Original Research

**Breast cancer detection using single-reading of breast tomosynthesis (3D-mammography) compared to double-reading of 2D-mammography: evidence from a population-based trial.**

Nehmat Houssami¹, Daniela Bernardi², Marco Pellegrini², Marvi Valentini², Carmine Fantò², Livio Ostillio², Paolina Tuffobene², Andrea Luparia², and Petra Macaskill¹.

[Running title: single-read 3Dmammography versus double-read 2Dmammography]

(1) Sydney School of Public Health, Sydney Medical School, University of Sydney, Sydney, Australia.

(2) U.O. Senologia Clinica e Screening Mammografico, Department of Diagnostics, Ospedale di Trento, Azienda Provinciale Servizi Sanitari, Trento, Italy

**Correspondence**

Professor Nehmat Houssami, School of Public Health (A27), Sydney Medical School, University of Sydney, Sydney 2006, Australia.

Email: nehmat.houssami@sydney.edu.au
Abstract

Background
Most population breast cancer (BC) screening programs use double-reading of 2D-mammography. We recently reported the Screening with tomosynthesis or standard mammography-2 (STORM-2) trial, showing that double-read tomosynthesis (pseudo-3D-mammography) detected more BC than double-read 2D-mammography. In this study, we compare screen-detection measures for single-reading of 3D-mammography with those for double-reading of 2D-mammography, to inform screening practice.

Methods
This is a secondary analysis based on STORM-2 which prospectively compared 3D-mammography and 2D-mammography in sequential screen-readings. Asymptomatic women ≥49 years who attended population-based screening (Trento, 2013-2015) were recruited. Participants recalled at any screen-read from parallel double-reading arms underwent further testing and/or biopsy. Single-reading of 3D-mammography, integrated with acquired or synthetized 2D-mammograms, was compared to double-reading of 2D-mammography alone for screen-detection measures: number of detected BCs, cancer detection rate (CDR), number and percentage of false-positive recall (FPR). Paired binary data were compared using McNemar's test.

Results
Screening detected 90, including 74 invasive, BCs in 85 of 9,672 participants. CDRs for single-reading using integrated 2D/3D-mammography (8.2 per 1000 screens; 95%CI 6.5-10.2) or 2Dsynthetic/3D-mammography (8.4 per 1000 screens; 95%CI: 6.7-10.4) were significantly higher than CDR for double-reading of 2D-mammography (6.3 per 1000 screens; 95%CI: 4.8-8.1), P<0.001 both comparisons. FPR% for single-read 2D/3D-mammography (2.60%; 95%CI: 2.29-2.94), or single-read 2Dsynthetic/3D-mammography (2.76%; 95%CI: 2.45-3.11), were significantly lower than
FPR% for double-read 2D-mammography (3.42%; 95%CI: 3.07-3.80), P<0.001 and P=0.002 respectively.

**Conclusions**

Single-reading of 3D-mammography (integrated 2D/3D or 2Dsynthetic/3D) detected more BC, and had lower FPR, compared to current practice of double-reading 2D-mammography alone – these findings have implications for population BC screening programs.

*Keywords*: breast cancer screening; digital breast tomosynthesis; mammography; population screening; screen-reading.
1- Background

The majority of population breast cancer (BC) screening programs, such those in Europe, the UK and Australia, provide breast screening using double-reading (interpretation by two readers) of standard two-dimensional (2D) digital mammography. Although double-reading of screening mammography is not a global practice, it was introduced into organized screening programs because it increases BC detection by an estimated 5%-15% of the proportion of detected cancers relative to single-reading (1-5). Mammography technology, however, has evolved considerably through development of digital breast tomosynthesis, an emerging pseudo-three-dimensional mammography technology, also referred to as 3D-mammography. Tomosynthesis has been examined in prospective trials (6-9) embedded within European population-based screening programs, and in retrospective studies conducted in North America (10-15) all of which demonstrate that 3D-mammography enhances detection measures. It is unknown whether integrating 3D-mammography technology into screening practice could potentially obviate the need for double-reading in the context of population BC screening programs.

We recently reported results of the screening with tomosynthesis or standard mammography-2 (STORM-2) trial in which breast tomosynthesis, or 3D-mammography, interpreted with either acquired 2D images or with reconstructed images (2Dsynthetic) was shown to detect more BC than 2D-mammography in double-reading of all mammographic examinations (9). In the present study, we compare screen-detection measures in STORM-2 for single-reading based on 3D-mammography (interpreted with either acquired 2D or synthesized 2D images) with those for ‘standard of care’ double-reading of 2D-mammography, to inform future BC screening practice.

2- Methods
STORM-2 is a prospective population-based screening study that compares mammography screen-reading in sequential phases in two parallel double-reading arms as shown in online-only Appendix 1: screen-reading in one double-reading used acquired 2D and 3D mammography, and another double-reading used 3D-mammography with reconstructed 2D-images (2Dsynthetic) whereby 2D is reconstructed from the 3D acquisition. The methodology, described by Bernardi et al(9), yields paired data for each screening examination (from the same participant) in each arm of the trial. Asymptomatic women aged ≥49 years attending biennial population-based screening through the Trento screening program, Italy, were recruited into the study between May 2013 and May 2015, and invited to have screening with 2D and 3D mammography. The study was granted institutional ethics approval, and consent was obtained from screening participants (9).

2.1 Mammography & screen-reading

Participants had digital mammography integrating both 2D and 3D mammography acquisitions (using Selenia® Dimensions Unit operated in COMBO© mode; Hologic, Bedford MA, USA) and using software that reconstructs 2D-mammographic images from 3D acquisitions (C-View™ 2D-software). All mammography acquisitions were obtained at the same screening examination with a single breast positioning per view: mediolateral oblique and cranio-caudal views were obtained for 2D and 3D acquisitions. Women who declined to participate in the trial had 2D-mammography. As reported in our earlier publication, the estimated mean glandular dose per view was 1.36mGy (SD 0.51) from 2D-mammography acquisition, 1.87mGy (SD 0.67) from 3D-mammography, and 3.22mGy (SD 1.16) from dual-acquisitions (2D+3D).

Online-only Appendix 1 shows the study’s schema: in one independently reported double-reading (readings A/B) screens were sequentially reported by radiologists viewing 2D-mammography alone, and then re-interpreted by the same radiologists (on the same day) using integrated 2D/3D-mammography; and in another independently reported double-reading (readings C/D) the same screens were interpreted sequentially by a different reader pair to that performing readings A/B using
2Dsynthetic and re-reported on the same day using integrated 2Dsynthetic/3D-mammography (9). Hence each screen was interpreted by two different reader pairs (a total of 4 readings) using 2D/3D or 2Dsynthetic/3D. Radiologists reported each screening mammogram independently of each other and according to the above-described sequence, and were asked to record whether or not to recall at each screen-reading phase. A screen was considered positive and the woman recalled to assessment (further investigations) if recalled by either screen-reader in either of the double-reading arms, based on recall at any screen-reading phase. Seven breast radiologists participated in screen-reading: all had experience in mammography screening and had previously used 3D-mammography (9). Previous mammograms were displayed, where available, at the time of screen-reading.

2.2 Outcome measures

Our primary outcome measures were the number of detected cancers and the cancer detection rate (CDR) per 1000 screens; the incremental CDR for single-reading using 2D/3D or 2Dsynthetic/3D screening compared to double-reading using 2D-mammography; and the number and percentage of false-positive recalls for these screen-reading strategies. Outcomes were ascertained on the basis of excision histology, or based on all investigations performed at assessment inclusive of additional imaging and histology from core needle biopsy (where performed) in recalled subjects. It is therefore possible that some cancers may have been missed by all screen-reads, and/or assessment of recalled women, however, this does not affect estimates of the present analysis which is comparative BC detection at screening.

2.3 Statistical Analysis

STORM-2 sample size was planned on the basis of the study’s primary end-point (comparison of CDR for double-readings) and has been described in our earlier publication (9). The present study reports a secondary analysis comparing single-reading using 2D/3D or 2Dsynthetic/3D against double-reading using 2D-mammography alone. Single-reading of 2D/3D or 2Dsynthetic/3D was
examined using the first read performed for each screening examination (in each double-reading arm of STORM-2) as this inherently represents the most independent screen-read. Two-way tables for paired binary data were used to compare the number of cancers detected exclusively at each screen-reading strategy (double-read 2D vs single-read 2D/3D; and double-read 2D vs single-read 2Dsynthetic/3D); the same analyses were performed for FPR. We calculated and compared CDRs per 1000 screens for these screen-reading strategies, and calculated the incremental CDR attributable to single-reading of integrated 2D/3D or 2Dsynthetic/3D screening against double-reading of 2D-mammography, taking account of the pairing of the binary outcomes for these strategies within women. We also calculated and compared the FPR% for these screen-reading strategies, again taking account of pairing of the binary outcomes within women. To provide additional context, we calculated the CDR and the FPR for single-reading of 2D-mammography.

Analyses were conducted using SAS/STAT® (9.4) (16), using exact methods to compute confidence intervals (CI) for CDR and FPR% estimates, and P-values for the McNemar test for paired binary data (17). StatsDirect software (v3.0.167) (18) was used to compute exact CIs for differences in paired proportions.

3- Results

9,672 screening participants with a median age of 58 years (IQR 53-63 years) were recruited from 10,255 invited into the study. Detailed characteristics of 90 screen-detected BCs (5 with bilateral BC) have been described in our earlier report of the trial (9): there were 74 invasive and 16 in-situ breast malignancies. Mean tumour size for invasive BC was 12.7 mm (SD 7.8) for 46 invasive BCs detected at 2D and also at integrated 2D/3D or 2Dsynthetic/3D screening, and 11.6 mm (SD 9.4) for 28 invasive BCs detected only at 2D/3D or 2Dsynthetic/3D screening.

3.1 Cancer Detection
Tables 1-2 present cancer detection data, based on 90 BCs (in 85 women with 5 affected by bilateral BC) detected at screening. As shown in Table 1(a), 60 BCs were detected at double-reading of 2D-mammography and also at single-reading of 2D/3D-mammography, 19 were detected only at single-reading of 2D/3D compared with one detected only at double-reading of 2D-mammography alone, and 10 cancers were not detected by any of the screen-reading strategies in this analysis (P<0.001). Comparison based on the alternate application of 3D-mammography technology (2Dsynthetic/3D, table 1(b)) showed that 58 BCs were detected at double-reading of 2D and also at single-reading of 2Dsynthetic/3D, 23 BCs were detected only at single-reading of 2Dsynthetic/3D compared with 3 detected only at double-reading of 2D-mammography alone, and 6 cancers were not detected by any of these screen-reading strategies (P<0.001).

CDRs are summarised in Table 2: CDR for each of single-reading using 2D/3D-mammography (8.2 per 1000 screens; 95%CI 6.5, 10.2) or using 2Dsynthetic/3D-mammography (8.4 per 1000 screens; 95%CI: 6.7, 10.4) were significantly higher than the CDR for double-reading of 2D-mammography (6.3 per 1000 screens; 95%CI: 4.8, 8.1), P<0.001 for both comparisons that took account of the pairing of the binary outcomes within women; this is also reflected in the corresponding estimates for incremental CDR shown in table 2. Estimated CDR for single-reading of 2D-mammography alone was 6.0 per 1000 screens (95%CI: 4.6, 7.7)

3.2 False-Positive Recall

Cross-tabulations (Table 3) and the associated estimates for FPR% (Table 4) report data for women classified as not having BC including FPR. FPR% for single-reading of 2D/3D-mammography (2.60%; 95%CI: 2.29, 2.94) was significantly lower than that for double-reading of 2D-mammography (3.42%; 95%CI: 3.07, 3.80), P<0.001; and the FPR% for single-reading of 2Dsynthetic/3D-mammography (2.76%; 95%CI: 2.45, 3.11) was also significantly lower than that for double-reading of 2D-mammography (3.42%; 95%CI: 3.07, 3.80), (P=0.002) taking account of pairing of the binary outcomes. The corresponding differences in FPR% (Table 4) indicate that
single-reading using 2D/3D or 2D synthetic/3D yields small but statistically significant reductions in FPR% compared to double-reading of 2D-mammography. Estimated FPR% for single-reading of 2D-mammography alone was 2.05% (95% CI: 1.78, 2.36).

4- Discussion

Cancer screening programs provide breast screening using 2D-mammography, the only breast imaging recommended for population screening based on evidence of BC mortality reduction (19, 20). To enhance cancer detection, 2D-mammography has been implemented using double-reading in the majority of organized BC screening programs. At present, there is an emerging cross-roads for mammography screening practice; digital breast tomosynthesis, or pseudo-3D-mammography, reduces overlapping breast tissue hence it improves cancer visibility and may also reduce false-positive recall (11, 12, 21, 22) relative to 2D-mammography. Although evidence that 3D-mammography improves screen-detection measures has accumulated in recent years (6, 7, 12-15, 23, 24), it remains unknown whether this mammography technology confers incremental screening benefit compared to 2D-mammography screening (19-21). Furthermore, 3D-mammography comes at a ‘cost’ both monetary and in terms of additional radiation to the breast depending on the technology used (and particularly whether used with acquired or with synthetised 2D images). In this complex cross-roads for BC screening practice, intensified by seemingly rapid adoption of 3D-mammography in some settings (21), we examined whether 3D-mammography could support more efficient (less resource or reader intensive) and potentially more effective population BC screening in the present study based on the STORM-2 trial (9). We compared detection measures for single-reading of 3D-mammography (interpreted alongside acquired 2D-mammograms or synthetic 2D-mammograms) with double-reading of 2D-mammography, and showed that a screening model integrating single-reading of 3D-mammography performs significantly better than the current screening standard of double-reading 2D-mammography.
Our analysis showed that the CDR for single-reading of integrated 2D/3D-mammography (8.2 per 1000 screens) was statistically significantly higher than that for double-reading of 2D-mammography (6.3 per 1000 screens), and the same findings were shown for single-reading of 2Dsynthetic/3D-mammography (CDR 8.4 per 1000 screens); hence single-reading of 2D/3D or 2Dsynthetic/3D yielded substantial incremental CDR (1.9 and 2.1 per 100 screens, respectively) compared to double-reading of 2D-mammography. Given that 2Dsynthetic/3D requires less radiation (less image acquisitions) than 2D/3D-mammography (9, 22), this screen-reading strategy may be an appropriate BC screening strategy for future. Because few studies have compared 2Dsynthetic/3D-mammography with 2D/3D-mammography, and allowing there are concerns that 2Dsynthetic images might have less capability for detection of micro-calcifications than acquired 2D-mammography due to lower spatial resolution (9, 22), we believe that integrated 2Dsynthetic/3D-mammography is worthy of further comparative evaluation in screening trials.

False recall is a harm of BC screening, so a further advantage of exploring a screening model based on single-reading of 3D-mammography (either 2D/3D or 2Dsynthetic/3D) is that it also reduces FPR by small but statistically (and clinically) significant proportions, as shown in Table 4. Here it should be noted that the lower recall from single-reading of 3D-mammography is mostly a result of applying ‘single-reading’ instead of ‘double-reading’ rather than a technology effect, and single-reading of 2D-mammography alone would similarly reduce FPR (compared to double-read 2D-mammography) as shown in our data (Table 4). However, single-reading of 2D-mammography would have the disadvantage of yielding approximately 5% lower CDR (CDR 6.0/1000 screens as shown in Results) than that for double-reading of 2D-mammography (CDR 6.3/1000). The direct effect of 3D-mammography technology on FPR has shown heterogeneous results with evidence that, in screening services where the FPR based on 2D-mammography is high, integrating 3D-mammography into screen-readings could lead to a reduction in FPR (8, 9, 12, 13, 21).
In this work, we explored opportunities to potentially streamline BC screening practice through 3D-mammography technology rather than apply this technology based on long-established screen-reading practices introduced to overcome the limitations of 2D-mammography. To put our findings into context, in terms of BC detection, it can be seen from the data in Table 1 that around 68% of screen-detected BC (61 from 90 cancers) were detected at standard double-reading of 2D-mammography. Single-reading of 2D/3D or 2Dsynthetic/3D detected an additional 21% to 26% of cancers (whereas single-reading of 2D-mammography alone would not have detected 3 cases, meaning that double-reading of 2D-mammography enabled detection of roughly 3% of all screen-detected cancers in participants). Hence single-reading integrating 3D-mammography has a better incremental BC detection yield than that achieved through double-reading of the standard mammography technology.

We emphasize that at present, 3D-mammography screening is not endorsed or applied for screening in population breast screening program, and there are major knowledge gaps on the long-term outcomes of population screening with integrated 2D/3D mammography. Specifically, we do not know whether the additional BC detection from 3D-mammography translates into screening benefit, meaning whether or not it confers incremental mortality reduction above that achieved with 2D-mammography screening (19-21), or whether 3D-mammography is augmenting the problem of BC overdiagnosis (25). However, we think that the evidence on screen-detection measures reported in this study justifies further research into a screening model based on single-reading integrating 3D-mammography, undertaken in population screening services, and underpinning comparison with current screening practice of double-reading 2D-mammography alone. This would also need to be complemented by cost-effectiveness research to inform screening policy decisions, factoring the cost of tomosynthesis technology and the additional time taken to read 3D-mammography.

There are limitations to this study, primarily the use of sequential screen-reading in the STORM-2 trial may have over-estimated the incremental CDR attributed to 3D-mammography (9). While that limitation is acknowledged, we point out that the additional BC detection estimated for 3D-mammography in STORM-2 is similar to that reported in other prospective trials of this technology.
It could also be argued that STORM-2, conducted in an Italian screening service, may not transfer more broadly to other breast screening programs. However, Italy’s population screening program is based on similar practice and screening standards as those in the majority of European countries. Furthermore, as outlined above, we think our findings should encourage screening program-based evaluation in new research studies of 3D-mammography and not a prompt change of BC screening practice.

Our work has provided evidence from a population-based screening study, indicating that single-reading of 3D-mammography (interpreted alongside acquired 2D-mammograms or synthetic 2D-mammograms) performed significantly better, in terms of BC detection and FPR measures, than the current screening standard of double-reading 2D-mammography alone. Our findings suggest that it may be time to rethink the BC screening model used in many population-based screening programs, by highlighting the potential for 3D-mammography technology to support less resource-intensive screen-reading practice that could also improve BC detection. The evidence we report supports evaluation of the proposed BC screening model, ideally in trials embedded within breast screening services, to ensure quality monitoring as part of the screening episode as well as data collection on longer-term outcomes in screening participants.

Conflicts of Interest (COI) statements: Dr Bernardi declares receipt of conference travel support from Hologic; all other authors (NH, MP, MV, CF, LO, PT AL, PM) have no COI to disclose.

Acknowledgements: N. Houssami receives research support through a National Breast Cancer Foundation (NBCF, Australia) Breast Cancer Research Leadership Fellowship.
Role of funding source: Funding sources did *not* have any role in the design of the study, in the collection, analysis, or interpretation of data, or in the writing of the manuscript.
Author Contributions

Houssami: study design and scientific direction, data analysis and interpretation, literature searches, manuscript writing and revision

Bernardi: study design and overall supervision, data collection and interpretation, contributed to manuscript writing and revision

Pellegrini: screen-reading, data collection, review of draft manuscript

Valentini: screen-reading, data collection, review of draft manuscript

Fantò: screen-reading, data collection, review of draft manuscript

Ostillio: screen-reading, data collection, review of draft manuscript

Tuttobene: screen-reading, data collection, review of draft manuscript

Luparia: screen-reading, data collection, review of draft manuscript

Macaskill: Statistical planning, data analysis and interpretation, contributed to manuscript writing and revision
Reference List

(1) Taylor P, Potts HW. Computer aids and human second reading as interventions in screening mammography: two systematic reviews to compare effects on cancer detection and recall rate. [Review] [60 refs]. *European Journal of Cancer* 2008 Apr,44(6), 798-807.


(21) Houssami N, Miglioretti DL. Digital Breast Tomosynthesis: A Brave New World of Mammography Screening. *JAMA Oncology* 2016 Jun 1, 2(6), 725-727.


**Table 1**: Two-way paired tables for breast cancer detection based on single-reading integrating 3D-mammography versus double-reading of standard 2D-mammography in the screening with tomosynthesis or standard mammography 2 (STORM-2) trial

<table>
<thead>
<tr>
<th>Mammography screen-reading strategy</th>
<th>Single-reading of 3D with 2D mammography positive</th>
<th>Single-reading of 3D with 2D mammography negative</th>
<th>Total</th>
<th>Mammography screen-reading strategy</th>
<th>Single-reading of 3D mammography with 2D synthetic† positive</th>
<th>Single-reading of 3D mammography with 2D synthetic† negative</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Double-reading of 2D mammography positive</td>
<td>60</td>
<td>1</td>
<td>61‡</td>
<td>Double-reading of 2D mammography 2D positive</td>
<td>58</td>
<td>3</td>
<td>61‡</td>
</tr>
<tr>
<td>Double-reading of 2D mammography negative</td>
<td>19</td>
<td>10§</td>
<td>29</td>
<td>Double-reading of 2D mammography 2D negative</td>
<td>23</td>
<td>6§</td>
<td>29</td>
</tr>
<tr>
<td>Total</td>
<td>79</td>
<td>11</td>
<td>90*</td>
<td>Total</td>
<td>81</td>
<td>9</td>
<td>90*</td>
</tr>
</tbody>
</table>

**P <0.001**  
**P <0.001**

† 2D synthetic refers to 2D mammography images reconstructed from the 3D (tomosynthesis) acquisition  
‡ Single-reading of 2D-mammography alone would have detected 58 of these cases  
* Total data for 90 breast cancers double-count 5 women (found to have bilateral breast cancer)  
§ Based on cancers detected in the study population at screening as part of the STORM-2 trial and does not include interval cancers  
** Exact P-value for McNemar’s test for paired binary data.
Table 2: Breast cancer detection rates and incremental cancer detection for single-reading of 3D-mammography *versus* double-reading of standard 2D-mammography in the STORM-2 trial

<table>
<thead>
<tr>
<th>Group</th>
<th>No. of detected cancers</th>
<th>Cancer detection rate (CDR) per 1000 screens (95% CI)</th>
<th>P*</th>
<th>Incremental CDR/1000 attributed to integrating 3D screening (95% CI) † versus 2D alone</th>
</tr>
</thead>
<tbody>
<tr>
<td>Standard double-reading of 2D-mammography</td>
<td>61‡</td>
<td>6.3 (4.8, 8.1)</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Single-reading of 3D with 2D mammography</td>
<td>79</td>
<td>8.2 (6.5, 10.2)</td>
<td>P &lt;0.001</td>
<td>1.9 (0.9, 2.9)</td>
</tr>
<tr>
<td>Single-reading of 3D with 2D synthetic mammography</td>
<td>81</td>
<td>8.4 (6.7, 10.4)</td>
<td>P &lt;0.001</td>
<td>2.1 (1.0, 3.3)</td>
</tr>
</tbody>
</table>

‡ Estimated CDR for 9672 screens are analyzed as 9677 to allow for 5 participants (in both the numerator and the denominator) found to have bilateral cancers with different detection results between breasts.

* Exact P-value for Mc Nemar’s test comparing paired binary data; P refers to comparison of CDR for screening based on single-reading of 3D-mammography *versus* double-reading of 2D mammography

† 95% confidence intervals take account of pairing of binary outcomes for each screen

‡ Single-reading of 2D-mammography alone would have detected 58 of these cases (estimated CDR 6.0 per 1000 screens)
Table 3: Two-way paired tables for screens classified as not having breast cancer, inclusive of false-positive recall, based on single-reading integrating 3D-mammography versus double-reading of standard 2D-mammography in the screening with tomosynthesis or standard mammography 2 (STORM-2) trial

<table>
<thead>
<tr>
<th>Mammography screen-reading strategy</th>
<th>Single-reading of 3D with 2D positive</th>
<th>Single-reading of 3D with 2D negative</th>
<th>Total</th>
<th>Mammography screen-reading strategy</th>
<th>Single-reading of 3D mammography with 2D synthetic† positive</th>
<th>Single-reading of 3D mammography with 2D synthetic† negative</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Double-reading of 2D mammography positive</td>
<td>148</td>
<td>180</td>
<td>328</td>
<td>Double-reading of 2D mammography positive</td>
<td>99</td>
<td>229</td>
<td>328</td>
</tr>
<tr>
<td>Double-reading of 2D mammography negative</td>
<td>101</td>
<td>9158</td>
<td>9259</td>
<td>Double-reading of 2D mammography negative</td>
<td>166</td>
<td>9093</td>
<td>9259</td>
</tr>
<tr>
<td>Total</td>
<td>249</td>
<td>9338</td>
<td>9587</td>
<td>Total</td>
<td>265</td>
<td>9322</td>
<td>9587</td>
</tr>
</tbody>
</table>

**P <0.001**

**Exaxt P-value for McNemar’s test for paired binary data

**P = 0.002**
Table 4: False-positive recall (FPR) proportion for single-reading of 3D-mammography *versus* double-reading of standard 2D-mammography in the STORM-2 trial

<table>
<thead>
<tr>
<th>Comparison based on all screens classified as not having BC (n=9587)</th>
<th>No. of FPR</th>
<th>FPR % (95% CI)</th>
<th>p*</th>
<th>Difference in FPR%** (95% CI)†</th>
</tr>
</thead>
<tbody>
<tr>
<td>Standard double-reading of 2D-mammography</td>
<td>328‡</td>
<td>3.42 (3.07, 3.80)</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Single-reading of 3D with 2D mammography</td>
<td>249</td>
<td>2.60 (2.29, 2.94)</td>
<td>&lt;0.001</td>
<td>-0.82 (-1.17, -0.48)</td>
</tr>
<tr>
<td>Single-reading of 3D with 2D synthetic mammography</td>
<td>265</td>
<td>2.76 (2.45, 3.11)</td>
<td>0.002</td>
<td>-0.66 (-1.07, -0.25)</td>
</tr>
</tbody>
</table>

* Exact P-value for Mc Nemar’s test for paired binary data

**Estimated difference in FPR% compared to standard double-reading of 2D mammography (minus sign indicates a reduction in FPR%)

†95% confidence intervals take account of pairing of binary outcomes for each screen

‡ Single-reading of 2D-mammography alone would have resulted in 197 false-recalls (estimated FPR 2.05%)