Physica Medica 57 (2019) 137–144

Contents lists available at ScienceDirect

Physica Medica

journal homepage: www.elsevier.com/locate/ejmp

Original paper

Experimental evaluation of seven quality control phantoms for digital breast tomosynthesis

Julie Sage[a,*], Karen L. Fezzani[a], Isabelle Fitton[b], Lama Hadid[c], Aurélie Moussier[d], Noëlle Pierrat[e], Antoine Martineau[f], Serge Dreuil[a], Loïc Heulers[a], Cécile Etard[a]

[a] Institut de Radioprotection et de Sûreté Nucléaire, PRP-HOM/SER/Unité d'Expertise en Radioprotection médicale, 31 avenue de la Division Leclerc, BP17, 92262 Fontenay-aux-Roses Cedex, France
[b] Hôpital Européen Georges Pompidou, 20 Rue Leblanc, 75015 Paris, France
[c] Hôpital Jean Verdier, Avenue du 14 Juillet, 93140 Bondy, France
[d] Institut Gustave Roussy, 114 Rue Edouard Vaillant, 94800 Villejuif, France
[e] Institut Curie, 26 Rue d'Ulm, 75005 Paris, France
[f] Hôpital Saint-Louis, 1 Avenue Claude Vellefaux, 75010 Paris, France

ARTICLE INFO

Keywords:
Breast
Diagnostic imaging
Digital breast tomosynthesis
Quality control
Screening

ABSTRACT

Purposes: The introduction of digital breast tomosynthesis (DBT) into the French breast cancer screening program is forecast by the authorities. The aim of the present study was to evaluate image quality phantoms to be used as internal quality controls.

Methods: Seven breast phantoms dedicated to quality control in mammography were evaluated on reconstructed DBT images: ACR Model 015, BR3D, DBT QC model 021, Mam/Digi-EPQC, MTM100, TOMOMAM® and TOMOPHAN®. Two representative image parameters of DBT images were studied: image score and z-resolution, when inserts were included in the phantom, on five DBT systems of three different brands. Three observers were involved.

Results: The MTM100, Mam/Digi-EPQC, BR3D, DBT QC model 021 phantoms' images presented artefacts affecting the image score. The ACR Model 015, TOMOMAM® and TOMOPHAN® phantoms appeared to be pertinent for DBT image score analysis. Due to saturation artefacts, Z-resolution results were not coherent with the theory for all phantoms except by using aluminium beads in the TOMOMAM® phantom.

Conclusions: Phantom manufacturers should be encouraged to collaborate with DBT system manufacturers in order to design universal phantoms suitable for all systems for more complete quality control. From our study we can propose several specifications for an ideal and universal phantom designed for internal quality control in DBT. Phantoms should allow sensitive image score measurements. The background structure should be realistic to avoid artefacts. Phantoms should have a standard breast-like shape and size.

1. Introduction

Tomosynthesis is a breast imaging technique that could be described as "pseudo-3D" [1,2]. It offers a gain in sensitivity and specificity in breast cancer detection compared to conventional full-field (2D) mammography [3–7], due to reduction in tissue overlap. Tomosynthesis is therefore of great interest to radiologists, even though the clinical indication for digital breast tomosynthesis (DBT) alone is not yet well defined [8,9]. DBT is already used in health centres in France and the number of facilities is increasing sharply. Although DBT appears to provide a benefit for medical diagnosis [3–7], it may deliver higher doses than 2D mammography, depending on the system [10–17]. Moreover, despite the recommendations of the European Reference Organisation for Quality Assured Breast Screening and Diagnostic Services (EUREF) [18], French DBT systems are still not well monitored, as the regulatory quality controls are not yet organised in France. Nevertheless, the introduction of this technique into the French breast cancer screening program is forecast by the authorities for the coming years. It is therefore necessary to organise the monitoring of the stability of the performance of DBT systems. In this context, the French

Abbreviations: ASF, artefact spread function; DBT, digital breast tomosynthesis; EUREF, European Reference Organisation for Quality Assured Breast Screening and Diagnostic Services; FWHM, full width at half-maximum; GEHC, general electric health care; SSP, slice sensitivity profile

* Corresponding author.
E-mail address: julie.sage@irsn.fr (J. Sage).

https://doi.org/10.1016/j.ejmp.2018.12.031
Received 4 May 2018; Received in revised form 6 November 2018; Accepted 21 December 2018
Available online 09 January 2019

Institute for Radiological Protection and Nuclear Safety (IRSN) led a study to evaluate the image quality of phantoms available on the market that could be used in a quality control programme for DBT systems to monitor the stability of their performance over time.

2. Materials and methods

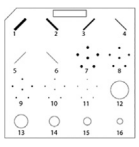
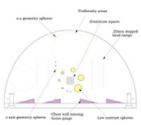
In France, regulatory quality control programmes distinguish internal quality controls, namely constancy tests done by users, from external quality controls done by accredited societies to quantify the absolute image quality and characteristics of DBT systems. This study is complementary to the EUREF protocol and focuses solely on internal quality control. The overall quality of reconstructed DBT images must be monitored over time by the system's users: radiographers or radiological technologists. The parameter usually used to assess global quality image in mammography is the image score [19], which is the French regulatory quality indicator used to monitor the stability of the whole 2D mammography chain. It corresponds to the number of visible inserts in a phantom simulating different sizes and shapes of breast lesions corresponding to masses (low-contrast detectability inserts), microcalcifications and fibres. The purpose is to apply this image score to reconstructed DBT images.

Seven specific phantoms available on the French market and containing a certain number of inserts were included in the study: MTM100 (CIRS, USA), TOMOMAM® (CIRS, USA) [20], Mam/Digi-EPQC [21] (QUART GmbH, Germany), TOMOPHAN® [22] (Phantom Laboratory, USA), BR3D Model 020 [23] (CIRS, USA), DBT QC model 021 [24], in a prototype version, and ACR Model015 [25] (CIRS, USA). Specifications of these phantoms are shown in Table 1. It should be noted that there are very few phantoms available on the market specifically designed for DBT, i.e. allowing measurements at different depths and with material, shape and structures that avoid, as far as possible, artefacts. Thus, phantoms originally designed for 2D mammography were also included in the study to check if they might be suitable for DBT. In particular, MTM100 and ACR Model 015, which are regulatory reference phantoms in France and the USA, were included in a comparative goal. When available, the phantom manufacturer's recommendations were applied.

2.1. Image score

To ensure the reliability of the image score, the phantoms must not generate any artefacts on reconstructed DBT images, otherwise the image score would be affected by the artefact. To check for artefacts, phantoms were imaged on a HOLOGIC Selenia® Dimensions® DBT system, in automatic mode. More generally, all images used in this study from all systems were checked for artefacts. The shape and strength of artefacts depend on the system design, especially the angular range and the number of projections as well as the image reconstruction algorithm and artefact reduction techniques. The images of artefact studies were visualised by three observers, under clinical conditions on the primary display system of the installation, looking for artefacts in the scoring zone that could be generated by the background, the material and the shape of the structures, the local contrast in the

Table 1
Principal specifications of the seven phantoms included in the study regarding image scoring.

Manufacturer	CIRS	CIRS	CIRS	CIRS	CIRS	PHANTOM LABORATORY	QUART
Model	ACR Mammographic Accreditation phantom Model 015	BR3D Breast Imaging Phantom Model 020	DBT QC model 021 (prototype version)	MTM 100	TOMOMAM®	TOMOPHAN® TSP004	Mam Digi/EPQC
Shape	Rectangular	Semi-circular	Semi-circular	Semi-circular	Semi-circular	Semi-circular	Semi-circular
Dimensions	108 mm × 102 mm	100 mm × 180 mm	100 mm × 180 mm	180 mm × 130 mm	Ø : 190 mm	100 mm × 180 mm	180 mm × 240 mm
Thickness (mm)	44	60 6 plates of 10 mm	60 6 plates of 10 mm	45	60	42 or 56[c]	45
Background	Homogenous – PMMA	Heterogeneous – Epoxy resin[a]	Homogenous	Homogenous – Epoxy resin[b]	Homogenous – Epoxy resin[b]	Homogenous	Homogenous
Micro-calcifications (groups)	5 (Al_2O_3) from 0.16 to 0.54 mm	6 ($CaCO_3$) From 0.13 to 0.40 mm	6 NC	7 From 0.09 to 0.30 mm	6 (glass) From 0.14 to 0.33 mm	/	12[d] From 0.18 to 0.60 mm
Masses	5 From 0.25 to 2 mm	6 From 1.8 to 6.3 mm	6 NC	7 From 1.8 to 6.3 mm	6 From 1.8 to 5.5 mm	8 (acrylic) From 1 to 10 mm	/
Fibres	6 (nylon) From 0.4 to 1.56 mm	7 From 0.15 to 0.60 mm	/	7 (nylon) From 0.30 to 1.25 mm	6 (nylon) From 0.15 to 0.41 mm	/	/
Drawings							
Other parameters available	No	No	Yes	No	Yes	Yes	Yes
Designed for DBT	No	No	Yes	No	Yes	Yes	Yes

[a] Two tissue-equivalent materials mimicking 100% adipose and 100% gland tissues "swirled" together in an approximate 50/50 ratio.
[b] Equivalent to 50% adipose and 50% gland tissues.
[c] Thickness used in this study.
[d] Each group is composed of six Landolt objects.

Table 2
Principal specifications of the five tomosynthesis systems included in the study.

Manufacturer	GEHC	HOLOGIC	SIEMENS	
Model	SenoClaire	Selenia® Dimensions®	Mammomat Inspiration	
Number of systems included	2	2	1	
Year of installation	2015	2014 and 2015	2010	
Tube motion	Step and Shoot	Continuous	Continuous	
Angular range (°)	25	15	50	
Number of projections	9	15	25	
Focal plane pixel size (μm)	100	95–117	85	
Spacing between focal planes used (mm)	1	1	1	
Workstation	GEHC IDI	HOLOGIC SV 10000X		
Carestream VUE Mammo	SIEMENS VB			
Monitors associated	BARCO MDNG-5121	BARCO MDCG-10130		
BARCO MDNG-5121	EIZO OFTD1083			
Range of settings given by AEC for phantom exposure	kV	29	29–33	27–29
	mAs	41–83	48–71	161–212
Filtration	Target/ Rh/Rh	W/Al	W/Rh	

phantom images or the phantom edge.

The image score analysis must also be discriminating enough to detect variations in image quality (lower or higher). Under usual conditions, the number of visible inserts on the focus plane should not be at the end of the range of inserts. Three DBT acquisitions of each phantom were performed in automatic mode on five DBT systems of three different brands. System specifications are described in Table 2. The range of settings given by each system's automatic exposure control (AEC) for phantom exposure is also given in Table 2. As the thickness and composition of phantoms vary substantially, the ranges of tension and charge are large. The image score was determined by three observers, at the inserts best focus plane, under clinical conditions on European Conformity-marked primary display systems with mammography dedicated monitors (cf. Table 2). Readers were allowed to vary contrast and/or zoom, and scroll through slices. The ambient light was checked to be between 10 and 20 lux, in order to ensure identical clinical conditions for the readings. The average score was calculated for all 3 readers and each system model.

2.2. Z-resolution

Regarding tomosynthesis specificities, it is a plus to monitor the spatial resolution in depth (Z-resolution). The measurement of Z-resolution was only available on TOMOMAM®, TOMOPHAN® and CIRS QC model 021 prototype.

With TOMOMAM®, the slice sensitivity profile in Z (SSP(z)) was determined on reconstructed images acquired in automatic mode for each system, by using the methodology described by Wu et al. [26] and adapted by Sechopoulos [1] on a 0.4 mm diameter steel bead, located at mid-thickness of the phantom and using the Eq. (1):

$$SSP(z) = \frac{Is(z) - IBG(z)}{Is(z_0) - IBG(z_0)} \quad (1)$$

where "z_0" is the in-focus plane location; "z" the current location; "Is" the mean pixel value of the bead signal; and "IBG" the background mean pixel value.

The SSP(z) was also determined on two 1 mm diameter aluminium beads located at mid-thickness of the phantom, at the left and right sides. It should be noted that these aluminium beads were not planned for that purpose by the designers of the phantom but for the purposes of geometrical tests. From SSP(z) curves thus determined, it is possible to estimate, among others, the full width at half-maximum (Z-FWHM), which is representative of Z-resolution.

The SSP(z) curve and its Z-FWHM were determined manually with ImageJ. The results were cross-checked with the automatic composite Z-FWHM measurements using the "geometric distortion tool" from the ImageJ plugin "NCCPM Tools for DBT QC" developed by the National Coordinating Centre for the Physics of Mammography: the stack of focal planes is re-sliced in the vertical direction and a composite plane is created using the maximum pixel value from all planes. This method is used to consider the spreading of bead signal on other planes.

With TOMOPHAN®, the SSP(z) was determined on reconstructed images acquired in automatic mode, using a technique of trigonometric projection of Z-axis depth onto the (X-Y) plane. The phantom includes two sets of two angled tungsten carbide bead ramps. The diameter of each bead is 0.18 mm, and beads are spaced vertically at 0.25 mm. Each pair of ramps covers a total height of 10 mm. The SSP(z) curve corresponds to the curve connecting the point of maximum intensity of each bead. The SSP(z) and its Z-FWHM were determined using the remote TOMOPHAN® QA software. Z-FWHM was also determined manually by plotting the intensity profile of each bead ramp. On the profile, the background value was subtracted from the maximum value. The number of beads with a maximum intensity over the half-maximum was then multiplied by 0.25 mm corresponding to the Z-axis spacing.

Maki et al. [27] have shown that the angular range of the X-ray tube has a strong influence on Z-resolution. When the angular range increases, the Z-resolution improves (i.e. FWHM value decreases). Thus, the SIEMENS system with a 50° angle would theoretically have a better resolution in Z (i.e. smaller FWHM value) than the other two systems with a smaller angular range (cf. Table 2). The Z-FWHM values thus determined had to be in agreement with this theory.

It should be noted that the DBT QC model 021 was evaluated in a prototype version and has not been evaluated in its current configuration. In the prototype version, several aluminium beads could have been used to check the SSP(z), but the images of these beads were affected by too many artefacts and were therefore not evaluated (cf. Fig. 1). These aluminium beads are no longer available with the current version of the phantom.

3. Results

3.1. Image score

The textured background of the BR3D is highly contrasted. This high contrast generated artefacts over the entire phantom (cf. Fig. 2(a)). These artefacts were troublesome in establishing an image score. Moreover, circular structures, in which the inserts are encrusted, were clearly visible in the reconstructed images (cf. Fig. 2(a)). These circular structures easily indicated the position of the inserts and consequently affected the image score analysis.

Analysis of the MTM100 phantom's images showed different kinds of artefacts generated by typography inserts (in-plane artefact of lower intensity); contrast scale (in-plane, lower and higher intensities) or holes in the upper part of the phantom (out-of-plane, lower intensity)

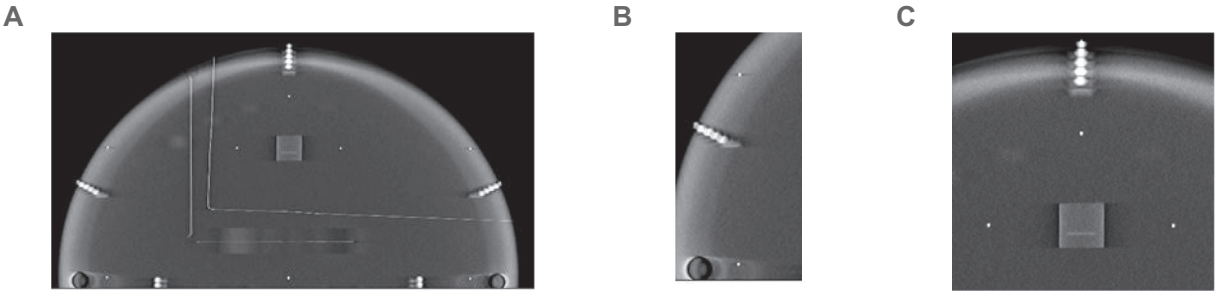


Fig. 1. (A) Image of the DBT QC model 021 prototype. Artefacts generated by the phantom edges (B) or the aluminium sheet and the missed tissue markers (C) near the aluminium beads did not allow evaluation of the SSP(z) parameters for the prototype version. These beads are not available with the current version of the phantom.

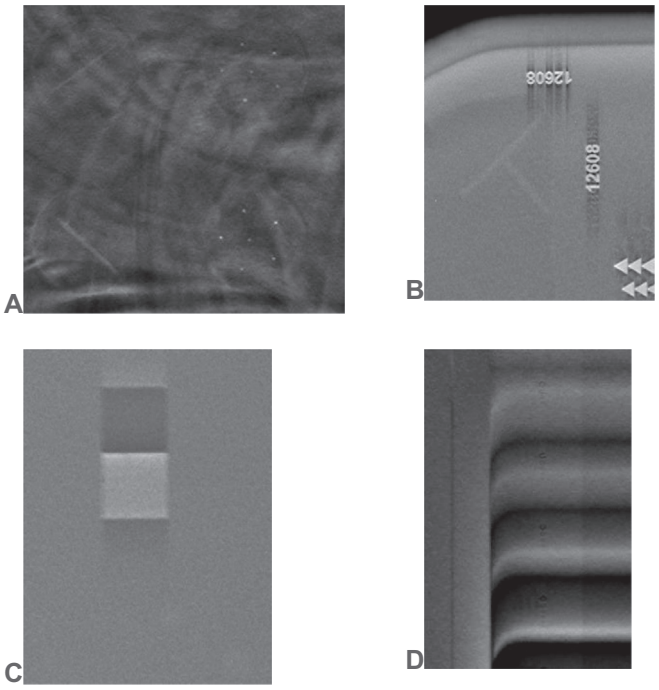


Fig. 2. Examples of artefacts generated by BR3D (A), MTM100 (B and C) and Mam Digi/EPQC (D) phantoms.

(cf. Fig. 2(b) and (c)). The microcalcifications and the fibres were obscured by some of the artefacts disturbing the image score analysis.

On the MAM/DIGI-EPQC phantom images, artefacts appeared on the entire step-wedge (cf. Fig. 2(d)). The inserts mimicking microcalcifications used to determine the image score were positioned on this step-wedge. Therefore, the image score analysis was affected by the generated artefacts. Analysis of the ACR, DBT QC model 021 prototype, TOMOMAM® and TOMOPHAN® phantom images showed that image scores were not impeded by artefacts. The mean image scores determined with the ACR, DBT QC model 021, TOMOMAM® and TOMOPHAN® on SIEMENS, GEHC and HOLOGIC systems are shown in Fig. 3.

The DBT QC model 021 prototype gave an almost identical score for the 3 DBT systems. In particular, regardless of the DBT system evaluated, 6 masses out of 6 were always visualised. The maximum score was always achieved regarding masses. This means that this phantom lacks sensitivity in detecting differences in image quality.

The results showed that with ACR, TOMOMAM® and TOMOPHAN® not all the inserts are visible. The highest number of fibres and microcalcifications was seen on ACR. The highest number of masses was seen on TOMOPHAN®. ACR and TOMOMAM® had a similar number of visible masses.

On TOMOMAM® and TOMOPHAN®, the fewest masses were seen with the SIEMENS system. On TOMOMAM®, the fewest fibres were also seen with the SIEMENS system. Whereas on the ACR, the highest number of fibres was seen with the SIEMENS system. There was no substantial difference between systems in the number of visible inserts mimicking microcalcifications, whatever the phantom considered.

3.2. Z-resolution

Fig. 4 presents the SSP(z) curves determined on the 3 systems with the TOMOMAM® 0.4 mm steel bead. On TOMOMAM®, a clipping occurred on the SSP(z) curves obtained for the 3 DBT systems with the steel bead for one to five consecutive slices. This clipping is particularly clear in the GEHC results. The SSP(z) curve of the GEHC system is also sharper than for other systems.

Table 3 presents the Z-FWHM values determined on the 3 systems with the TOMOMAM® 0.4 mm steel bead and 1 mm aluminium beads, as well as with TOMOPHAN®.

TOMOMAM® phantom image analysis gave the best Z-resolution with the steel bead for HOLOGIC systems and the worst one for the GEHC system. The results were not as expected.

By using the two aluminium beads, the Z-FWHM results were in agreement with the theory: the smallest Z-FWHM value was obtained with the SIEMENS system and the maximum Z-FWHM value with the HOLOGIC system. A discrepancy was noticed between the FWHM values calculated manually and using the NCCPM software. In particular, for aluminium beads on HOLOGIC systems the difference between the manual and automatic calculation was large.

Fig. 5 presents the SSP(z) curves determined on the 3 systems with the TOMOPHAN®. The results were not in agreement with the theory as the SIEMENS system showed the highest Z-FWHM (cf. Table 3).

4. Discussion

Periodic quality controls of DBT systems are not defined in French regulations and a national quality control programme is needed. The EUREF protocol [18] is well adapted to tests done by medical physicists or qualified staff. It could be applied to the so-called French external quality control programme. Nevertheless, very few tests are addressed to radiographers or radiological technologists in the scope of an internal quality control programme. The EUREF protocol recommends a daily/weekly image of a homogenous block of PMMA, but does not recommend monitoring of reconstructed image scores. Several image quality phantoms with different specifications are available on the market to monitor DBT systems in this context, but no studies have previously evaluated these phantoms on different DBT systems.

Table 4 summarises the phantom parameters and their advantages and drawbacks in the present study. Other parameters not studied here might be worth checking in the setting of a quality control programme, such as in-plane resolution, signal difference-to-noise ratio (SDNR), homogeneity, geometry, and missed tissue.

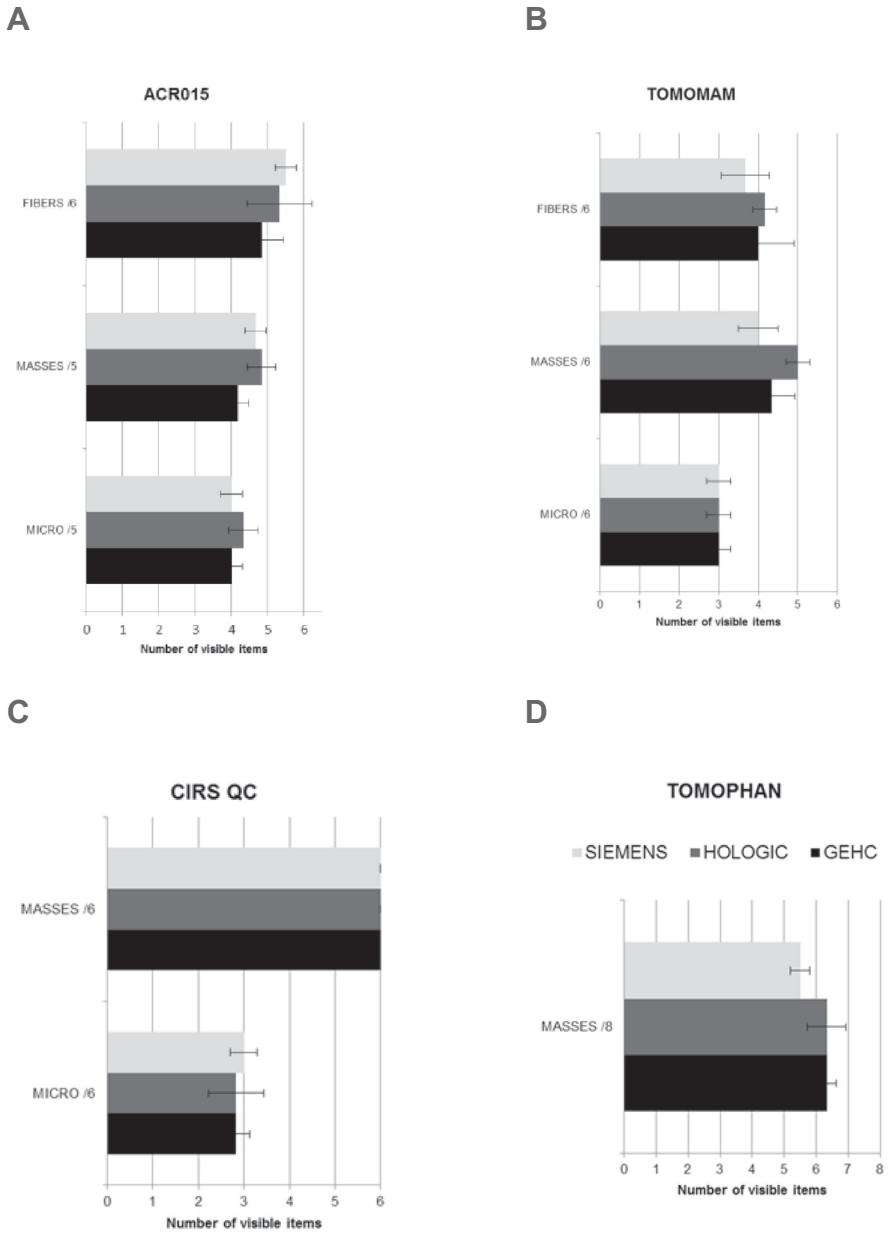


Fig. 3. Image scores (with the standard deviation) determined on SIEMENS, GEHC and HOLOGIC systems in automatic mode with ACR (A), TOMOMAM® (B), DBT QC model 021 (C) and TOMOPHAN® (D) phantoms. The number of each type of insert is indicated on the bar graph: e.g. "Masses /6" means there are 6 masses in the phantoms.

In our study, four phantoms could not be recommended for quality control of internal image scores. BR3D, MTM100 and MAM / DIGI EPQC did not appear to be adapted to DBT specificities because of their design. Indeed, sharp edges (i.e. MAM / DIGI EPQC phantom's step-wedge) or high-contrast structures (i.e. MTM100 typography and BR3D textured background) generated artefacts on reconstructed DBT images that may disturb the image score analysis.

Although the BR3D generates artefacts, it should be remembered that it is of substantial use in research and development work on DBT systems. Its textured background is of particular interest for evaluation of DBT reconstruction algorithms or for comparison of DBT with 2D mammography [28,29]. However, the artefacts generated by the background showed that the structured background might not be sufficiently realistic. As the image reconstruction/processing is adjusted for true breast images, it could operate differently with such a background compared with breast images. This makes comparisons difficult.

Moreover, the image score on the prototype version of DBT QC model 021 showed that this prototype version always gave the maximum score for masses, leading to a lack of sensitivity to image quality variations. The current version has not been evaluated. The manufacturer may have made changes in this regard.

In our study, the ACR model 015, TOMOMAM® and TOMOPHAN® appeared to be pertinent for periodic monitoring of DBT image scores. It is interesting to note that these phantoms all have a uniform background, which seems better suited to periodic internal quality control. It should be mentioned that phantoms with a uniform background might generate images quite different from true breast images as the reconstruction/processing is, as explained above, adjusted for true breast images. This might influence the phantom's sensitivity in detecting clinically relevant image quality problems and should be addressed in a complementary study to check phantom sensitivity over time. For the same reasons, comparison between systems regarding image scores is not possible.

The sizes of microcalcifications and fibres on the ACR model 015 are

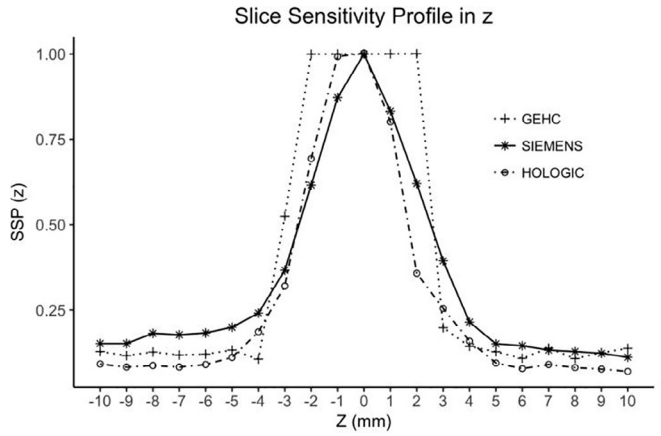


Fig. 4. Slice sensitivity profiles determined manually on the SIEMENS, GEHC and HOLOGIC systems with the 0.4 mm diameter steel bead of the TOMOMAM® phantom.

greater than with TOMOMAM®. This could explain the highest number of fibres and microcalcifications seen on ACR. Masses in TOMOMAM® were bigger than in ACR, but the two systems had a similar number of visible masses. This can be explained by the fact that glandular/adipose ratios and shapes of masses differ between these two phantoms.

The very similar results in terms of microcalcifications for the three systems regardless of the phantom raise questions. There could be a variety of explanations regarding the systems and phantom design. This could mean that systems have the same detection limit for microcalcifications, taking into account the restriction mentioned above about the impossibility of comparing system scores. This could also be due to the specifications of groups of microcalcifications and the visualisation limit of human eyes. For instance, if steps between sizes of microcalcification groups are too large, with one group always visible and one group too small to be seen by human observers, systems will then all have the same score. This phenomenon was not identified with masses.

This study needs to be completed by an evaluation of phantom scores at various dose levels (mAs). Indeed, the sensitivity of the phantom should allow for the detection of variations in image quality over time. This will be addressed in a complementary study.

The smallest phantom evaluated in this study, the ACR model 015, only permits the measurement of image score. Nevertheless, it is an international reference phantom used for regulatory purposes in the USA in the scope of the Mammography Quality Standard Act [30]. A new larger ACR digital mammography phantom has been designed for use with the ACR Digital Mammography QC Manual published on July 29, 2016 [31]. It could be interesting to evaluate this larger phantom in the future.

The TOMOMAM® and TOMOPHAN® image score results were not affected by artefacts. Image scores appeared to be more sensitive to image quality modification on TOMOMAM® than on TOMOPHAN®.

It is important to mention that TOMOMAM® contains fibres, microcalcifications and masses, in line with the 2D mammography quality control programme in France, whereas TOMOPHAN® only contains masses.

The TOMOMAM and TOMOPHAN phantoms are still of particular interest because they allow measurement of SSP(z). With TOMOMAM®, the fact that SSP(z) curves with the steel bead reached the maximum value of the grey scale means that these systems saturated on the steel bead. Therefore, the calculation of Z-FWHM was distorted by normalisation of the SSP(z) curves with this truncated maximum value (cf. Eq. (1)) and the results were considered invalid. Thus, the use of the steel bead for the determination of the SSP(z), and other parameters arising from it, cannot be recommended for internal quality control of DBT systems. It is interesting to note that this saturation was more pronounced on GEHC systems (cf. Fig. 4). Indeed, the clipping well visible on the GEHC curve might be due to an image reconstruction process, like enhancement processing, specific to the GEHC system. This specific reconstruction process and the iterative reconstruction algorithm might cause the sharpness of the SSP(z) curves on the GEHC system [17].

With the aluminium bead, SSP(z) curves also showed saturation on HOLOGIC and SIEMENS for four to six slices. Nevertheless, the composite Z-FWHM results were as expected in theory, i.e. the largest tomosynthesis angle has the best z-resolution and the smallest angle the lowest z-resolution. Results are also in agreement with the results shown in NHSBSP evaluation [12–14]. This means that the saturation effect has less impact on the truncated maximum pixel value and SSP(z) curves with aluminium than with steel beads. TOMOMAM®'s aluminium beads might be used to replace the steel bead more specific to the measurement of the artefact spread function. Nevertheless, it should be verified in a complementary study that the aluminium truncated maximum pixel value does not affect the sensitivity of this test over time.

There may be several reasons for the difference between manual and composite Z-FWHM values. The aluminium beads are situated at the left and right sides, near the edges of the phantom, resulting in a slight non-uniformity of the background around the beads. This effect is more pronounced on HOLOGIC. The background taken into account for the manual calculation might be different from that taken into account by the NCCPM software. Moreover, the manual calculation did not take into account possible geometric distortion or differences in pixel size regarding focal plans, as can be the case for the HOLOGIC system [13]. The aluminium bead signal spreading was more pronounced than with the steel bead, due to saturation.

The Z-resolution determined with SSP(z) curve shapes (cf. Fig. 5) using TOMOPHAN® is better on the GEHC system than on the HOLOGIC system whose scanning angles are respectively 25° and 15° (cf. Table 1). These results were expected. Z-resolution determination on the TOMOPHAN® appeared satisfactory for the evaluation of the GEHC and HOLOGIC systems. But SSP(z) curves on the SIEMENS system have an abnormal very large oscillating shape (cf. Fig. 5) giving the highest Z-FWHM. This shape was confirmed by manual analysis. It might be partly explained by the fact that the measurement includes the impact of varying distance from the chest wall. Moreover, the trigonometric

Table 3
Z-FWHM results on the GEHC, HOLOGIC and SIEMENS systems with TOMOMAM® and TOMOPHAN® phantoms.

	GEHC Z-FWHM (mm)		HOLOGIC Z-FWHM (mm)		SIEMENS Z-FWHM (mm)	
	Manual	Automatic	Manual	Automatic	Manual	Automatic
TOMOMAM® 0.4 mm steel bead	6.6	6	4.8	4.2	5.8	5
TOMOMAM® 1 mm aluminium beads Average	7.8	7.6	8.6	11.9	5.2	5.7
TOMOPHAN®	1.5	1.9	2.25	2.5	4.5	4.2

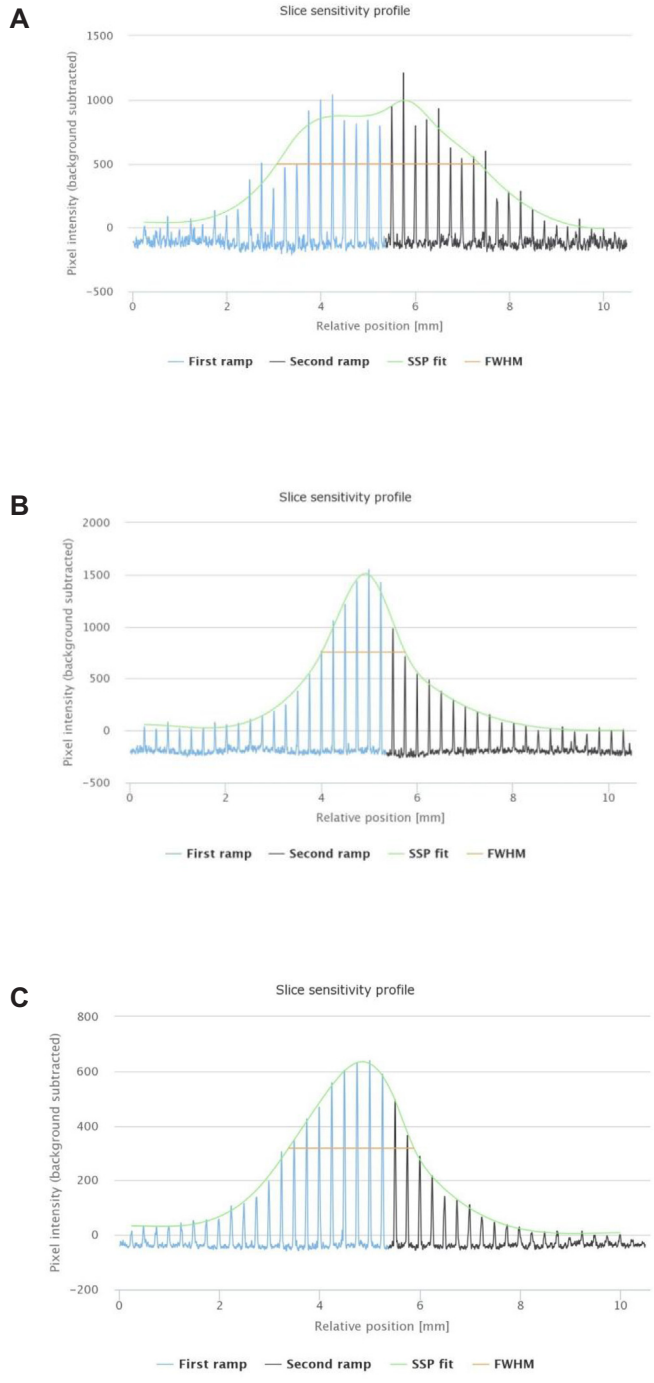


Fig. 5. Slice sensitivity profile curve determined on the SIEMENS (A), GEHC (B) and HOLOGIC (C) systems with the TOMOPHAN® phantom and reported by the TOMOPHAN® QA software.

projection of Z-axis depth onto the (X-Y) plane assumes that the focal planes are parallel to the Bucky surface. For the Siemens system, the focal planes are reconstructed with a slight angle compared to the Bucky surface [12]. The measurement is certainly biased by this angle. Thus the trigonometric projection of Z-axis depth onto the (X-Y) plane is not valid with this system. This method is not universal.

5. Conclusion

The visual image score analysis is an easy way to check the constancy of the whole DBT quality chain and to motivate system users to

Table 4
Overview of phantom parameters evaluated in this study.

Manufacturer	CIRS	CIRS	CIRS	CIRS	CIRS	PHANTOM LABORATORY	QUART
Model	ACR Mammographic Accreditation phantom Model 015	BR3D Breast Imaging Phantom Model 020	DBT QC model 021 (prototype version)	MTM 100	TOMOMAM®	TOMOPHAN® TSP004	Mam Digi/EPQC
Artefacts on image score zone	No	Yes	No	Yes	No	No	Yes
Image score	Well adapted	Affected by artefact	No margin in structure visualisation	Affected by artefact	Well adapted	Only low contrast detail available Low sensitivity to image quality variations	Affected by artefact
Image score measurements at different depths	No	Yes, with different plates	Yes, with different plates	No	Yes, with reversible position	No	Yes, Landolt objects at different depths
SSP(z)	NA[a]	NA[a]	Not evaluated for the current version of the phantom	NA[a]	Consistent with theory on aluminium beads. Not recommended for steel beads.	Not recommended for all the systems	NA[a]

[a] NA: not available.

achieve internal quality control. Before recommending its implementation, the sensitivity of the phantom needs to be checked to ensure that image scores allow for the detection of variations in DBT image quality over time. This will be addressed in a complementary study.

Analysis of visual phantom parameters should remain simple in internal quality control. This is already implemented for 2D mammography and could be implemented for the tomosynthesis system if the test is confirmed to be sensitive enough to identify image quality problems over time in a complementary study. More complete parameters should also be monitored by qualified persons. This study focused only on Z-resolution. More generally, this study showed that before choosing a specific phantom, it is recommended to validate it on the DBT systems it will be associated with. Manufacturers of DBT phantoms should be encouraged to collaborate with system manufacturers in order to design universal phantoms suitable for all systems.

Our study shows that it is possible to propose several specifications for an ideal and universal phantom designed for internal quality control in DBT. Phantoms should contain fibres, microcalcifications and masses to allow the image score measurements. Insert sizes should be suitable for all DBT system performances, such that the resulting image score is sensitive enough to identify image quality problems over time (higher and lower quality) on all systems. This means that the number of visible inserts on the focus plane should not be at the end of the range of inserts. Their design should be realistic enough to avoid artefacts that could be generated by the background, the material and the shape of the structures, and the local contrast in the phantom images or the breast edge. They should permit image score measurements at different depths and/or for different thicknesses. The phantoms should have a standard breast-like shape and size. Finally, ideal and universal phantoms could also permit the measurement, on all kinds of DBT systems, of other parameters recommended by EUREF [18], such as SSP(z), in-plane resolution, SDNR, homogeneity, geometry and missed tissue check.

Acknowledgements

The authors are very grateful to the suppliers CIRS, MEDITEST, ORION, PHANTOM LABORATORY and QUART GmbH who were willing to lend us their phantoms as well as Hôpital Européen Georges Pompidou, Hôpital Jean Verdier, Institut Curie, Institut Gustave Roussy and Hôpital Saint-Louis for access to their DBT systems.

Declaration of interest

One of the authors of this manuscript declares a relationship with ORION, the French supplier of TOMOMAM®. Indeed, Isabelle FITTON is one of the co-inventors of the TOMOMAM® phantom.

References

[1] Sechopoulos I. A review of breast tomosynthesis. Part I. The image acquisition process. Med Phys 2013. https://doi.org/10.1118/1.4770279.
[2] Sechopoulos I. A review of breast tomosynthesis. Part II. Image reconstruction, processing and analysis, and advanced applications. Med Phys 2013. https://doi.org/10.1118/1.4770281.
[3] Ciatto S, Houssami N, Bernardi D, Caumo F, Pellegrini M, Brunelli S, et al. Integration of 3D digital mammography with tomosynthesis for population breast-cancer screening (STORM): a prospective comparison study. Lancet Oncol 2013;14(7):583–9.
[4] Skaane P, Bandos AI, Gullien R, Eben EB, Ekseth U, Haakenaasen U, et al. Comparison of digital mammography alone and digital mammography plus tomosynthesis in a population-based screening program. Radiology 2013;267(1):47–56.
[5] Haas BM, Kalra V, Geisel J, Raghu M, Durand M, Philpotts LE. Comparison of tomosynthesis plus digital mammography and digital mammography alone for breast cancer screening. Radiology 2013;269(3):694–700.
[6] Rose SL, Tidwell AL, Bujnoch LJ, Kushwaha AC, Nordmann AS, Sexton Jr. R. Implementation of breast tomosynthesis in a routine screening practice: an observational study. AJR Am J Roentgenol 2013;200(6):1401–8.
[7] McDonald Elizabeth S, et al. Effectiveness of digital breast tomosynthesis compared with digital mammography outcomes analysis from 3 years of breast cancer screening. JAMA Oncol 2016;2(6):737–43. https://doi.org/10.1001/jamaoncol.2015.5536.
[8] Houssami N, Skaane P. Overview of the evidence on digital breast tomosynthesis in breast cancer detection. The Breast 2013;22:101–8.
[9] Vedantham S, Karellas A, Vijayaraghavan GR, Kopans DB. Digital breast tomosynthesis: state of the Art. Radiology 2015;277(3).
[10] Dance DR, Strudley CJ, et al. Comparison of breast doses for digital tomosynthesis estimated from patient exposures and using PMMA breast phantoms. In: Maidment ADA, Bakic PR, Gavenonis S, editors. Breast Imaging 11th International Workshop, IWDM 2012, Philadelphia, PA, USA, July 8-11, 2012, Proceedings. Verlag Berlin Heidelberg: Springer; 2012. p. 316–21.
[11] Feng SSJ, Sechopoulos I. Clinical digital breast tomosynthesis system: dosimetric characterization. Radiology 2012;263:35–42.
[12] National Health Service Breast Screening Program (2015) Technical evaluation of Siemens Mammomat Inspiration digital breast tomosynthesis system. NHSBSP Equipment Report 1306. Available via https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/505808/nhsbsp-equipment-report-1306v2_uploaded_070316.pdf.
[13] National Health Service Breast Screening Program (2014) Technical evaluation of Hologic Selenia® Dimensions® digital breast tomosynthesis system. NHSBSP Equipment Report 1307. Available via https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/505800/nhsbsp-equipment-report-1307v2.pdf_uploaded_070316.pdf.
[14] National Health Service Breast Screening Program (2016) Technical evaluation of GE Healthcare SenoClaire digital breast tomosynthesis system. NHS Breast Screening Programme Equipment Report 1404 available via https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/604430/Tech_Eval_GE_SenoClaire_tomo_1404_FINAL.pdf.
[15] Cavagnetto F, Taccini G, Rosasco R, Bampi R, Calabrese M, Tagliafico A. 'In vivo' average glandular dose evaluation: one-to-one comparison between digital breast tomosynthesis and full-field digital mammography. Radiat Prot Dosim 2013. https://doi.org/10.1093/rpd/nct120.
[16] Shin Sung Ui, Chang Jung Min, et al. Comparative evaluation of average glandular dose and breast cancer detection between single-view digital breast tomosynthesis (DBT) plus single-view digital mammography (DM) and two-view DM: correlation with breast thickness and density. Eur Radiol 2014. https://doi.org/10.1007/s00330-014-3399-z.
[17] Maldera A, De Marco P, Colombo PE, Origgi D, Torresin A. Digital breast tomosynthesis: dose and image quality assessment. Phys Med 2017;33:56–67. https://doi.org/10.1016/j.ejmp.2016.12.004.
[18] European Reference Organisation for Quality Assured Breast Screening and Diagnostic Services (2018) Protocol for the Quality Control of the Physical and Technical Aspects of Digital Breast Tomosynthesis Systems. European Reference Organisation for Quality Assured Breast Screening and Diagnostic Services, Nijmegen. Available via http://www.euref.org/european-guidelines/physico-technical-protocol#bottom.
[19] National Health Service Breast Screening Program (2013) Routine quality control tests for full-field digital mammography systems. Equipment report 1303: fourth edition. Available via https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/442720/nhsbsp-equipment-report-1303.pdf.
[20] TOMOMAM description available at http://www.orion-france.com/en/medical-imaging/mammography_tomomam.html.
[21] QUART Digi EPC technical data sheet available at https://quart.de/sites/default/files/2017-06/Quart-mam-digi-epqc-eng.pdf.
[22] The Phantom Laboratory (2016) Tomophan™ TSP004 Manual. Copyright © 2016. available at phantomlab.com.
[23] CIRS BR3D data sheet available at http://www.cirsinc.com/file/Products/020/020%20DS%20102915(1).pdf.
[24] CIRS DBT QC model 021 Data sheet available at http://www.cirsinc.com/file/Products/021%20DS%20121517(1).pdf.
[25] BR3D data sheet available at http://www.cirsinc.com/file/Products/020/020%20DS%20102915(1).pdf.
[26] Wu T, Moore RH, Rafferty EA, Kopans DB. A comparison of reconstruction algorithms for breast tomosynthesis. Med Phys 2004;31:2636–47.
[27] Maki A, Mainprize J, Mawdsley G, Martin Y. Quality control of breast tomosynthesis for a screening trial: preliminary experience. In: Tingberg A, Lång K, Timberg P, editors. Breast Imaging 13th International Workshop, IWDM 2016, Malmö, Sweden, June 19-22, 2016, Proceedings. Verlag Berlin Heidelberg: Springer; 2016. p. 11–9.
[28] Meyblum E, Gardavaud F, Dao TH, Fournier V, Beaussart P, Pigneur F, et al. Breast tomosynthesis: dosimetry and image quality assessment on phantom. Diagn Interv Imaging 2015. https://doi.org/10.1016/j.diii.2014.12.010.
[29] Hye-Suk P, Ye-Seul K, Hee-Joung K, Young-Wook C, Choi Jae-Gu. Optimization of configuration parameters in a newly developed digital breast tomosynthesis system. J Radiat Res 2013;55(3):589–99.
[30] U.S. FOOD AND DRUG ADMINISTRATION (FDA) (2004) The Mammography Quality Standards Act. Available via https://www.fda.gov/Radiation-EmittingProducts/MammographyQualityStandardsActandProgram/Regulations/ucm110823.htm.
[31] American College of Radiology (ACR) (2016) Digital Mammography Quality Control Manual. Available via http://www.acraccreditation.org/Modalities/Mammography.