



## Research article

# Reversal of the hanging protocol of Contrast Enhanced Mammography leads to similar diagnostic performance yet decreased reading times



Koos van Geel<sup>a,\*</sup>, Ellen M. Kok<sup>b</sup>, Jorian P. Krol<sup>a</sup>, Ivo P.L. Houben<sup>a,c</sup>, Fabienne E. Thibault<sup>d</sup>, Ruud M. Pijnappel<sup>e</sup>, Jeroen J.G. van Merriënboer<sup>f</sup>, Marc B.I. Lobbes<sup>a,c</sup>

<sup>a</sup> Department of Radiology, Maastricht University Medical Center, PO box 5800, 6202, AZ, Maastricht, the Netherlands

<sup>b</sup> Department of Education, Utrecht University, Heidelberglaan 1, Room E3.34, 3584, CS, Utrecht, the Netherlands

<sup>c</sup> Maastricht University, School for Oncology and Developmental Biology (GROW), Universiteitsingel 40, 6229, ER, Maastricht, the Netherlands

<sup>d</sup> Department of Radiology, Institut Curie, 26 rue d'Ulm, 75248, Paris Cedex 05, France

<sup>e</sup> Department of Radiology, Universitair Medisch Centrum Utrecht, Heidelberglaan 100, 3584, CX, Utrecht, the Netherlands

<sup>f</sup> Maastricht University, School of Health Professions Education, Maastricht University, PO Box 616, 6200, MD, Maastricht, the Netherlands

## ARTICLE INFO

## Keywords:

Contrast Enhanced Mammography  
Full-Field Digital Mammography  
Hanging protocols  
Eye-tracking methodology  
Visual expertise

## ABSTRACT

**Objectives:** Contrast-enhanced mammography (CEM) was found superior to Full-Field Digital Mammography (FFDM) for breast cancer detection. Current hanging protocols show low-energy (LE, similar to FFDM) images first, followed by recombined (RC) images. However, evidence regarding which hanging protocol leads to the most efficient reading process and highest diagnostic performance is lacking. This study investigates the effects of hanging-protocol ordering on the reading process and diagnostic performance of breast radiologists using eye-tracking methodology. Furthermore, it investigates differences in reading processes and diagnostic performance between LE, RC and FFDM images.

**Materials and methods:** Twenty-seven breast radiologists were randomized into three reading groups: LE–RC (commonly used hangings), RC–LE (reversed hangings) and FFDM. Thirty cases (nine malignant) were used. Fixation count, net dwell time and time-to-first fixation on malignancies as visual search measures were registered by the eye-tracker. Reading time per image was measured. Participants clicked on suspicious lesions to determine sensitivity and specificity. Area-under-the-ROC-curve (AUC) values were calculated.

**Results:** RC–LE scored identical on visual search measures,  $t(16) = -1.45$ ,  $p = .17$  or higher- $p$  values, decreased reading time with 31%,  $t(16) = -2.20$ ,  $p = .04$ , while scoring similar diagnostic performance compared to LE–RC,  $t(13.2) = -1.39$ ,  $p = .20$  or higher  $p$ -values. The reading process was more efficient on RC compared to LE. Diagnostic performance of CEM was superior to FFDM;  $F(2,26) = 16.1$ ,  $p < .001$ . Average reading time did not differ between the three groups,  $F(2,25) = 3.15$ ,  $p = .06$ .

**Conclusion:** The reversed CEM hanging protocol (RC–LE) scored similar on diagnostic performance compared to LE–RC, while reading time was a third faster. Abnormalities were interpreted quicker on RC images. A RC–LE hanging protocol is therefore recommended for clinical practice and training. Diagnostic performance of CEM was (again) superior to FFDM.

## 1. Introduction

Contrast-enhanced mammography (CEM) has been shown to be superior to Full-Field Digital Mammography (FFDM) for both the detection of breast cancer and the evaluation of disease extent [1–4]. A typical CEM exam consists of a low-energy (LE) image, which is comparable to FFDM [5,6], and a post-contrast recombined (RC) image, which shows areas of contrast uptake [1]. Prior studies found that all diagnostic parameters of CEM were significantly higher when

compared to FFDM [2]. It even matched the diagnostic accuracy of breast MRI, which in general is considered to be the most accurate breast imaging modality [7–9]. At present, all vendors present CEM-cases on their workstations using a hanging protocol (the order in which the images are presented to the radiologist) showing the LE images first, followed by the RC images, either as overlay or as separate image [3,10,11]. However, evidence on what hanging protocol is most effective is lacking [12], as there is no knowledge on how radiologist read CEM exams in clinical practice.

\* Corresponding author.

E-mail address: [k.vangeel@maastrichtuniversity.nl](mailto:k.vangeel@maastrichtuniversity.nl) (K. van Geel).

<https://doi.org/10.1016/j.ejrad.2019.05.013>

Received 3 March 2019; Received in revised form 11 May 2019; Accepted 14 May 2019

0720-048X/ © 2019 Elsevier B.V. All rights reserved.

Eye-tracking methodology allows us to investigate how the sequence of LE images and RC images of a hanging protocol affects the reading process. An eye-tracker measures where, when and for how long a radiologist looks during the reading process [13]. A particular image area, such as a lesion visible on one of the images, could draw the radiologist's attention, moving the eyes to this area. Eye movements thus reflect the radiologist's directed attention [14,15]. Eye-tracking can objectively measure whether radiologists find lesions faster and fixate longer on lesions in a certain hanging protocol.

This paper aims to investigate the effects of hanging protocols on the reading process and diagnostic performance of breast radiologists. Also, it aims to investigate the differences in the reading process and diagnostic performance between CEM and FFDM. We hypothesized that: [1] participants of a (reversed) RC-LE hanging protocol will be more efficient and will score higher on diagnostic performance in the reading of cases compared to participants using the (regular) LE-RC hanging protocol [2]; participants will be more efficient in their reading and will score higher on diagnostic performance of RC images compared to LE images [3]; participants using any CEM protocol will be more efficient and score higher on diagnostic performance compared to participants using conventional FFDM.

## 2. Material and methods

### 2.1. Participants

Breast radiologists, fellows in breast radiology and residents in an advanced rotation of breast radiology were eligible for participation in this experiment. To acquire a diverse set of radiologists, members of the European Society of Breast Imaging (EUSOBI) were invited by e-mail to participate. To accommodate participation of members from abroad, part of the data collection took place in a dedicated room on-campus during the 2018 congress of the European Society of Radiology (ESR).

Participants were randomly allocated to one of three experimental groups; the FFDM group, which evaluated the FFDM case only; the LE-RC group, which evaluated the LE image first, followed by the RC image of each CEM-case, similar to current used hanging protocols; and the RC-LE group, which evaluated the RC image first followed by the LE-image of each CEM-case (i.e., the 'reversed' hanging protocol).

### 2.2. Materials

#### 2.2.1. Images

For this study, the CEM-cases of 30 patients were used. As all patient cases were anonymized our certified ethical committee waived the need to obtain informed consent from patients. The images originated from our hospital's database consisting of CEM-cases acquired between 2012 and 2016 [4]. The CEM principle and its imaging protocol were described earlier [3,16]. In summary, a LE and a high-energy (HE) image are obtained of both breasts in the standard mediolateral oblique (MLO) and craniocaudal (CC) views two minutes after intravenous administration of 1.5 mL/kg body weight iodine-based contrast medium (Iopromide 300 mg/ml) with a flow rate of 3 mL/s followed by saline flush. The LE and HE image are recombined to create the RC image which visualizes areas of contrast uptake.

Due to the resolution of the non-diagnostic LCD screen (specifications regarding the reading set-up are delineated under *Apparatus*) the images of only the right or left breast were used in the experiment. LE images and RC images of a patient case were shown after each other during the experiment whereas sequence (RC-LE or LE-RC) varied as described before. A typical CEM exam used in the experiment is visualized in Fig. 1.

Out of all CEM-cases nine cases were selected that were considered typical malignant cases selected by a breast radiologist with four years of experience in CEM. Each malignant case contained one malignant lesion. All (histologically proven) malignancies were invasive

carcinomas of no special type (NST). The size of the abnormalities on the images ranged from 0.9 x 0.6 cm (31 x 21 pixels) to 1.8 x 1.4 cm (65 x 50 pixels). Additionally, 21 CEM-cases with only negative findings were selected. Images that contained benign lesions such as simple cysts or fibroadenomas, artefacts [17] or (micro) calcifications were excluded. The malignant-benign ratio of the selected cases was similar to the percentage malignancies of our CEM database, which is 28% [4]. Four cases had a breast density category A, twelve cases B, ten cases C and three cases were considered to have a breast density category D.

#### 2.2.2. Apparatus

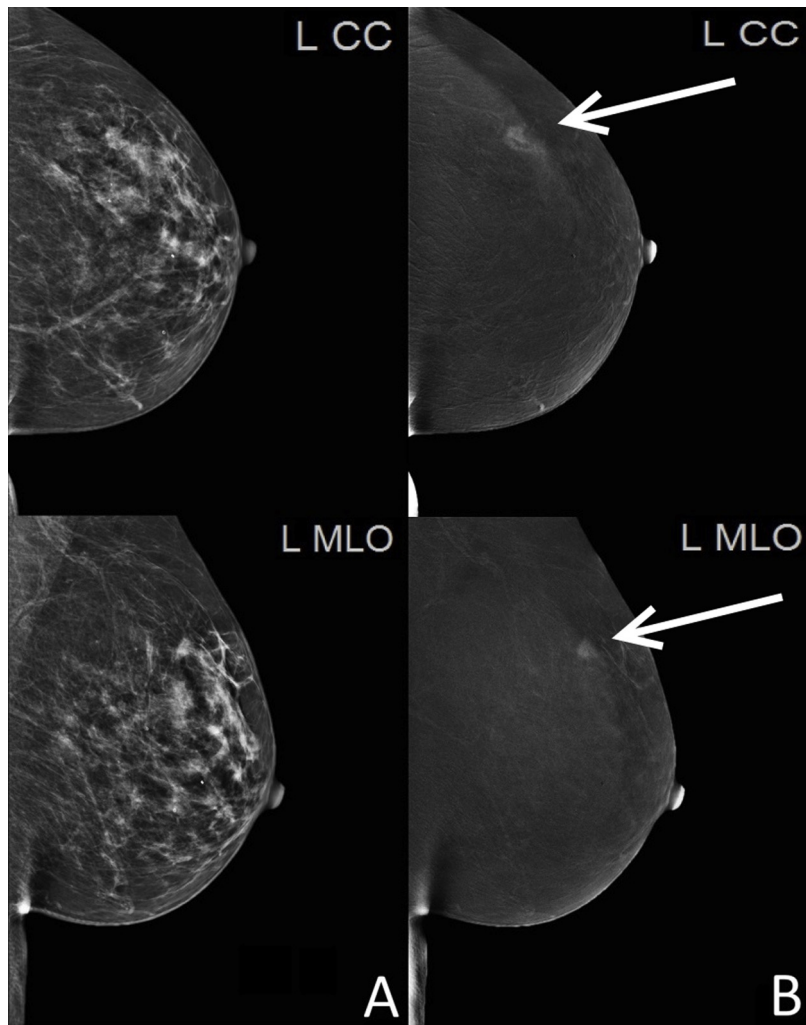
Eye movements were measured using a SensoMotoric Instruments (SMI, Teltow, Germany) 250 Hz remote eye tracker. Participants' head movements were not physically restricted, although they were instructed to avoid head movements as much as possible. As the right eye is generally dominant, the eye movements of the participants' right eyes were used [13]. The stimuli were shown on a Dell 22" liquid screen display with a resolution of 1080 x 1650 pixels in portrait set-up. The distance between the participant and monitor was approximately 70 centimeters and the visual angle  $\theta$  is therefore 41°. A saccade-based detection algorithm was used and the minimal fixation duration was set to 22 ms. The eye-tracker was calibrated using a five-point calibration prior to and halfway through the experiment. Calibration was repeated until a deviation smaller than 1° of visual angle on the x- and y-axis was obtained. Eye-tracking data of one participant was excluded from the analysis as the eye-tracking deviations were greater than 1.0° visual angle.

Participants used the mouse of the experiment computer to click on lesions suspicious for malignancy they identified on an image and the space bar to navigate to the next image or case. The experimental set-up is visualized in Fig. 2.

#### 2.3. Procedure

The experiment was carried out by each participant individually. Participants were first instructed that they were going to evaluate 30 patient cases of women recalled from a breast cancer screening program. They were instructed to search for malignant masses but were informed that not all images contained masses. Participants were instructed to click on all areas which they deemed malignant (BIRADS 4-5) on the MLO- as well as CC-part of every image; participants of the FFDM group were instructed to click on the FFDM image and participants of the CEM groups on the LE as well as the RC image. They were informed that the images would not contain any technical artefacts, architectural disturbances, micro calcifications or benign masses. When participants of the CEM groups finished the reading of the first MLO- and CC-images of a patient case (respectively LE images or RC images) they should press the space bar for the second MLO- and CC-images. It was not possible to return to the previous images of a patient case and to reevaluate the previous images.

After receiving the instructions participants subsequently wrote down their age, sex, hospital where they were employed, number of years licensed as radiologist, number of workdays per week working as a breast radiologist, fellow or resident and, if applicable, number of years of experience with evaluating CEM images. Participants then evaluated a practice case to check if all instructions were clear, followed by the five-point calibration procedure for the eye tracker after which they started with the reading of the patient cases. After 15 patient cases participants had a short break of two minutes followed by a recalibration before they continued with the last 15 patient cases. Participants received no feedback on their performance throughout the experiment.



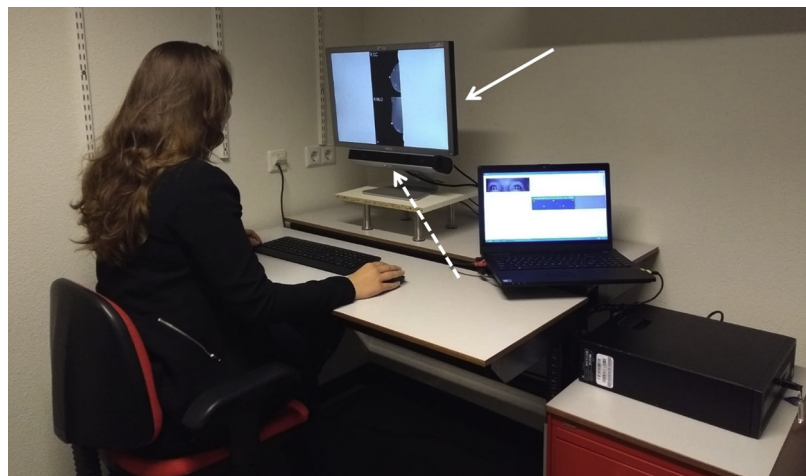
**Fig. 1.** Typical example of a CEM exam, with only the left breast shown. On the left (A), CC and MLO views of the low-energy (LE) images are shown, which are similar to conventional mammography. On the right (B), CC and MLO views of the recombined (RC) images are shown. An area of contrast uptake can be seen (arrows), suspicious for breast cancer. Biopsy results confirmed an invasive breast cancer at this site. A and B were shown after each other during the experiment.]

2.4. Statistical Analysis

2.4.1. Reading process measures

To investigate the reading process efficiency, the following eye-

tracking measures were used: Average fixation count, average net dwell time, and average reading time. Average fixation count was defined as the average number the participants' eyes stood still (fixated) on a malignant area, which were our areas of interest (AOIs). Average



**Fig. 2.** Picture of the experimental set-up. A participant is reading a CEM image on the stimulus monitor (arrow) while the eye movements are registered by the eye-tracker (dashed arrow).]

fixation duration was defined as the average time participants fixated on AOIs. Average net dwell time was defined as the total time participants fixated on the AOIs, averaged per case [13]. Furthermore, average reading time was measured per image (LE/FFDM and RC image) and per case (LE/FFDM and RC image reading time combined). Average reading time was defined as the total time that participants needed to evaluate a case.

#### 2.4.2. Diagnostic performance

Sensitivity and specificity were used as measures of diagnostic performance. Sensitivity was defined as the number of malignant areas a participant clicked on divided by the total number of nine malignant areas of the experiment. Specificity was defined as the number of images where a participant did *not* click on benign areas divided by the total number of 30 images of the experiment. The mouse clicks on the last image (LE or RC for the two CEM groups and FFDM for the FFDM group) of each of the 30 patient cases participants evaluated were used to calculate sensitivity and specificity. To analyze differences in diagnostic performance on separate LE and RC images of a patient case, sensitivity per image category and specificity per image category were calculated.

Furthermore, the participants worked in different hospitals and potentially had different criteria to call a lesion malignant. “Over”calling would result in high sensitivity yet low specificity. An aggregate measure of diagnostic performance was therefore necessary. Participants’ sensitivity and specificity were used to calculate individual receiver-operating characteristic (ROC) curves and participants’ value of area under the ROC curve (AUC) was used as an aggregate measure of diagnostic performance.

#### 2.4.3. Statistics

To compare the RC-LE group with the LE-RC group (hypothesis 1), independent *t*-tests were used with average fixation count, average net dwell time, average time to first fixation, and average reading time as dependent variables of the reading process and sensitivity, specificity, and AUC-value as dependent variables of diagnostic performance. To compare participants’ reading process (average fixation count, average net dwell time, and average reading time) on RC images and LE images of the same patient cases (hypothesis 2) paired sample *t*-tests were used. Finally, to compare the reading process (average net dwell time, average time to first fixation and average evaluation time) and diagnostic performance (sensitivity, specificity, and AUC-value) of the two CEM-groups versus the FFDM group (hypothesis 3) one-way ANOVAs were used with CEM-groups contrasted to the FFDM group. Data analysis was performed using SPSS (IBM SPSS Statistics for Windows, Version 24.0. Armonk, NY: IBM Corp. Released 2016).

### 3. Results

#### 3.1. Demography of participants

The participants ( $N = 27$ , mean age = 42.1 years,  $SD = 10.4$ , 67% female) originated from eighteen different hospitals in seven European countries (the Netherlands, Belgium, United Kingdom, France, Ireland, Italy, and Spain). Two participants were residents in an advanced rotation of breast radiology, the other 25 were radiologists with a mean career span of 10 years ( $SD = 9.0$ ). The participants worked as breast radiologists for an average of 3.9 days per week ( $SD = 1.2$ ). Nine participants also had prior experience with evaluating CEM (mean span = 3.15 years,  $SD = 1.62$ ). The three study groups did not significantly differ in any of the following demographic factors: age, gender, country, career span, experience with CEM, and average breast radiology working hours per week; in one-way ANOVAs comparing the three groups the highest *F*-value was  $F(2, 26) = 1.91$ ,  $p = .19$ .

The reading process and diagnostic performance measures per group can be found in Table 1. The results of the independent *t*-tests on

reading process and diagnostic performance measures can be found in Table 2.

#### 3.2. Influence of hanging protocol

The reading process and diagnostic performance measures of the two CEM groups are found in the first two columns of Table 1. The eye movement measures of the two CEM groups are generally similar. These findings indicate a similar reading process for the two CEM groups. Furthermore, the RC-LE group took 6.3 seconds to read the RC image and 7.3 to read the LE image while the LE-RC group respectively took an average of 5.8 and 14 seconds to read the RC and LE images, respectively. Finally, in the RC-LE group the diagnostic performance measures are equally high between the two CEM groups with an AUC value of .93 for the RC-LE group and .96 for the LE-RC group.

Furthermore, the results of the independent *t*-tests concerning hypothesis 1 are found in Table 2. The *t*-tests on the eye movement measures are all non-significant. Thus participants of both the RC-LE and LE-RC group had similar eye movement measures on the RC images and LE images. However, a significant effect of hanging protocol was found on reading time per case, indicating that participants of the RC-LE group on average needed 6.20 seconds less to evaluate a patient case. The participants of both groups needed a similar amount of time to evaluate the RC images. A significant effect of hanging protocol on reading time of the LE-images was found, indicating that participants of the RC-LE group needed less time to evaluate the LE-image compared to the LE-RC group. On overall, the *t*-tests on diagnostic performance were non-significant, indicating a similar sensitivity, specificity, and AUC-value for both the CEM-groups.

#### 3.3. Differences between RC images and LE images on reading process and diagnostic performance measures

The reading process and diagnostic performance measures of the RC and LE images are found in the first four columns of Table 1. The average fixation count is lower on the RC images; 3.24 and 2.75 fixations on the RC images and 3.58 and 4.15 fixations on the LE images. This difference is significant with a *p*-value of .003. The average net dwell time is also lower on RC images, 1280 and 1025 ms compared to 1301 and 1444 ms for the LE images. This difference is also significant,  $p = .029$ . Time to first fixation is lower on RC images, 773 and 1107 ms compared to 888 and 1295 ms for the LE images yet this difference is non-significant,  $p = .22$ . Furthermore, participants read RC images faster compared to LE-images,  $p < .001$ . Finally, participants scored higher on sensitivity on RC images, 99% and 98% compared to 100% and 80%,  $p = .004$  and higher on specificity on RC images, 90% and 94% compared to 85% and 80%,  $p = .019$  compared to LE-images. The complete results of the paired sample *t*-tests of RC compared to LE images are found in Table 3.

#### 3.4. Comparison of CEM-groups to FFDM-only group

For hypothesis 3, on the LE and FFDM-images no effect of group was found on fixation count,  $F(2,25) = 1.09$ ,  $p = .35$ , contrast  $t[23] = -1.39$ ,  $p = .18$ . A tendency towards lower net dwell time for the CEM-groups was found,  $F(2, 23) = 2.79$ ,  $p = .08$ , contrast  $t(8.74) = -1.86$ ,  $p = .10$ . A tendency towards higher total reading time was found for the CEM-groups,  $F(2,25) = 3.15$ ,  $p = .06$ ,  $t(24) = 1.44$ ,  $p = .16$ . On sensitivity a positive effect in favor of CEM-groups was found,  $F(2,26) = 17.2$ ,  $p < .001$ , contrast  $t(8.09) = .003$ . No effect was found on specificity,  $F(2,26) = 2.01$ ,  $p = .16$ , contrast  $t(24) = 1.65$ ,  $p = .11$ . Finally, a positive effect in favor of CEM-groups was found on AUC,  $F(2,26) = 16.1$ ,  $p < .001$ , contrast  $t(24) = 5.62$ ,  $p < .001$ . In summary, participants in the CEM-groups were not more efficient during the reading process but scored higher on diagnostic performance compared to FFDM-only.

**Table 1**  
Reading process and diagnostic performance measures per group and per image category.

		RC-LE (N = 10)		LE-RC (N = 9)		FFDM-ONLY (N = 8)	
		RC	LE	RC	LE	/	FFDM
		μ (SD)	μ (SD)	μ (SD)	μ (SD)	/	μ (SD)
Reading process measures	Unit						
Average fixation count	#	3.24 (1.24)	3.58 (1.74)	2.75 (1.74)	4.15 (2.23)		5.26 (2.94)
Average net dwell time	ms	1280 (356)	1301 (495)	1025 (502)	1444 (780)		2243 (1248)
Average time to first fixation	ms	773 (333)	888 (481)	1107 (635)	1295 (691)		1197 (580)
Average reading time per image	s	6.30 (2.44)	7.31 (4.07)	5.82 (2.76)	14.0 (4.26)		12.8 (7.19)
Diagnostic performance measures	Unit	μ (SD)	μ (SD)	μ (SD)	μ (SD)	/	μ (SD)
Sensitivity overall	%	100 (0)		98 (7.4)		79 (13)	
Sensitivity per image	%	99 (4.0)	100 (0)	98 (7.0)	80 (11)		
Specificity overall	%	87 (15)		94 (7.2)		80 (17)	
Specificity per image	%	90 (10)	87 (15)	94 (7.2)	80 (8.4)		
Area under the curve		.93 (.075)		.96 (.048)		.80 (.055)	

**Table 2**  
Independent t-tests on reading process and diagnostic performance measures with RC-LE versus LE-RC group as independent variables.

Reading process measure	t	df	p	Mean difference (95%-CI)
Fixation count (RC)	.68	16	.51	.48 (-1.02 - 1.99)
Fixation count (LE)	-.58	16	.57	-.56 (-2.61 - 1.49)
Net dwell time (RC)	1.24	16	.23	205 (-180 - 689)
Net dwell time (LE)	-.47	16	.65	308 (-796 - 509)
Time to first fixation (RC)	-1.40	16	.18	-334 (-841 - 173)
Time to first fixation (LE)	-1.45	16	.17	-407 (-1008 - 194)
Reading time per case	-2.20	16	.04	-6.2 (-12.2 - 2.65)
Reading time per image (RC)	.49	16	.63	5.98 (-2.00 - 3.20)
Reading time per image (LE)	-3.38	16	.004	-6.56 (-10.7 - 2.45)
Diagnostic performance measure	t	df	p	Mean difference (95%-CI)
Sensitivity overall	1.00	8.00	.35	.25 (-.032 - .082)
Sensitivity (RC)	.52	17	.61	.026 (-.042 - .069)
Sensitivity (LE)	5.49	8.00	.001	.20 (.13 - .27)
Specificity overall	-1.39	13.17	.20	.053 (-.19 - .041)
Specificity (RC)	-.94	17	.36	-.040 (-.12 - .046)
Specificity (LE)	1.24	17	.23	.056 (-.050 - .19)
Area under the curve	-.83	17	.42	.030 (-.086 - .037)

**Table 3**  
Paired sample t-tests of reading process and diagnostic performance measures on RC and LE images.

Variable	t	df	p	Mean difference (95%-CI)
Fixation count	-3.47	17	.003	-.87 (-1.40 to -.33)
Net dwell time	-2.38	17	.029	-220 (-415 to -25)
Time to first fixation	-1.27	17	.22	-152 (-405 - 101)
Reading time per image	-4.30	17	< .001	-4.60 (-6.86 to -2.34)
Sensitivity	3.32	18	.004	.088 (.032 - .14)
Specificity	2.58	18	.019	.076 (.014 - .14)

**4. Discussion**

In general, dedicated CEM workstations are configured with a LE-RC hanging protocol [18]. Hence, radiologists view the exam by first starting with the LE images, evaluating them as regular FFDM images [5]. The RC image is then used to check whether lesions (which were observed on the LE image) enhance or not. Radiologists with experience in the reading of CEM-cases reported that lesions are more salient on RC images than on the LE images. By reversing the current hanging protocol the radiologist’s attention could be immediately drawn to conspicuous areas. Radiologists can therefore find lesions up to 31% (6 seconds) faster and can potentially reach higher diagnostic performance if they would evaluate the RC image prior to the LE image. An average decrease in case reading time of 6 seconds may sound small. However, the difference will add up as radiologists may read perhaps tens or even hundreds of cases on a daily basis. Moreover, these differences in reading times were found in radiologists who are experts in terms of

accuracy as well as speed. The differences may even be more pronounced in less experienced radiologists.

For this study, three hypothesis were tested: [1] participants of a (reversed) RC-LE hanging protocol will be more efficient and will score higher on diagnostic performance in the reading of cases compared to participants using the (regular) LE-RC hanging protocol [2]; participants will be more efficient in their reading and will score higher on diagnostic performance of RC images compared to LE images [3]; participants of using any CEM protocol will be more efficient and score higher on diagnostic performance compared to participants using conventional FFDM.

With respect to the order of hanging protocols no difference was found on eye movement measurements nor on diagnostic performance. Nevertheless, a difference on average reading time per case was observed. Consequently, hypothesis [1] was not supported. We assumed that the higher saliency of malignancies on RC images would direct the radiologists’ attention towards these areas faster, which would be reflected by more efficient eye movements. However, eye movements in our study proved to be similar in both groups. While eye movements can provide invaluable information about visual search and attention, it does not provide definite answers on how abnormalities are interpreted [19,20].

The interpretation process after the detection of an abnormality is to some extent reflected by the average reading time [13]. The difference found in average reading time between the two protocols was mainly caused by the lower average reading time of the LE images, since the reading times of RC images did not differ between the two protocols. In addition, participants of the RC-LE protocol showed identical diagnostic performance compared to the LE-RC protocol. The combination of

shorter reading time and equally high diagnostic performance indicates that abnormalities were interpreted more easily by radiologists using the RC-LE protocol compared to the LE-RC protocol.

Second, it was found that participants of CEM-groups fixated less often and shorter on malignancies on RC images compared to LE images. Hypothesis 2 was therefore supported. Less fixations and less time needed for fixating on abnormalities to evaluate a patient case can indicate that the interpretation process of RC images was more efficient [13]. Participants did not fixate earlier on malignant lesions on RC images compared to LE images as time to first fixation did not differ. However, considering the substantial standard deviations, a potential yet small effect between the two image categories may not have been discernible. Furthermore, participants scored higher on diagnostic performance on RC images compared to LE images. However, this observation is less relevant, because in clinical practice RC and LE images are evaluated together.

Third, it was found that the reading process on LE images was similar to the reading process of FFDM-images, confirming findings of previous studies that they are diagnostically equal [5,6]. In line with many previous studies, our study also showed that diagnostic performance was superior in CEM groups when compared to conventional FFDM only [1,2,5,21]. Hypothesis 3 was therefore partly supported. LE images are comparable to FFDM images [5] and this similarity presumably caused a similar reading strategy.

The findings of this study may have clinical implications. In the imaging community, there is some concern about the increased reading time of CEM exams, as it consists of double the number of images per patients. This might result in substantial increases in workload for radiology staff. Also, Lebron-Zapata et al. [22] showed that CEM might even be considered as screening tool for women with high risk for developing breast cancer. When high volumes of CEM exams are produced, such as screening settings, an increase in reading time is not desirable. This investigation shows that these arguments may be less of a concern as long as a RC-LE hanging protocol is used. Moreover, the results indicate that sensitivity could increase with 25% (from 79% to 99%) and specificity with 13% (from 80% to 90%) with similar reading times, when an RC-LE hanging protocol is adopted in a screening setting instead of the current FFDM standard.

To the best of our knowledge, this is the first study on the effects of hanging protocol modification on radiologists' reading process and diagnostic performance. It is shown that the time to evaluate a case is influenced and is one-third shorter for a particular sequence, while diagnostic performance was not influenced. Modification of hanging protocols could therefore impact radiologists' workflow. In most Picture Archiving and Communication Systems (PACS) it is possible to modify hanging protocols, yet it is unknown if radiologists do this and what the effects are. More research on the influence of modification of hanging protocols in different radiologic examinations is therefore advised.

Perhaps slightly counterintuitive, the findings of the first two research questions indicate that the main value of the RC-image may not lie in the *detection* of abnormalities, but more in the *interpretation* of abnormalities. Considering that radiologists in this research needed less fixations and less time to evaluate RC images while scoring higher on diagnostic performance the RC images can be considered the less complex component of CEM. For training purposes, it is generally recommended to start with less complex material and to gradually increase the complexity as the learner advances [23,24]. CEM cases in which the RC image is shown prior to the LE image may help (breast) radiologists in training to learn to read mammograms. In the same manner, it is also recommended to start with RC images followed by LE images when learning to read CEM exams.

This study has some limitations. First, due the small sample size, some small effects may not have been detectable, such as a potential effect on average time to first fixation. In eye-tracking research generally groups with a different level of expertise such as novices and

experts are compared to each other [25]. Eye-tracking studies like ours with groups of a comparable expertise level are scarce [19,25,26]. Difference in eye movement measures between groups of a comparable expertise level might be smaller than differences between groups of large difference in expertise [26]. Moreover, in eye-tracking research sample sizes are generally small [19,25] and many efforts have been taken to create a sample size as large as possible.

Second, participants were asked to only look for suspicious lesions, while the task in the clinical workplace is much broader than that, in example, the additional need for detecting suspicious calcifications. However, previous studies have shown that the added value of CEM in the reading of calcifications is limited [27]. Although this limitation may have biased our results, it is similar in all different groups.

Another limitation concerns suspicious lesions which do not enhance on contrast-enhanced images, such as mucinous carcinoma or microcalcifications [5,27,28]. Users of CEM should always be aware of other lesions that do not enhance, and check extra for these. However, it could be the case that readers overlook not enhancing findings. Cases containing such lesions were not used in this experiment. The impact of such lesions on the reading process and diagnostic performance of radiologists using a RC-LE or LE-RC hanging protocol can therefore not be deciphered with this experiment. A follow-up study with such suspicious yet unenhancing lesions is warranted.

Finally, in this experiment participants could not go back to a previous image of a patient case as this could blur the findings of the investigation on the order of the hanging protocol. In the clinical workplace, however, radiologists are able to alternate between LE and RC image as often as they like. Therefore, a follow-up study in a more clinical and ecologically valid setting is warranted.

## 5. Conclusion

Reversal of a CEM hanging protocol from the commonly used LE-RC order to the RC-LE order lowers the case reading time, while diagnostic performance is maintained for breast cancer detection. Furthermore, the reading process is more efficient. Like other studies, we showed that diagnostic accuracy of CEM is superior to FFDM. Based on our observations, we would recommend to use a RC-LE hanging protocol in everyday clinical practice, but also in training.

## Author Notes

Koos van Geel is a Doctor of Medicine and PhD candidate; Ellen Kok, is a PhD and post-doctoral researcher; Jorian Krol is a Doctor of Medicine and resident in radiology; Ivo Houben is a Doctor of Medicine and PhD candidate; Fabienne Thibault is a Doctor of Medicine, PhD and radiologist; Ruud Pijnappel is a Doctor of Medicine, full professor, EBBI and radiologist; Jeroen van Merriënboer is a full professor; Marc Lobbes is a Doctor of Medicine, PhD, EBBI and radiologist.

## Conflicts of interest

We have one source of funding to declare: General Electric Company, Healthcare division (Chicago, Illinois, USA) reimbursed the costs to rent a meeting room on the venue of the European Radiology Conference of 2018. This meeting room was necessary for our data acquisition. The research plan, data acquisition, data analysis, data interpretation and dissemination of our manuscript was not influenced by General Electric Company and was unconditionally the sole responsibility of the authors. The authors have no other sources of funding to declare.

## References

- [1] B.K. Patel, M.B.I. Lobbes, J. Lewin, Contrast Enhanced Spectral Mammography: A Review, Seminars in ultrasound, CT, and MR. 39 (1) (2018) 70–79.

- [2] A.S. Tagliafico, B. Bignotti, F. Rossi, et al., Diagnostic performance of contrast-enhanced spectral mammography: Systematic review and meta-analysis, *Breast (Edinburgh, Scotland)*. 28 (2016) 13–19.
- [3] I.P. Houben, P. Van de Voorde, C.R. Jeukens, et al., Contrast-enhanced spectral mammography as work-up tool in patients recalled from breast cancer screening has low risks and might hold clinical benefits, *Eur J Radiol*. 94 (2017) 31–37.
- [4] U.C. Lalji, I.P. Houben, R. Prevos, et al., Contrast-enhanced spectral mammography in recalls from the Dutch breast cancer screening program: validation of results in a large multireader, multicase study, *European radiology*. 26 (12) (2016) 4371–4379.
- [5] U.C. Lalji, C.R. Jeukens, I.P. Houben, et al., Evaluation of low-energy contrast-enhanced spectral mammography images by comparing them to full-field digital mammography using EUREF image quality criteria, *European radiology* 25 (10) (2015) 2813–2820.
- [6] M.A. Francescone, M.S. Jochelson, D.D. Dershaw, et al., Low energy mammogram obtained in contrast-enhanced digital mammography (CEDM) is comparable to routine full-field digital mammography (FFDM), *Eur J Radiol*. 83 (8) (2014) 1350–1355.
- [7] M. Lobbes, *Comparison Between Breast MRI and Contrast-Enhanced Digital Mammography*, (2018), pp. 47–56.
- [8] S. Jochelson M, K. Pinker, D. David Dershaw, et al., Comparison of screening CEDM and MRI for women at increased risk for breast cancer: A pilot study, *Eur J Radiol*. 97 (December) (2017) 37–43.
- [9] E.M. Fallenberg, F.F. Schmitzberger, H. Amer, et al., Contrast-enhanced spectral mammography vs. mammography and MRI - clinical performance in a multi-reader evaluation, *European radiology* 27 (7) (2017) 2752–2764.
- [10] V. Sorin, Y. Yagil, A. Yosepovich, et al., Contrast-Enhanced Spectral Mammography in Women With Intermediate Breast Cancer Risk and Dense Breasts, *AJR Am J Roentgenol*. (2018) W1–w8.
- [11] N. Nishikawa, K. Yanagisawa, K. Naoi, et al., Possibility of Exposure Dose Reduction in Contrast Enhanced Spectral Mammography Using Dual Energy Subtraction Technique : A Phantom Study, Springer International Publishing, Cham, 2014.
- [12] M.K. Markey, *Physics of mammographic imaging*, CRC press, 2012.
- [13] K. Holmqvist, M. Nyström, R. Andersson, et al., *Eye Tracking: A Comprehensive Guide to Methods and Measures*, Oxford University Press, 2011.
- [14] H. Jarodzka, K. Holmqvist, H. Gruber, *Eye tracking in Educational Science: Theoretical frameworks and research agendas*, *Journal of Eye Movement Research*. 10 (1) (2017).
- [15] E.M. Kok, H. Jarodzka, Before your very eyes: the value and limitations of eye tracking in medical education, *Med Educ*. 51 (1) (2017) 114–122.
- [16] C. Bhimani, D. Matta, R.G. Roth, et al., Contrast-enhanced Spectral Mammography: Technique, Indications, and Clinical Applications, *Academic radiology*. 24 (1) (2017) 84–88.
- [17] Y. Yagil, A. Shalmon, A. Rundstein, et al., Challenges in contrast-enhanced spectral mammography interpretation: artefacts lexicon, *Clin Radiol*. 71 (5) (2016) 450–457.
- [18] A.-K. Carton, S. Saab-Puong, M. Suminski, *SenoBright Contrast Enhanced Spectral Mammography Technology*, General Electric Company, Wauwatosa, WI, 2012, p. 7.
- [19] A. Gegenfurtner, E. Kok, K. van Geel, et al., The challenges of studying visual expertise in medical image diagnosis, *Med Educ*. 51 (1) (2017) 97–104.
- [20] E.M. Kok, K. van Geel, J.J.G. van Merriënboer, et al., What We Do and Do Not Know about Teaching Medical Image Interpretation, *Frontiers in Psychology*. 8 (309) (2017).
- [21] X. Zhu, J.M. Huang, K. Zhang, et al., Diagnostic Value of Contrast-Enhanced Spectral Mammography for Screening Breast Cancer: Systematic Review and Meta-analysis, *Clinical breast cancer* (2018).
- [22] L. Lebron-Zapata, M.S. Jochelson, Overview of Breast Cancer Screening and Diagnosis, *PET clinics* 13 (3) (2018) 301–323.
- [23] J.J. Van Merriënboer, R.E. Clark, M.B. De Croock, Blueprints for complex learning: The 4C/ID-model, *Educational technology research and development* 50 (2) (2002) 39–61.
- [24] J.J. van Merriënboer, J. Sweller, Cognitive load theory and complex learning: Recent developments and future directions, *Educational Psychology Review*. 17 (2) (2005) 147–177.
- [25] A. Gegenfurtner, E. Lehtinen, R. Säljö, Expertise differences in the comprehension of visualizations: a meta-analysis of eye-tracking research in professional domains, *Educational Psychology Review*. (2011) 1–30.
- [26] K. van Geel, E.M. Kok, J. Dijkstra, et al., Teaching Systematic Viewing to Final-Year Medical Students Improves Systematicity but Not Coverage or Detection of Radiologic Abnormalities, *Journal of the American College of Radiology : JACR*. 14 (2) (2017) 235–241.
- [27] Y.C. Cheung, Y.H. Juan, Y.C. Lin, et al., Dual-Energy Contrast-Enhanced Spectral Mammography: Enhancement Analysis on BI-RADS 4 Non-Mass Microcalcifications in Screened Women, *PloS one* 11 (9) (2016) e0162740.
- [28] I.P. Houben, S. Vanwetswinkel, V. Kalia, et al., Contrast-enhanced spectral mammography in the evaluation of breast suspicious calcifications: diagnostic accuracy and impact on surgical management, *Acta Radiologica: Acta Radiol* 24 (2019) 284185118822639.