Incremental effect from integrating 3D-mammography (tomosynthesis) with 2D-mammography: Increased breast cancer detection evident for screening centres in a population-based trial

Francesca Caumo, Daniela Bernardi, Stefano Ciatto, Petra Macaskill, Marco Pellegrini, Silvia Brunelli, Paola Tuttobene, Paola Bricolo, Carmine Fantò, Marvi Valentinii, Stefania Montemezzi, Nehmat Houssami

*Corresponding author. School of Public Health (A27), Sydney Medical School, University of Sydney, Sydney 2006, Australia. Tel.: +61 0419 273510. E-mail addresses: nehmath@med.usyd.edu.au, nehmah.houssami@sydney.edu.au (N. Houssami).

1 Post-humous (deceased May 2012).

Article history:
Received 20 October 2013
Accepted 15 November 2013

Keywords:
Population cancer screening
Mammography
Digital breast tomosynthesis

ABSTRACT

Background & objectives: Three-dimensional (3D)-mammography (tomosynthesis) may improve breast cancer detection. We examined centre-specific effect of integrated 2D mammography based on the STORM (screening with tomosynthesis or standard mammography) trial.

Methods: Asymptomatic women who attended population-based screening through Trento and Verona screening centres were recruited into STORM, a prospective comparison of screen-reading in two sequential phases: 2D-mammography only and integrated 2D/3D mammography. Outcomes were the number and rates of detected cancers and of false positive recalls (FPR), and incremental cancer detection rate (CDR). Paired binary data were compared using McNemar’s test.

Results: Of 33 cancers detected in Trento, 21 were detected at both 2D and 2D/3D screening, 12 cancers were detected only with integrated 2D/3D screening compared with none detected at 2D-only screening (P < 0.001). Of the 26 cancers detected in Verona, 18 were detected at both 2D and 2D/3D screening, 8 cancers were detected only with integrated 2D/3D screening compared with none detected at 2D-only screening (P = 0.008). There were no differences between centres in baseline CDR, and incremental CDR attributable to 3D-mammography was similar for Trento (2.8/1000 screens) and for Verona (2.6/1000 screens). Trento had 239 FPR (5.7% of screens): 103 FPR at both screen-readings, 93 FPR only at 2D-mammography compared with 43 FPR only at 2D/3D-mammography (p < 0.001). Verona had 156 FPR (5.2% of screens): 78 FPR at both screen-readings, 48 FPR only at 2D-mammography compared with 30 FPR only at 2D/3D-mammography (p = 0.054). Estimated reduction in FPR proportion had recall been conditional to 2D/3D-mammography-positivity differed between centres (21.0% versus 11.5%; P = 0.02).

Conclusion: Integrated 2D/3D-mammography significantly increased cancer detection for both screening services; potential reduction in FPR is likely to differ between centres with those experiencing relatively higher FPR most likely to benefit from 2D/3D-mammography screening.

© 2013 Elsevier Ltd. All rights reserved.

Introduction

The evolution of three-dimensional (3D)-mammography, or breast tomosynthesis, from digital mammography technology has progressed to clinical application [1–4]. Until recently, most evaluations of 3D-mammography have been based on selected series or image-sets enriched with cancer cases [2]. In 2013, two prospective screening trials [5,6] reported their findings, the Oslo trial (interim report) [6] and the Italian trial [5], both indicating that breast screening using conventional (2D) mammography with 3D-mammography significantly improves breast cancer detection relative to screening with 2D-mammography alone. These landmark studies [5,6] are anticipated to transform future screening practice however, at present, there is very limited evidence from individual screening services or centres on the effect of integrated 2D/3D mammography on detection measures. Such information
methods and population have been described by Ciatto and colleagues in a report of the trial. In the present paper, we evaluate screening detection measures for the two centres based in Trento and Verona that participated in STORM. We focus on estimating and comparing the incremental effect of integrating 3D with 2D-mammography on centre-specific breast cancer detection measures and false positive screening.

Methods

This study examines centre-specific detection for two screening centres that participated in the STORM trial. The STORM study methods and population have been described by Ciatto and colleagues in a report of the trial’s primary outcomes. In brief, a prospective trial was conducted in two population-based screening services in Trento and Verona (August 2011–June 2012), in Italy. STORM compared sequential mammography screen-readings: screen-reading with conventional 2D-mammography only, and screen-reading with integrated 2D/3D mammography.

Asymptomatic women aged 48–71 years who had screening mammography through the participating centres were invited to have 2D/3D mammography screening, and informed consent was obtained from all women; those declining to have 2D/3D mammography as part of the study received standard 2D-mammography. The study was granted institutional ethics approval in both Trento and Verona. The Trento and Verona breast screening services have used digital mammography since 2005 and 2007 respectively, and each of these services monitors and reports screening outcomes in accord with European standards.

Digital mammography was performed with a Selenia Dimensions Unit allowing 2D/3D mammography to be acquired as the COMBO procedure (Hologic, Inc. Bedford MA, USA): this acquires 2D and 3D images at the same screening examination and with a single breast compression per view; each of the 2D and 3D acquisitions consisted of bilateral two-view (mediolateral oblique and cranio-caudal) mammography.

Screen-reading and recall rules

The screen-reading protocol in STORM has been detailed in our previous publication, and was based on double-reading whereby recall was decided based on recall by either radiologist. Mammography was interpreted sequentially by radiologists, initially based on 2D-mammography alone, and re-interpreted by the same radiologists based on integrated 2D/3D mammography. Radiologists were required to record whether or not to recall at each screen-reading phase before continuing to the next phase: recall was based on recall at either screen-reading phase. Eight radiologists were involved in screen-reading, five in Trento and three in Verona—all were experienced in breast imaging and had mammography screening experience. They had received some training in 3D-mammography before participating in STORM. Previous mammograms, where available, were displayed to the radiologist at the time of screen-reading.

Detection measures & outcomes

We defined detection measures in terms of the number of breast cancers detected at each screen-reading phase (2D-mammography, and integrated 2D/3D mammography), as well as the cancer detection rate (CDR), the incremental CDR, and the number and rates of false positive recalls (FPR).

Outcomes were ascertained on the basis of excision histology in those who had surgery, or based on the complete assessment outcome inclusive of any work-up imaging or histology from core needle biopsy, in all recalled subjects. Because our study focuses on the difference in detection by the two screen-reading methods, some cancers will have been missed by both 2D and 2D/3D mammography (these represent future interval cancers). This outcome is not evaluated in the present study and does not affect our estimates for comparative detection for 2D-only versus integrated 2D/3D mammography.

Statistical analysis

The statistical planning of the STORM trial has been described by Ciatto et al. For the present evaluation, we examined the detection measures defined above for each screening service, and compared these for 2D-mammography and integrated 2D/3D mammography using Mc Nemar’s test for paired binary data. We also compared CDR and incremental CDR, as well as measures of FPR, between the two services using the chi-squared test for independent proportions.

FPR measures for each service were examined based on the overall FPR percentage (using the recall rule applied in the trial), and the FPR proportions contributed by both, or each of, 2D and 2D/3D screening (for the latter, this means the FPR caused only at 2D-mammography or only at integrated 2D/3D-mammography). We also estimated the conditional FPR using integrated 2D/3D mammography positivity as a condition to recall — under this condition, cancers recalled at 2D-mammography only would not be recalled. All analyses were conducted using SAS. All confidence intervals and P-values were computed using exact methods.

Results

Amongst screening participants (median age 58 years) 59 breast cancers were detected at screening in 57 subjects; the analysis is simplified by double-counting the 2 subjects with bilateral cancer (hence we report 7294 screens for total). Cancer characteristics including histology distribution have been described in our primary report of STORM — 52 invasive cancers and 7 ductal carcinoma in-situ (DCIS) were detected at screening. Centre-specific breast cancer detection data are reported in Tables 1 and 2. Cross-classified results (Table 1) show that integrated 2D/3D screening resulted in significantly more cancers detected compared to screening with 2D-mammography for each centre. Of the 33 cancers detected in Trento, 21 were detected at both 2D and 2D/3D screening, and 12 cancers were detected only with integrated 2D/3D screening compared with none detected at 2D-only screening (P < 0.001). Of 26 cancers detected in Verona, 18 were detected at both 2D and 2D/3D screening, and 8 cancers were detected only with integrated 2D/3D screening compared with none detected at 2D-only screening (P = 0.008).

Table 2 shows that centre-specific CDR at 2D-mammography screening was similar for Trento (based on 4254 screens) and Verona (based on 3040 screens), and that CDR for integrated 2D/3D-screening did not differ between the two services (P = 0.79). The incremental CDR attributable to 3D-mammography was very similar for both centres: 2.8/1000 screens for Trento and 2.6/1000 screens for Verona (Table 2).

Centre-specific measures of FPR are shown in Tables 3 and 4. For the Trento centre, there were 239 FPR (5.7% of screens classified as not having cancer) comprised of the following: 103 FPR at both
Discussion

Overviews of the breast cancer screening trials confirm that mammography screening confers benefit whilst noting potential harms [10,11]. Methods that improve the precision of population breast screening, allowing better differentiation between true and false screen-detection, would therefore be worthy of investigation, and presently none seem more promising than 3D-mammography (digital breast tomosynthesis) [2,5,6,12]. In this evaluation, we report detection measures for two screening centres that participated in STORM [5], a population-based breast screening trial. Our intention was to elucidate consistencies and/or differences in the effect of integrating 3D with 2D-mammography for breast screening, to inform population screening programs of the potential impact and transferability of this screening strategy at the level of the individual screening service. We found that the increased breast cancer detection attributable to adding 3D-mammography as part of an integrated screen-reading strategy was evident and was also significant for each of the services that participated in the trial, and the estimated incremental CDR attributable to 3D-mammography was similar at 2.8/1000 screens for Trento and 2.6/1000 screens for Verona. It is important to interpret these results with the understanding that ‘baseline’ CDR (for standard 2D-mammography screening) for each of these screening services did not significantly differ between services, nor did the CDR achieved by integrated 2D/3D-mammography screening. Hence the estimated incremental CDR indicates a consistent and true reflection of the effect that is reasonably attributable to 3D-mammography (and not an artificial effect that could have resulted from a low CDR at baseline screening for any individual screening service).

Interpretation of the estimated incremental CDR evident for the Trento and Verona screening centres should factor some of the potential study limitations, as outlined in our report of STORM [5], namely that we used a sequential screen-reading design. Although this confers the methodological strength of maintaining the same screening mammogram and the same screen-reader (providing paired data for each subject), it also means that the 2D-mammography screen-read was compared to an integrated screen-read of the 2D with 3D mammograms. Therefore our results cannot be taken as a comparison of 2D-mammography versus 3D-mammography and cannot be extrapolated as such. In addition, we defined outcomes in terms of detection measures at screening, hence our cancer detection estimates did not include the additional follow-up required to ascertain interval cancers [5] – this does not invalidate our comparative detection measures because interval cases will have been missed by both screen-reading methods [5]. Interpretation of our results should also factor that both screening centres used double-reading as the standard practice in European programs, and the incremental CDR from integrated 2D/3D-mammography on a single-reading strategy is not established (and is the focus of an ongoing evaluation).

An important finding of this study is that the effect of 2D/3D-mammography on false positive recalls (FPR) somewhat differed at the service level. The findings across both centres were consistent in that approximately half of FPR were falsely recalled at 2D-mammography and also at integrated 2D/3D-mammography screening, with evidence that significantly more FPR were caused

---

**Table 1**

Breast cancer detection for screening centres participating in the STORM trial [5].

<table>
<thead>
<tr>
<th>Breast cancers detected at screening</th>
<th>Breast cancers detected at screening</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Trento</strong></td>
<td><strong>Verona</strong></td>
</tr>
<tr>
<td>Mammogram</td>
<td>Mammogram</td>
</tr>
<tr>
<td>2D/3D positive</td>
<td>2D/3D positive</td>
</tr>
<tr>
<td>2D positive</td>
<td>2D positive</td>
</tr>
<tr>
<td>21</td>
<td>18</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>21</td>
<td>18</td>
</tr>
<tr>
<td>2D negative</td>
<td>2D negative</td>
</tr>
<tr>
<td>12</td>
<td>8</td>
</tr>
<tr>
<td>0[a]</td>
<td>0[a]</td>
</tr>
<tr>
<td>12</td>
<td>8</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>Total</strong></td>
</tr>
<tr>
<td>33</td>
<td>26</td>
</tr>
<tr>
<td>0[a]</td>
<td>0[a]</td>
</tr>
<tr>
<td>33</td>
<td>26</td>
</tr>
</tbody>
</table>

**P < 0.001**

**<sup>a</sup> Exact P-value for McNemar’s test for paired binary data.**

**<sup>b</sup> Based on cancers detected in the study population at screening and does not include follow-up for interval cancers, hence this cell has a value of 0 – this does not affect the comparative detection data shown in this cross-tabulation (see Methods).**

---

**Table 2**

Comparison of breast cancer detection rates for two screening centres participating in the STORM trial [5].

<table>
<thead>
<tr>
<th>Breast cancers detected at screen-reading</th>
<th>Number of screening examinations</th>
<th>No. of detected cancers</th>
<th>Cancer detection rates (CDR) per 1000 screens (95% CI)</th>
<th>P&lt;sup&gt;a&lt;/sup&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>2D mammography</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Trento</td>
<td>4254</td>
<td>21</td>
<td>4.9 (3.1–7.5)</td>
<td>0.63</td>
</tr>
<tr>
<td>Verona</td>
<td>3040</td>
<td>18</td>
<td>5.9 (3.5–9.3)</td>
<td>0.79</td>
</tr>
<tr>
<td><strong>Integrated 2D/3D mammography</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Trento</td>
<td>4254</td>
<td>33</td>
<td>7.8 (5.3–10.9)</td>
<td>1.0</td>
</tr>
<tr>
<td>Verona</td>
<td>3040</td>
<td>26</td>
<td>8.6 (5.6–12.5)</td>
<td></td>
</tr>
<tr>
<td><strong>Incremental CDR attributed to adding 3D mammography in screen-reading</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Trento</td>
<td>4254</td>
<td>12</td>
<td>2.8 (1.5–4.9)</td>
<td></td>
</tr>
<tr>
<td>Verona</td>
<td>3040</td>
<td>8</td>
<td>2.6 (1.1–5.2)</td>
<td></td>
</tr>
</tbody>
</table>

**<sup>a</sup> P compares detection between the two screening centres participating in the trial.**
only at 2D-mammography compared with FPR caused only at integrated 2D/3D-mammography (Tables 3 and 4). On the basis of these data, we estimated that integrated 2D/3D-mammography screening would effectively reduce FPR if a conditional recall rule was applied (instead of the trial protocol which recalled based on either screen-reading phase). However, because FPR caused only at 2D-mammography was lower for Verona than Trento (Table 4), the potential reduction in the FPR proportion by using recall conditional to integrated 2D/3D-mammography-positivity was more evident for Trento (the centre with higher FPR at 2D-mammography). This would suggest that, in terms of potential reduction in FPR from integrated 2D/3D-mammography, screening services with relatively higher FPR at 2D-mammography would stand to gain more by using 2D/3D-mammography, than those with relatively lower FPR at 2D-mammography, because there is less scope for integrated 2D/3D-mammography to further lower the FPR proportion for the latter. This is reflected in our estimated reduction in the proportion of FPR that had conditional recall been applied, which differed between the two centres (21.0% for Trento versus 11.5% for Verona; \( P < 0.02 \)). Although the estimated reductions in the proportion of FPR represent (hypothetical) calculations, they nonetheless highlight a differential effect on potential reduction in FPR from integrated 2D/3D-mammography. This is not to say that some centres will not benefit through a reduction in FPR by using integrated 2D/3D-mammography, we only point out that the expected improvement (reduction) in terms of FPR is likely to vary across screening settings and services, in part driven by the existing FPR using standard 2D-mammography screening. Therefore services that are experiencing high FPR rates in breast screening would stand to benefit the most from introducing 2D/3D-mammography as a harm (false recall) reduction strategy. This conclusion would need to factor the extent to which trial findings would be reproducible in screening settings that do not use double-reading since our trial was implemented in programs practicing double-reading. We have explicitly stated that this evaluation provides evidence to inform population screening programs of the potential impact of integrated 2D/3D-mammography screening at the level of the individual screening service. Our findings should not be taken to advocate a change from established mammography screening practice—we propose that this should only occur with consideration of evidence from several screening trials [4,5,12]. For now, the data we report in this study facilitate better understanding of the effect of integrated 2D/3D-mammography on both true-positive and false-positive detection measures, at the centre level, and can be used to guide decisions about breast screening practice and the planning of future trials.

In conclusion, our evaluation of screening detection measures for two centres that participated in a prospective population-based screening trial shows that integrating 3D-mammography with 2D-mammography resulted in consistent and significant improvement in breast cancer detection for both centres. In addition, there was consistent evidence that more FPR were caused only at 2D-mammography compared with FPR caused only at integrated 2D/3D-mammography, however the extent to which using integrated 2D/3D-mammography will reduce FPR will be partly dependent on the level (magnitude) of FPR at 2D-mammography. Based on our study findings, services experiencing a relatively high FPR at conventional 2D-mammography screening would be expected to gain the most by a transition to integrated 2D/3D-mammography screening.

### Role of funding source
The funding sources did not have any role in the design of the study, in the collection, analysis, or interpretation of the data, in the writing of the report, or in the decision to submit the paper.

### Conflict of interest statement
None declared.
Acknowledgments

Centres and investigators in Italy participating in the STORM trial declare receiving assistance from Hologic (USA) and Technologic (Italy) in the form of tomosynthesis technology, and travel support to attend collaborators’ meetings. Assoc/Prof Houssami receives research support through a National Breast Cancer Foundation (NBCF Australia) Practitioner Fellowship. Professor Macaskill receives research support through Australia’s National Health and Medical Research Council Program grant 633003 to the Screening & Test Evaluation Program.

References