European Radiology (2019) 29:3802–3811
https://doi.org/10.1007/s00330-018-5978-x

BREAST

Comparing two visualization protocols for tomosynthesis in screening: specificity and sensitivity of slabs versus planes plus slabs

Valentina Iotti[1] · Paolo Giorgi Rossi[2] · Andrea Nitrosi[3] · Sara Ravaioli[1] · Rita Vacondio[1] · Cinzia Campari[4] · Vanessa Marchesi[1] · Moira Ragazzi[5] · Marco Bertolini[3] · Giulia Besutti[1,6] · Carlo Alberto Mori[1] · Pierpaolo Pattacini[1] · the RETomo Working Group

Received: 19 April 2018 / Revised: 30 October 2018 / Accepted: 17 December 2018 / Published online: 8 February 2019

Abstract

Objectives Tomosynthesis (DBT) has proven to be more sensitive than digital mammography, but it requires longer reading time. We retrospectively compared accuracy and reading times of a simplified protocol with 1-cm-thick slabs versus a standard protocol of slabs + 1-mm-spaced planes, both integrated with synthetic 2D.

Methods We randomly selected 894 DBTs (including 12 cancers) from the experimental arm of the RETomo trial. DBTs were read by two radiologists to estimate specificity. A second set of 24 cancers (8 also present in the first set) mixed within 276 negative DBTs was read by two radiologists. In total, 28 cancers with 64 readings were used to estimate sensitivity. Radiologists read with both protocols separated by a 3-month washout. Only women that were positive at the screening reading were assessed. Variance was estimated taking into account repeated measures.

Results Sensitivity was 82.8% (53/64, 95% confidence interval (95%CI) 67.2–92.2) and 90.6% (95%CI 80.2–95.8) with simplified and standard protocols, respectively. In the random screening setting, specificity was 97.9% (1727/1764, 95%CI 97.1–98.5) and 96.3% (95%CI 95.3–97.1), respectively. Inter-reader agreement was 0.68 and 0.54 with simplified and standard protocols, respectively. Median reading times with simplified protocol were 20% to 30% shorter than with standard protocol.

Conclusions A simplified protocol reduced reading time and false positives but may have a negative impact on sensitivity.

Key Points

- *The adoption of digital breast tomosynthesis (DBT) in screening, more sensitive than mammography, could be limited by its potential effect on the radiologists' workload, i.e., increased reading time and fatigue.*
- *A DBT simplified protocol with slab only, compared to a standard protocol (slab plus planes) both integrated with synthetic 2D, reduced time and false positives but had a negative impact on sensitivity.*

Keywords Breast neoplasms · Mass screening · Mammography · Workflow · Sensitivity and specificity

✉ Valentina Iotti
valentina.iotti@ausl.re.it

[1] Radiology Unit, AUSL Reggio Emilia, IRCCS, Reggio Emilia, Italy

[2] Epidemiology Unit, AUSL Reggio Emilia, IRCCS, Reggio Emilia, Italy

[3] Medical Physics Unit, AUSL Reggio Emilia, IRCCS, Reggio Emilia, Italy

[4] Screening Coordinating Unit, AUSL Reggio Emilia, IRCCS, Reggio Emilia, Italy

[5] Pathology Unit, AUSL Reggio Emilia, IRCCS, Reggio Emilia, Italy

[6] Clinical and Experimental Medicine PhD Program, University of Modena and Reggio Emilia, Modena, Italy

Abbreviations

ASMN	Arcispedale Santa Maria Nuova
AUSL	Azienda Unità Sanitaria Locale
BIRADS	Breast imaging-reporting and data system
CAD	Computer-aided detection
CAM	Initials of reader 1
CC	Craniocaudal
CI	Confidence interval
DBT	Digital breast tomosynthesis
DCIS	Ductal carcinoma in situ
DM	Digital mammography
FDA	Food and Drug Administration
IQR	Interquartile ranges

IRCCS	Istituto di Ricovero e Cura a Carattere Scientifico (Research Hospital)
MLO	Mediolateral oblique
PACS	Picture archiving and communication system
RV	Initials of reader 2
SR	Initials of reader 3
Sy-2D	Synthetic 2D

Introduction

In western countries, breast cancer screening is recommended by most scientific societies and governmental agencies as an evidence-based intervention to reduce mortality [1–5]. Digital mammography (DM) is now the standard for screening; it is strongly recommended in women aged 50–69 and conditionally recommended for women aged 45–49 [1]. Nevertheless, mammography sensitivity is not optimal, which can limit screening efficacy [6], particularly in dense breasts [7].

Digital breast tomosynthesis (DBT) is a new imaging technology that provides a 3D reconstruction of the breast from a limited angle scan involving a series of low-dose mammographic exposures [8].

In trials comparing the two technologies on the same women, i.e., double-testing studies measuring accuracy, DBT has proven to be more sensitive than DM alone in detecting cancers [9–13]. On the contrary, relative specificity of DBT versus DM varies in different studies, mostly because DM specificity varies across countries and programs [11–14]. Overall, this makes DBT a very promising technology for screening. The FDA has approved DBT as an adjunct to DM or synthetic 2D (Sy-2D) [15] for screening, and this combination has now been adopted in some screening facilities in the USA [16]. The addition of DBT to DM means doubling the radiation dose and, with many systems, also increases the discomfort during imaging acquisition. For this reason, Sy-2D images elaborated from the DBT acquisition have been developed. DBT + Sy-2D has shown better sensitivity than DM in most studies [13, 14, 17], while its non-inferiority compared to that of DBT + DM has not yet been clearly demonstrated. Regarding specificity, DBT + Sy-2D has shown worse performance than DBT + DM [14]. Thus synthetic 2D is probably the easiest way now available to reduce harms in the introduction of tomosynthesis in screening [18].

The new European guidelines suggest screening with either DBT with Sy-2D or DM alone, in the context of an organized screening program (conditional recommendation); the recommendation is supported by evidence of very low certainty; although data on cross-sectional performance are very promising, there is no evidence on how this gain in sensitivity can translate into better prognosis [1]. One of the main concerns about the adoption of DBT in screening is its potential effect on the radiologists' workload: depending on breast thickness, each single DBT view consists of an average of 60 sections (1-mm-spaced planes), increasing interpretation time, and possibly errors due to reader fatigue [19].

Existing studies show that reading times for DBT are 50 to 100% longer than those for DM [10, 20–22], with a net increase in screening readings of about 25–30 s per case. In our analyses of the RETomo [20] reading times, differences almost disappear in cases sent to assessment. Our interpretation is that the longer the reading times, the smaller the difference, suggesting that DBT's longer reading time is mostly due to the multiple images themselves and not to the interpretation of findings [20].

Therefore, reducing reading times for DBT + Sy-2D is critical to introducing this new technology in screening programs, particularly in those countries with a National Health Service providing screening for free and aiming to reach high coverage levels in the whole population.

Aim

Here, we present the preliminary results of a study nested in the RETomo trial [20] comparing accuracy and reading times of a simplified protocol with slab only versus a standard protocol of slab plus planes, both integrated with Sy-2D. We also compare the simplified (DBT slabs + Sy-2D) and standard protocols (DBT slabs + planes + Sy-2D) with the DBT + DM (DBT slabs + planes + standard digital mammography) readings; the latter is the protocol currently investigated in the RETomo trial [20] and re-used in the same setting in this retrospective reader study.

Materials and methods

Setting

The province of Reggio Emilia has a population of around 550,000. The population-based breast cancer screening program has been active since 1994. The target population includes women aged 45–74 years old. The screening interval is 1 year for women aged 45–49 and 2 years for women aged 50–74. Independent double reading is used for all mammograms, with arbitration in the event of disagreement [23].

Study design

This is an accuracy and concordance retrospective study nested in a large trial comparing DBT + DM versus DM. The trial protocol is described elsewhere [20] and is briefly reported in the box (Fig. 1).

Two sets of DBTs were drawn from women participating in the experimental arm of the RETomo trial in the period from January 2015 to December 2016. All images were acquired

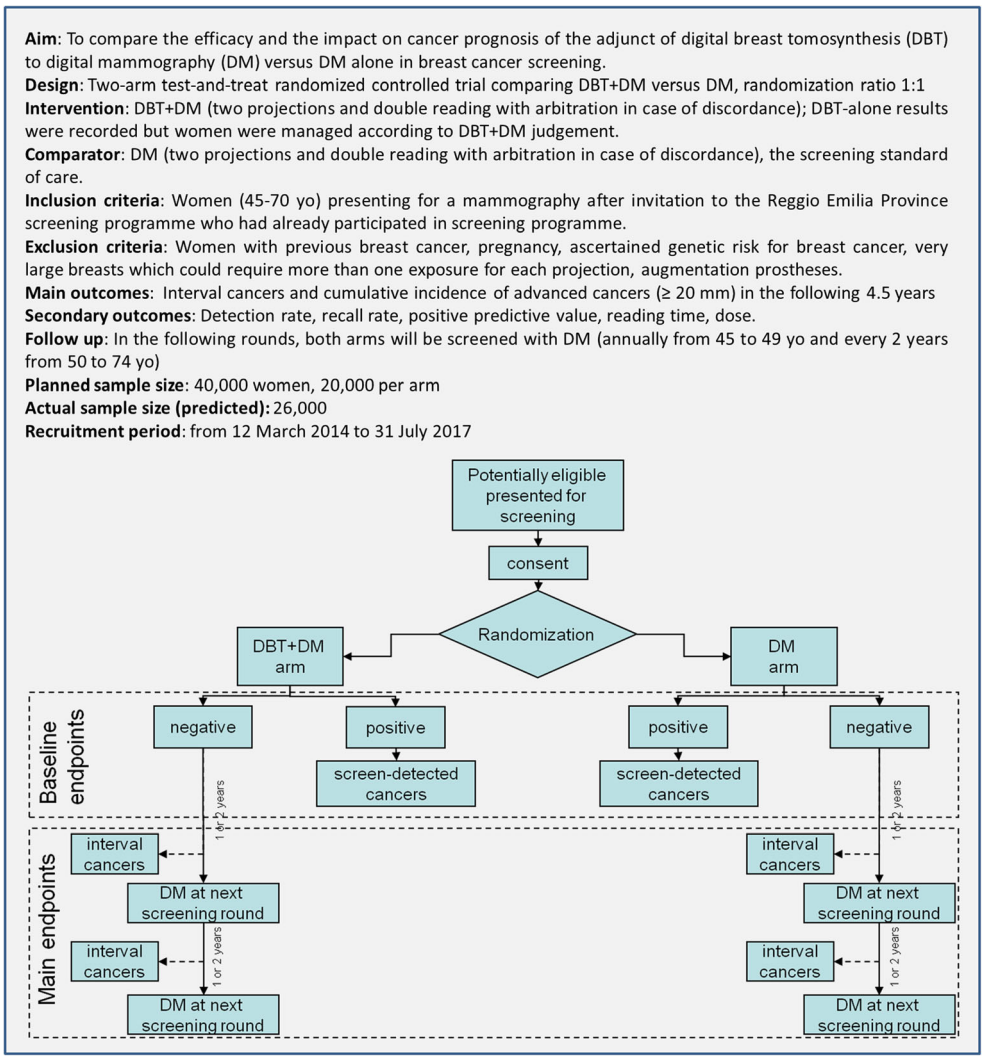

Fig. 1 RETomo study design—trial registration: NCT02698202

with Senographe Essential digital mammography systems (GE Healthcare), equipped with tomosynthesis. DBT image sets were read according to three different protocols by two independent readers after a 3-month washout period between readings; during this period, the readers were also involved in routine screening readings, thus favoring washout. In this preliminary analysis, sensitivity and specificity were calculated considering as true positives only cancers detected at baseline screening (i.e., detected within 9 months after a positive screening examination so as to include BIRADS-3 referred to early recall). The final reference will be the result of assessment for all the women that resulted positive at screening and the follow-up for all women resulting negative at screening; thus, true-positive cases will be all screen-detected cancers, all interval cancers, and all cancers screen-detected at the following round.

The first set (specificity set) was defined to measure the specificity of the Sy-2D + slabs (simplified protocol), Sy-2D + slabs + planes (standard protocol), and DBT + DM. It included 894 DBT examinations randomly selected from the RETomo trial participants. Of these, 32 were referred for an assessment and 12 had a cancer. Thus, 882 non-cancer cases with two readers (CAM and RV) each were available for specificity analysis.

The second set (sensitivity set) was defined to estimate the sensitivity: 276 negatives randomly drawn from the screening population, 20 cancers not included in the first sample, and 8 cancers included in the first set. Finally, 28 malignant lesions with 64 readings were available for sensitivity estimates (including 2 ductal carcinomas in situ [DCIS]). These readings came from 8 cancers with three readings, i.e., the ones present in the both sets were read by CAM, RV, and SR (RV actually read these cases twice, but only the first reading was considered), and 20 with two readings: 4 from the first set were read by CAM and RV and 16 from the second set were read by RV and SR.

Results of this retrospective study were only used for research purposes and women were not informed, nor managed consequently.

Images were classified according to a screening modified breast imaging-reporting and data system (BIRADS) classification: BIRADS 1 and 2 are grouped together as "negative,"

and "positive" readings are classified as BIRADS 4 or 5; BIRADS 3 was not used in the first level test since shortened follow-up interval can only be decided after the assessment.

Sample size and study power

For the sensitivity, it was estimated that about 48 true cancers were needed to be read with each protocol (i.e., invasive cancers and DCIS) to obtain an 80% power for a non-inferiority test excluding a relative sensitivity lower than 0.9, with alpha = 0.05 [24]. The non-inferiority threshold has been defined arbitrarily, in analogy to the requirements defined for other screening tests [25]. We included all the 28 available cancers. Considering that we had at least two readings for each cancer, we had at least 56 readings. This number should partially counter balance the non-independence of observations: with a design effect of about 1.2, the power would be 80%; if the design effect was 1.5, the power would be 67%, and in the hypothesis of perfect correlation, i.e., design effect 2 equal to cluster size, the power would be 55% [25].

For the specificity, it was estimated that about 800 true-negative cases were needed with each protocol to obtain 80% power for a non-inferiority test excluding a relative specificity lower than 0.98. We conservatively sampled about 894 cases, including an initially unknown number of cancers.

Reading protocols

The sets were read according to three protocols with the following workflows (Fig. 2):

A) DBT + DM protocol: the radiologist read the DBT image set in sequence, scrolling continuously from slabs (1-cm-thick, 0.5 cm overlap) to 1-mm-spaced planes; all the images are always visualized. The workstation then presented DM in a standardized hanging protocol; the radiologist could immediately visualize any DBT projection simultaneously on the screen.
B) Standard DBT protocol (Sy-2D + slabs + planes): the radiologist read DBT image sets in sequence, scrolling continuously from the Sy-2D (GE V-Preview), slabs (1-cm-thick, 0.5 cm overlap) to planes (1-mm-spaced); all the images are always visualized; the radiologist could immediately visualize any DBT projection simultaneously on the screen.
C) Simplified DBT protocol (Sy-2D + slabs): the radiologist read the DBT image sets only with the Sy-2D (GE V-Preview) and slabs (1-cm-thick, 0.5 cm overlap); the radiologist could immediately visualize any DBT projection simultaneously on the screen.

For all protocols, previous mammograms were available for comparison.

All three readers (CAM, RV, and SR) were specialized breast radiologists, with 3 years of experience with DBT in screening and clinical setting. Readers were blinded to the cancer frequency in the set, i.e., whether they were reading the set for specificity or for sensitivity assessment.

Reading time was automatically measured by the radiology information system as the difference between opening and closing times for each examination. Reading times longer than 5 min were excluded as these were extreme outliers, usually due to external reasons (e.g., answering telephone, lack of connection).

Randomization

To minimize the bias due to incomplete washout after first and second reading, we randomized the protocol sequence. We prepared 6 subsets of about 150 cases each from the specificity set and 3 subsets of about 100 cases from the sensitivity set. Then, we randomized the subsets to have balanced combinations of protocols: A-B-C, B-C-A, C-A-B.

Statistical analyses

Sensitivity was evaluated on 28 lesions and 64 independent readings: 8 cancers had three readings and 20 had two readings. The non-inferiority tests comparing sensitivity and specificity of the simplified protocol versus the standard protocol were performed with McNemar's test corrected for repeated measures [26].

Specificity was evaluated on 882 DBTs of women who were negative at screening.

Sensitivity and specificity 95%CI were computed taking into account the intra-cluster correlation within different readings of the same case using the survey estimate module for proportions in STATA 13.0 in which the standard errors are estimated with Taylor's linearization method [27].

We present Cohen's kappa values [28] with relative 95%CI for the concordance between the different protocols using pooled data from the two sets. We report kappa values for cancers, screening negative women, and for the total. We also report the kappa values, and relative 95%CI, for inter-rater agreement with each reading protocol using pooled data from the two sets; for the 8 cancers with three readings, only the first two readings were considered.

Mean, median, and interquartile ranges (IQR) are reported for reading time; *t* test was used to compare means. All analyses were performed with STATA 13.0 [29].

Funding and ethical approval

The RETomo trial was approved by the provincial Ethics Committee (11/11/2013, ASMN 2013/0029304) and has been registered at clinicaltrials.gov: NCT02698202. It has been partially funded by the Regione Emilia Romagna Public Health System and sustained by the institutional funds of the

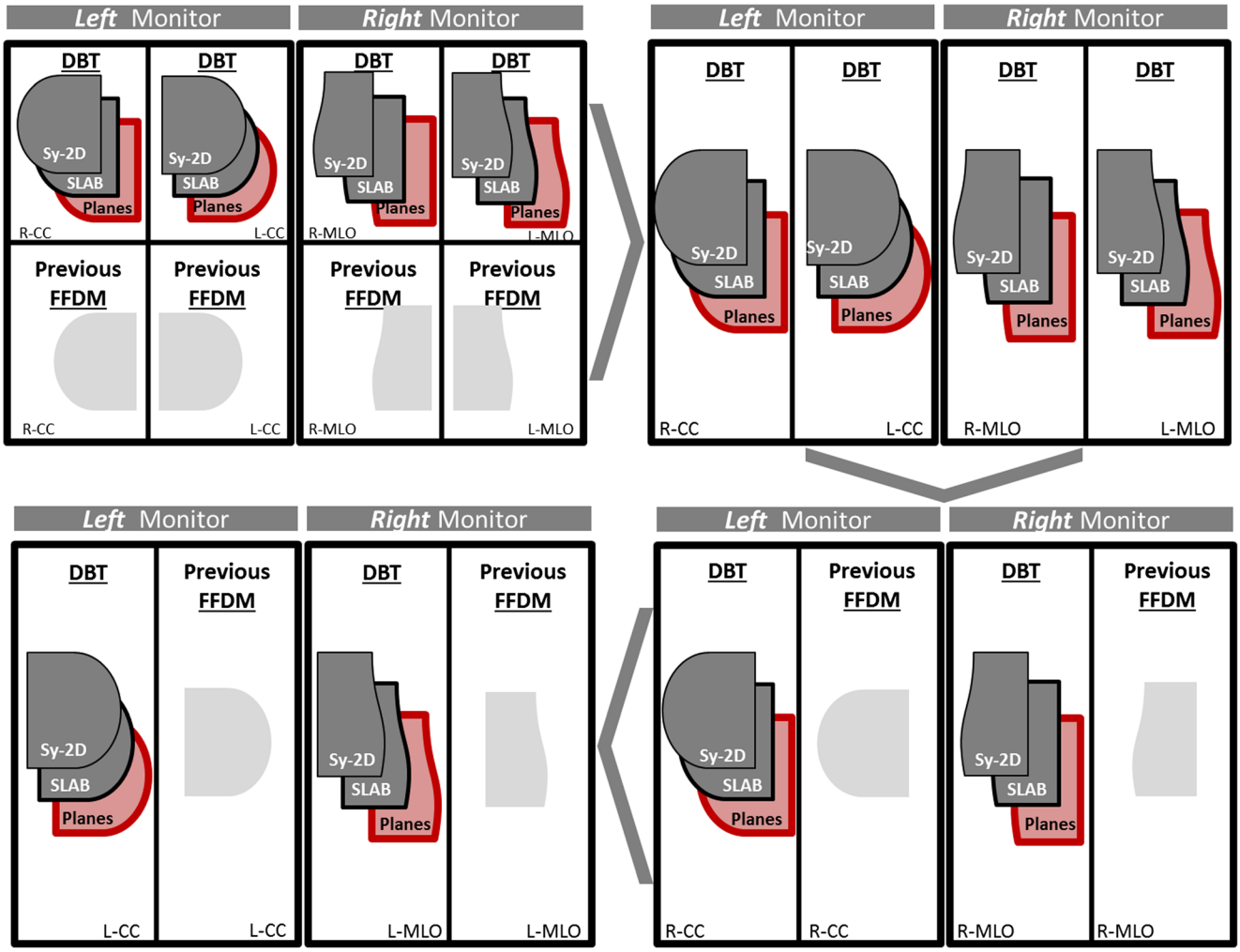

Fig. 2 Hanging protocols adopting Sy-2D. The standard protocol included Sy-2D, slabs, and planes (in red), displayed in sequence by scrolling the mouse; clicking (gray arrowhead) switched to the four different main screens. The simplified protocol included only Sy-2D and slabs

Reggio Emilia Local Health Authority. GE Healthcare loaned one tomosynthesis option for our study period, provided the software for simplified protocol, and partly funded part of the additional readings made for this substudy; GE's support of our study was unconditional and was approved after the definition of the protocol. The Reggio Emilia Local Health Authority is the data owner and has the full responsibility for publishing the data. All participants signed a written informed consent.

Results

From the specificity set, including 882 non-cancer cases with two readings each, specificity was 97.9% (1727/1764, 95%CI 97.1–98.5) and 96.3% (95%CI 95.3–97.1) with the simplified (Sy-2D + slabs) and standard (Sy-2D + slabs + planes) protocols, respectively; DBT + DM specificity was 96.4% (1700/1764 95%CI 95.4–97.2) (Table 1). The results of this study for specificity, with the non-inferiority threshold set as having 95% probability of less than 2% lower specificity, show non-inferiority of the simplified protocol on specificity compared to standard protocol ($p < 0.0001$) and to DBT + DM ($p < 0.0001$).

Considering the two sets, we had 28 cancers, with 64 readings made by different readers, i.e., 20 cancers with two readings and 8 with three readings. Sensitivity was 82.8% (53/64, 95%CI 66.2–92.2) and 90.6% (58/64, 95%CI 80.2–95.8) with simplified and standard protocols, respectively. DBT + DM sensitivity was 92.2% (59/64 95%CI 82.3–96.8) (Table 1). The probability that the sensitivity of the simplified protocol is lower the non-inferiority threshold of less 10% is 0.33 when we compare to the standard protocol and 0.59 when we compare to DBT + DM ($p = 0.59$).

With the simplified protocol, 2 cases were missed by both readers, despite the fact that the images do not show any appreciable difference of the radiological signs in slabs and planes (Fig. 3).

Table 1 Sensitivity and specificity of Simplified (Sy-2D + slabs), Standard (Sy-2D + slab + planes) and DBT + DM protocol

	True lesions (invasive cancers and DCIS)				No screen-detected lesion			
Readings	Positive	Negative	Sensitivity	95%CI	Positive	Negative	Specificity	95%CI
Simplified protocol (Sy-2D + slabs)	53	11	82.8	66.2–92.2	37	1727	97.9	97.1–98.5
Standard protocol (Sy-2D + slabs + planes)	58	6	90.6	80.2–95.8	65	1699	96.3	95.3–97.1
DBT + DM protocol	59	5	92.2	82.3–96.8	64	1700	96.4	95.4–97.2

With the standard protocol and DBT + DM, every malignant lesion was identified by at least one reader (Table 2).

Cohen's kappa value for the intra-reader agreement between the simplified and standard protocols was 0.66 (95%CI 0.59–0.73), while the agreement of simplified with DBT + DM was 0.61 (95%CI 0.53–0.68) and the agreement of standard with DBT + DM was 0.70 (95%CI 0.64–0.76) (Table 3).

Cohen's kappa values for the inter-reader agreement was 0.68 (95%CI 0.58–0.79) and 0.54 (95%CI 0.44–64) with simplified and standard protocols, respectively; for DBT + DM kappa, it was 0.52 (95%CI 0.42–0.62) (Table 4).

Median reading times with simplified protocol ranged between 35 and 69 s among readers; for the standard protocol, the range was from 60 to 97 s and for DBT + DM, from 62 to 109 s. Looking at the 75th percentile, differences between simplified and standard protocols decreased, ranging from 65 to 90s for the simplified protocol and from 68 to 109 s for the standard protocol; the DBT + DM showed longer reading times, ranging from 89 to 127 s (Table 5). The mean was lower for readers 1 and 2 in both sets (< 0.00005), but similar for readers 2 and 3 in the sensitivity set ($p = 0.59$).

Discussion

Main findings

In this study, we measured the accuracy of a new simplified protocol for reading DBTs in a screening setting. The simplified protocol was effective in reducing reading times, particularly for the shortest time. It at least maintained the same specificity than the standard protocol, if not improving it. Furthermore, the simplified DBT protocol showed the highest reproducibility compared to the standard DBT protocol and DBT + DM, even if all are within the values reported in previous studies (i.e., from 0.45 to 0.7) [30–32]. On the other hand, the simplified protocol showed lower sensitivity, and the non-inferiority test clearly indicates that there is a high probability that the sensitivity is more than 10% inferior to that with the standard protocol.

Limits

For logistical reasons, the workstation was separated from the daily screening practice. The connection with picture

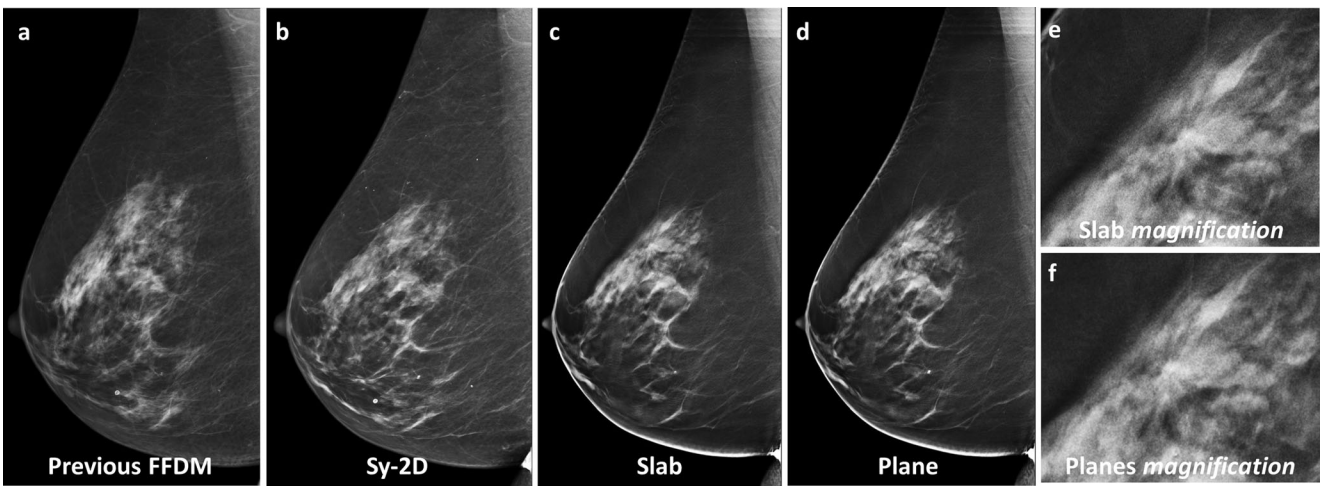

Fig. 3 Cancer missed by both readers. M.R., 68 years old, parenchymal distortion in the upper quadrants of the right breast. The patient was thus recalled. **a** Previous (2 years before) full-field digital mammography, negative. **b** The distortion is difficult to detect on the Sy-2D but equally visible both in **c** slabs and **d** planes [**e** slab magnification; **f** plane magnification]. However, reading only Sy-2D and slabs, the few images could limit its perception. Histology: diagnosis of IDC G2 (6 mm)

Table 2 Inter-rater agreement for tomosynthesis of screen-detected cancers only

	Simplified protocol	Standard protocol	DBT + DM
Cases with 2 readings (20)			
Both readers positive	18	18	16
1 pos/1 neg	0	2	4
Both neg	2	0	0
Cases with 3 readings (8)			
3 readers positive	4	5	7
2 pos/1 neg	2	2	1
1 pos/2 neg	2	1	0
3	0	0	0

archiving and communication system (PACS) was not optimized, and times to download and visualize the images were thus longer than for routine screening activity. The DBT + DM median reading times in this study ranged from 47 to 91 s, while in the trial, median reading times ranged from 35 to 81 s. Nevertheless, our results give an indication of at least which protocol is less time consuming.

There may be a learning curve, and certainly, the newer procedure (the simplified DBT protocol) was at a disadvantage. However, it must be remembered that DBT is quite new for all radiologists. Although all the authors involved in this study had experience with DBT + DM thanks to the RETomo trial, Sy-2D was completely new for everyone.

As in all retrospective studies, readers did not express their judgment under the pressure of real practice as no woman was in fact affected by this reading.

As the power of the study for sensitivity is limited, the result of the non-inferiority test should be considered with caution.

The reference we used in this paper to assess true cases is not the one planned in the final analyses. For the moment, we only have results from the baseline assessment for women who were positive at screening. Any cancers in women who were negative will be detected only in the next 2 years. This ascertainment bias, at least for sensitivity, favors the DBT + DM protocol. In fact, in the RETomo trial, the actual screening procedure was DBT + DM. Therefore, only women that were positive to DBT + DM in the screening readings were assessed. Any lesion rated malignant with Sy-2D, with the simplified or standard protocol, which was not visible at DBT + DM, was considered as false positive at this point. Only the follow-up will permit correctly classifying these as true positive.

Finally, given the high variability observed between readers, including more readers in future studies on this topic may be opportune.

Table 3 Agreement between reading protocols, for DBT of screen-detected cases, no lesions found, and total

	Simplified protocol (Sy-2D + slabs)				Simplified protocol (Sy2D + slabs)				Standard protocol (Sy-2D + slabs + planes)			
True lesions (invasive cancers + DCIS)	Standard protocol (Sy-2D + slabs + planes)	Positive	Negative	Total	DBT + DM	Positive	Negative	Total	DBT + DM	Positive	Negative	Total
	Positive	51	6	58	Positive	51	8	59	Positive	56	3	59
	Negative	1	5	6	Negative	2	3	5	Negative	2	3	5
	Total	53	11	64	Total	53	11	64	Total	58	6	64
	Kappa value 0.531 95%CI 0.234–0.829				Kappa value 0.300 95%CI 0.013–0.613				Kappa value 0.503 95%CI 0.126–0.880			
No screen detected lesions	Standard protocol (Sy-2D + slabs + planes)	Positive	Negative	Total	DBT + DM	Positive	Negative	Total	DBT + DM	Positive	Negative	Total
	Positive	37	54	91	Positive	38	61	93	Positive	52	41	93
	Negative	23	2200	2223	Negative	28	2193	2221	Negative	39	2182	2221
	Total	60	2254	2314	Total	60	2254	2314	Total	91	2223	2314
	Kappa value 0.474 (0.374–0.573)				Kappa value 0.399 (0.299–0.499)				Kappa value 0.547 (0.459–0.636)			
Total	Standard protocol (Sy-2D + slabs + planes)	Positive	Negative	Total	DBT + DM	Positive	Negative	Total	DBT + DM	Positive	Negative	Total
	Positive	89	60	149	Positive	83	69	152	Positive	108	44	152
	Negative	24	2205	2229	Negative	30	2196	2226	Negative	41	2185	2226
	Total	113	2265	2378	Total	113	2265	2378	Total	149	2229	2378
	Kappa value 0.661 (0.593–0.729)				Kappa value 0.605 (0.534–0.676)				Kappa value 0.699 (0.638–0.759)			

Table 4 Inter-rater agreement with the three reading protocols

	Simplified protocol (Sy-2D + slabs)			Simplified protocol (Sy-2D + slabs + planes)				DBT + DM			
	Reader 1				Reader 1				Reader 1			
Reader 2	Positive	Negative		Reader 2	Positive	Negative		Reader 2	Positive	Negative	
Positive	38	26	64	Positive	41	57	98	Positive	40	56	96
Negative	7	1115	1122	Negative	6	1082	1088	Negative	11	1079	1090
	45	1141	1186		47	1139	1186		51	1135	1186
Kappa 0.683 (0.581–0.785)				Kappa 0.541 (0.442–0.640)				Kappa 0.517 (0.417–0.617)			

Interpretation and comparison with previous studies

To reduce DBT reading times, two main approaches have been studied in the literature: the use of a computer-aided detection (CAD) system and reducing the number of DBT images to evaluate and simplify visualization protocols.

A CAD system for DBT has recently been developed mainly for the detection of masses [33, 34]. A retrospective study by Morra et al [19] showed that the DBT-CAD system can detect a large percentage (89%, 99 of 111) of breast cancers manifesting as masses and microcalcification clusters, with an acceptable false-positive rate (2.7 per breast view). Balleyguier et al [35] reported a similar study design, but analyzed only bilateral craniocaudal (CC) DM and mediolateral oblique (MLO) DBT images for each patient and concluded non-inferiority of radiologist interpretation performance using concurrent CAD, with a reduction in reading time of 23.5% (from an average of 48.2 s without CAD to 39.1 s with CAD). Benedikt et al [36] achieved similar results in their retrospective multi-reader multi-case crossover design; analyzing bilateral CC DM and MLO DBT images for each patient, they determined that concurrent use of CAD with DBT resulted in 29.2% faster reading time, while maintaining reader interpretation performance.

With the second approach, i.e., reducing the number of DBT images to evaluate, Dustler et al [37] found that increasing slab thickness, using 2-mm slabs instead of 1-mm planes, can reduce reading time of normal cases by 20%.

Our results are in line with this second approach: we adopted thicker slabs (10 mm) and obtained a 20–30% reduction of the median times. On the other hand, if we look at reading times over the 75th percentile, the advantage of the simplified protocol was much smaller than that for median time, suggesting that more difficult cases need longer times, regardless of the protocol. Further studies might be necessary to define the optimal balance between the thickness of slabs, with the consequent number of images and reading times, and the sensitivity for the simplified protocol.

Surprisingly, radiological signs of the two cancers completely missed with the simplified protocol were almost identical in slabs and planes (Fig. 3). Both readers were astounded that they could have missed such clearly visible lesions. There might be different possible explanations for the lower sensitivity of slabs, for example, the perception of the same abnormalities is different when presented in several planes instead of few slabs. Also, presenting a lot of similar images, as in a series

Table 5 Median, interquartile range (IQR), and mean reading time per reader and DBT set

		Simplified protocol (Sy-2D + slabs)	Simplified protocol (Sy-2D + slabs +planes)	DBT+DM
Specificity set		s	s	s
Reader 1	Median (IQR)	64 (48–82)	88 (72–109)	100 (78–127)
	Mean	67	97	109
Reader 2	Median (IQR)	61 (46–83)	78 (60–100)	77 (60–99)
	Mean	68	89	85
Sensitivity set		s	s	s
Reader 2	Median (IQR)	69 (54–90)	86 (70–108)	87 (71–113)
	Mean	76	97	97
Reader 3	Median (IQR)	35 (24–65)	42 (30–68)	45 (27–89)
	Mean	58	60	62

of 1-mm-spaced planes, could reduce the probability of a careless mistake. Further, DBT sets with slabs only—the newest protocol—could require a learning period. Indeed, the radiologist (RV) who read both the sets had 5 false negatives in the specificity set whereas none read after the specificity set.

Friedewald [38] described how scrolling through the DBT images could be challenging because of the tendency for distraction as different portions of the breast come in and out of focus, suggesting reviewing one portion of the image while scrolling and then to repeat moving the focus. This approach, which we can define "cine-mode," is more suitable for DBT reading rather than reviewing the entire image and then moving on to the next slice, which would be an adjustment of the standard 2D reading. However, limiting the visualization to slabs alone could make the cine-mode suboptimal.

The sustainability of introducing DBT into routine screening programs depends on by how much reading times can be reduced. Most of the European population–based screening programs are struggling with resource restrictions due to economic constraints and the reduction of the public health budget. The introduction of a new technology with higher resource consumption risks making the programs unsustainable. We need a careful technology assessment that measures the balance between the possible benefits of higher specificity and sensitivity and the increase in reading time and technology costs.

To conclude, this study described how a simplified protocol, without the presentation of planes, maintained similar DBT specificity, increased reproducibility, and decreased reading times but may have a negative impact on sensitivity.

Acknowledgments Many thanks to all the personnel for their committed work during data collection and to all women who participated in the study for their fundamental contribution. We want to sincerely thank Carlo Alberto Mori, MD, for his precious support, devoted work for this study, and inestimable experience in breast diagnosis. The following are also members of the RETomo working group. Screening readers and post-recall assessment: Coriani C, MD; Pescarolo M, MD; Stefanelli G, MD; Tondelli G, MD; Beretti F, MD; Caffarri S, MD. Screening Coordinating Center: Paterlini L, MD. Radiographers Coordinator: Canovi L, Colli M, Boschini M. Scientific direction: Cavuto S; Braglia L. We thank Jacqueline Costa for editing the text. The results of this study were presented orally at a scientific session at the RSNA Annual Meeting in Chicago 2017 and at the ECR in Vienna 2018.

Funding It has been partially funded by the Regione Emilia-Romagna (Public Health System) and sustained by the institutional funds of the Reggio Emilia Local Health Authority (AUSL)-IRCCS.

Compliance with ethical standards

Guarantor The scientific guarantor of this publication is Paolo Giorgi Rossi, PhD.

Conflict of interest VI, AN, RV, and PP have received speakers' fees and travel grants from GE Healthcare. CAM received financial support from GE Healthcare to allow the conclusion of the study after his retiring. PGR, SR, CC, VM, MR, and MB disclosed no relevant relationships.

Statistics and biometry One of the authors (Paolo Giorgi Rossi, PhD) has significant statistical expertise.

Informed consent Written informed consent was obtained from all subjects (patients) in this study.

Ethical approval Institutional Review Board approval was obtained.

Methodology
- Retrospective
- Diagnostic or prognostic study
- Performed at one institution

References

1. European Commission Initiative on Breast Cancer. Recommendations on breast cancer screening. Available via http://ecibc.jrc.ec.europa.eu/recommendations/list/3 Access 24 May 2017
2. Oeffinger KC, Fontham ETH, Etzioni R et al (2015) Breast cancer screening for women at average risk: 2015 guideline update from the American Cancer Society. JAMA 314(15):1599–1614
3. Siu AL, U.S. Preventive Services Task Force (2016) Screening for breast cancer: U.S. Preventive Services Task Force recommendation statement. Ann Intern Med 164:279–296
4. Perry N, Broeders M, de Wolf C et al (2008) European guidelines for quality assurance in breast cancer screening and diagnosis. Fourth edition–summary document. Ann Oncol 19(4):614–622
5. Council of the European Union (2003) Council recommendation of 2 December 2003 on cancer screening (2003/878/EC). OJ L 327, Office for Official Publications of the European Communities; 34–38
6. Dibden A, Offman J, Parmar D et al (2014) Reduction in interval cancer rates following the introduction of two-view mammography in the UK breast screening programme. Br J Cancer 110:560–564
7. Mandelson MT, Oestreicher N, Porter PL et al (2000) Breast density as a predictor of mammographic detection: comparison of interval- and screen-detected cancers. J Natl Cancer Inst 92(13): 1081–1087
8. Vedantham S, Karellas A, Vijayaraghavan GR et al (2015) Digital breast tomosynthesis: state of the art. Radiology 277:663–684
9. 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
10. 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(1):47–56
11. Skaane P, Bandos AI, Gullien R et al (2013) Prospective trial comparing full-field digital mammography (FFDM) vs. combined FFDM and tomosynthesis in a population-based screening programme using independent double reading with arbitration. Eur Radiol 23:2061e71
12. Lång K, Andersson I, Rosso A et al (2016) Performance of one-view breast tomosynthesis as a stand-alone breast cancer screening modality: results from the Malmo Breast Tomosynthesis Screening Trial, a population-based study. Eur Radiol 26:184–190

13. 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(3): 697–706. https://doi.org/10.1148/radiol.2015142566
14. Bernardi D, Macaskill P, Pellegrini M et al (2016) Breast cancer screening with tomosynthesis (3D mammography) with acquired or synthetic 2D mammography compared with 2Dmammography alone (STORM-2): a population-based prospective study. Lancet Oncol 17(8):1105–1113
15. US Food and Drug Administration. Premarket approval. SENOCLAIRE. Available via http://www.accessdata.fda.gov/scripts/cdrh/cfdocs/cfPMA/pma.cfm?id=P130020S001 Access 24 May 2017
16. McDonald ES, Oustimov A, Weinstein SP et al (2016) Effectiveness of digital breast tomosynthesis compared with digital mammography: outcomes analysis from 3 years of breast cancer screening. JAMA Oncol 2:737–743
17. Caumo F, Zorzi M, Brunelli S et al (2017) Digital breast tomosynthesis with synthesized two-dimensional images versus full-field digital mammography for population screening: outcomes from the Verona screening program. Radiology 170745. https://doi.org/10.1148/radiol.2017170745
18. Houssami N (2018) Evidence on synthesized two-dimensional mammography versus digital mammography when using tomosynthesis (three-dimensional mammography) for population breast cancer screening. Clin Breast Cancer 18(4):255–260.e1. https://doi.org/10.1016/j.clbc.2017.09.012
19. Morra L, Sacchetto D, Durando M et al (2015) Breast cancer: computer-aided detection with digital breast tomosynthesis. Radiology 277(1):56–63
20. 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. Article in press
21. Bernardi D, Ciatto S, Pellegrini M et al (2012) Application of breast tomosynthesis in screening: incremental effect on mammography acquisition and reading time. Br J Radiol 85:e1174–e1178
22. Dang PA, Freer PE, Humphrey KL et al (2014) Addition of tomosynthesis to conventional digital mammography: effect on image interpretation time of screening examinations. Radiology 270: 49–56
23. Campari C, Giorgi Rossi P, Mori CA et al (2016) Impact of the introduction of digital mammography in an organized screening program on the recall and detection rate. J Digit Imaging 29(2): 235–242
24. Tang ML, Tang NS, Chan IS et al (2002) Sample size determination for establishing equivalence/noninferiority via ratio of two proportions in matched-pair design. Biometrics 58(4):957–963
25. Meijer CJ, Berkhof J, Castle PE et al (2009) Guidelines for human papillomavirus DNA test requirements for primary cervical cancer screening in women 30 years and older. Int J Cancer 124(3):516–520. https://doi.org/10.1002/ijc.24010
26. Eliasziw M, Donner A (1991) Application of the McNemar test to non-independent matched pair data. Stat Med 10:1981–1991
27. Wolter KM (2007) Introduction to variance estimation, 2nd edn. Springer, New York
28. Cohen J (1960) A coefficient of agreement for nominal scales. Educ Psychol Meas 20:37–46
29. StataCorp (2013) Stata: Release 13. Statistical Software. StataCorp LP, College Station
30. Galati F, Marzocca F, Bassetti E et al (2017) Added value of digital breast tomosynthesis combined with digital mammography according to reader agreement: changes in BI-RADS rate and follow-up management. Breast Care (Basel) 12(4):218–222. https://doi.org/10.1159/000477537
31. Choi WJ, Kim HH, Lee SY et al (2016) A comparison between digital breast tomosynthesis and full-field digital mammography for the detection of breast cancers. Breast Cancer 23(6):886–892
32. Carbonaro LA, Di Leo G, Clauser P et al (2016) Impact on the recall rate of digital breast tomosynthesis as an adjunct to digital mammography in the screening setting. A double reading experience and review of the literature. Eur J Radiol 85(4):808–814. https://doi.org/10.1016/j.ejrad.2016.01.004.
33. van Schie G, Wallis MG, Leifland K et al (2013) Mass detection in reconstructed digital breast tomosynthesis volumes with a computer-aided detection system trained on 2D mammograms. Med Phys 40(4):041902
34. Samala RK, Chan HP, Hadjiiski L et al (2016) Mass detection in digital breast tomosynthesis: deep convolutional neural network with transfer learning from mammography. Med Phys 43(12):6654
35. Balleyguier C, Arfi-Rouche J, Levy L et al (2017) Improving digital breast tomosynthesis reading time: a pilot multi-reader, multi-case study using concurrent computer-aided detection (CAD). Eur J Radiol 97:83–89
36. Benedikt RA, Boatsman JE, etal SCA (2017) Concurrent computer-aided detection improves reading time of digital breast tomosynthesis and maintains interpretation performance in a multireader multicase study. AJR Am J Roentgenol 24:1–10
37. Dustler M, Andersson M, Fornvik D et al (2013) A study of the feasibility of using slabbing to reduce tomosynthesis review time. Proc. SPIE 8673 Medical Imaging 86731L
38. Friedewald SM (2017) Breast tomosynthesis: practical considerations. Radiol Clin North Am 55(3):493–502. https://doi.org/10.1016/j.rcl.2016.12.004