Screening outcome for consecutive examinations with digital breast tomosynthesis versus standard digital mammography in a population-based screening program

Tone Hovda[1] · Siri H. B. Brandal[2] · Sofie Sebuødegård[3] · Åsne S. Holen[3] · Hilde Bjørndal[1] · Per Skaane[2] · Solveig Hofvind[3,4] 🆔

Abstract

Objectives To retrospectively investigate early performance measures of digital breast tomosynthesis (DBT) versus standard digital mammography (DM) for consecutive screening rounds.

Methods We included information about 35,736 women screened in BreastScreen Norway, 2008–2016, with at least two consecutive screening examinations. The pair of two consecutive screening examinations was the unit of analysis, and results from the subsequent examination were the measure of interest. Screening technique changed during the study period, resulting in four study groups: DM after DM, DBT after DM, DM after DBT, and DBT after DBT. We compared selected early performance measures between the study groups.

Results Recall for DM after DM was 3.6% and lower for all other study groups ($p < 0.001$). The rate of screen-detected breast cancer was 4.6/1000 for DM after DM; for DBT after DM and DBT after DBT, it was 9.9/1000 and 8.3/1000, respectively ($p < 0.001$ relative to DM after DM), and for DM after DBT 4.3/1000. The rate of tubular carcinoma was higher for DBT after DBT or after DM compared with DM after DM ($p < 0.01$). The rate of histologic grade 1 tumors was higher for DBT after DM compared with DM after DM ($p < 0.001$). We did not observe any statistical difference in the interval cancer rates.

Conclusions Lower recall and higher cancer detection rates for screening with DBT were sustainable over two consecutive screening rounds. Positive predictive values were higher for DBT than DM. There were no differences in the interval cancer rates between the study groups.

Key Points

• *There is limited knowledge about early performance measures for screening with digital breast tomosynthesis beyond one screening round.*

• *A decline in recall rate and an incline in the rate of screen-detected breast cancer were observed for women screened with DBT compared with DM, irrespective of prior screening technique. The interval breast cancer rate did not differ statistically for women screened with DBT versus DM.*

• *Tumor characteristics tended to be prognostic favorable for DBT compared with DM with no differences in rates of more advanced cancers. The clinical significance of increased cancer detection and the potential for future mortality reduction remain unknown.*

Keywords Screening · Breast neoplasm · Digital breast tomosynthesis · Mammography · Digital mammography

✉ Solveig Hofvind
solveig.hofvind@kreftregisteret.no

[1] Department of Radiology, Vestre Viken Hospital, Drammen, Norway

[2] Division of Radiology and Nuclear Medicine, Oslo University Hospital, Oslo, Norway

[3] Cancer Registry of Norway, P.O 5313, Majorstuen, 0304 Oslo, Norway

[4] Oslo Metropolitan University, Faculty of Health Science, Oslo, Norway

Abbreviations

CC	Craniocaudal
DBT	Digital breast tomosynthesis
DCIS	Ductal carcinoma in situ
DM	Digital mammography
ER	Estrogen receptor
FN	False negative
FP	False positive
HER2	Human epidermal growth factor receptor 2
IDC	Invasive ductal carcinoma
ILC	Invasive lobular carcinoma
ITC	Invasive tubular carcinoma
MLO	Mediolateral oblique
OTST	Oslo Tomosynthesis Screening Trial
OVVV-study	Oslo-Vestfold-Vestre Viken study
PPV	Positive predictive value
PR	Progesterone receptor
SM	Synthetic mammogram
STORM	Screening with Tomosynthesis OR standard Mammography
TN	True negative
TP	True positive

Introduction

Digital breast tomosynthesis (DBT) has been claimed to be a more sensitive screening tool than standard digital mammography (DM) due to an increased rate of screen-detected breast cancer [1–5]. The effect of DBT on recall rates has varied, probably due to the different screening regimes and logistics [2–7]. The limited numbers of studies reporting interval breast cancer have not been able to show any substantial difference between the two techniques [8, 9]. In addition, the increase of screen-detected breast cancer has raised the question about tumor characteristics and clinical significance related to screening with DBT [8, 10].

In most screening studies, DBT has been used in combination with DM (DM+DBT) [11], which has led to concerns about the radiation dose, as the dose is almost double that of standard DM [1]. To overcome this problem, vendors have developed an algorithm for reconstructing synthetic two-dimensional images from the raw DBT data (SM). SM seems to equal DM in combination with DBT regarding cancer detection [5, 12–14].

Most studies on screening using DBT have been performed as paired or unpaired studies [2] with only one, a prevalent screening examination with DBT, per woman. However, in a screening setting, early performance measures in subsequent screening rounds with DBT are of substantial interest. McDonald et al [15] evaluated the effect of DBT in subsequent rounds and reported sustainable rates of recalls and screen-detected breast cancer, as well as a reduction in interval breast cancer. The impact of DBT on recall rate is less obvious in studies from Europe versus the USA, due to a generally lower rate, different study designs, as well as screening techniques [1–4, 7, 16]. Further studies, including subsequent screening examinations with DBT, also from Europe, are thus needed.

In BreastScreen Norway, DBT has been used in two different study settings in the Oslo region: the Oslo Tomosynthesis Screening Trial [4] and Oslo-Vestfold-Vestre Viken study (OVVV-study) [17]. Taking advantage of these studies, we were able to provide novel information on the impact of DBT used in consecutive screening rounds and for different combinations of prior and subsequent screening techniques. The primary objective of our study was to compare early performance measures as recall, screen-detected, and interval breast cancer in consecutive screening rounds with DBT and DM, including results of prognostic and predictive tumor characteristics.

Materials and methods

This retrospective study was based on data from BreastScreen Norway. The screening program is administered by the Cancer Registry of Norway [18] and thus covered by the Cancer Registry Regulation, giving approval with waiver of informed consent to perform studies based on the data collected in the program [19]. The OTST and OVVV-study were covered by approvals from the Regional Ethical Committee for Medical and Health Research Ethics and the Institutional Data Protection Officer, respectively.

BreastScreen Norway is described in detail elsewhere [18]. In short, the program is population-based, offering women aged 50–69 two-view mammographic screening every second year. The screening procedure includes independent double reading with consensus or arbitration. In this study, screen reading was performed by radiologists with varying experiences in breast imaging and screen reading, from beginners to more than 20 years of experience (Appendix 1).

The study population included 35,736 women, all screened in Oslo as part of BreastScreen Norway, within the study period, from 2008 to 2016. All women screened with at least two consecutive screening examinations during this period were eligible for inclusion. One observation/unit of analysis included a pair of two consecutive screening examinations. The screening result of the first examination in the pair was either negative or positive with a negative assessment, while the screening result of the consecutive examination was the outcome of interest. Some women (n = 19,498) were screened more than twice during the study period and may thus contribute with more than one observation. In total, 69,624 observations were included in the study.

The screening techniques varied during the study period. Standard DM was used as a screening technique in 2008, 2009, 2010, 2013, and 2016, DM+DBT in 2011 and 2012, and SM+DBT in 2014 and 2015. GE Senographe and GE Seno Advantage Workstation were used from 2008 to August 2010 (DM), and Hologic Selenia Dimensions and Secureview Hologic Workstation from 2010 to the end of the study period (for DM, for DM+DBT, and for SM+DBT).

Four study groups were established. We conditioned on the prior screening technique when creating the four groups, while results from the subsequent screening examination were the measures of interest in this study. The "DM after DM" group consisted of the pair of two consecutive screening exams with DM. A pair of two consecutive screening exams including a prior DM and a consecutive DBT was the group "DBT after DM." A pair of two consecutive screening exams with DBT was the "DBT after DBT" group, while the pair of a prior DBT and a consecutive DM represented the group "DM after DBT" (Fig. 1).

All screening examinations included two views, craniocaudal (CC) and mediolateral-oblique (MLO), of both breasts. For those with a prior DBT, images were available at screen reading and/or consensus in addition to prior DM images. We analyzed the data on screening examination level and assumed independence between the observations. To test this assumption, Pearson's correlation coefficients for recall rate and rates of screen-detected and interval breast cancer were estimated.

Recall rate was defined as the percentage of screening examinations resulting in a call back for further assessment due to abnormal mammographic findings among all screening examinations performed in the actual study group. Rates of screen-detected and interval breast cancer were estimated as the number of screen-detected or interval breast cancer, invasive or ductal carcinoma in situ (DCIS), per 1000 screening examinations, for each study group. Rates were given for DCIS and invasive breast cancer separately and in total. Interval breast cancer was defined as breast cancer diagnosed after a negative screening examination or > 6 months after a false-positive screening result and within 2 years after the last screening examination [20]. Due to the requirement of 2-year follow-up, interval breast cancer rates will be given for women subsequently screened in 2014 or earlier ($n = 44,515$ screening examinations). Accordingly, women screened with SM+DBT after DM and with DM after SM+DBT are not included in the analyses of interval breast cancer (Table 1).

Women diagnosed with screen-detected breast cancer were considered true-positives (TP) while women diagnosed with interval breast cancer were considered false-negative (FN). Women screened and recalled, but not diagnosed with breast cancer, were false-positive (FP) while women with a negative screening examination without interval breast cancer were true-negative (TN). Sensitivity was defined as TP/(TP+FN) and specificity as TN/(TN+FP). Positive predictive value for recall (PPV-1) was estimated as the number of exams resulting in a diagnosis of screen-detected breast cancer divided by the total number of recalls within each study group.

We performed descriptive analyses and present prognostic (histological type and grade, tumor diameter, and lymph node involvement) and predictive (molecular subtypes) tumor characteristics as rates per 1000 women screened for screen-detected and interval breast cancers.
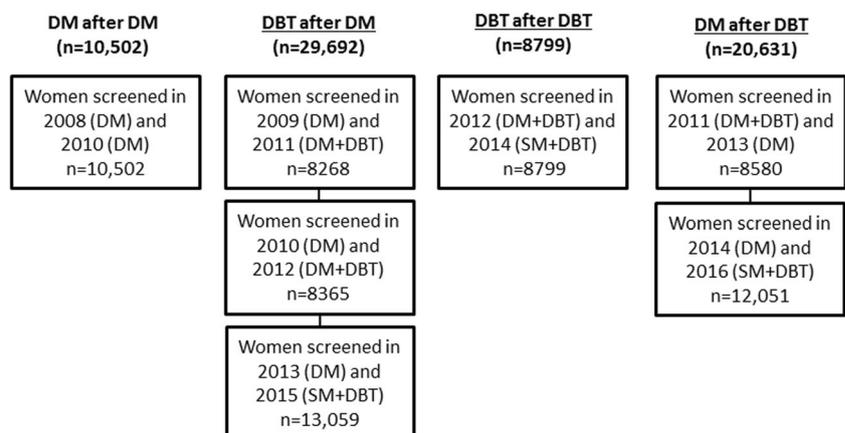
All results are given separately for SM and for DM in combination with DBT—six groups—in Appendices 2 and 3.

The differences between groups were tested using t tests, chi-square test, and test of proportions. A p value below 0.05 was considered statistically significant. We used the Bonferroni correction when appropriate. Statistical analyses were performed using IBM SPSS Statistics version 25.

Results

The mean and median age of the women at subsequent screening examination was 60.5 (range 60.3–60.7) and 60.0 years (range 60.0–61.0), respectively (Table 1).

Fig. 1 Four study groups showing pairs with different combinations of prior and subsequent screening technique. The subsequent examination in each pair occurs 2 years after the prior examination

DM after DM (n=10,502)
Women screened in 2008 (DM) and 2010 (DM) n=10,502

DBT after DM (n=29,692)
Women screened in 2009 (DM) and 2011 (DM+DBT) n=8268
Women screened in 2010 (DM) and 2012 (DM+DBT) n=8365
Women screened in 2013 (DM) and 2015 (SM+DBT) n=13,059

DBT after DBT (n=8799)
Women screened in 2012 (DM+DBT) and 2014 (SM+DBT) n=8799

DM after DBT (n=20,631)
Women screened in 2011 (DM+DBT) and 2013 (DM) n=8580
Women screened in 2014 (DM) and 2016 (SM+DBT) n=12,051

Table 1 Age (mean and median), recalls (%), screen-detected and interval breast cancer (per 1000 screened), positive predictive value of recalls (PPV-1, %), sensitivity (%), and specificity (%) among women subsequently screened with different combinations of screening techniques in Oslo as part of BreastScreen Norway, 2010–2016

Subsequently screened 2010–2016										
	DM after DM $n = 10,502$		DBT after DM $n = 29,692$		DBT after DBT $n = 8799$		DM after DBT $n = 20,631$		Total $n = 69,624$	
Age at screening (years)										
Mean (SD)	60.5	5.1	60.7	5.1	60.3	5.4	60.2	5.2	60.5	5.2
Median (IQR)	61	(56–65)	60	(56–65)	60	(55–65)	60	(56–65)	60	(56–65)
Recalled (%)	373	3.6	801	2.7[1]	168	1.9[1,2]	444	2.2 [1,2]	1786	2.6
Screen-detected cancer (per 1000 screened)										
DCIS	9	0.9	47	1.6[5]	16	1.8[5]	10	0.5	83	1.2
Invasive	39	3.7	248	8.4	57	6.5	79	3.8	422	6.1
Total	48	4.6	295	9.9[1]	73	8.3[1]	89	4.3[2,3]	505	7.3
PPV-1 (%)		12.9		36.4[1]		43.5[1]		20.0[2,3,4]		28.1
Subsequently screened 2010–2014 with 2-year follow-up										
	DM after DM $n = 10,502$		DBT after DM $n = 16,633$		DBT after DBT $n = 8799$		DM after DBT $n = 8580$		Total $n = 44,514$	
Interval breast cancer (per 1000 screened)										
DCIS	1	0.1	1	0.06	0	0.0	3	0.4	5	0.1
Invasive	22	2.1	31	1.9	18	2.1	23	2.3	94	2.1
Total	23	2.2	32	1.9	18	2.1	26	3.0	99	2.2
Sensitivity (%)		67.6		83.7		80.2		62.9		76.9
Specificity (%)		96.9		98.0		98.9		98.6		98.0
Total 2-year cancer incidence (screen-detected and interval breast cancer)										
	DM after DM $n = 10,502$		DBT after DM $n = 16,633$		DBT after DBT $n = 8799$		DM after DBT $n = 8580$		Total $n = 44,514$	
SDC (rates per 1000)	48	4.6	164	9.9	73	8.3	44	5.1	329	7.4
IBC (rates per 1000)	23	2.2	32	1.9	18	2.1	26	3.0	99	2.2
Total (rates per 1000)	71	6.8	196	11.8[1]	91	10.3[4]	70	8.2[6]	428	9.6

[1] $p = <0.001$ compared with DM after DM

[2] $p = <0.001$ compared with DBT after DM

[3] $p < 0.001$ compared with DBT after DBT

[4] $p < 0.05$ compared with DM after DM

[5] $p < 0.001$ compared with DM after DBT

[6] $p < 0.05$ compared with DBT after DM

Recall for women screened with DM after DM was 3.6%, higher than for all other study groups ($p < 0.001$) (Table 1). The lowest recall, 1.9%, was observed among women with two consecutive DBT examinations. The recall for DM after DBT of 2.2% was significantly lower than for DBT after DM (2.7%, $p < 0.001$).

The rate of screen-detected breast cancer was 4.6/1000 screens for DM after DM (Table 1). For DBT after DM, it was 9.9/1000 screens, and for two consecutive DBTs, 8.3/1000 screens ($p < 0.001$ for both, compared with DM after DM). For those screened with DM after DBT, the rate of screen-detected breast cancer was 4.3/1000 screens ($p < 0.001$ compared with DBT after DM or DBT). The detection rate of DCIS was higher for both DBT after DM (1.6/1000) and DBT after DBT (1.8/1000) compared with DM after DBT (0.5/1000 screens, $p < 0.001$ for both).

PPV-1 ranged from 12.9% (DM after DM) to 43.5% (DBT after DBT, $p < 0.001$) (Table 1). PPV-1 was higher for DM after DBT compared with DM after DM ($p < 0.05$).

The rates of interval breast cancer ranged from 1.9 to 3.0/1000, with no statistically significant differences between groups (Table 1). The sensitivity ranged from 63% (DM after DBT) to 84% (DBT after DM) and specificity ranged from 97% (DM after DM) to 99% (DBT after DBT).

The rate of invasive ductal carcinoma was 6.4/1000 for women screened with DBT after DM, compared with 2.5/1000 for those screened with DM after DM, $p < 0.001$ and statistically significant after the Bonferroni correction. The

rate of IDC in the DBT after DBT group was 4.8/1000 ($p < 0.01$ compared with DM after DM), and 3.2/1000 in the DBT after DM group. As for invasive tubular carcinoma, the highest rates were observed for DBT after DM (0.7/1000) and DBT after DBT (0.8/1000), compared with none for DM after DM, $p < 0.01$ for both. No differences in mean and median tumor diameter were observed between groups. The rate of tumors < 10 mm was 3.3/1000 in the DBT after DBT group compared with 1.5/1000 in the DM after DM group ($p < 0.01$) and the rates for tumors 10–19 mm were 3.7/1000 and 1.1/1000, respectively ($p < 0.001$). For DBT after DBT, the rates were 2.8/1000 and 2.7/1000, respectively ($p < 0.01$ and $p < 0.05$ compared with DM after DM). No differences in rates of tumors > 20 mm were observed between groups (Table 2). Further, for DBT after DM, the rates of grade 1 invasive cancer were 3.2/1000 and of grade 2 3.7/1000, compared with 1.0/1000 and 1.2/1000, respectively, for DM after DM ($p < 0.001$ for both). For DBT after DBT, the rates were 3.0/1000 ($p < 0.01$ compared with DM after DM) and 2.5/1000 ($p < 0.05$ compared with DM after DM). No differences in the rates of lymph node–positive disease or human epidermal growth factor receptor 2 (HER2)–positive or triple-negative tumors were observed between groups (Table 2).

For interval breast cancers, no statistical differences in rates of histological type, tumor diameter grade, or molecular subtypes were observed when the Bonferroni correction was applied (Table 3).

Recalls and screen-detected breast cancer, and PPV-1 for DM/SM+DBT after DM, were in accordance with the combined DBT after DM group (Appendix 2). This also applied to DM after DM/SM+DBT and the combined DM after DBT group. The groups including SM+DBT tended to be of lower grade compared with groups including DM+DBT; otherwise, no statistical differences regarding histopathological characteristics were observed (Appendix 3).

The estimated Pearson correlation coefficients were at most 0.06 for recall and 0.01 for screen-detected and interval breast cancer (results not shown in table).

Discussion

We found a lower recall among women screened with DBT after DM compared with DM after DM, but even lower among those with a prior DBT (DM or DBT after DBT). The highest rates of screen-detected breast cancer were observed among women screened with DBT after DM or consecutive DBTs. PPV-1 was highest for DBT after DBT and lowest for DM after DM. Further, higher rates of invasive tubular carcinomas, tumors < 20 mm, grade 1 and grade 2 tumors, and luminal A and luminal B HER2-negative subtypes were observed for women screened with DBT after DM or after two consecutive DBTs. No statistical differences

in the rate of interval breast cancer were observed between the study groups.

A significantly lower recall for those with a prior DBT than for those with a prior DM may indicate that the availability of prior DBT examinations at screen reading and/or consensus lowers the probability of recall, possibly due to a large amount of image information available. Additionally, apparent suspicious findings on screening images could be explained by the superimposition of overlapping tissue when compared with prior DBT images, resulting in a negative screening interpretation. A learning effect as the radiologists were becoming more experienced with DBT during the study period may also be considered.

The highest rates of screen-detected cancer were observed in the groups screened with DBT after a prior DM. This increase is in line with the results from other studies [3, 5, 7] and might be partly attributed to the increased conspicuity of lesions and reduced masking of overlapping tissues by DBT [1]. Further, tumors may be diagnosed at an earlier stage than in the absence of DBT, thereby prolonging the lead-time. Based on this, the question whether the increased detection rate is only transient may be raised. In our study, the increased rate of screen-detected breast cancer sustained two screening rounds. However, the detection rate for DBT after DBT tended to be lower than for DBT after DM, though not statistically significant. Follow-up through several subsequent screening rounds is needed to evaluate whether the increased cancer detection rate observed for DBT is transient and due to extended lead-time.

An interesting question when screening with DBT is do the excess in the number of tumors represent "killing cancers," or small and less aggressive tumors of inferior clinical significance? Studies are showing conflicting results regarding that topic [4, 7, 17, 21]. We found increased rates of invasive tubular carcinomas, tumors < 20 mm, histologic grade 1 and 2 tumors, and tumors with favorable lymph node status and molecular subtypes for women screened with DBT compared with DM, while no differences in rates of tumors > 20 mm, histologic grade 3 tumors, lymph node–positive, HER2-positive, or triple-negative tumors were observed. The detection rate of DCIS was also statistically higher among those screened with DBT compared with DM after DBT, but not compared with DM after DM, though the numbers are small. This study was not designed to investigate overdiagnosis, but our findings show that the excess cancers diagnosed with DBT seem to be less aggressive tumors. We cannot definitely conclude about the clinical implications of the increased rate of cancers diagnosed by DBT. However, having the debate of overdiagnosis in mind, that is the detection of small, less aggressive tumors which would possibly not have been diagnosed or come to a clinical appearance in the women's lifetime in the absence of screening, DBT may contribute in that direction. Further studies in this regard are needed. Additionally, studies on even more stratified

Table 2 Screen-detected cancer. Histopathological characteristics of invasive tumor among women screened with different combinations of screening techniques in Oslo as part of BreastScreen Norway, 2010–2016. Rates per 1000 screened women

	DM after DM		DBT after DM		DBT after DBT		DM after DBT		Total	
	n = 39		n = 248		n = 57		n = 79		n = 423	
	n	per 1000	n	per 1000	n	per 1000	n	per 1000	n	per 1000
Histopathological type										
IDC	26	2.48	191	6.43[1]	42	4.77[2]	65	3.15	324	4.65
ILC	12	1.14	28	0.94	7	0.80	6	0.29[2]	53	0.76
ITC	0	0	22	0.74[2]	7	0.80[2]	3	0.15	32	0.46
Other inv	1	0.10	7	0.24	1	0.11	5	0.24	14	0.20
Tumor diameter (mm)										
Mean (SD)	15.2	(10.7)	13.3	(10.1)	12.2	(8.7)	12.2	(10.7)	13.1	(10.1)
Median (IQR)	12	(7–24)	11	(7–15)	11	(7–14)	10	(6–15)	11	(7–15)
< 10 mm	16	1.52	99	3.33[2]	25	2.84[3]	34	1.65	174	2.50
10–19 mm	11	1.05	111	3.74[1]	24	2.73[2]	34	1.65	180	2.58
20–29 mm	8	0.76	22	0.74	2	0.23	6	0.29	38	0.55
30 mm+	4	0.38	13	0.44	6	0.68	3	0.15	26	0.37
Not available	0		3		0		2		5	
Grade										
1	10	0.95	94	3.17[1]	26	2.95[2]	21	1.02	151	2.17
2	13	1.24	111	3.74[1]	22	2.50[3]	42	2.04	188	2.70
3	16	1.52	38	1.28	9	1.02	16	0.78	79	1.13
Not available	0		5		0		0		5	
Lymph node										
Negative	33	3.14	202	6.80[1]	43	4.89	64	3.10	342	4.91
Positive	6	0.57	38	1.28	11	1.25	13	0.63	68	0.98
Not available	0		8		3		2		13	0.19
Subtype										
Luminal A	22	2.09	180	6.06[1]	35	3.98[3]	56	2.71	293	4.21
Luminal B, HER2−	5	0.48	37	1.25[3]	16	1.82[2]	14	0.68	72	1.03
Luminal B, HER2+	3	0.29	7	0.24	1	0.11	1	0.05	12	0.17
HER2	2	0.19	2	0.07	0	0	3	0.15	7	0.10
Triple negative	7	0.67	14	0.47	1	0.11	5	0.24	27	0.39
Not available	0		8		4				12	

IDC, invasive ductal carcinoma; ILC, invasive lobular carcinoma; ITC, invasive tubular carcinoma. p-values calculated with DM after DM as reference

[1] $p < 0.001$ compared with DM after DM, statistically significant after Bonferroni correction

[2] $p < 0.01$ compared with DM after DM

[3] $p < 0.05$ compared with DM after DM

histopathological analyses and treatment options will bring valuable information into this debate.

In our study, PPV-1 was significantly higher among those screened with DBT; thus, the rate of false-positives was lower. In addition to a lower recall in general among those screened with DBT, this is beneficial both to the screened women and to society due to a reduced burden of false-positive screening results [1]. The highest value of PPV-1 was observed among those with two consecutive DBTs, mainly due to the very low recall rate in this study group.

Interval breast cancer is essential when evaluating the effectiveness of a new screening technique [1, 15]. The results from the OTST (Oslo Tomosynthesis Screening Trial) and the STORM (Screening with Tomosynthesis OR standard Mammography) trial did not show any reduction in the interval breast cancers among those screened with DBT compared with DM [8, 9]. Our results support these findings. However, an obvious limitation of the results is the small number of cancer cases. Further studies including analyses of interval breast cancer after screening with DBT are thus needed.

Table 3 Interval breast cancer. Histopathological characteristics of invasive tumor among women screened with different combinations of screening techniques in Oslo as part of BreastScreen Norway, screened 2010–2014, with 2-year follow-up. Rates per 1000 screened women

	DM after DM		DBT after DM		DBT after DBT		DM after DBT	
	$n = 22$		$n = 31$		$n = 18$		$n = 23$	
	n	per 1000	n	per 1000	n	per 1000	n	per 1000
Histopathological type								
IDC	18	1.71	26	1.56	18	2.05	15	1.75
ILC	4	0.38	4	0.24	0		4	0.47
ITC	0		0		0		0	
Other inv	0		1	0.06	0		4	0.47[2]
Tumor diameter (mm)								
Mean (SD)	21.3	(11.8)	26.0	(16.6)	18.0	(6.6)	18.2	(8.4)
Median (IQR)	19	(12–25)	23.5	(15–35)	16.5	(13–24)	15	(12–25)
< 10 mm	2	0.19	5	0.30	2	0.23	2	0.23
10–19 mm	9	0.86	7	0.42	8	0.91	11	1.28
20–29 mm	6	0.57	9	0.54	7	0.80	2	0.23
30 mm+	4	0.38	9	0.54	1	0.11	4	0.47
Not available	1		1		0		4	
Grade								
1	7	0.67	1	0.06[1]	2	0.23	2	0.23
2	13	1.24	17	1.02	7	0.80	12	1.40
3	2	0.19	12	0.72	9	1.02[2]	8	0.93[2]
Not available	0		1		0		1	
Lymph node								
Negative	12	1.14	18	1.08	14	1.59	17	1.98
Positive	10	0.95	12	0.72	4	0.45	4	0.47
Not available	0		1		0		2	
Subtype								
Luminal A	11	1.05	13	0.78	11	1.25	10	1.17
Luminal B, HER2−	6	0.57	6	0.36	2	0.23	7	0.82
Luminal B, HER2+	0		3	0.18	0		1	0.12
HER2	1	0.10	6	0.36	2	0.23	3	0.35
Triple negative	0		3	0.18	2	0.23	0	
Not available	4		0		1		2	

IDC, invasive ductal carcinoma; *ILC*, invasive lobular carcinoma; *ITC*, invasive tubular carcinoma. *p*-values calculated with DM after DM as reference

[1] $p < 0.01$ compared with DM after DM, not significant after Bonferroni correction

[2] $p < 0.05$ compared with DM after DM, not significant after Bonferroni correction

The combination of DM after DM/SM+DBT into one study group and DM/SM+DBT after DM into one group creates some heterogeneity within the groups. However, this combination had a minor influence on the results for recall rates, cancer detection rates, and PPV (Appendix 2). This is in accordance with the results from other studies, showing comparable results for DM and SM with regard to early performance measures [5, 12]. Tumors detected with SM+DBT tended to be of lower histologic grade compared with DM+DBT; otherwise, no differences regarding histopathological characteristics were observed between SM+DBT and DM+

DBT (Appendix 3). As SM+DBT is now preferred to DM+DBT in screening because of radiation concerns [5, 17, 22], these differences are not considered of major importance in the interpretation of the overall results.

Our study had some limitations. The total time-span of the study was 8 years (2008–2016) resulting in a high number of radiologists with different experience in screen reading, particularly with DBT (Appendix 1). Further, the study groups were screened at different time periods, which might influence the results, among others, because of the change of mammographic equipment and reading environments. There was also some

heterogeneity regarding the women comprising the groups as some contributed with more than one observation in the study population. However, the correlation coefficients were estimated to be at most 0.06 for recalls and 0.01 for screen-detected and interval breast cancer; thus, the dependency between observations was considered minor. Furthermore, the number of prior screening examinations available at screen reading and consensus may vary. Lastly, the study was performed in Oslo, which represents a small geographical area with a relatively large population—45,000 women in the target group of the screening program. The women represent a heterogeneous population including women who have moved to Oslo from all parts of Norway, as well as about 20% immigrants [23].

To conclude, our study showed that recall remained lower and screen-detected breast cancer higher for consecutive screening examinations with DBT versus standard DM. We observed higher rates of tumors with smaller tumor diameter and more favorable histological grade and molecular subtype, for tumors detected with DBT compared with DM, while no differences in rates of larger and histologic grade 3 tumors were observed. This questions the clinical implications of the increased detection rate. Whether our results remain stable over several more screening rounds is yet to be investigated. Both a long-term increased rate of screen-detected breast cancer, as well as a reduction in the rate of interval breast cancer, is crucial when considering DBT as a sustainable and more effective tool for mammographic screening than standard DM alone.

Compliance with ethical standards

Guarantor The scientific guarantor of this publication is Prof. Solveig Hofvind.

Conflict of interest The authors of this manuscript declare no relationships with any companies whose products or services may be related to the subject matter of the article. SH is the Head of BreastScreen Norway, but has permanent employment as a researcher at the Cancer Registry of Norway, independent of her job as administrative leader of BreastScreen Norway.

Statistics and biometry One of the authors has significant statistical expertise.

Informed consent Written informed consent was waived by the Institutional Review Board.

The Cancer Registry Regulation gives approval with waiver of informed consent to perform surveillance, quality assurance, and studies based on data collected as a part of invitation to and participation in the program.

Ethical approval Institutional Review Board approval was obtained. Data was collected from BreastScreen Norway and thus covered by the Cancer Registry Regulation. Data from the Oslo Tomosynthesis Study Trial, as a part of BreastScreen Norway, was also approved by the regional ethical committee and data from the Oslo-Vestre-Viken-Vestfold study was approved by the institutional Data Protection Officer.

Study subjects or cohorts overlap Some study subjects or cohorts have been included in previously published papers—the Oslo Tomosynthesis Trial and the Oslo-Vestfold-VestreViken study.

Skaane P, Bandos AI, Gullien R et al (2013) Comparison of digital mammography alone and digital mammography plus tomosynthesis in a population-based screening program. Radiology 267:47–56.

Hofvind S, Hovda T, Holen ÅS et al (2018) Digital Breast Tomosynthesis and Synthetic 2D Mammography versus Digital Mammography: Evaluation in a Population-based Screening Program. Radiology. https://doi.org/10.1148/radiol.2018171361:171361

Methodology
• Retrospective
• Observational
• Performed at one institution

References

1. Gilbert FJ, Tucker L, Young KC (2016) Digital breast tomosynthesis (DBT): a review of the evidence for use as a screening tool. Clin Radiol 71:141–150
2. Marinovich ML, Hunter KE, Macaskill P, Houssami N (2018) Breast cancer screening using tomosynthesis or mammography: a meta-analysis of cancer detection and recall. J Natl Cancer Inst 110:942–949
3. Friedewald SM, Rafferty EA, Rose SL et al (2014) Breast cancer screening using tomosynthesis in combination with digital mammography. JAMA 311:2499–2507
4. Skaane P, Bandos AI, Gullien R et al (2013) Comparison of digital mammography alone and digital mammography plus tomosynthesis in a population-based screening program. Radiology 267:47–56
5. Bernardi D, Macaskill P, Pellegrini M et al (2016) Breast cancer screening with tomosynthesis (3D mammography) with acquired or synthetic 2D mammography compared with 2D mammography alone (STORM-2): a population-based prospective study. Lancet Oncol 17:1105–1113
6. Ciatto S, Houssami N, Bernardi D et al (2013) Integration of 3D digital mammography with tomosynthesis for population breast-cancer screening (STORM): a prospective comparison study. Lancet Oncol 14:583–589
7. Zackrisson S, Lång K, Rosso A et al (2018) One-view breast tomosynthesis versus two-view mammography in the Malmo Breast Tomosynthesis Screening Trial (MBTST): a prospective, population-based, diagnostic accuracy study. Lancet Oncol. https://doi.org/10.1016/S1470-2045(18)30521-7
8. Skaane P, Sebuødegård S, Bandos AI et al (2018) Performance of breast cancer screening using digital breast tomosynthesis: results from the prospective population-based Oslo Tomosynthesis Screening Trial. Breast Cancer Res Treat 169:489–496
9. Houssami N, Bernardi D, Caumo F et al (2018) Interval breast cancers in the 'screening with tomosynthesis or standard mammography' (STORM) population-based trial. Breast 38:150–153
10. Morris E, Feig SA, Drexler M, Lehman C (2015) Implications of overdiagnosis: impact on screening mammography practices. Popul Health Manag 18(Suppl 1):S3–S11

11. Gilbert FJ, Tucker L, Gillan MG et al (2015) Accuracy of digital breast tomosynthesis for depicting breast cancer subgroups in a UK retrospective reading study (TOMMY trial). Radiology 277:697–706

12. Skaane P, Bandos AI, Eben EB et al (2014) Two-view digital breast tomosynthesis screening with synthetically reconstructed projection images: comparison with digital breast tomosynthesis with full-field digital mammographic images. Radiology 271:655–663

13. Mariscotti G, Durando M, Houssami N et al (2017) Comparison of synthetic mammography, reconstructed from digital breast tomosynthesis, and digital mammography: evaluation of lesion conspicuity and BI-RADS assessment categories. Breast Cancer Res Treat. https://doi.org/10.1007/s10549-017-4458-3

14. Zuckerman SP, Conant EF, Keller BM et al (2016) Implementation of synthesized two-dimensional mammography in a population-based digital breast tomosynthesis screening program. Radiology 281:730–736

15. McDonald ES, Oustimov A, Weinstein SP, Synnestvedt MB, Schnall M, Conant EF (2016) Effectiveness of digital breast tomosynthesis compared with digital mammography: outcomes analysis from 3 years of breast cancer screening. JAMA Oncol 2: 737–743

16. Aase HS, Holen ÅS, Pedersen K et al (2018) A randomized controlled trial of digital breast tomosynthesis versus digital mammography in population-based screening in Bergen: interim analysis of performance indicators from the To-Be trial. Eur Radiol. https://doi.org/10.1007/s00330-018-5690-x

17. Hofvind S, Hovda T, Holen ÅS et al (2018) Digital breast tomosynthesis and synthetic 2D mammography versus digital mammography: evaluation in a population-based screening program. Radiology. https://doi.org/10.1148/radiol.2018171361:171361

18. Hofvind S, Tsuruda K, Mangerud G et al (2017) The Norwegian Breast Cancer Screening Program, 1996-2016: Celebrating 20 years of organised mammographic screening. In: Cancer in Norway 2016 - Cancer incidence, mortality, survival and prevalence in Norway. Oslo: Cancer Registry of Norway

19. Ministry of Health and Care Services (2001) Forskrift om innsamling og behandling av helseopplysninger i Kreftregisteret (The Cancer Registry Regulation) Available via https://lovdata.no/dokument/SF/forskrift/2001-12-21-1477. Accessed 10 April 2019

20. Hofvind S, Sagstad S, Sebuødegård S, Chen Y, Roman M, Lee CI (2018) Interval breast cancer rates and histopathologic tumor characteristics after false-positive findings at mammography in a population-based screening program. Radiology 287:58–67

21. Pattacini P, Nitrosi A, Giorgi Rossi P et al (2018) Digital mammography versus digital mammography plus tomosynthesis for breast cancer screening: the Reggio Emilia Tomosynthesis Randomized Trial. Radiology 288:375–385

22. Bernardi D, Li T, Pellegrini M et al (2018) Effect of integrating digital breast tomosynthesis (3D-mammography) with acquired or synthetic 2D-mammography on radiologists' true-positive and false-positive detection in a population screening trial: a descriptive study. Eur J Radiol 106:26–31

23. Bhargava S, Tsuruda K, Moen K, Bukholm I, Hofvind S (2018) Lower attendance rates in immigrant versus non-immigrant women in the Norwegian Breast Cancer Screening Programme. J Med Screen 25:155–161