accuracy of mammography (Mandelson et al, 2000). Second, the peak age of breast cancer diagnosis in Chinese women is between 45 and 55 years old, about 10–20 years younger than that in Caucasian women (Li et al, 2011; Fan et al, 2014), and mammography has also been shown to be less effective in younger as compared with older women (Checka et al, 2012; Shen et al, 2012). Finally, many studies have demonstrated that breast ultrasound has the potential of detecting small, node-negative, mammographically occult, invasive breast cancers in young women or women with dense breast, thus improving the effectiveness of screening among these women (Kolb et al, 2002; Crystal et al, 2003; Benson et al, 2004; Berg et al, 2008; Corsetti et al, 2008). Based on these concerns, the Chinese government decided to use ultrasound as the primary method for the ‘Two Cancer Screening’ campaign (Ministry of Health of China and All-China Women’s Federation, 2009). However, so far there is no clear guideline for breast cancer screening in China because of a lack of evidence from randomised trials comparing the effectiveness of ultrasound vs mammography in Chinese women.

We, therefore, conducted a multi-centre randomised trial to compare the performance of ultrasound and mammography for screening breast cancer in Chinese women. We focused this study on high-risk population because breast cancer incidence in China is still relatively low (Fan et al, 2014). To be more cost-effective, we recruited high-risk women by using a questionnaire-based risk-assessment model (Peking Union Medical College Hospital (PUMCH) model) (Shen et al, 2012; Xu et al, 2012), established through our previous case-control study (Xu et al, 2012) and adjusted from the evidence of other risk models (Costantino et al, 1999; Rockhill et al, 2001; Anothaisintawee et al, 2012). We aim to find a breast cancer screening strategy that suits the demographic and socioeconomic conditions in China.

**PATIENTS AND METHODS**

**Study sites and participant eligibility.** This study was a multi-centre randomised trial conducted in 14 breast centres across eight provinces in China (Beijing, Chongqing, Guangdong, Hebei, Jiangsu, Shanghai, Shanxi, and Tianjin) (Supplementary Figure 1). The lead centre was PUMCH.

Between 1 November 2008 and 1 November 2010, a total of 47 709 women aged 30–65 years both from urban and rural areas were evaluated for study eligibility at the 14 sites. Breast cancer risk was assessed based on the PUMCH risk-assessment model, established from a previous case-control study of 416 breast cancer patients and 1156 age- and region-matched controls (Xu et al, 2012). Clinicians conducted in-person interviews with the subjects to collect information on factors including demographic characteristics, life style, reproductive history, and family history. Age, body mass index (kg m$^{-2}$), age of the first menarche, stress anticipation, oral contraceptive use, menopausal status, history of benign breast disease biopsy, and family history of breast cancer in the first-degree relatives were considered as risk factors, after adjusting from other risk models (Costantino et al, 1999; Rockhill et al, 2001; Anothaisintawee et al, 2012). Each factor was given a coefficient score, and the sum of scores of all risk factors constituted the risk score (range: 0–71.8). A woman with a risk score of 30 or higher, which counted for about top 30% of all the women evaluated, was considered as high risk and thus eligible as a participant for the study.

In addition to a PUMCH risk score below 30, exclusion criteria also included pregnancy, lactation, known metastatic cancer, signs or symptoms of breast disease, presence of breast implants, breast surgery within prior 12 months, and those who had a mammography or ultrasound exam within the past 12 months.

**Study design.** As shown in Figure 1, at the initial screening, a total of 13 339 eligible women were randomised into three groups: mammography alone (group 1), ultrasound alone (group 2), or the combined methods (group 3). In the combined group, the participants completed both mammography and ultrasound exams in a randomised order. A screening result was defined as positive if either mammography or ultrasound result was positive. Participants with a positive result underwent biopsy for definitive diagnosis. The physical exam was done after a mammography and/or ultrasound exam, and the physician was not masked to the results of the mammography or ultrasound. If the results of imaging were negative but the physical examination considered suspicious malignancy, a biopsy was also triggered and the results were recorded. Woman with a negative screening result or benign biopsy was invited to be screened again 1 year later. The number of interval cancers was obtained from the interviews with participants and review of medical records. For those who did not come back for the following screening, telephones, mails, E-mails, or face-to-face interviews were used to follow-up. The study database was closed on 1 December 2011. The study was reviewed and approved by the institutional review board of PUMCH. All recruited subjects provided a written informed consent.

This trial was registered at clinicaltrials.gov (identifier: NCT01880853).

**Screening methods and quality assurance.** A standard two-view mammography was performed by using digital mammography. Screening ultrasound was performed by using colour Doppler and high-resolution transducers with maximum frequency of at least 12 MHz, scanning both transverse and sagittal planes by experienced physicians of each centre. The Breast Imaging-Reporting and Data System (BI-RADS) (Liberman and Menell, 2002; Costantino et al, 2006) lexicon was used to organise the interpretation and reporting of both screening modalities. Assessment for each lesion was recorded according to BI-RADS categories: 1, negative; 2, benign; 3, probably benign; 4, suspicious malignancy; and 5, highly suggestive of malignancy.

For quality and consistency among all study centres, experienced radiologists from each centre were further trained to ensure the adequate and unambiguous use of BI-RADS. Each centre was also inspected by lead-centre radiologists for interpretation consistency of at least 95% by 200 random records of mammography and/or ultrasound exams. All mammography and ultrasound examinations were performed and interpreted separately by different physicians, masked to the results of the alternate method.

**Determination of reference standard.** For breast lesions screened more than once, the most severe imaging assessments on mammography or ultrasound were used as the primary end points. BI-RADS scores of the categories 1, 2, and 3 were...
considered negative screening results, whereas scores 4 and 5 were considered positive results (Berg et al, 2008). For those in the combined group, a screening result was defined as positive if either mammography or ultrasound result was positive. A BI-RADS score of greater than 3 was considered a positive test result; a score of 3 or less, negative.

**Cost analysis.** The cost per breast cancer detected was calculated by dividing the total cost of a screening programme by the number of cancers detected (Feig, 2011). The programme cost included consultation, physical exam, one of the three screening modalities, biopsy, and histopathology. Since the screening cost varies among different regions because of diverse economic status in China, for simplicity, we adopted the estimated cost of the China’s ‘Two Cancer Screening’ campaign from the perspective of health care provider, which was 4 RMB (Chinese Yuan) for consultation and physical exam, 70 RMB for ultrasound, and 200 RMB for mammography (Ministry of Health of China and All-China Women’s Federation, 2009). Since core needle biopsy costs at least twice as much as that of open excision biopsy (OEB), OEB is a common practice in China. We therefore estimated the average cost of 400 RMB for biopsy confirmation (300 RMB for OEB plus 100 RMB for histopathology). Our estimated costs will be the lower threshold of the total costs. The final costs were converted to both US dollar and International dollar (measured by Purchasing Power Parity). We used the exchange rate 6.31:1 to convert RMB to US dollar, and 3.5:1 to convert it to International Dollar (World Bank, http://wdi.worldbank.org/table/4.16).

**Statistical consideration.** We calculated the sample size based on the hypothesis that breast ultrasound was expected to improve sensitivity compared with mammography among Chinese women. Assuming that the sensitivity increased from 50 to 90%, 4167 high-risk subjects were needed to make it 5% statistical significance (two-sided) with 80% power, while allowing for 25% missing data.

The primary unit of analysis was the person-year, with the most severe breast imaging assessments by mammography alone, by ultrasound alone, or by either mammography or ultrasound in the combined group used as the primary end points. The screening cancer yield was defined as the proportion of all malignant cases among all person-year screenings. Sensitivities and specificities in the combined group were determined by the positive results of biopsies and 1-year follow-ups as the ‘gold standard’. The positive predictive value (PPV) was the malignancy rate among cases that test positive by one of the screening modalities. The McNemar test was used to compare the sensitivity and specificity for the natural pairing of assessments within a participant. Analysis of Variance and $\chi^2$-test was used to estimate the statistical differences unless otherwise specified. Area under the receiver operating characteristic (ROC) curve (AUC) was used to compare the diagnostic accuracy of ultrasound with that of mammography. Pairwise comparison of AUC was done according to DeLong et al (DeLong et al, 1988). Statistical analyses were performed using the SPSS software (version 17.0: SPSS Inc., IL, USA), with P-value <0.05 as the threshold of significance. All statistical tests were two-sided.

**RESULTS**

Participant characteristics and screening method preference. Among the 13 339 eligible women who were randomised into the three screening groups (Figure 1), 12 519 (94%) participants finished the initial screening and 8692 (69%) participants came back to complete the second screening. These women were eligible for analysis.

The demographics or risk factors were comparable among the three randomised groups both in the initial and the second screening (Table 1 and Supplementary Table 1). The mean age at enrolment was 46.4 years. Approximately 65% of women were younger than 50 years at the enrolment, and more than two-third were premenopausal.

Although the number of women randomised into the three method groups were virtually identical (Figure 1), among the 12 519 participants who finished the initial screening, the ultrasound group had a significantly higher follow-up rate ($n = 4214, 94.8\%$) than those in the mammography group ($n = 4170, 93.8\%$) or women in the combined group ($n = 4135, 93.0\%$) ($P<0.01$). Similarly, among the 8692 women who came
Breast cancer screening in high-risk Chinese women

BRITISH JOURNAL OF CANCER

back 1-year later, significantly more women in the ultrasound group came back (n = 3082, 73.1%) than those in the mammography group (n = 2815, 67.5%) or in the combined group (n = 2795, 67.6%) (P < 0.01). These data suggested that ultrasound was more acceptable among the participants.

Screening findings. Among a total of 47 subjects with positive screening findings based on the dichotomised BI-RADS score (1–3 as negative and 4–5 as positive), 30 women with breast cancers and 17 women with benign lesions were confirmed by biopsy (Figures 1 and 2). No additional biopsy was triggered by physical examination. With the two-round screenings, the mammography group only identified 4 invasive ductal carcinomas, 1 DCIS, and 2 benign lesions. The ultrasound group detected 11 cancers which were all invasive tumours and 6 benign tumours. The combined group found 12 invasive cancers, 2 DCIS, and 9 benign lesions. The 30 cancer patients had a significantly higher PUMCH model risk scores (44.6 ± 8.0) than the women with negative screening results and benign diseases (39.5 ± 6.2, P < 0.01) (Tables 1 and 2), indicating the utility of this risk model in identifying high-risk population.

Characteristics of subjects with cancer. Table 2 shows the characteristics of the 30 women detected with cancer; all of which had unilateral cancer. The mean age of these women was 48 years, about 2 years older than the study participants. The cancer yields were 0.42 per 1000 (2/4810) in the 30–39 age group, 1.97 per 1000 in the 40–49 age group (18/9157), and 1.38 per 1000 in the 50–65 age group (10/7244) (P = 0.068). Among these cancers, 3 (10%) were DCIS (stage 0), 12 (40%) were stage I, 14 (47%) were stage II, and only one (3%) cancer was stage III. Twenty-three (76.7%) of them were node-negative local disease and 18 (60%) had tumour size ≤ 2 cm.

Nineteen of the 30 cancers were detected at the initial screening (1.52 per 1000 of 12519), and 11 were detected at 1 year later (1.27 per 1000 of 8692) (P = 0.63). Among the 19 cancers detected in the initial screening, 6 (32%) were stage 0 or stage I tumours, whereas 9 of the 11 cancers (82%) detected at the second screening were in this category (P = 0.02). There were no interval cancers found during each of the 12 months follow-up after the initial or the second screening.

Ultrasound vs mammography. The screening results and pathological outcomes of the two-round screenings in the combined group were shown in Figure 2. In the initial screening, 9 (2.18/1000) cancers were detected among the 4135 participants in this group; and 5 (1.19/1000) cancers were diagnosed among the 2795 women in the second round of screening. Of the total 14 cancers detected in both screenings, 8 cancers were found suspicious by both ultrasound and mammography and 6 cancers were detected by ultrasound alone. Overall, the McNemar test showed the sensitivity of ultrasound (14/14, 100%) was significantly higher

Table 1. Characteristics of the initial screened participants

<table>
<thead>
<tr>
<th></th>
<th>MMG group (N = 4170)</th>
<th>US group (N = 4214)</th>
<th>MMG + US group (N = 4135)</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age at enrolment, y</td>
<td>0.99</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean ± s.d.</td>
<td>46.3 ± 8.0</td>
<td>46.4 ± 8.0</td>
<td>46.4 ± 8.0</td>
<td></td>
</tr>
<tr>
<td>Age group at enrolment, y</td>
<td>0.98</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>30–39</td>
<td>21.2</td>
<td>20.9</td>
<td>21.0</td>
<td></td>
</tr>
<tr>
<td>40–49</td>
<td>43.6</td>
<td>44.1</td>
<td>44.2</td>
<td></td>
</tr>
<tr>
<td>50–65</td>
<td>35.2</td>
<td>35.0</td>
<td>34.9</td>
<td></td>
</tr>
<tr>
<td>Body mass index</td>
<td>0.54</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean ± s.d.</td>
<td>25.0 ± 2.9</td>
<td>25.1 ± 2.9</td>
<td>25.1 ± 2.9</td>
<td></td>
</tr>
<tr>
<td>Menarche age, y</td>
<td>0.42</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>&lt;14</td>
<td>53.8</td>
<td>54.9</td>
<td>55.2</td>
<td></td>
</tr>
<tr>
<td>&gt;14</td>
<td>46.2</td>
<td>45.1</td>
<td>44.8</td>
<td></td>
</tr>
<tr>
<td>Stress anticipation (0–9)</td>
<td>0.54</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>6.3</td>
<td>6.1</td>
<td>6.4</td>
<td></td>
</tr>
<tr>
<td>1–4</td>
<td>36.1</td>
<td>37.2</td>
<td>35.4</td>
<td></td>
</tr>
<tr>
<td>5–9</td>
<td>57.5</td>
<td>56.8</td>
<td>58.2</td>
<td></td>
</tr>
<tr>
<td>Oral contraceptive use</td>
<td>0.99</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>13.8</td>
<td>13.7</td>
<td>13.7</td>
<td></td>
</tr>
<tr>
<td>No</td>
<td>86.2</td>
<td>86.3</td>
<td>86.3</td>
<td></td>
</tr>
<tr>
<td>Menopausal status</td>
<td>0.66</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Premenopausal</td>
<td>67.8</td>
<td>68.7</td>
<td>68.1</td>
<td></td>
</tr>
<tr>
<td>Postmenopausal</td>
<td>32.2</td>
<td>31.3</td>
<td>31.9</td>
<td></td>
</tr>
<tr>
<td>Prior benign breast diseases</td>
<td>0.70</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>11.8</td>
<td>11.2</td>
<td>11.6</td>
<td></td>
</tr>
<tr>
<td>No</td>
<td>88.2</td>
<td>88.8</td>
<td>88.4</td>
<td></td>
</tr>
<tr>
<td>Family history of breast cancer</td>
<td>0.63</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>8.2</td>
<td>7.6</td>
<td>8.0</td>
<td></td>
</tr>
<tr>
<td>No</td>
<td>91.8</td>
<td>92.4</td>
<td>92.0</td>
<td></td>
</tr>
<tr>
<td>Risk score</td>
<td>0.84</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean ± s.d.</td>
<td>39.5 ± 6.2</td>
<td>39.6 ± 6.2</td>
<td>39.6 ± 6.2</td>
<td></td>
</tr>
</tbody>
</table>

Abbreviations: MMG = mammography; US = ultrasound; y = years.
than that of mammography (8/14, 57.1%; \( P = 0.04 \); Table 3). The specificity of ultrasound was 6910 of 6916 (99.9%), similar to the specificity of mammography, which was 6913 of 6916 (100.0%) \( (P = 0.51) \); Figure 2 and Table 3). There were 6 false-positive cases identified by ultrasound and 3 cases identified by mammography; however, the PPVs between ultrasound (14/20, 70.0%) and mammography (8/11, 72.7%) were similar \( (P = 0.87 \); Figure 2 and Table 3). The diagnostic accuracy (AUC) of ultrasound was 0.999 (95% CI, 0.999–1.000), significantly higher than that of mammography (0.766, 95% CI, 0.591–0.941; \( P = 0.01 \); Table 3).

In the mammography group, 3 cancers were detected out of 4170 screened participants in the initial screening, and 2 cancers were diagnosed out of 2815 screenings in the second screening (Figure 1). In the ultrasound group, the corresponding numbers was 7 out of 4214 and 4 out of 3082, respectively. The cancer yield of the ultrasound group (11/7296, 1.51/1000) was numerically higher than that of the mammography group (5/6985, 0.72/1000), though the difference did not reach significance \( (P = 0.16) \). The cancer yield of the combined group (14/6930, 2.02/1000) was higher than that of the mammography group \( (P = 0.04) \), but there was no difference between the combined group and the ultrasound group \( (P = 0.47) \).

**Cost analysis.** Costs of the screening modalities are shown in Table 4. Overall, finding one cancer would need to screen 1397 women by ultrasound at the cost of 49 700 RMB ($7876), and 495 women by both screening methods at the cost of 136 287 RMB ($21 599). Therefore, ultrasound is the least costly screening modality for breast cancer, costing only 17.4% of mammography or 36.5% of combination screening.

**DISCUSSION.** Early detection is particularly important in China and many other countries since chemotherapy and radiation therapy may not be available in many areas. In the present study, 82% of the cancers found by the annual screening were in stages 0 and I, which were probably curable by surgery alone. This multi-centre randomised trial demonstrates, for the first time, that ultrasound examination is a sensitive, specific, and the least costly screening modality to detect breast cancer in high-risk Chinese women. With high sensitivity (100%), specificity (99.9%), diagnostic accuracy (0.999), and PPV (70.7%), ultrasound performed superior to or at least as good as mammography. The ultrasound group also showed a higher cancer yield (11/7296, 1.51/1000) than that of the mammography group (5/6985, 0.72/1000), though the difference did not reach significance \( (P = 0.16) \). One important feature of
using ultrasound for breast cancer screening in China is the lower per-cancer-finding cost, which is only 17.4% of mammography or 36.5% of combined methods. In addition, ultrasound is not only widely accessible across China but also more acceptable among our participants.

We attribute the low sensitivity of mammography in this study mainly to the effects of younger age of the study population with a mean age of 46 years and 68% of the women were premenopausal who are more likely to have denser breast tissue. Chinese women are commonly found to have dense breasts, with 66% classified as heterogeneous or extremely dense (Zulfiquer et al, 2011). There is also an inverse relationship between age and breast density; the younger the woman, the denser the breast (Berg et al, 2008; McCawley et al, 2009; Cheka et al, 2012). Previous studies and this trial showed that the peak age of breast cancer in ethnically Chinese women was significantly younger than that of Caucasian women (DeSantis et al, 2011; Li et al, 2011; Shen et al, 2012; Fan et al, 2014). Although age hardly affects the ultrasound exam, mammography appears to be more sensitive for patients over 50 years of age than those younger (McCawley et al, 2009; Shen et al, 2012). A retrospective analysis of the Chinese government ‘Two Cancer Screening’ campaign also showed ultrasound was more sensitive than mammography in Chinese women, especially in premenopausal patients (Wang et al, 2013).

Cost-effectiveness is one of the crucial elements in cancer screening, especially with consideration of the massive population in China. Ultrasound is expensive in the Western countries and usually not covered by insurance for routine breast cancer screening (Feig, 2011). However, it costs only one-third to one-fifth of mammography and thus cost much less out-of-pocket expenses in China.

Since the incidence of breast cancer in China is relatively low (Fan et al, 2014), it would cost much more to detect one cancer if average-risk Chinese women were all screened. Therefore, a suitable risk-prediction model is needed to make the nationwide screening practical. Although screening will temporarily increase the incidence of breast cancer, the cancer incidence of our screened high-risk participants reached 217.7/10^5 initially and 178.9/10^5 at 1-year follow-up (calculated by the combined group). Therefore, the PUMCH risk model for identifying high-risk women seems working, though further work needs to be done to improve it.

Only 69.4% of the initial screened participants accepted the invitation of the next year screening. The reasons for dropout are likely complex, which may include lack of awareness of breast cancer that is common among Chinese women (Wu et al, 2012), inadequate community or family support, financial constraints, social inconvenience, and reluctance to visit doctors.

Although the lower referral rate in the mammography group may have a negative effect on cancer yield, the most important factor making the difference of yield between ultrasound and mammography is the sensitivity. In the combined group, the participants were screened by both ultrasound and mammography; however, the cancer yield of ultrasound is higher than that of mammography (14/6930, 2.02/1000).

Breast cancer screening has been criticised for making over-diagnosis because a substantial percentage of cancers mainly DCIS detected by aggressive screening will never evolve into more life-threatening metastatic cancer (Bleyer and Welch, 2012). In our study, all the 11 cancers detected in the ultrasound group were invasive tumours, whereas 3 DCIS cases were detected in either the mammography group or in the combined group. Whether ultrasound screening could decrease over-diagnosis needs to be further studied.

A few limitations of this study need to be noted. Firstly, mortality was not assessed in the present study because of the lack of long-term follow-up and relatively small number of participants. It was believed that, surrogate end points, such as the diagnostic performance of a screening modality or smaller size and earlier cancer stages, may be used at present. Studies have shown a correlation between those surrogates with better mortality outcomes (Michaelson et al, 2003; Smith et al, 2004). Secondly, the risk-prediction model, PUMCH model, has not been evaluated in other countries, and it is difficult to confirm that the Chinese women included in the screening programme is suitable risk-prediction model. PUMCH model, has not been evaluated in other countries, and it is difficult to confirm that the Chinese women included in the screening programme is suitable risk-prediction model.
modalities since all the eligible women were randomised into the three screening groups. Thirdly, the drop-out may underestimate the number of interval cancers, which may overestimate the performance of screening methods. However, this limitation has less effect on the comparison between ultrasound and mammography because of the natural pairing of assessments within a participant.

In summary, data from this multi-centre randomised study suggest that ultrasound could be both sensitive and cost-effective for breast cancer screening in high-risk Chinese women aged 30–65 years. Whether this method of screening is best performed every year or every 2 years, and whether it is also suitable for low-to-intermediate risk women remain to be further evaluated.

ACKNOWLEDGEMENTS

We thank all the study participants and the medical teams who took care of these women. This work was supported by the Eleventh Five-year Key Programs for Science and Technology Development of China (grant number 2006BAI02A09); and the Beijing Municipal Science and Technology Key Development Program (grant number D090507043404009).

CONFLICT OF INTEREST

The authors declare no conflict of interest.

REFERENCES


This work is licensed under the Creative Commons Attribution-Non-Commercial-Share Alike 4.0 International License. To view a copy of this license, visit [http://creativecommons.org/licenses/by-nc-sa/4.0/](http://creativecommons.org/licenses/by-nc-sa/4.0/)

Supplementary Information accompanies this paper on British Journal of Cancer website ([http://www.nature.com/bjc](http://www.nature.com/bjc))