DOCTOR OF MEDICINE

Contrast Enhanced Spectral Mammography

Exploring applications for monitoring response to neoadjuvant chemotherapy and textural analysis of images

Savaridas, Sarah Louise

Award date:
2022

Contrast Enhanced Spectral Mammography: Exploring applications for monitoring response to neoadjuvant chemotherapy and textural analysis of images

Sarah Louise Savaridas MBChB, FRCR
Submitted for the degree of Doctor of Medicine by Research

University of Dundee
March 2022

Table of Contents

Figures

Tables

Abbreviations

ABUS	Automated breast ultrasound
ANC	Axillary node clearance
ANN	Artificial neural network
ANOVA	One-way analysis of variance
AUC	Area under the curve
AUTO	Auto-regressive
CAD	Computer-aided diagnosis
CC	Cranio-caudal
CE-DBT	Contrast enhanced digital breast tomosynthesis (CESM with DBT)
CE-MRI	Contrast enhanced magnetic resonance imaging
CESM	Contrast enhanced spectral mammography
CESM(CE)	Contrast enhancement
CESM(CE+calc)	Contrast enhancement plus microcalcification
CET	Contrast-enhanced tomosynthesis
CEUS	Contrast enhanced ultrasound sound
CI	Confidence interval
CI	Confidence interva/
CR	Complete response
CT	Computed tomography
DBT	Digital breast tomosynthesis
DCIS	Ductal carcinoma in situ
DICOM	Digital imaging and communications in medicine
DOST	Discrete orthonormal Stockwell transform
Ellip	Ellipsoid
ER	Oestrogen receptor
FFDM	Full field digital mammography
FH	Freehand
FN	False negative
FP	False positive
GFB	Gabor filter bank
GLCM	Grey length co-occurrence matrix
GLDM	Grey level dependence matrix
GLSZM	Grey level size zone matrix
GRAD	Absolute gradient
HER2	Human epidermal growth factor receptor 2
HG	High grade
HIST	First order histogram
HR	Hormone receptor
IDC	Invasive ductal carcinoma
ILC	Invasive lobular carcinoma
ITC	Isolated tumour cells
ITS	Invasive tumour size
IV	Intravenous
LCIS	Lobular carcinoma in situ
kV	kiloVoltage
LE	Low-energy

LG	Low grade
MDT	Multi-disciplinary team
MeSH	Medical subject headings
MLO	Mediolateral oblique
NACT	Neoadjuvant chemotherapy
NPV	Negative predictive value
NST	No special type
pCR	Pathological complete response
PIS	Patient information sheet
PPV	Positive predictive value
PR	Progesterone receptor
RC	Recombined subtracted image
RCB	Residual cancer burden
RD	Relative enhancement difference
RF	Radiofrequency
RLE	Relative lesion enhancement
RLM	Run length matrix
ROC	Receiver operating characteristic
ROI	Region of interest
SD	Standard deviation
SRD	Subtracted relative enhancement difference
TA	Textural analysis
TDLU	Terminal ductal lobular unit
TIC	Time-signal intensity curves
TN	True negative
TP	True positive
US	Ultrasound
VAS	Visual analogue scale
WAV	Haar wavelets
WI	Washout index
WTS	Whole tumour size

Acknowledgements

There are many people who have helped me to complete this work. I would like to thank my supervisor, Professor Andy Evans for his guidance and support. I am also grateful to members of the Breast Imaging Research Group, both past and present, Ms Patsy Whelehan, Dr Violet Warwick, Dr Kulsam Ali and Dr Sarah Vinnicombe, for their patience and knowledge.

I would like to thank the whole breast imaging team, for their enthusiasm to learn a new technique, and pathologists Dr Colin Purdie and Dr Dawn Fleming for their work calculating RCB scores, despite significant workload pressures.

I wish to acknowledge collaborators from other institutions, Dr Sarah Tennant for tirelessly collating images from Nottingham and Prof. Colin McCowan, Dr Utkarsh Agrawal and Dr Adeniyi Fagbamigbe for their assistance building classification models.

I recognise the all the patients who participated in CONDOR, and the funding bodies that supported my work: TENOVUS, the British Society of Breast Radiologist, Symposium Mammographicum and the Mackintosh Foundation.

Finally, I am enormously grateful to my family for their support over the last four years. Without them this would not have been possible.

Publications arising from this work

Savaridas, S.L. Misinterpretation of raw data: fundamental flaws in 'The diagnostic performance of CESM and CE-MRI in evaluating the pathological response to neoadjuvant therapy in breast cancer: a systematic review and meta-analysis'**.** BJR. Published Online: 2021 https://doi.org/10.1259/bjr.20210741

Savaridas, S.L. Tennant, S.L. Quantifying lesion enhancement on Contrast Enhanced Mammography: A review of published data. Clinical Radiology. Jan 2022: https://doi.org/10.1016/j.crad.2021.12.010

Savaridas, S.L. Whelehan, P. Warwick, V. Vinnicombe, S.J. Evans, A.J. Contrast-enhanced digital breast tomosythesis and breast MRI to monitor response to neoadjuvant chemotherapy: Patient tolerance and preference. BJR. Published online: 2022 http://doi.org/10.1259/bjr.20210779

Declaration

I declare that I, Sarah Louise Savaridas, am the author of this thesis. I have consulted all references cited. The work of which the thesis forms a record has been done by myself, except for the generation of machine learning models the details of which are clearly stated in section 5.4 of the methods chapter. This work has not previously been accepted for a higher degree.

Sarah L Savaridas

Summary of Contents

Introduction: Contrast-enhanced digital breast tomosynthesis (CE-DBT) is a novel imaging technique, combining contrast enhanced mammography (CESM) with tomosynthesis, which may offer an alternative to MRI for monitoring response to neoadjuvant chemotherapy (NACT). Textural analysis (TA) may increase CESM accuracy, generating quantitative functional data.

Methods: CONDOR was a prospective feasibility study in which women undergoing NACT were imaged with CE-DBT and MRI. Patient experience was investigated using questionnaires. The separate and combined components of CE-DBT were assessed; ie. Low energy (LE) mammogram, tomosynthesis (DBT), CESM contrast-enhancement (CESM(CE)), combined CESM contrast and microcalcifications seen on LE (CESM(CE+calc) and the fully combined CE-DBT. Imaging techniques were directly compared for accuracy identifying multifocality, complete pathological response (pCR), residual whole tumour size (WTS) and invasive tumour size (ITS). Additional patients' CESM images were acquired using either Hologic or GE equipment. Lesions were segmented with freehand-ROI and ellipsoid-ROI on craniocaudal (CC) and mediolateral oblique (MLO) views and textural features were extracted. Machine learning was used to build models for the classification of lesions according to malignancy, tumour grade, receptor status and pCR-status. CESM-washout features were calculated by comparing enhancement on initial and delayed-MLO views. These were correlated to MRI-curves.

Results: Eighteen patients with 24 cancers were enrolled. Response prediction accuracy for pCR was 81.25% (95%CI: 54.35-95.95) for CESM(CE) vs 62.5% (95%CI: 35.43-84.80) for MRI; ITS concordance coefficient was 0.70 for CESM(CE) vs 0.66 for MRI; WTS concordance coefficient was 0.69 for CESM(CE+calc) vs 0.87 for MRI. Accuracy for identifying multifocality was 94.6% (95%CI:81.8-99.3) for CESM(CE) vs 89.2%(95%CI:74.6-97.0) for MRI. No benefit from the addition of DBT was identified for any measure. On 77% of occasions patients preferred CE-DBT, with significantly better overall experience (p=0.008).

All TA models for classifying malignant lesions were highly accurate. The Ellipsoid_ROI model performed better than FH_ROI; sensitivity 0.998 vs 0.953,

specificity 0.916 vs 0.891, $p<0.05$. Both-view model produced most consistently good results, followed by CC-view. There was no difference in AUC; 0.987 vs 0.988 vs 0.985 for both-view vs CC-view vs MLO-view. Promising results were seen for prediction of ER-status. Model accuracy was poorer for prediction of tumour grade, PR-status, HER-status and pCR status, potentially due to small subset numbers. CESM-washout features generated from ellipsoid_ROI were significantly different between the MRI curve types.

Conclusion: I have demonstrated that CESM has similar accuracy to MRI for identifying multifocality and predicting response to NACT, and is preferred by patients. The combination of CESM and low-energy mammogram – CESM(CE+calc) – increases the sensitivity for residual *in situ* disease. No additional benefit was afforded by combining CESM with tomosynthesis.

I have demonstrated the first multivendor CESM classification model using textural analysis data to identify malignant lesions. Promising results were seen for prediction of ER-status. With larger datasets this may translate to more accurate pCR classification models. Successful quantification of CESM-washout was achieved with significant differences between MRI curve types. I recommend further work utilises segmentation Ellipsoid_ROIs on single-view CESM images.

Chapter 1. Introduction

Breast cancer remains the commonest malignancy in women in Scotland, accounting for 29% of all female cancers in 2015, and 13.5% of female cancer deaths in 2016.[1] An understanding of the underlying pathogenesis is critical for early diagnosis and distinguishing cancer from benign disease, to allow best possible treatment outcome.

1.1 Histopathology

Breast cancer is a heterogeneous group of malignant diseases of the breast, primarily arising from the terminal ductal lobular unit (TDLU) of the breast. This encompasses pre-invasive *in situ* disease to invasive cancer with varying hormone receptor profiles and histological grades.

1.1.1 Ductal carcinoma *in situ* (DCIS)

Ductal carcinoma in situ (DCIS) is a monoclonal proliferation of epithelial cells within the breast ducts. Whilst these cells have the cytological features of malignancy within the breast ducts, there is no stromal invasion across the basement membrane. In the majority of cases it involves a single duct system and is most frequently considered to arise from the TDLU. DCIS is considered a non-obligate precursor of invasive disease with variable malignant potential according to the histological grade. This grading system is based on the degree of cellular atypia, with high grade DCIS having the greatest invasive potential and therefore the poorest prognosis if left untreated. In high grade DCIS cells are pleomorphic and irregularly spaced. The nuclei are usually large, approximately three times size of an erythrocyte, with variation in size and irregular contours. Mitoses are frequent. It is often a solid lesion with comedo-type central necrosis and polarisation of cells is rare. By contrast, in low grade DCIS cells are monomorphic, evenly spaced with rounded, centrally-placed small nuclei (1-2 x erythrocyte). A micropapillary or cribriform pattern is common with polarisation of cells covering micropapillae or lining the intercellular lumina. Intermediate grade DCIS cannot readily be assigned to high or low grade; it

demonstrates moderate nuclear pleomorphism with variable growth pattern and usually some degree of cell polarisation.[2 3]

1.1.2 Invasive carcinoma

By definition, for a malignant lesion to be classified as invasive, there is invasion of malignant cells across the basement membrane. There are a number of morphologically distinct invasive cancers. Of these, the most common is **invasive carcinoma of no special type (NST)**, frequently described as **invasive ductal carcinoma (IDC).** Up to 75% of invasive breast tumours fall into this category which demonstrates less than 50% of special type morphology. **Invasive lobular carcinoma** (**ILC)** is the second most common histological subtype of invasive breast cancer, accounting for 5-10% of all breast cancer cases. The key histological feature of ILC is the loss of E-cadherin expression. E-cadherin is a calcium-dependent transmembrane protein that maintains tissue integrity, cell-to-cell adhesion and prevents tissue invasion. Thus, the absence of e-cadherin results in increased capacity for tissue invasion. Lobular tumours tend to be larger with ill-defined margins and multicentricity and bilaterality is more common.[4] These tumours typically show oestrogen and progesterone receptor (ER/PR) positivity and lack HER2 receptors.[5]

The remainder of invasive tumours is a diverse group of rare sub-types including **tubular carcinoma**, **medullary carcinoma, invasive cribriform carcinoma, mucinous carcinoma and papillary carcinoma.** These tend to be less aggressive , and therefore have a better prognosis that IDC.[6]

1.1.3 Histological tumour grade

Tumour grade is recognised as an important prognostic factor and is used in treatment planning, such as the decision to offer the patient neoadjuvant chemotherapy (NACT), chemotherapy prior to definitive surgery. The method used by pathologists to assess grade was originally described by Bloom and Richardson in 1957[7] and subsequently modified into the Nottingham Histologic Score System which is in routine use today.[8] This method allows objective scoring of three tumour characteristics: tubule formation/glandular differentiation, nuclear features (atypia/pleomorphism) and mitotic activity. Each characteristic is scored on a three-

point scale, and the sum of all characteristics provides the overall grade; scores 3-5 equate to Grade 1, scores 6-7 to Grade 2 and the highest scores 8-9 to Grade 3. Provisional histological grade is increasingly calculated using tissue from diagnostic core biopsy. This is particularly important when alternatives to primary surgery, such as neoadjuvant chemotherapy are being considered. Meta-analysis data of concordance between histological grade calculated on core biopsy and subsequent surgical excision specimen shows moderate correlation (pooled agreement: 71.1, kappa: 0.54), with core grade under-estimating the true grade substantially more frequently.[9] This discrepancy is primarily due to tumour heterogeneity and the fact that a core biopsy samples a relatively small proportion of the overall tumour.

1.1.4 Oestrogen and progesterone receptors

Oestrogen receptor (ER) is an oestrogen-activated nuclear transcription factor that regulates development and proliferation of breast tissue.[10] Establishing ER status for invasive tumours is essential, and is used to determine whether a patient will benefit from endocrine therapy.[11 12] The significance of progesterone receptor (PR) status is more controversial; although there is evidence that in combination with ER positivity, PR positivity confers a better outcome and likelihood of response to hormone therapy than ER positivity alone.[13]

Hormone receptor status is assessed using immunohistochemistry as the method of choice.[2 12] The most widely used method (the Allred score) is based on the sum of the assessment of intensity (0-3) and proportion staining (0-5) with the cut-off for positivity of >/=3. Studies indicate that oestrogen receptor status of invasive carcinomas can be reliably assessed on core biopsy, although may be less reliable for progesterone receptor status.[14-16]

1.1.4 Human epidermal growth factor receptor 2 (HER2)

The human epidermal growth factor receptor-2 (HER2) is a tyrosine kinase receptor and plays a key role in one of the most important growth factor pathways in breast cancer. Over-expression of the HER2 protein is associated with aggressive histological features and poor prognosis.[17 18] However, anti-HER2 targeted therapy confers benefits in neoadjuvant, adjuvant and metastatic settings.[17 19-21] Thus, establishing the HER2 receptor status is both essential for prognostication and

treatment decisions and can be accurately assessed on core biopsy specimens.[15 16]. Her2 status is routinely assessed using immunohistochemistry; a score of 0/1 is considered negative, a score 2+ is equivocal and a score of 3+ is positive. Fluorescent *in situ* hybridization (FISH) is performed when initial analysis is equivocal.

1.1.5 Classification of breast cancer

More recently, gene profiling has been used to classify breast cancer. Initially the following three sub-types were identified; luminal-type, basal-type and HER2 positive.[22] Subsequent work demonstrated that the luminal-type could be further divided into luminal A, B and C subgroups. Whilst Luminal-C is not a routinely used subtype; luminal-A and luminal-B are accepted subtypes.[23] As it is not practicable to perform gene profiling on all breast cancers in routine clinical practice, broadly equivalent subtypes can be identified according to receptor status as demonstrated in the following table.[24]

SUBTYPE	ER	PR	HER 2
LUMINAL A	Positive	Positive	Negative
LUMINAL B	Positive	Negative	Negative
HER2 +	Negative	Negative	Positive
BASAL	Negative	Negative	Negative

Table 1: Subtypes with corresponding recep tor status

1.1.6 Lymph node status

The presence, or absence, of lymph node metastases is an important prognosticator and as such effects the stage of disease. Nodal disease is classified as isolated tumour cells (ITC), micrometastases and macrometastases. Both micro- and macrometastases are considered to be involved nodes; a cluster of tumour cells measuring 0.2-2mm is described as a micrometastasis and a cluster of over 2mm, a macrometastasis. ITC is the presence of single tumour cells or small clusters of

tumour cells <0.2mm in diameter and is not considered a positive node for staging purposes.

Typically, lymph nodes that are considered suspicious on imaging due to the thickness of the cortex are biopsied. The majority of patients with pre-operatively proven lymph node metastases will proceed to axillary node clearance (ANC). However, for patients receiving neoadjuvant chemotherapy (NACT) there an expanding body of evidence that when complete nodal response is seen on imaging an extended sentinel node biopsy (to remove at least three lymph nodes), is safe and confers a reduced morbidity.[25] It is important to note that in the neoadjuvant setting, micrometastases and ITCs may represent larger metastases that have responded to neoadjuvant therapy, and are therefore are considered positive. Nodes previously containing tumour often show fibrosis and scarring which may be useful for prognostication and treatment planning.

1.1.7 Breast cancer staging staging[26]

Breast cancer staging incorporates measures of disease involvement in the breast (primary tumour), regional lymph nodes and distant metastases. Two grading systems may be used; TNM staging as explained in table 2,[26] and a numeric systemic explained in table 3. As patients with stage 3 disease have a significantly high likelihood of distant metastases, cross-sectional imaging staging is recommended.

	Stage	Definition
Primary tumour (T)	T0	No evidence of primary tumour
	Tis	DCIS and Paget's disease
	T1	Tumour ≤ 20mm
	T2	Tumour > 20mm ≤ 50mm
	T3	Tumour > 50mm
	T4	Direct invasion to chest wall and/or skin
Lymph node (N)	N0	No regional lymph node metastasis / ITC only
	N1	Micrometastases or metastases in 1-3 axillary lymph nodes
	N2	Metastases in 4-9 axillary lymph nodes, or ipsilateral internal mammary lymph nodes in the absence of axillary lymph node metastases
	N3	Metastases in ≥ 10 axillary / infraclavicular lymph nodes; or metastases in ipsilateral internal mammary chain nodes + axillary nodes; or metastases in ipsilateral supraclavicular lymph nodes
Distant metastases (M)	M0	No clinical or radiological evidence of distant metastases

	M1	Metastases in distant organs or non-regional nodes

Table 2: Summary of TNM staging

Stage 1	Tumour ≤ 20mm +/- axillary micrometastases
Stage 2	Tumour ≤ 20mm ≤ 50mm + 1-3 axillary or IMN metastases; or Tumour ≤ 50mm without lymph node metastases
Stage 3	Tumour > 50mm; or Metastases in ≥ 4 axillary or IMN lymph nodes Local invasion of skin or chest wall, including inflammatory cancer
Stage 4	Distant metastases

Table 3: Numeric system of breast cancer staging

1.1.8 Benign disease

The priority of breast imaging and histological assessment is to identify patients with malignant lesions or those with malignant potential. However, differentiating solid benign lesions from malignant disease is frequently challenging on imaging alone, thus biopsy is usually necessary for histological confirmation. Exceptions to this include classical appearance of a fibroadenoma in a young patient and typical appearance of fat necrosis with a clear history of trauma.[27] Common solid benign breast lesions include; fibroepithelial tumours (fibroadenomas, phyllodes tumours), intra-ductal papillomas, hamartomas and fat necrosis.

Fibroadenomas are benign lesions composed of a mixture of epithelium and connective tissue. Chronic lesions may demonstrate calcification and can present radiologically as areas of indeterminate microcalcification. **Phyllodes tumours** range from benign through borderline to malignant lesions. Benign phyllodes have more cellular stroma than fibroadenomas and are often larger in size. However, distinguishing the two lesion types on core biopsy alone can be challenging as the characteristic stromal features may only be present in parts of the lesion, therefore excisional biopsy is performed if there is any doubt to the diagnosis. Borderline lesions may be locally infiltrative with increasing numbers of mitoses present, and mild to moderate stromal cell atypia. Malignant phyllodes tumours, which are rare, have highly atypical sarcoma-like stroma. **Papillomas** are intraductal lesions with a central fibrovascular core with an overlying myoepithelial and epithelial layer. They may occur as solitary lesions, typically in larger central ducts close to the nipple, or as multiple lesions (papillomatosis) predominantly involving the peripheral TDLU.[2]

Hamartomas are pathologically indistinct lesions composed of varying amounts of benign epithelium, fibrous tissue and fat.[28] Diagnosis is invariably made by

correlating radiological and pathological findings. **Fat necrosis** is a sterile inflammatory process, usually preceded by trauma although this may not be recalled by the patient. Thin capsules or lipid cysts may be present in areas of fibrosis. Chronic lesions may present as oil cysts frequently with ring-like calcification present in the cyst wall.[29] Benign breast changes such as sclerosis, inflammatory change, fibrocystic change and fibrosis can also mimic a malignant lesion on imaging.

1.2 Imaging of breast cancer

Breast imaging is used for screening, diagnosis, local staging of breast cancer and monitoring of response to neoadjuvant therapies. Conventional imaging, which has established long-term use includes mammography and ultrasound.

1.2.1 Mammography

Mammography is a radiographic technique designed specifically for imaging the breast. It is required to demonstrate both microcalcifications that may be 100nm or less in size with high inherent contrast, and soft tissue abnormalities with much lower intrinsic contrast than in general radiography. Therefore, low-energy x-rays are utilised with an exposure time of up to 2 seconds. Initially single view mammography was used but after a large randomised trial demonstrated an increase in cancer detection rate of 24% with two view mammography, this became standard practice.[30] The standard views now performed are craniocaudal (CC) and medial lateral oblique (MLO). Additional views are used for problem solving, including lateral, rolled, magnification and focal compression views. For each view, the breast is positioned on the support platform enclosing the digital detector and held in place by the compression paddle. Compression reduces breast thickness, dose, patient movement, composite shadowing and improves unsharpness.
Mammography has been improved by the conversion from analogue (film-screen) mammography to full field digital mammography (FFDM). The overall diagnostic accuracy using digital mammography is similar to film-screen but increased accuracy is seen in younger women and those with a higher proportion of fibroglandular tissue and therefore denser breasts.[31]

Whilst mammography has a population based sensitivity of between 75% and 80%, this is significantly reduced in women with dense breasts, dropping below 50% in those with extremely dense breasts.[32 33] Furthermore, mammography does not provide any functional information, thus differentiating between treated fibrotic tissue or scarring and recurrent or residual tissue is challenging.

1.2.2 Ultrasound

Handheld grey-scale ultrasound is predominantly used for targeted imaging to further characterise clinical and imaging abnormalities and to guide tissue biopsies. It is possible to diagnose certain benign lesions, such as simple cysts, using ultrasound alone.[34] In 1995, Stavros[35] published strict diagnostic criteria using ultrasound to identify benign solid masses with a negative predictive value of 99.5%. This work has been used to inform guidance regarding the diagnosis of classical fibroadenomas in young patients on imaging alone.[27] However, with a positive predictive value of only 38% and a specificity of 67.8% the majority of solid masses continue to require biopsy. Grey-scale ultrasound has limited use in non-targeted imaging, is heavily operator dependent and does not provide functional information. Colour Doppler imaging and power Doppler imaging are established, widely available techniques which provide an indication of lesion vascularity. Doppler features suggestive of malignancy include; central vascularity, hypervascularity and penetrating or branching vessels. However there remains a significant overlap between the Doppler appearances of benign and malignant lesions.[36 37]

1.2.3 Advanced nuclear medicine imaging techniques

To overcome some of the limitations of the aforementioned conventional imaging, advanced ultrasound, radiographic and nuclear-medicine imaging techniques and breast magnetic resonance imaging have been developed.
Scintimammography or molecular breast imaging (MBI) is a nuclear medicine technique in which patient have intravenous administration of tecnicium-99m (Tc-99m) sestamibi. Imaging of the breast is then performed using a gamma camera, with an acquisition lasting approximately 10 minutes. Rounded foci of abnormal enhancement within the breast are considered suspicious for malignancy.[38]
Regarding detection of malignancy, scintimammography has high sensitivity,

equivalent to MRI (89-96%) which is maintained even in dense breasts. Indications for this technique include supplementary screening in dense breasts, high-risk screening, assessing response to neoadjuvant chemotherapy.[39] However, the anatomical detailed acquired at MRI is not achieved with scintimammography. Furthermore, there may be logistical difficulties integrating the use of this technique into a breast imaging pathway, as it usually requires a dedicated room and access to an accredited nuclear physicist.[40]

Positron emission tomography (PET) is a whole body imaging technique, which uses a positron-emitting radiopharmaceutical, fluorodeoxyglucose (FDG), to identify foci of abnormal metabolic activity, as seen in malignancy. This technique can be combined with computed tomography (CT) images, which demonstrate structural and morphological characteristics, to produce PET/CT images.[41] With respect to breast imaging, PET/CT has utility for staging of local, regional and distant disease and may be useful for monitoring response to neoadjuvant chemotherapy and response to chemotherapy in metastatic disease.[42]

1.2.4 Advanced ultrasound imaging techniques

Contrast-enhanced ultrasonography (CEUS) is a relatively new technique which utilises intravenous microbubble contrast agents which result in enhanced backscatter of ultrasound waves and therefore increased amplitude of flow signals. This provides an indication of the microvascularity of a lesion. Use of early contrast agents was largely unsuccessful due to the fragility of the microbubbles when the high frequency probes required for breast ultrasound were used. Greater success has been seen with second generation contrast agents. Features suggestive of malignancy include; peripheral or heterogeneous enhancement, penetrating vessels, early intense wash-in and fast wash-out. The addition of CEUS to grey-scale ultrasound has been demonstrated to improve diagnostic accuracy; however, there is a wide range of both reported sensitivities and specificities; 64-100% and 38%-97% respectively.[37 43] Shear wave elastography is a method that quantifies lesional and perilesional stiffness. This technique has been demonstrated to aid differentiation of benign and malignant lesions as the latter demonstrate increased stiffness. Emerging evidence suggests it may also be useful for differentiating

histological types of breast cancer and predicting outcome of neoadjuvant chemotherapy.[44-46]

1.2.5 Magnetic Resonance Imaging (MRI)

MRI utilises a superconductor magnetic field to polarise and align the protons within the bodys tissues. When radiofrequency (RF) pulses are passed through the body the protons spin out of equilibrium. The MRI sensors detect the energy released as the protons 'relax' and realign with the magnetic field. Due to the variation in the relaxation times of protons within different tissues, high resolution images can be produced.[47]

Contrast-enhanced MRI (CE-MRI), characterises the vascularity of lesions through the use of intravenous (IV) gadolinium-based contrast medium. Prior to MRI image acquisition the patient is cannulated to allow administration of contrast. The patient is then required to lie prone within the MRI scanner, with the breast immobilised with a breast-specific RF coil. The scan can last up to 40 mins whilst various pre- and post-contrast sequences are performed. Neo-angiogenesis, the formation of new blood vessels, is a recognised feature of tumour growth.[48] Analysis of lesion enhancement is indicative of lesion vascularity thus allowing acquisition of functional data. Rapid intense enhancement is associated with malignant lesions, higher grade tumours and residual active tumour within the tumour bed.[49 50] Time-signal intensity curves (TIC) are routinely generated from breast MRI studies by categorising the dynamic lesion enhancement pattern. In a seminal paper, Kuhl *et al* classified time-signal intensity curves as type I, steady enhancement; type II, plateau of signal enhancement and type III, washout of signal intensity. Type III curves were demonstrated to be more frequently associated with malignant lesions with a sensitivity of 91% and specificity of 83%.[49] Subsequent work quantified the relative signal intensity change for the respective curve types; a signal change of 10% or less is considered plateau enhancement (type II); with an increase or decrease of over 10% considered steady enhancement (Type I) or washout (Type III) respectively. Where there is heterogeneity in a lesion the most suspicious curve should be reported.[51-54] In addition to predicting malignancy, enhancement characteristics have been shown to predict tumour grade.[50 55] In the context of NACT, a change to a more benign TIC at mid-treatment has been shown to correlate with tumour response.[56]

Diffusion weighted imaging (DWI) is a form of MRI imaging that exploits the random Brownian motion of water molecules within tissues. Highly cellular tissues, such as tumours, typically demonstrate lower apparent diffusion coefficient (ADC) values or 'restricted diffusion'.[57] Potential applications include enhanced breast cancer screening and monitoring response to NACT.[57 58] Furthermore, the addition of DWI sequences to CE-MRI protocols has been shown improve diagnostic accuracy for detecting breast cancer.[53]

CE-MRI is frequently considered the 'gold standard' breast imaging technique due to high sensitivity for breast cancer detection with a pooled sensitivity of 93%.[59] Multiple studies have also demonstrated that MRI provides a more accurate assessment of tumour size and identification of multifocal disease than conventional imaging.[60]

However, MRI also has several disadvantages. The relatively low specificity (71%) and therefore relative high false positive rate may lead to more extensive surgery,[61] or to delays in treatment whilst second look ultrasound and biopsies are performed. This may be further hampered by lack of access to MRI guided biopsy. Furthermore, MRI is an expensive technique with limited availability. It may be contraindicated (e.g. metal foreign bodies) or poorly tolerated by patients due to claustrophobia and discomfort. Finally, there are concerns regarding the long-term effects of gadolinium deposition following multiple administrations of gadolinium-based contrast agent.[62]

Two promising imaging modalities are emerging as possible alternatives to MRI for the diagnosis, staging and surveillance of breast cancer. Contrast-enhanced spectral mammography (CESM) and digital breast tomosynthesis (DBT).

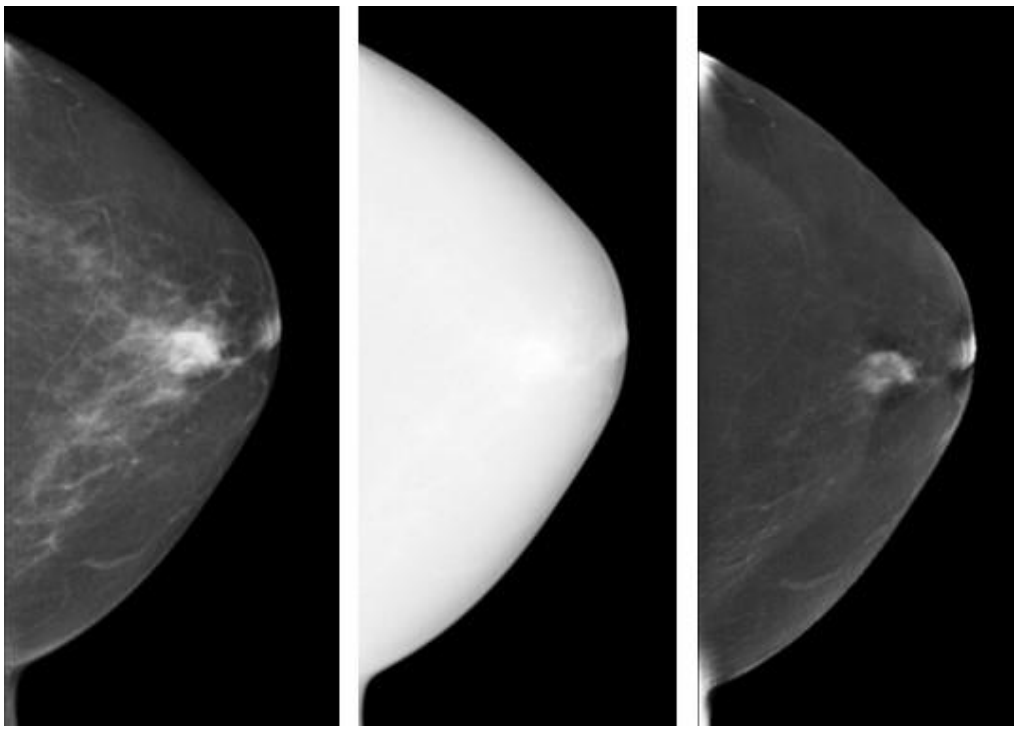
1.2.6 Contrast-enhanced spectral mammography (CESM)

CESM is a functional imaging technique, which in a similar manner to CE-MRI, is able to demonstrate the neoangiogenesis of tumours through the use of IV contrast media. Contrast mammography was first developed in the early 2000s. Initially the technique involved acquiring a pre-contrast 'mask' mammogram which was digitally subtracted from subsequent post-contrast mammograms.[63] Whilst showing promising early results this technique was hampered by motion artefact and long duration of breast compression.[64]

Lewin *et al* published a feasibility study of dual-energy contrast mammography in 2003.[65] Since then there have been multiple improvements in both hardware and

software, although the technique remains similar. Iodinated contrast is given intravenously at least two minutes prior to breast compression to ensure perfusion of the breast. Following this, two acquisitions: a low kiloVoltage (kV) image (below the k-edge of iodine) and a high kV image, are performed in each mammography position. Two views of each breast are taken (CC and MLO), as with a standard mammogram and a total of eight diagnostic images are produced.

As the low-energy (LE) image is below the k-edge of iodine, the image produced is equivalent to a standard mammographic view. Using digital post-processing the LE image is then subtracted from the high-energy image to produce a recombined (RC) image which demonstrates only the areas of enhancement. Thus, the RC image demonstrates the vascularity / neoangiogensis of breast lesions. Below is an example of a CESM study in the left CC position.



a. b. c.

Figure 1: The diagnostic images; the low energy image (a) and the subtracted recombined CESM image (c) are transferred to PACS for reporting. The raw high-energy image (b) is not used for clinical purposes

Published studies have demonstrated that CESM has both a higher sensitivity and diagnostic accuracy for breast cancer detection than FFDM.[66] It has been demonstrated to be comparable to MRI for the detection of breast cancer and assessment of lesion size[67 68] and it is better tolerated by patients.[69]

1.2.7 Digital breast tomosynthesis (DBT)

DBT is a modified mammographic pseudo-3D imaging technique. Multiple low-dose images are acquired in an arc over the breast, then reconstructed into a stack of multiple slices, usually 0.5-1mm thick.[70 71] Variations in the acquisition are present between vendors both in the tube motion; stop and shoot versus continuous tube motion, and the angular range. The overall glandular dose of DBT is similar or slightly higher than FFDM.[72]

The pseudo-3D nature of tomosynthesis images helps to eliminate overlapping fibroglandular breast tissue thus improving conspicuity of lesions. As a result, DBT has demonstrated increased cancer detection rates, especially in mixed and dense breasts (although not very dense breasts) when compared with FFDM.[71 73]

Furthermore, DBT improves visualisation of lesion margins. This is useful both for identifying malignant structural features such as spiculation and distortions[74 75] and the reassuring smooth margins present in benign lesions.[76]

1.2.8 Contrast-enhanced Digital Breast Tomosynthesis (CE-DBT)

CE-DBT is a new technique combining CESM and DBT. Within the same breast compression, a LE mammogram, RC image and unenhanced tomosynthesis images are produced. There is a dearth of evidence in the literature due to the novel nature of the CE-DBT modality, but it has the potential to combine the functional data acquired at CESM with the improved structural information acquired at DBT.

1.3 Neoadjuvant chemotherapy

With advances in oncological treatment, increasing numbers of women are receiving pre-surgical neoadjuvant chemotherapy (NACT). This strategy is indicated in a number of clinical scenarios. This includes patients with initially inoperable disease, due the locally advanced status with skin or chest wall involvement, when the aim is to shrink the tumour sufficiently to allow definitive surgical intervention. Other situations include disease which would necessitate a mastectomy or where a lumpectomy would have a poor cosmetic result if treated with primary surgery, when

NACT may reduce the tumour burden sufficiently to allow cosmetically acceptable breast conservation surgery. Finally, it may downstage the axilla in patients with positive axillary lymphadenopathy and allow a sentinel node procedure as opposed to a lymph node clearance which carries significantly increased morbidity.[77]

1.3.1 Neoadjuvant chemotherapy regimes

Standard neoadjuvant chemotherapy regimens for breast cancer combine an anthracycline, cyclophosphamide and a taxane.[78] This is frequently given as three cycles of epirubicin plus cyclophosphamide followed by three cycles of Docetaxel. For patients with HER2 positive disease, anti-HER2 agents such as trastuzumab and pertuzumab are given alongside docetaxel.[19 79 80] Chemotherapy regimens may be modified for less fit patients, for example four cycles of a combination of docetaxel and cyclophosphamide may be used.

1.3.2 Monitoring response to neoadjuvant chemotherapy

Imaging monitoring of treatment response is necessary throughout the course of NACT to enable decisions regarding the aforementioned surgical options to be made. It also allows early surgery for those women with progressive disease. In the future, it may be possible to avoid surgery altogether in patients with a complete radiological response, instead confirming the pathological response with multiple image guided biopsies.

The current gold-standard staging and monitoring technique is contrast-enhanced magnetic resonance imaging, as it the most sensitive and specific test to demonstrate response following NACT.[81 82] Unfortunately in the UK, due to pressures on imaging services it can be difficult to access in a timely fashion, and it is an expensive and time-consuming technique. Furthermore, in some patients, it is either contraindicated (e.g. in patients with pacemakers) or not possible (because of body habitus), or poorly tolerated (e.g. due to claustrophobia).[83] Mammography and ultrasound are widely available but lack the ability to provide functional data and are therefore less reliable indicators of disease response.[84]

1.3.3 Assessing pathological response to neoadjuvant chemotherapy

The surgical specimen is processed and assessed. Key features reported include; **whole tumour size** (WTS): this is the largest diameter including invasive and *in situ* disease (with the exception of classical LCIS); **invasive tumour size** (ITS): the maximum dimension of the invasive tumour, not including satellites or foci of lymphovascular invasion; and **in situ carcinoma size:** this is reported for pure DCIS or pleomorphic lobular *in situ* lesions where no invasive disease is seen. **Disease extent:** is reported for multiple foci of disease.

Residual invasive tumour size can be challenging to measure. If there are multiple scattered foci of disease it is recommended to measure overall extent of residual foci. However, if clear islands of cancer are present they are measured individually. The remaining tumour volume relative to tumour bed size indicates amount of residual tumour.[2]

Response evaluation in solid tumours (RECIST)

RECIST is a method of quantifying change in tumour burden, primarily used for clinical trials for assessment of treatment outcomes. [85] Tumour lesions may be assessed radiologically or clinically using calipers; lymph nodes are measured in short axis on CT. Response is classified as

- Complete response (CR): Disappearance of all target lesions. Short-axis of pathological lymph nodes reduced to <10mm
- Partial Response (PR): ≥30% decrease in sum diameter of target lesions
- Stable disease (SD): A change in sum diameter of <30% decrease and <20% increase.
- Progressive disease: ≥20% increase in sum diameter of target lesions where the absolute increase ≥5mm OR new lesion(s)

1.3.3.1 Complete pathological response

There is no standardised definition for pCR, but the following four definitions are found in the literature, *ypT* refers to residual tumour in the breast, *ypN* refers to residual tumour in lymph nodes:[86]

- *ypTO, ypN0*: No invasive or in situ disease in breast or nodes

- *ypT0/is, ypN0*: No invasive residual disease in breast or nodes, residual *in situ* disease in the breast may be present.
- *ypT0/is ypN0/+*: No invasive residual disease in breast but *in situ* disease in the breast may be present. Infiltrated lymph nodes allowed.
- *ypT≤1mic ypN0/+*: No gross invasive residual disease in breast or nodes. However focal invasive and *in situ* breast residuals and infiltrated lymph nodes may be present.

The most frequently used definition for pCR is *ypT0/is, ypN0,* used by the MD Anderson Cancer Center, Neo-Breast International group and Breast and Colorectal Cancer Study Group.[86] As the presence of residual in-situ disease following NACT, in the absence of invasive disease, has been shown to have no adverse effect on survival or local recurrence rate,[87] this is the definition used for this study.

1.3.3.2 Residual cancer burden (RCB)

The residual cancer burden (RCB) index is a method of quantifying residual disease following NACT that was described by Symmans in 2007 as an independent predictor of risk of recurrence.[88] Estimated from pathological sections of the **primary tumour bed** and **lymph nodes** following neoadjuvant chemotherapy, the RCB score includes six variables;

Primary tumour bed

- Size of the primary tumour bed area in two dimensions.
- Overall cancer cellularity of the largest cross-sectional area of residual tumour.
- Percentage of cancer that is *in situ* disease

Lymph nodes

- Number of positive lymph nodes
- Diameter of largest metastasis

These variables are entered into the online RCB calculator provided by the MD Anderson centre (http://www.mdanderson.org/breastcancer_RCB) and the RCB index is generated. This index is divided into four classes reflecting increasing amounts of residual disease; RCB 0 (pCR), RCB-I, RCB-II and RCB-III. The RCB index and RCB class has subsequently been validated by a number of prospective trials, and demonstrated to predict 10-year disease free survival in all immunophenotype classes.[89-91]

1.4 Textural analysis

Textural analysis (TA) is a technique for evaluating the appearance, position and pattern of elements of an object within an image. The majority of medical images are now digitalised, composed of multiple pixels of varying grey-level intensities. Texture features are mathematical parameters which can be calculated from pixels according to their grey-level intensity and distribution within a 2D digital image. Texture analysis has been successfully applied to breast imaging, including breast MRI, breast DWI and PET/CT.[92-95]

MaZda (Technical University of Lodz, Poland) is a software package that can be applied to 2D and 3D medical images to generate textural parameters. Using this software, it is possible to manually segment images using either a freehand or oval region of interest (ROI) and select which texture features are to be analysed.[96-98] There are multiple parameters that can be applied, the main ones are considered below.[97 99]

1.4.1 Histogram (HIST):

The histogram of an image is the count of how many pixels within an image demonstrate a specific grey-level value. Various parameters can be calculated from the histogram, including the mean, variance and percentiles. This has the potential to quantify the degree of enhancement demonstrated by a lesion on CESM imaging.

1.4.2 Absolute gradient (GRAD):

This a measure of the variation in grey-level across an image. For example, a gradual change from a high intensity (light grey) to low intensity (dark grey) will have a low gradient, whereas an abrupt change would have a high gradient.

1.4.3 Run-length matrix (RLM):

This is a method of searching an image for ‘runs’ of adjacent pixels with the same grey-level value in a specified direction. Parameters generated from this include short-run emphasis (the proportion of runs occurring that are short in length) and

fraction of image in runs (the proportion of the image pixels including in any runs within the matrix programing).

1.4.4 Grey level co-occurrence matrix (GLCM):

This is a second-order histogram and produces statistical information regarding the distribution of pixel pairs. It is computed by defining both a direction (0°, 45°, 90°, 135°) and a distance between which the pixels are separated. As no directional variation in textural parameters is expected in breast imaging the output can be averaged.[100] Coarse and fine features can be defined by the distance separating the pixels in question.[93 101] Examples of parameters generated include entropy (a measure of heterogeneity) and contrast (how much difference there is between grey level values).

1.4.5 Auto-regressive (AUTO):

This model is a way of describing shapes within an image, through analysing the relationship between neighbouring pixels. It assumes that the pixel intensity, with reference to the mean value for the image pixel intensity, can be predicted by the neighbouring pixels.

1.4.6 Haar Wavelets (WAV):

Wavelets represent the frequency of one-dimensional signal change, rapid signal change equating to a high frequency and slower signal change a lower frequency. Wavelets are scaled up and the image transformed into 2D frequency channels before generating textural parameters.

1.4.7 Textural analysis applied to breast disease

Textural analysis parameters have been successfully applied to breast MRI images, to differentiate benign and malignant lesions[102] and breast cancer subtypes,[92] and to predict response to neoadjuvant chemotherapy[93 100 103-105].

Chapter 2. CONDOR: CONtrast-enhanced Digital breast tomosynthesis for monitoring Of Response to neoadjuvant chemotherapy – a literature review

2.1 Introduction

Breast cancer remains the commonest malignancy in women in Scotland, accounting for 29% of all female cancers in 2015, and 13.5% of female cancer deaths in 2016.[1] Neoadjuvant chemotherapy (NACT), may be given as initial treatment to down-stage disease in the breast and/or axilla to allow or improve the cosmetic outcomes of surgical interventions. Furthermore, it provides an *in vivo* indication of tumour chemosensitivity. Treatment response is monitored with imaging throughout the course of NACT to guide surgical and oncological treatment decisions. Contrast-enhanced digital breast tomosynthesis is an entirely novel technology may offer an alternative to the current gold-standard monitoring technique: contrast-enhanced magnetic resonance imaging (MRI)

2.2 Methods

Literature searches were conducted in MEDLINE using Medical Subject Headings (MeSH) where possible, supplemented by ‘keyword’ searches. As CE-DBT is any entirely novel technique a literature search returned no results for its use in the context of NACT. Therefore, the question was broken down into individual components to allow broader, all-encompassing searches to be performed. All searches were limited to the English language.

2.2.1 CE-DBT in breast imaging

Search terms were as follows: [Contrast enhanced tomosynthesis OR (tomosynthesis OR digital breast tomosynthesis OR DBT)] AND [(contrast enhanced mammography OR CESM)].
All articles concerning the use of CESM and DBT for image interpretation of breast cancer in the same patients contemporaneously were included. All review articles considering these techniques separately, for dose calculation or alternative technologies only were excluded.

2.2.2 CESM or CE-DBT for monitoring response to NACT in breast cancer patients

Search terms were as follows: [Contrast enhanced tomosynthesis OR tomosynthesis OR digital breast tomosynthesis OR DBT OR contrast enhanced mammography OR CESM] AND [neoadjuvant chemotherapy OR NACT OR neoadjuvant therapy] AND [Breast cancer OR breast neoplasm].
All articles concerning the use of CE-DBT, CESM or DBT for the monitoring of breast cancer patients undergoing NACT were included. Articles which considered these techniques separate to NACT, the use of alternative technologies only or other aspects of NACT were excluded.

2.3 Results and Discussion

2.3.1 CE-DBT in breast imaging

The largest prospective study which considers the parallel use of CESM and DBT included 185 women with suspicious breast lesion (BIRADS 4 and 5)[106]. All patients underwent full field digital mammography (FFDM), DBT, CESM, MRI and contrast enhanced digital tomosynthesis (CET) using a prototype machine. CET was performed over a 15-degree arch acquiring 22 images alternating between low and high energy. Comparative performance analysis of diagnostic accuracy between various modalities was performed using a parametric receiver operating characteristic (ROC) method. The studies demonstrated that the step-wise addition of

DBT and subsequently CESM to FFDM and DBT resulted in a progressively significant increased diagnostic accuracy. MRI had a non-significant inferior sensitivity for diagnosis of invasive and non-invasive malignancy compared to the combination of FFDM, DBT and CESM; average sensitivity for each of three readers ranged from 86-93% vs 93-98% respectively. Interestingly, as discussed in subsequent correspondence there was no significant increase in accuracy with the further addition of CET, suggesting that there was no gain in the use of CET in addition to the combination of DBT and CESM.[107 108] Comparison of CESM combined with DBT versus the prototype CET is not possible due to the methodology of the paper.

One further prospective study investigates the role of CESM and DBT for local staging of breast cancer in patients with biopsy-proven breast cancer considered at increased risk of locally advanced or multicentric carcinomas.[109] DBT is reported as the most sensitive technique for calculation of size and CESM the most accurate for identifying multiplicity. However, the methods of this study are questionable. With regards to size, the tolerance margin allowed between imaging technique and surgical specimen is not defined and it is not reported whether the imaging techniques were over or under-estimating size. Regarding multiplicity, sensitivities of FFDM, DBT and CESM are reported as 53.8%, 77% and 92.3% respectively. Unfortunately, neither the absolute numbers of tumours with satellites on pathology, nor the false positive rate is provided. CESM is described as the 'modality of choice' for early breast cancer and FFDM + DBT more accurate for locally advanced disease. However, the presence or absence of statistical significance in the accuracy of T-staging is not reported.

2.3.2 CESM or DBT for monitoring response to NACT in breast cancer patients

There is no published data on the use of CE-DBT for monitoring NACT. Therefore, the use of DBT and CESM for this purpose is considered separately. Only one paper could be identified that investigated the use of DBT in monitoring response to NACT[110]. Indeed, whilst the American College of Radiology guidelines include DBT as an appropriate modality for monitoring response to NACT this is entirely inferred from studies investigating DBT as a screening or diagnostic tool.[77] Park *et al*

conducted a prospective study of 50 women with stage II-III breast cancer treated with NACT. All patients underwent pre-operative FFDM, DBT, automated breast ultrasound (ABUS) and MRI. Pre-treatment and interim treatment imaging were not considered. Analysis was based solely on the greatest diameter of suspicious abnormality on imaging and pathological specimen. In the context of a spiculate mass the full extent of spicules was included, though it is known that a proportion of spiculation is due to fibrosis. It is therefore perhaps not surprising that DBT overestimated residual tumour size by 5mm in 33.4% and 10mm 21.6%. Overall there was no significant difference in size prediction between MRI and DBT. Both techniques out performed ABUS and FFDM.

Three studies using CESM for this purpose show promising results.[111-113] One study of 46 patients compared CESM and MRI maximum lesion size measurements taken before, during and after NACT. Agreement between the measurements on the two imaging techniques is reported as 0.96, 0.94 and 0.6 at the respective time-points. Critically, post-NACT CESM measurements had stronger agreement with histology than did MRI, though both methods tended to underestimate disease extent. The distinction of responders and non-responders (RECIST criteria) was identical for both techniques in 44/45 patients. CESM had a higher sensitivity and specificity for pathological complete response (pCR) than MRI, 100% and 84% vs 87% and 60% respectively.[111] In a second, larger study of 65 patients, CESM and MRI imaging were performed before and after NACT.[112] In this study CESM was comparable to MRI for assessment of pCR with a sensitivity of 95% vs 95% and specificity of 66.7% and 68.9% respectively. CESM accurately showed tumour size to within 10mm in 72.3% of patients, with an average underestimation of 5mm (SD = 16.8mm). By comparison MRI accurately demonstrated tumour size in 69.2% patients, with an average underestimation of 5.2mm (SD = 15.5mm). The tolerance used of +/-10mm, is a substantial margin in the context of the average pathological size being only 19.6mm (range 0-75mm).[112] In both studies the patient was considered to have achieved a pCR when there was absence of residual invasive or in situ (DCIS) disease. It is worth noting that the most commonly accepted definition of pCR is the absence of invasive disease, with residual in situ disease permitted. Finally, a very small retrospective study of 8 patients demonstrated that CESM had a sensitivity and specificity of 83.3% and 100% respectively.[113]

2.4 Key Research Questions

This literature review has generated the following key research questions to address the study aim of evaluating the performance of CE-DBT for the monitoring of response to NACT:

- How accurate is CE-DBT in assessing pathological response to NACT measured by pathological complete response (pCR) and residual cancer burden (RCB)?
- How well does CE-DBT compare with MRI for monitoring response to NACT?
- Is the additional imaging (CE-DBT) feasible within the workflow of the breast unit?
- Is CE-DBT considered an acceptable imaging technique by patients?

Chapter 3. TACESM: Textural Analysis of breast lesions on Contrast Enhanced Spectral Mammography, a feasibility study – a literature review

3.1 Introduction

CESM is a functional imaging technique, which is becoming more widespread in clinical practice. It utilises a dual energy subtraction method following administration of intravenous contrast agent to produce 2D images that demonstrate the vascularity and neo-angiogenesis of breast lesions.

The benefits of functional breast imaging techniques have previously been demonstrated with contrast-enhanced magnetic resonance imaging (MRI), currently accepted as the gold-standard for local staging of breast cancer, due to its high sensitivity (pooled sensitivity of 93%)[59]. Both the intensity of enhancement and the wash-in/out characteristics aid discrimination between benign and malignant lesions and quantification of response to neoadjuvant chemotherapy (NACT) - intense enhancement is associated with malignant lesions and residual active tumour within the tumour bed.[114 115] Time-signal intensity curves (TICs) are routinely generated from breast MRI studies by categorising the washout pattern of the gadolinium contrast agent. Type 1 curves (progressively enhancing) more common in benign lesions, type 2 curves (plateau) of intermediate concern for malignancy, and type 3 (washout curves) suspicious for malignancy.[114]

Textural analysis is a technique for evaluating the appearance, position and pattern of pixels according to their grey-level intensity within a 2D digital image.[99] This method allows quantitative analysis of images and may improve discrimination between benign and malignant lesions as well as between malignant lesion subtypes. Furthermore, it has been demonstrated that lesion heterogeneity as measured by textural analysis of breast MRI studies may be useful in predicting response to NACT in breast cancer patients.[103]

Published studies have demonstrated that CESM has similar accuracy to MRI for the detection of breast cancer[67] and is often better tolerated by patients.[69]

It has been demonstrated that lesion heterogeneity as measured by textural analysis of breast MRI studies may be useful to predict the response to NACT in breast cancer patients. The majority of studies extract textural features using grey-level co-occurrence matrix (GLCM), a method of quantifying inter-pixel spatial relationships.[103-105 116] With the exception of one study[104] measures of increased lesion heterogeneity have been shown to be related to more aggressive malignancies with poorer responses to NACT. Michoux *et al*[104] found reduced texture heterogeneity (measured using GLCM parameters) in the more aggressive tumours. However, unlike other studies they excluded necrotic areas. Such necrotic areas are more frequently seen in aggressive tumours and would inevitably increase the heterogeneity of a lesion as they do not enhance.

Recent work has considered temporal changes in lesion heterogeneity between baseline and interim MRI scan.[93 117] In the largest of these studies, including 88 patients, Henderson *et al*[93] considered the change in T2 heterogeneity between baseline and interim MRI scan. Entropy, or lesion heterogeneity was derived from a GLCM with inter-pixel distances of n=2 and n=5 for fine and coarse texture respectively. Response to NACT was assessed using residual cancer burden (RCB) scores. They reported that heterogeneity changes are associated with response to NACT across all immunophenotypes with coarse features being the most accurate. A further study of 35 patients considered textural changes within tumour subregions. It found that the tumour subregion with most aggressive phenotype i.e. the subregion that is the most heterogeneous on initial MRI, plays the dominant role in determining clinical outcomes.[118] As CESM also uses contrast agent and the degree of CESM enhancement has been shown to correlate with the density of blood vessels, it may be possible to demonstrate neoangiogenesis associated with tumours and produce similar functional data to that of MRI.[119]

It must be noted that this literature search has been refreshed following the start of the study. When the TACESM protocol was designed in 2017 there was a dearth of published data pertaining to the use of textural analysis or radiomics with CESM images, therefore the methods were based on existing knowledge derived from MRI radiomics.

3.2 Methods

Literature searches were conducted in MEDLINE and PUBMED. Due to the relatively novel nature of CESM and thus sparsity of published information pertaining to functional data, the Medical Subject Headings (MeSH) used were intentionally broad; [(Textural analysis) OR (radiomics) OR (computer-assisted image processing) OR (computer-assisted image interpretation) OR (CAD) OR (enhancement)] AND [(Contrast enhanced mammography) OR (contrast enhanced spectral mammography) OR (contrast mammography) OR (CESM)]. Searches were limited to the English language.

Following removal of duplicate references all titles and abstracts were reviewed. The following inclusion / exclusion criteria were used.

Inclusion criteria:

- Contrast-enhanced spectral (dual energy) mammography
- Quantitative or qualitative measure of intensity or pattern of lesion enhancement
- Quantitative analysis of lesion enhancement through CAD software, radiomics or textural analysis
- Correlation to histology of lesion
- Original research

Exclusion criteria:

- Contrast-enhanced temporal mammography
- Ex-vivo studies (phantom, modelling, animal studies)
- Quantitative or qualitative measure background parenchymal enhancement
- CAD software used solely to identify potential lesions
- Case reports, case series, reviews, commentaries and guidelines

Papers meeting the aforementioned criteria were then read in full. Additional papers identified due to references or citations ‘snow-balling’ were also included.

3.3 Results

Literature searches were conducted on 27th April 2021. There were 754 initial search results. Following removal of duplicates (73 papers), 681 abstracts were screened. Of these, the majority did not meet inclusion criteria (646 papers). 35 articles were included and these papers read in full. One further paper was identified through ‘snow-balling’. Of the 36 papers initially included; a further thirteen were excluded following full review of the manuscript. Twenty-three papers were ultimately included; 13 pertained to assessment of intensity or pattern of lesion enhancement, and 10 considered textural analysis for lesion assessment, including those using CAD software. Details are provided in figure 2.

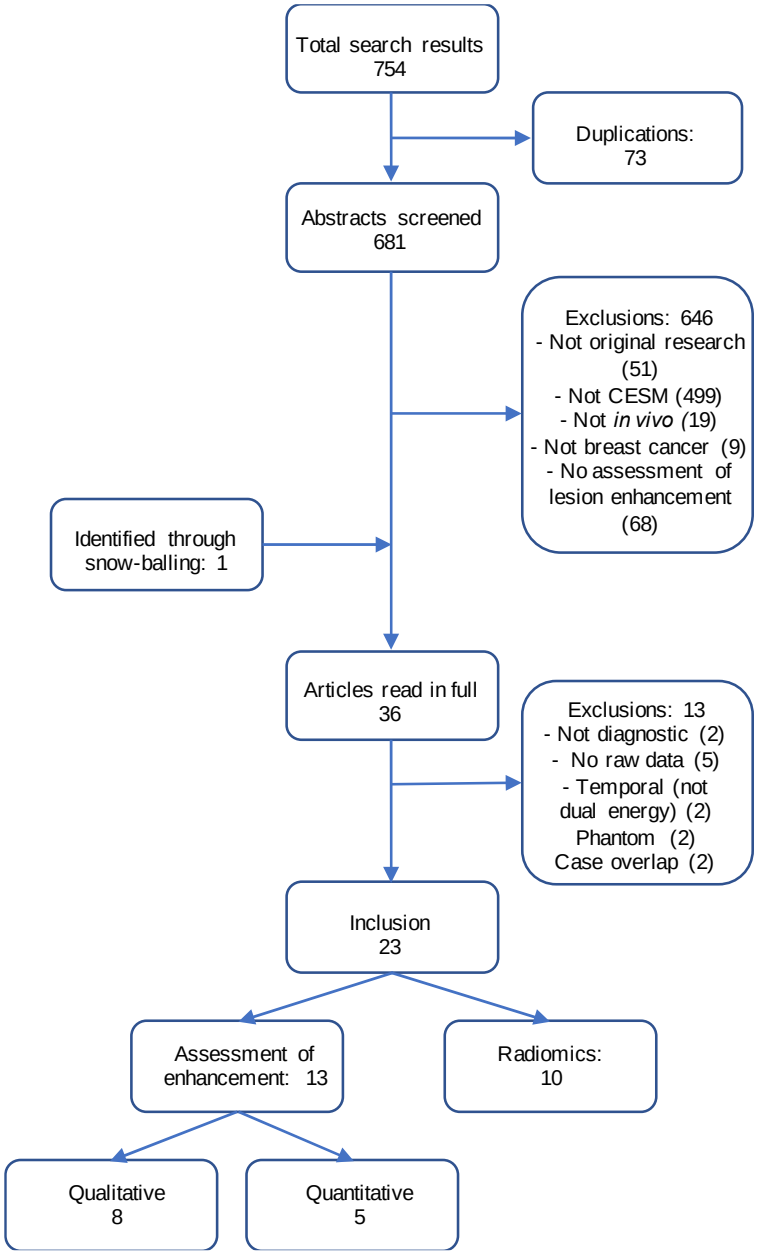


Figure 2: Flowchart illustrating included papers

3.3.1 Lesion Enhancement

Details of the studies pertaining to assessment of lesion enhancement are shown in table 4.

Paper	Institution	Study period	n	Type	Quantitative / Qualitative	Benign vs malignant	Malignant subtypes	Includes In situ	Vendor	Pattern / degree enhancement
Luczynska, Med Sci Mon, 2015	University hospital Krakow, Poland	NR	193 patients	Retrospective	Qualitative	Yes	No	Yes – separate data	GE	Degree
Luczynska, Anticancer Res 2018	University hospital Krakow, Poland	2011-2012	84 lesions (82 patients)	Retrospective	Qualitative	No	Yes	No	GE	Degree
Rudnicki, Eur Rad 2019	University hospital Krakow, Poland	NR	167 patients	Retrospective	Quantitative	Yes	No	Yes – separate data	GE	Degree
Kamal, Eur J Rad, 2015	Kasr ElAiny Hospital, Cairo, Egypt	2012-2013	99 lesions	Retrospective	Qualitative	Yes	No	Yes	GE	Degree & pattern
Tsigginou, BJR, 2016	Alexandra Hospital, Athens, Greece	2014-2015	226 lesion (216 patients)	Prospective	Qualitative	Yes	No	Yes	GE	Degree
Lobbes, Eur J Rad 2018	Maastricht University Medical Centre, Netherlands	2013-2014	85 lesions	Retrospective	Quantitative	Yes	No	No	GE	Degree
Deng, BJR, 2018	Chang Gung Memorial Hospital, Taiwan	2012-2015	152 lesions	Retrospective	Quantitative and Qualitative	Yes	No	Yes	GE	Degree
Travieso-Aja, Radiologia, 2019	Grupo Hospitalario San Roque, Spain	2013-2017	353 lesions	Prospective	Qualitative	Yes	No	Yes	GE	Degree & pattern
Van Nijnatten, BJR, 2019	Maastricht University Medical Centre & MSKCC	2010-2017	44 patients	Retrospective	Qualitative	No	Yes	No	GE	Degree
Chi, Medicine, 2020	Yantai Yuhuangding Hospital, China	2018-2019	312 lesions	Retrospective	Qualitative	Yes	No	Yes	GE	Degree & Pattern
Lv, J Comp Assist Tomogr, 2020	Yantai Yuhuangding Hospital, China	2017-2018	283 patients	Retrospective	Quantitative	Yes	No	Yes	GE	Degree
Liu, Scientific Reports, 2020	Sichun University West China Hospital	2017-2019	145 lesions (131 patients)	Retrospective	Quantitative	Yes	Yes	Yes – separate data	GE	Degree
Goh, Eur Rad, 2021	Kaohsing Veterans Hospital	2012-2019	94 lesions	Retrospective	Qualitative	Yes	No	Yes	Hologic	Degree

Table 4: Characteristics of studies pertaining to assessment of CESM lesion enhancement

The majority, 11/13 were retrospective studies. One study used a commercially available full-field digital mammography system manufactured by Hologic [120], the remainder used equipment manufactured by GE Healthcare. Most specified that 1.5ml/kg intravenous iodinated contrast agent was administered at 3ml/sec.[121-125] There was variation in the image acquisition protocols between studies, however this is felt to reflect normal clinical practice and should not invalidate results.
For further analysis, these studies were separated into those providing qualitative assessment 'eye-balling' enhancement properties, and those using quantitative analyses.

3.3.2 Qualitative Assessment of enhancement characteristics

Eight studies were identified which reported qualitative assessment [67 120-126]. Of these, six considered intensity of lesion enhancement between benign and malignant lesions, and two papers explored the difference between invasive ductal and invasive lobular carcinomas. When data for multiple readers was reported for individual lesions this was averaged to the nearest whole number.

3.3.2.1 Intensity of lesion enhancement; benign vs malignant lesions

All six studies included both invasive and ductal carcinoma *in situ* (DCIS) within the malignant group. Intensity of enhancement was measured on a categorical scale, either a three-point scale[67 120 122], or two-point scale[121 123 125]. To allow comparison, mild, faint and weak; moderate and medium; and marked, intense and strong were considered synonymous terms. A summary of results is displayed in table 5.

Study	Benign lesions			Malignant lesions			*p* values
	Mild (%)	Mod (%)	Marked (%)	Mild (%)	Mod (%)	Marked (%)	
Luczynska, Med. Sci. Monit ,2016	25 (51.0)	16 (32.7)	8 (16.3)	35 (34.4)	45 (31.4)	63 (44.1)	Chi^2; $p<0.001$
Tsigginou, BJR, 2016	-	25 (67.6)	12 (32.4)	-	21 (22.8)	71 (77.2)	NR
Chi, Medicine, 2020	35 (32.1)	50 (45.8)	24 (22.0)	7 (3.4)	46 (22.7)	150 (73.9)	spearman; r=0.533, p=0.000
Travieso-Aja, Radiologia, 2019	58 (80.6)	-	14 (19.4)	121 (43.1)	-	160 (56.9)	Chi^2; $p<0.001$
Kamal, Eu J Rad, 2015	15 (83.3)	-	3 (16.7)	18 (22.2)	-	63 (77.8)	Chi^2; $p<0.001$
Goh, Eur Rad, 2021	21 (60.0)	12 (34.3)	2 (5.7)	7 (21.2)	9 (27.3)	17 (51.5)	OR 22.6; p=0.002
Range	32 – 83%	32 – 68%	6 – 32%	3 – 43%	23 – 31%	52 – 78%	

Table 5: Qualitative assessment of degree of lesion enhancement

Malignant lesions had a consistently higher proportion of marked enhancement, ranging from 52-78% compared to mild enhancement, 3-43%. The converse was true for benign lesions with 6-32% and 32-83% demonstrating marked and mild enhancement respectively. However, whilst 86-92% of all lesions demonstrating marked enhancement were malignant, 17-68% of those demonstrating mild enhancement were also malignant.

One study quantified degree of correlation, reporting a moderate correlation between enhancement intensity and benignity; r= 0.533, p = 0.000.[122] A further study reported marked enhancement was a predictor of malignancy on multivariate analysis OR, 22.6; 95%CI; 3.1, 166.6; p=0.002.[120]

3.3.2.2 Intensity of enhancement: Invasive ductal carcinoma vs invasive lobular carcinoma

Van Nijnatten *et al* [124] presented a comparison of degree of lesion enhancement by molecular subtype. They showed the findings for 22 cases of both IDC and ILC, assessed by three independent readers. A higher proportion of ILC was shown to demonstrate mild enhancement as compared to IDC, however despite good inter-observer agreement, k=0.723 (0.584-0.862), this only reached statistical significance for one reader. Conversely, IDC tended to demonstrate a greater proportion of marked enhancement, but only reached statistical significance for a different solitary reader. A further paper found no significant correlation between histological type and enhancement intensity although a slightly higher proportion of IDC demonstrated marked enhancement (57% vs 50%).[126]

3.3.2.3 Qualitative pattern of Enhancement

Three papers considered the pattern of enhancement; defined as homogenous, heterogeneous or ring enhancement as per the BIRADS MRI lexicon [121-123]. Chi *et al*[122] calculated malignant odd ratios (OR) and found lesions with heterogeneous enhancement tended to be malignant OR 3.228 (95% CI: 1.986 – 5.247), whereas those demonstrating homogenous enhancement tended to be benign OR 0.287 (95% CI: 0.175 – 0.471). No correlation was seen between ring enhancement and malignancy. These findings are consistent with two other papers which reported that heterogeneous enhancement tended to be characteristic for malignant lesions

(p<0.05) and that ring enhancement did not discriminate between benign and malignant lesions.[121 123] Results are shown in table 6.

Paper	Benign lesions (%)			Malignant lesions (%)			*p* value
	Homogenous	Heterogeneous	Ring	Homogenous	Heterogeneous	Ring	
Chi, Medicine, 2020	58 (53.2)	40 (36.7)	11 (10.1)	50 (24.6)	131 (64.5)	22 (10.8)	
Travieso-Aja. Radiologia, 2019	38 (52.8)	31 (43.1)	3 (4.2)	104 (37.0)	168 (59.8)	9 (3.2)	Chi^2; p=0.038
Kamal, Eu J Rad. 2015	6 (33.3)	2 (11.1)	10 (55.6)	18 (22.2)	56 (69.1)	7 (8.6)	Chi^2; p<0.001

Table 6: Qualitative assessment of pattern of lesion enhancement

3.3.3 Quantitative Enhancement Characteristics

Five papers were identified that considered quantitative assessment of lesion enhancement. For each analysis, a region of interest (ROI) was drawn over the lesion. One group used semi-automated image segmentation; [127] the remainder used manual segmentation with either a FH_ROI or ellipsoid_ROI.[128-131] Two groups considered *in situ* and invasive malignancy separately, three groups all malignant lesions together. Methods of analysis varied between a direct measure of lesion pixel grey level intensity,[127 129] comparison of grey level intensity between lesion ROI with that of a background ROI,[128 130-132] or both[130]. Unlike with computed tomography (CT) there is no recognised grey-scale unit for CESM. Therefore, although methods of analysis were similar, the values generated vary dramatically between studies. This precludes meta-analysis of the data. Details of the studies are given in table 7.

Paper	View	ROI	Measure	Benign	All malignant	*in situ*	Invasive
Rudnicki, 2019	CC & MLO	Manual, ellipsoid	%RS_MLO %RS_CC SDNR_MLO SDNR_CC	3.3% ± 2.2 2.6% ±1.7 5.62 ±3.74 5.12 ± 3.15	-	3.3% ± 1.9 2.6% ± 2.0 5.69 ± 3.63 5.21 ± 3.6	5.5% ± 3.0 4.8% ± 3.1 9.31 ± 5.15 8.44 ± 4.14
Lobbes, 2018	CC only	Manual, ellipsoid	CGV	2030 AU ± 21.5	2065 AU ± 11.4	-	-
Deng, 2018	Either	Semi-automatic segmentation	Maximum (pixel value) 95th percentile 75th percentile mean	261.25 159.72 114.74 85.79	1022.20 744.57 588.78 475.05	-	-
Lv, 2020	NR	Manual, freehand	ROI rROI1 rROI2 rROI3	86.737 ± 21.900 38.453 ± 33.921 41.786 ± 35.864 51.705 ± 40.999	114.33 ± 29.918 82.120 ± 42.342 87.158 ± 45.943 95.101 ± 56.646	-	-
Liu, 2020	CC & MLO	Manual, ellipsoid	CNR1_Maximum CNR1_75th percentile CNR1_Mean CNR2_Maximum CNR2_75th percentile CNR2_Mean RSD_Maximum RSD_75th percentile RsD_Mean	10.1% 4.8% 2.7% 13.2% 5.2% 3.0% 215.1% 11.4% 11.0%	 48.5% 6.0% -9.7%	11.8% 8.1% 6.5% 10.4% 8.5% 6.2%	27.2% 9.2% 7.6% 26.0% 7.5% 6.4%

Table 7: Quantitative assessment of lesion enhancement intensity (%RS: percentage signal difference between enhancing lesion and background, SDNR: signal-difference-to-noise ratio, CGV: CESM Grey Value, ROI: lesion ROI, rROI: relative grey value, ROI1 – tissue surrounding lesion, ROI2 – tissue away from lesion, ROI3 – pectoralis, CNR: contrast to noise ratio, early image (CNR1) and late image (CNR2), RSD: relative signal difference)

3.3.3.1 Degree of enhancement: benign vs malignant

All studies concluded that malignant lesions enhance more avidly than benign lesions (p<0.05). However, it must be noted that two papers included non-enhancing lesions which were assigned a zero-value. In both instances these were disproportionately benign which will bias results.[127 131] Four of the papers analysed with ROC curves reporting area under the curve (AUC) values ranging from 0.700 – 0.877, and three calculated threshold values for discriminating benign from malignant lesions according to Youden's index.[127 128 130 131] Details are displayed in table 8.

Paper	Measure	Benign	Malignant	Threshold value	Sensitivity	Specificity	AUC	CI 95%	P value
Rudnicki, 2019	%RS_MLO %RS_CC SDNR_MLO SDNR_CC	3.3% ± 2.2 2.6% ±1.7 5.62 ±3.74 5.12 ± 3.15	NR	3.8% 3.4% 7.222 6.512	NR	NR	0.713 0.725 0.710 0.700	0.64; 0.79 0.65; 0.80 0.63; 0.79 0.62; 0.78	<0.0001 <0.0001 <0.0001 <0.0001
Deng, 2018	Pixel value	NR	NR	220.94	0.755	0.886	0.875	0.811; 0.940	<0.0001
Lv, 2020	ROI rROI1 rROI2 rROI3	86.737 38.453 41.786 51.705	114.33 82.120 87.158 95.101	87.815 42.519 55.225 71.079	0.663 0.713 0.762 0.772	0.798 0.838 0.732 0.612	0.795 0.833 0.812 0.741	0.745; 0.840 0.786; 0.873 0.763; 0.854 0.687; 0.790	NR
Liu, 2020	CNR1 CNR2	NR	NR	NR	0.814 0.930	0.755 0.529	0.843 0.755	0.773; 0.898 0.677; 0.822	<0.0001 <0.0001

Table 8: Quantitative assessment of lesion enhancement; ROC analysis with threshold values where provided.

Results for *in situ* disease are inconsistent; Rudnicki *et al* reported significant difference in parameters between invasive and *in situ* $p<0.008$, but no significant difference between benign and *in situ* lesions.[128] By contrast Liu *et al* described a significant difference in enhancement intensity between benign lesion and *in situ* disease p=0.001, with no significant difference between *in situ* and infiltrating disease p=1.000.[131]

3.3.3.1 Temporal enhancement characteristics

Two articles[127 131] compared the signal intensity of lesions between initial view and subsequent view. Methodology between the papers was similar. Deng *et al* consistently performed CC before MLO, Liu *et al* performed CC first in the majority (79/96) of cases. The interval time between views was 102 seconds and 104 seconds respectively. A change of less than 10% was considered stable enhancement in both papers. The majority of malignant lesions demonstrated decreasing or washout characteristics (67-71%), whereas benign lesions tended to demonstrate increasing or wash-in characteristics (48-58%). In both instances this trend reaches statistical significance ($p<0.05$). Whilst up to a third of benign lesions show wash-out characteristics, less than 20% of malignant lesions demonstrate wash-in.

3.3.4 Textural analysis

Ten papers from five research groups were included for review. A range of approaches have been taken. Some used textural features from RC images alone,[133-135] others extracted textural information from both RC and LE images and in four cases, the textural feature statistics were combined with additional clinico-radiological data.[135-138] The textural features extracted and machine learning methods varied. Therefore, whilst all papers report a measure of diagnostic accuracy, the heterogeneity of methodology precludes meta-analysis.

3.3.4.1 Benign vs malignant

Interest in the application of textural analysis and radiomics software to CESM images has increased over the last couple of years with a variety of approaches taken by differing research groups. The majority of the literature, seven papers from four groups, considered the differentiation of benign and malignant lesions. Case numbers for each study were small, mean=83.7 (range:49-139).

There was no consistency with regards to the textural features included in each model. Details are provided in five papers.[137-140] Lin *et al* provided the greatest detail, reporting 19 significant features, of which ten are from RC images: two STAT features, five Grey Level Size Zone Matrix (GLSZM) features and solitary features for GLCM, RLM and grey level dependence matrix (GLDM).[138] Patel *et al* reported ten top features, of which seven were textural features on RC images; three were GFB features, two GLCM features, a single DOST feature, and a HIST feature.[137] By contrast Fanizzi *et al* described two RC textural features; a WAV feature (variance) and STAT feature (relative smoothness).[139] This inconsistency in selected significant features is further emphasised by the work of Losurdo *et al*. They utilised two different techniques to evaluate feature importance, embedded and wrapper methods, and there was little overlap in the significant features identified by the respective methods.[140]

Measures of diagnostic accuracy ranged from 0.87-0.9 with sensitivity and specificity of 0.70-1.00 and 0.66- 0.92 respectively. ROC curve AUC values ranged from 0.76-0.95. Two papers compare CESM-models to human readers. Both reported an improved specificity but decreased sensitivity with the CESM-model.[137 139]

3.3.4.2 Histology of breast cancer

Three papers were identified that considered the possibility of using radiomics analysis to predict immunohistological features of breast cancers, including invasiveness, grade and hormone receptor (HR) status. Fanizzi *et a*l[139] reported highest accuracies for discriminating HER2+/HER2- (90.87%), ER+/ER- (83.79%) and Ki67+/Ki67- (84.80%). With respect to RC images, discriminating textural features identified were variation coefficient, relative smoothness (Ki67, ER, tumour grade); variation range (Ki67, ER) and entropy (Ki67, tumour grade). Marino *et al* reported accuracies for differentiating HR-positive and HR-negative subgroups (73.9-78.4%) in addition to accuracies of 79.4-87.4% for differentiating invasive and non-invasive tumours.[134] Further work by this group comparing CESM and MRI derived radiomics suggested that similar accuracies may be achieved in assessment of invasiveness, HR status and grade.[133]

3.4 Discussion

There is an increasing body of evidence that qualitative and quantitative assessment of enhancement characteristics of breast lesions seen on CESM is possible and may have important clinical applications.

Malignant lesions tend to demonstrate more avid enhancement, whether assessed by the reader ‘eye-balling’ the image whilst reporting, or assessed quantitatively using computer software. The difficulty however, is applying this evidence to clinical practice. For example, whilst the majority of malignant lesions demonstrate moderate-marked enhancement, on some series up to 68% of mildly enhancing lesions were also proven to be malignant.[121] This may be partly explained by the proportion of ILC and DCIS within this series (23% of malignant lesions), as these subtypes may be more challenging to identify and demonstrate weaker CESM-enhancement than IDC, similar to findings at MRI.[124 126 128 141] Further difficulties arise due to inter-reader variability in subjectively classifying enhancement intensity, which whilst described as ‘good’, resulted in significantly different findings for each reader.[124] Significant correlation between qualitative and quantitative assessment has been shown[132] though emerging evidence indicates that quantitative assessment may be more reliable and diagnostic thresholds for enhancement are

proposed.[128 130] This does not solve the fundamental issue that whilst a benign biopsy may be temporarily unpleasant, a missed cancer diagnosis can be catastrophic. Therefore, any method designed to reduce benign biopsy rates must not result in an increased number of missed cancers. Difficulties with work in this area include the small data-sets and the imbalance between benign and malignant enhancing lesions. This is partly because CESM tends to be used clinically when there is a higher index of suspicion for malignancy, compounded by the fact that benign lesions are less likely to enhance at all. Possibly as a result of this, some investigators include unenhanced lesions in their datasets.[127 131] This inevitably skews, and in some situations invalidates, the results. For example, when the data from Deng *et al* are analysed with unenhanced lesions removed, the mean grey level for enhancing benign lesions increases from 85.79 to 314.56, and for enhancing malignant lesions from 479.05 to 514.134.[127] In light of this, the suggested diagnostic threshold of 220.94 for differentiating benign and malignant lesions is obsolete.

Two further publications which included only enhancing lesions[128 130] proposed diagnostic thresholds using a ratio of the grey-level intensity values between the lesion and background ROIs. In both cases the proposed thresholds have relatively low sensitivities of 49% and 71.3% with higher specificities of 88% and 83.8% for Rudnicki *et al* and Lv *et al* respectively. Therefore, using these proposed diagnostic thresholds would result in an unacceptable number of false negatives or missed cancers. For this reason, whilst mild enhancement may decrease the readers' suspicion score it should not dissuade them from performing a biopsy.

Assessment of dynamic enhancement intensity may be better at discriminating benign lesions.[127 131] The limited evidence to date simply compares enhancement intensity on the initial view (usually CC) and subsequent view (usually MLO), with potential for confounding due to differing enhancement intensity by reason of the position of the breast. This could be developed to use a true delayed image which would potentially mirror wash-out characteristics seen on MRI.

Assessing pattern of enhancement is likely to further increase reliability, with heterogeneous lesions more likely to be malignant. Ring enhancement was not shown to be a useful discriminator. This could be because it was not sub-classified as nodular or smooth. Kamal *et al* postulate that this may be due to the presence of cysts; unlike with MRI, where cysts can be reliably identified on unenhanced

sequences due to central T2 hyper-intensity, it is not possible to gain equivalent information on CESM.[123]

Radiomics provides the possibility of combining quantifiable assessment of both intensity and pattern of enhancement. An increasing body of evidence suggests that this may be helpful to differentiate benign from malignant lesions. Unfortunately, individual study numbers are small and due to variations in methodology cannot be combined for meta-analysis. Furthermore, no research group to date has published a method that can be applied to CESM images generated by different vendors. Attempts have been made to establish a diagnostic cut-off point to differentiate benign from malignant lesions, however these have all concentrated on 'accuracy' according to Youden's index and failed to recognise that the requirement for high sensitivity outweighs the need for specificity in this instance. In both studies that compared human readers to CESM-models, the readers out-performed the model in terms of sensitivity. A further confounding factor is that whilst malignant lesions tend to be more heterogeneous in architecture than specific benign lesions types, for example fibroadenomas, classifying all benign lesions together introduces heterogeneity due to the wide range of possible benign lesions. However, to analyse benign lesions according to sub-type will require significantly larger CESM image databases than are currently available.

The greatest clinical application of radiomics may be in the non-invasive assessment of the whole tumour for prediction of tumour immunohistology. At present initial treatment decisions are based on core-biopsy, which only provides a small proportion of tumour for pathological assessment. Breast cancers are known to be heterogeneous, both in grade and HR status. If the least aggressive part of a tumour is biopsied the treatment offered may not be appropriate. Radiomics may offer a non-invasive alternative assessment of the whole tumour.

There is consistent evidence that benign and malignant lesions tend to demonstrate different enhancement characteristics - benign lesions tending to demonstrate weaker, more homogenous enhancement. Limited evidence suggests that the temporal enhancement profile also varies between malignant and benign lesions, with malignant lesions tending to demonstrate 'wash-out' or decreasing pattern of enhancement and benign a progressively enhancing one. The application of radiomics to CESM images is a promising technique for differentiating benign and malignant lesions, with potential to predict immunohistological features.

3.5 Summary

Future work is required to establish how this information can be applied in clinical practice. Any technique designed to reduce the number of benign biopsies performed requires a high sensitivity to avoid false negative assessments and therefore missed cancers - diagnostic threshold models should be adjusted accordingly. A radiomics model that can be applied to clinical images will require a much larger dataset and multi-vendor images. To allow future meta-analysis and strengthen the evidence base, fundamental methodological features should be assessed, such as method of segmentation, images used and textural features that are consistently discriminatory.

3.6 Key Research Questions

There was a dearth of research regarding textural analysis and CESM, and little published evidence regarding quantitative measures of enhancement characteristics at the inception of this research project. However, during the course of the research there have been several promising papers published which have been discussed above. Following this recent literature review, the following research questions have been generated

- Is it possible to build a model to discriminate benign from malignant lesions using HIST and GLCM texture features only?
- Is it possible to build a model to discriminate the immunophenotype of malignant tumours using HIST and GLCM texture features only?
- Is it feasible to apply the same textural analysis software to CESM images generated by different vendor mammography equipment (Hologic and GE)?
- What is the optimal method of segmentation: freehand or ellipsoid?
- Are both mammographic views, CC and MLO, required to build an accurate model to discriminate benign from malignant lesions
- Where delayed images are available is it possible to assess washout characteristics and correlate with malignancy, immunophenotype or response to NACT?

Chapter 4: Methods (CONDOR)

4.1 Study overview

This is an ethically-approved prospective, paired imaging comparison study of female patients, aged over 18 years, receiving NACT to treat invasive breast cancer (IRAS project ID 244564).

Alongside standard-of-care imaging with MRI, study participants were imaged using CE-DBT before, during and after completing NACT.

Potential candidates were identified at the breast multidisciplinary team (MDT) meeting when the decision to commence NACT was made. The clinician / nurse pre-screened with regard to the safety of administering iodinated contrast media and other exclusion criteria (see below) at a clinical appointment and eligible women were offered the patient information sheet (PIS). Women who accepted the PIS were contacted via telephone after a minimum of 24 hours to establish whether they wished to take part in the study. Formal written consent was acquired prior to the initial CE-DBT study.

Participants underwent imaging with CE-DBT alongside the gold standard 3-Tesla MRI. The CE-DBT and MRI were performed within the shortest interval possible, ideally within seven days of each other, in whichever order was logistically preferable on an individual basis. Following initial and end-of-treatment imaging patients were asked to complete a questionnaire regarding their experience of the imaging investigations.

The decision regarding final surgical plan was made by the MDT after reviewing all imaging and clinical findings. Pathological data was recorded from the core biopsies at diagnosis and the surgical specimen. In addition to routine analysis, the RCB score was calculated on the resection specimen. The study design is illustrated in the flow chart below;

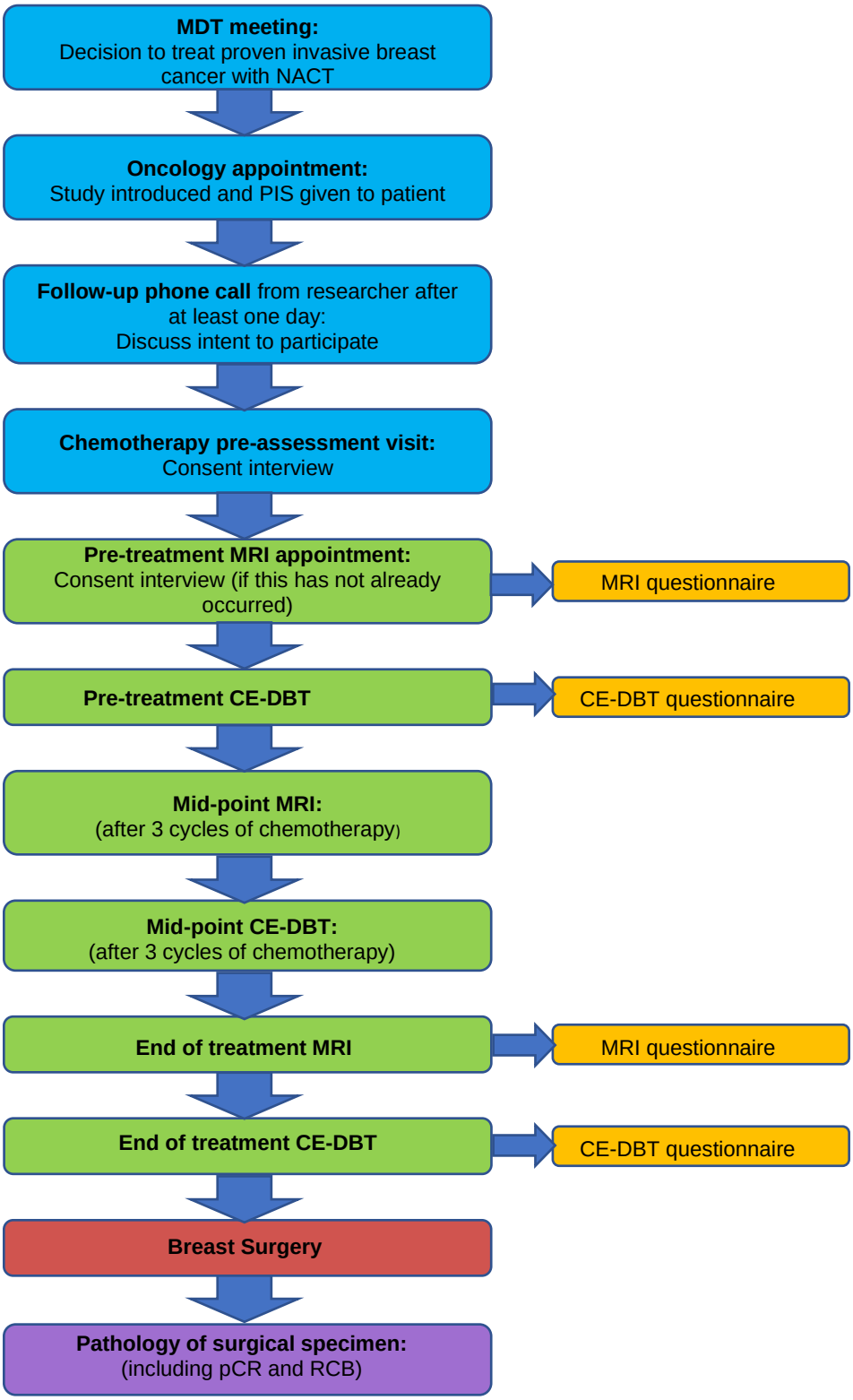


Figure 3: CONDOR study flowchart

4.1.1 Inclusion criteria

- Age 18 years or older
- Female
- Requiring imaging monitoring of response to NACT for biopsy proven invasive breast cancer
- Symptomatic and screen-detected breast cancer

4.1.2 Exclusion criteria

- Unable to give informed consent
- Patients with a contraindication to CE-DBT contrast agent (e.g. previous contrast reaction, iodine allergy, severe asthma, renal impairment)
- Previous breast cancer surgery or implants
- Pregnancy or lactation
- Contraindication to breast MRI

4.2 CE-DBT

For each CE-DBT study the participant had an IV cannula inserted into a dorsal hand vein or an antecubital vein by a trained practitioner. If the patient already had a venous cannula or peripherally inserted central catheter *in situ* for clinical reasons, this was used instead. Prior to contrast injection, patients were informed about sensations (e.g. a feeling of warmth throughout the body, a metallic taste in the mouth). Iodinated contrast material, omnipaque 300 (GE) 1.5mg/kg was then injected intravenously using a power injector at a rate of 2-3ml/sec.
All CE-DBT images were acquired using the commercially available Selenia Dimentions system (Hologic). After 3 mins, to allow time for contrast perfusion, imaging in the CE-DBT unit was commenced. The delayed MLO image was acquired 9 mins after injection. See figure 4.

For the pre-treatment CE-DBT the order of image acquisition was as follows:

- **Index breast: MLO CE-DBT (CESM then DBT), CC CE-DBT**
- **Contralateral breast: CC CE-DBT, MLO CE-DBT**
- **Index breast: MLO CESM ONLY***

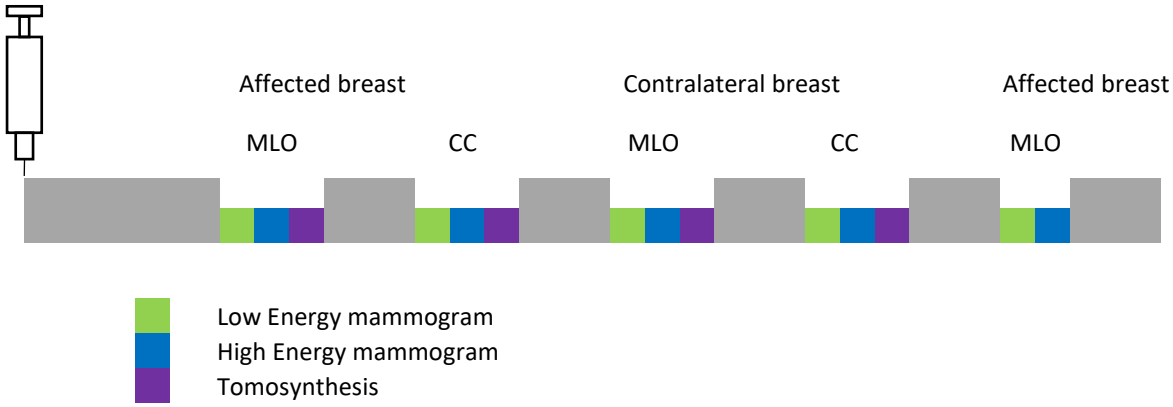


Figure 4: Diagram to illustrate pre-treatment image acquisition. *Following a protocol amendment, if the patient had known bilateral cancer delayed MLO images were performed on both breasts.

For the subsequent mid-treatment and end-of-treatment CE-DBT the contralateral breast was not imaged, except in cases of bilateral cancer. The participants were observed for 20 minutes following the procedure with the cannula left in case of delayed contrast reactions.

4.2.1 Reading protocol (CE-DBT):

Prior to commencing the study I attended an accredited course for CESM and DBT interpretation. I read the three components of the CE-DBT studies in strict sequential order. Lesions were scored according to the Royal College of Radiologists classification system ranging from 1 (malignant), 2 (benign) through to 5 (highly suspicious of malignancy).[142]

- The LE images were read first. The findings documented included size and location of lesion(s), presence and dimension of suspicious calcification and total disease extent.
- The DBT images were read next; any lesions identified had features, size and location recorded. Total disease extent was also measured.
- The RC images were read last and findings documented. For all lesions; shape, border, enhancement pattern were recorded. Change in lesion

enhancement between the initial and delayed MLO was subjectively assessed, and classified as increasing, stable or decreasing.

4.3 CE-MRI

In addition to CE-DBT the patient received standard-of-care imaging. This included a full field digital mammogram at diagnosis in addition to MRI before, during and after treatment. The patients did not require the standard-of-care mammogram after treatment as this was replaced by the low energy 2D image performed as part of the CE-DBT (as described above).

All CE-MRI were performed on Siemens 3T Prisma Fit scanner using a dynamic contrast-enhanced protocol. The sequences included; T1 2D axial high resolution, T2 axial turbo spin echo, diffusion sequences, T1 3D dynamic sequnces (2 pre-contrast and 7 post-contrast) and a delayed T1 axial high resoltion sequence.

4.3.1 Reading protocol (CE-MRI)

MRI scans were read by an experienced radiologist blinded to CE-DBT findings, but aware of the clinical and FFDM findings, (i.e. identical conditions to those under which CE-DBT was read). Lesion position, size and enhancement characteritics were recorded. As with CE-DBT, lesions were scored according to the Royal College of Radiologists classification.[142]Total 2D disease extent was documented. Lesion signal intensity was measured on the workstation for generation of TICs.

4.4 Pathological assessment

Histology data was recorded from the diagnostic core biopsy and surgical excision specimen. With regards to the diagnostic biopsy; the grade of invasive and, where present, *in situ* disease, tumour type and receptor staus were recorded. In addition the surgical pathology included tumour size (WTS and ITS), presence of vascular invasion, multifocality and lymph node status. The RCB score was subsequently calculated and RCB class assigned.

4.5 Patient Questionnaires

To assess patient acceptibility, participants were asked to complete questionnaires regarding their experience of CE-DBT and MRI, both following pre-treatment imaging and post-treatment imaging. The questions were identical on both questionnaires, as shown in the appendix. There were eight questions with a four-point categorical response scale. Two questions regarding anxiety and breast pain were assessed using a visual analogue scale (VAS); participants were asked to place a mark on a linear scale from 0-100. Finally, at both imaging time-points, patients were asked to indicate a preference. To capture preference based on patient experience, rather than expectation of the test accuracy, the preference question was prefaced with 'assuming CE-DBT and MRI provided equivalent diagnostic information'. Several free-text boxes were provided to allow the participants to expand on responses.

4.6 Outcome analysis

Statistical analyses were performed using SPSS (SPSS for Windows. 2017, v25. Armonk. NY: IBM Corp) and MedCalc (MedCalc for Windows, v20.011). Ostend, Belgium: MedCalc Software). Graphical analysis was performed on Microsoft Excel (Microsoft Excel for Mac. 2017, v15.30. Redmond, WA. Microsoft Corp.) and SPSS. Software was chosen according to availability of required functionality.

4.6.1 Background and feasibility

The study was necessarily exploratory given the novelty of the CE-DBT technique. The intention was to perform CE-DBT and MRI on 25 patients. As this was a feasibility study a formal sample size calculation was not performed. Feasibility assessment of likely recruitment rate for a future trial and the practicalities of performing the additional imaging within the clinical work flow were considered. Patients had both interventions (MRI and CE-DBT) therefore it was not necessary to compare background patient characteristics. Time between imaging techniques at each time-point and end-of-treatment imaging and surgery were compared using Wilcoxon Sign Rank test for non-parametric data.

4.6.2 Primary outcomes

4.6.2.1 Prediction of pCR status on end-of-treatment images

Patients with matched CE-DBT and MRI end-of-treatment imaging were included and maximum suspicious disease dimensions were recorded for all components of CE-DBT and MRI in each breast. Pathological results; ITS and WTS were considered the 'ground truth', and imaging findings were correlated to these. Analysis was conducted at 'breast level', in other words for patients with bilateral cancers the response in each breast was considered separately but in cases of multifocal, unilateral disease the combined maximum disease extent was measured. The components of CE-DBT; low energy (LE) mammogram, DBT and recombined CESM image showing contrast enhancement (CE) were first considered separately. To establish the benefit of the additional combined technique the maximum measurement for CESM(CE) plus microcalcification on LE mammogram: CESM(CE+calc) and maximum measurement for all components: CE-DBT was then considered.

Pathological complete response was defined as the absence of residual invasive disease within the breast (ypT0/is) as the presence of residual in-situ disease in the absence of invasive disease has been shown to have no adverse effect on survival or local recurrence rate.[87] The accuracy of imaging for assessing complete response was conducted at breast level, i.e. for women with bilateral cancers the response in each breast was considered separately. Consistent with previous work evaluating the diagnostic accuracy of MRI for identifying pCR, sensitivity was defined as the proportion of lesions demonstrating pCR at surgical excision with a corresponding imaging complete response; and specificity the proportion of lesions with residual invasive disease (non-pCR) with an incomplete response on imaging.[82] Diagnostic accuracy, sensitivity, specificity, positive predictive value (PPV) and negative predictive value (NPV) were calculated.

4.6.2.2 Prediction of WTS and ITS on end-of-treatment images

Concordance of residual WTS and ITS with size of residual disease as predicted on each imaging modality was considered. As in previous analysis both the individual components of CE-DBT and the combined scores were recorded. Analysis was

conducted at ‘lesion level’; the size of individual lesions, in cases of pathological multifocality, were considered separately. With respect to lesion size, wherever possible the largest tumour diameter on imaging (to nearest mm) was used for analysis. In a few instances of complex extensive disease, total disease extent on imaging was used. For example, when multifocal disease was described differently on the various imaging modalities, or when apparently multifocal disease was demonstrated to be contiguous on final pathology.
Results are displayed in scatterplots and Bland Altman plots. Concordance of residual WTS and ITS with size of residual disease as predicted on each imaging modality was calculated using Lins concordance coefficient[143].

4.6.3 Secondary Outcomes

4.6.3.1 Diagnostic accuracy for detection of multifocality on initial imaging

Pre-treatment imaging assesses local disease extent and can identify multifocal and bilateral lesions. Identifying additional disease, not seen on conventional imaging is important to ensure accurate monitoring and subsequent treatment. However false positives, when a suspicious area is identified on imaging but is subsequently proven to be benign, results in additional biopsies with associated morbidity and cost. When considering the diagnostic accuracy for detection of multifocality every additional lesion was considered a *positive*, all breasts with no additional lesions were considered a *negative*. A *false positive (FP)* was any ‘lesion’ with a suspicion score of 3 or above identified on imaging but subsequently proven to be benign or normal breast tissue. A *false negative (FN)* was defined as a lesion not identified or called benign on an imaging technique but subsequently proven to be malignant, either having been identified on alternative imaging or at final surgery. A *true positive (TP)* was any lesion correctly identified on imaging as malignant. A *true negative (TN)* was defined as a lesion correctly identified as benign on an imaging technique. Diagnostic accuracy, sensitivity, specificity, PPV and NPV were calculated.

4.6.3.2 Change in size between initial and interim imaging: predicting response

Change in maximum disease extent within each breast as measured on each imaging technique was compared to pCR-status and RCB-class to explore whether change in size by mid-treatment could predict a good pathological outcome. Both

absolute change in size and percentage change in size were considered. Due to small sample size RCB classes were grouped into RCB 0-I vs RCB II-III. Statistical analysis of categorical outcomes (pCR-status and RCB-class) was performed using the Mann Whitney U test for non-parametric data.

4.6.4 Assessment of patient acceptability and preference of CE DBT vs MRI

Acceptability and preference was assessed using patient questionnaires. The Wilcoxon sign rank test for related samples was used to assess for significant differences between the modalities on those questions using categorical response scales. Non-parametric VAS data was also analysed using a Wilcoxon signed rank test as recommended by Heller *et al*[144]. Binary outcome data was analysed using a McNemar test for related samples. The content of the free-text responses was summarised according to the subject.

Chapter 5: Methods (TACESM)

This is an ethically-approved, hypothesis generating, retrospective, multicentre image analysis study. CESM images were from two sites, Dundee and Nottingham. All images were of women aged 18 years and over who consented to the CESM procedure. (IRAS project ID: 266560)

5.1 CESM technique

CESM images were acquired at two centres. On the first site, Dundee, they were acquired as part of ethically approved imaging studies, CONDOR and CONTEST. Details of CONDOR are provided in the Chapter 4. CONTEST is a prospective imaging study comparing the diagnostic accuracy of CESM and MRI for local staging of patients with known breast cancer. Imaging for both CONDOR and CONTEST was performed using the Hologic protocol described in Chapter 4.

On the second site, Nottingham, all CESM images were acquired as part of routine clinical care. Images were acquired using the commercially available Senobright system (GE healthcare). Prior to image acquisition, 100ml of iodinated contrast (Niopam 300) was given at a rate of 3ml/sec via a pump injector. Following an interval of 2 minutes the initial image was acquired, order of image acquisition was usually index breast CC, contralateral CC, then bilateral MLO views. However, as they were not acquired as part of a clinical study a strict protocol was not followed.[145] The similarities and differences between the protocols are shown in the table 9 below.

Site	Vendor	Contrast medium	Vol. of contrast	Time to imaging	Order of image acquisition	Delayed MLO
Dundee	Hologic	Omnipaque 300	1.5mg/kg	3 mins	Index MLO Index CC Contralateral MLO Contralateral CC	Yes
Nottingham	GE	Niopam 300	100ml	2 mins	Variable	No

Table 9: Imaging protocols for CESM image acquisition

5.1.1 Image identification

Consecutive CESM studies were reviewed from both sites. Those that met the inclusion criteria below were allocated a study number and the raw DICOM (Digital Imaging and Communications in Medicine) CESM images pseudonymised and labelled accordingly.

5.1.2 Inclusion criteria

- Women aged 18 years or older
- Available pathology; either from core biopsy or surgical specimen
- Visible mass lesion on CESM images

5.1.3 Exclusion criteria

- No abnormal enhancement on CESM
- Non-mass enhancement only on CESM

5.2 Pathology

In the case of multifocal or bilateral disease, all lesions with corresponding pathology were included. The relevant pathology records were then reviewed and the immunohistology of the corresponding core biopsy and surgical specimen (where available) were documented. The images were subsequently linked to the pathology records for each individual according to the study number. The following details were recorded;

- Manufacturer of mammographic equipment (GE or Hologic)
- Biopsy suspicion score (B1-5)
- Final pathological diagnosis
- Where available subjective assessment of wash-in / out was recorded

Where the diagnosis was of breast malignancy the following additional details were recorded

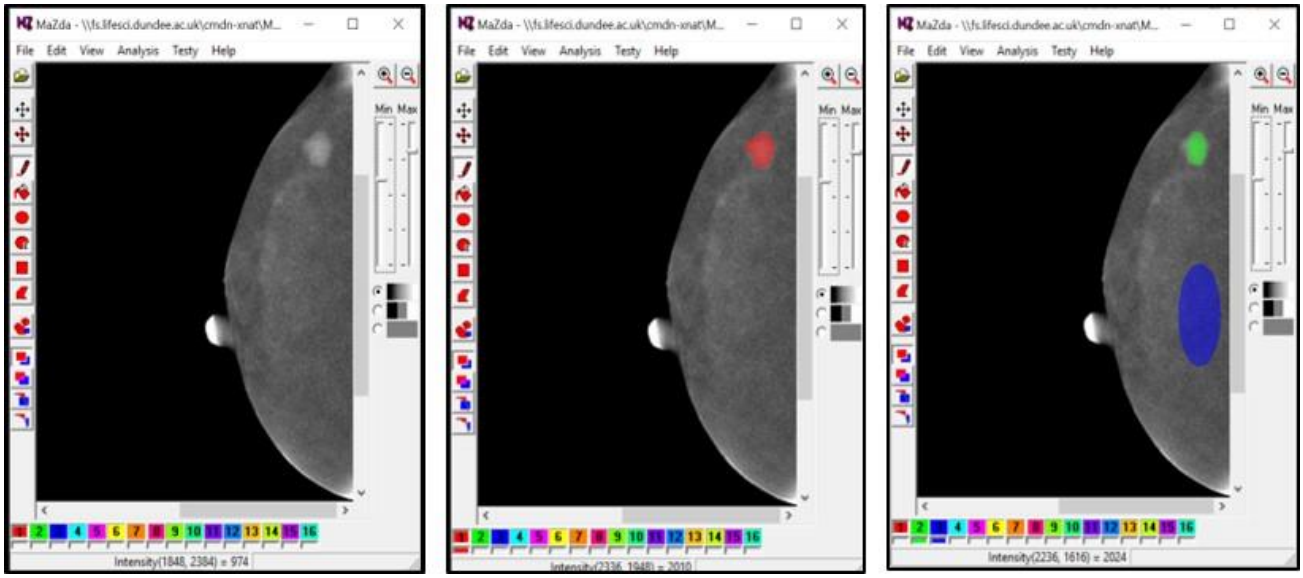
- Whether the patient received NACT
    - If yes; whether they had a complete pathological response
- Tumour histological subtype
- Tumour grade (from the surgical specimen)
- Receptor status (ER, PR and HER2)

5.3 Textural analysis

The pseudonymised raw DICOM images and immunohistological data were transferred to the University of Dundee, using an encrypted drive, for textural and statistical analysis.

Textural analysis (TA) was performed on the raw DICOM images using MaZda analysis software[146]. Where multiple lesions were present all lesions with available corresponding pathology were analysed. Lesions were identified on the pathology case report form by their position within the breast to ensure the pathology-radiology correlation. Whilst most lesions were visible on both views (CC and MLO), some were only visible on an individual view due to the position of the lesion within the breast. All views of lesions were included for TA.

As illustrated below (figure 5), for each lesion, on each view, two separate ROIs were drawn; a free-hand ROI (FH_ROI) encompassing the entire lesion, and an oval ROI (oval_ROI) covering the largest possible area of the lesion. Foreign bodies such as marker clips were excluded from the ROI. A further ROI representative of the grey-level signal of the normal breast tissue was drawn, labelled 'background'. This process was repeated for the second mammographic view. Figure 6 illustrates the hierarchy of data; from patient-level to ROI-level



1a. **1b.** **1c.**

Figure 5(a) RCC image loaded onto the MaZda software programme, (b) red freehand ROI (c) ellipsoid ROI (green) and background ROI (blue).

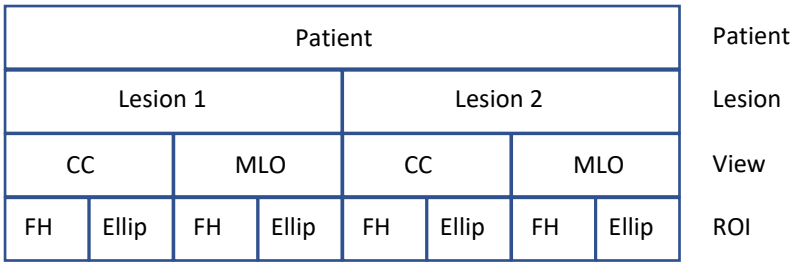


Figure 6: Hierarchy of data.

Despite extensive literature searches, no published data pertaining to textural analysis applied to CESM images was identified when the study protocol was prepared in 2017. Therefore this model was proposed based on textural analysis work conducted on breast MRI images to differentiate benign from malignant lesions, and malignant histological subtype classifications.[92] Prior to conducting textural analysis, grey-normalisation and re-binning was conducted, in-line with previous research.[93] Grey-scale normalisation was carried out by rescaling the data to fit within three standard deviations of the grey-level mean. For analysis of image intensity, it was necessary to re-bin the data to 6 bits/pixel. This was to reduce the number of grey-levels from the 2^{16} grey-levels present within a DICOM image, thus increasing the chances of identifying co-occurring pixel intensities. With regards to GLCM features the distance, or separation between pixel pairs was set at n=2 for fine texture and n=5 for coarse texture, as per published data on textural analysis in breast disease.[93] For each ROI first-order statistics based on histogram analysis were calculated, primarily to quantify the degree of enhancement, and GLCM statistics were generated to assess lesion heterogeneity.[96]

5.4 Machine Learning

Machine learning was undertaken in collaboration with Health Data Scientists, Utkarsh Agrawal (UA) and Adeniyi Francis Fagbamigbe (AFF), supervised by Prof. Colin McCowan, from the University of St. Andrews. SS gave information and advice on the different aspects of the data but UA and AF were responsible for the independent generation of the models. Models were developed using descriptive characteristics for results of the textural analysis and then discussed with SS who was solely responsible for clinical interpretation of the results. The following outcomes were investigated

- Lesion classification as Benign or Malignant
    - Model accuracy according to type of ROI (Freehand vs Ellipsoid)
    - Model accuracy according to mammographic view (CC vs MLO vs both)
    - Model accuracy according to vendor (Hologic vs GE)
- Invasive lesion classification according to tumour grade
- Invasive lesion classification according to hormone (oestrogen and progesterone) receptor (ER and PR) status and HER2 receptor status
- Pre-treatment lesion classification according to response to neoadjuvant chemotherapy (pCR vs non-pCR)

5.4.1 Classification technique

The classification technique adopted in this study is a machine learning algorithm technique used to determine the group in which a new object belongs by assigning class labels.[147] A training set where the final class labels are known, for example benign/malignant, was used to construct the classifier. We employed a four-layered Artificial Neural Network (ANN) classification algorithm, also known as multi-layer perceptron, a technique based on a back-propagation (experiential) learning method,[148] as illustrated in figure 7 below. The information received by the input layer consisted of the 22 TA features (input) in addition to the specific output measure of interest (final class label). This information was then forwarded to two hidden layers via synapses, and finally to the output layer. In the training set the final class label is known. The predicted output is compared with known outputs and the error is calculated. The network then learns by backpropagating the error and readjusting the weights. The learning continues until the error is below the 'error threshold'.

Using the classification rules, the goal of the ANN is to accurately predict the final class label of the test objects i.e. classifying outcome of interest. A technical description of the model was created by Health Data Scientists, AU and AF, and states that 'with a sigmoid activation function and adam optimisation with binary cross-entropy loss function to model the weighted sum, achieved the best results'. There is no overlap between the training and test sets.

The number of input values was fixed for each model, however the number of nodes (illustrated as circles in figure 7) within the hidden layers varied. These nodes were randomly assigned ‘weights’ at the beginning of each run. Information is transferred from each input value through the nodes in the hidden layers to the output layer. The relative weights of each node were then adjusted through backpropagation according to the accuracy of the output classification, this pathway (illustrated with red arrows) is referred to as an epoch. One run is comprised of 300 epochs, the number of passes of the entire training dataset the machine learning algorithm has completed.

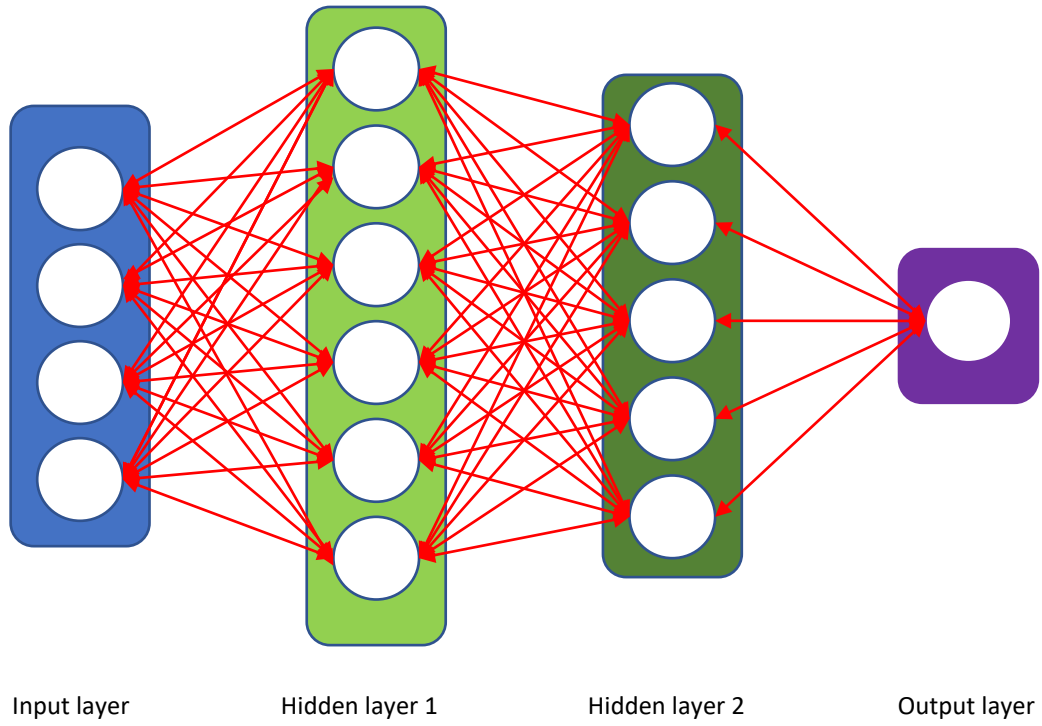


Figure 7: Graphical representation of an ANN. The white circles represent nodes; total red arrows represent one epoch.

5.4.2 Feature Engineering

The input layer consisting of individual TA outputs generated from the GLCM and HIST were considered features. These features have different ranges of values which can skew the results while classifying. The features were updated using a process known as z-score normalisation, by subtracting the mean in the numerator (zero-mean) and dividing by the standard deviation (unit-variance), to ensure that all features used the same scale; 0-1.

Correlation between features was checked to assess the need for feature reduction. A high correlation indicates that the features are related indicating the need for feature reduction, while a low correlation indicates features are not related. It must be noted that unlike similar studies[133 134 137 139] the number of TA input features was optimised to 22 based on previous MRI.[92] Correlation among the features was not high i.e. features were less dependent on each other. Therefore, following advice from the data scientists, feature reduction was not performed as doing so would likely result in poorer model accuracy. The possibility of feature reduction to identify the key TA features for each model could be explored in future work with a larger dataset.

5.5 ANN models: Statistical Analysis

For each analysis, the classification of one outcome was compared with another for example benign vs malignant, using the extracted textural features. The technique for generating the ANN model remained consistent for each analysis. However, the models produced were inevitably different with varying numbers of nodes in each hidden layer. The proportion of data randomly selected for training and testing also differed between models.

The results were reported using parameters model accuracy, F1 score, sensitivity, specificity, PPV and NPV. F1 score is a measure of model accuracy for a dataset based on precision and recall relative to a specific positive class. The results show average accuracy with standard deviation and 95% confidence interval (CI) for 30 runs. Python 3.7 on Jupyter notebooks was used for all the coding analysis.

5.5.1 Benign vs Malignant

Due to the small number of benign lesion samples, oversampling was performed to balance the number of samples to aid the classification. Random samples for benign lesions were generated to balance the class distribution at 2:1 (malignant:benign).

Subset analysis was conducted using separate models to compare accuracy for:

- Type of ROI: FH_ROI vs Ellipsoid_ROI: During each run 75% of the data was randomly selected for training, and the remaining 25% for testing. The

model had 100 and 50 hidden nodes for the two hidden layers respectively, with 300 epochs during each run.

- Mammographic view: CC vs MLO vs combined (CC+MLO): During each run 75% of the data was randomly selected for training, and the remaining 25% for testing. The model had 100 and 300 hidden nodes for the two hidden layers respectively, with 300 epochs during each run.
- Vendor: Hologic vs GE: During each run 75% of the data was randomly selected for training, and the remaining 25% for testing. The model had 100 and 50 hidden nodes for the two hidden layers respectively, with 300 epochs during each run.

A test of equality of proportions was used to establish whether there was a significant difference in accuracy between the models for each subset analysis.

5.5.2 Tumour Grade

Using the same technique, models were developed to classify lesions according to tumour grade. Due to extremely small numbers of grade 1 tumours it was not possible to use this as a separate category. Therefore, grade 1 & 2 tumours were combined to form a 'low grade' (LG) group and compared to grade 3 tumours which were considered high grade (HG).

Since the ratio of class distribution was balanced, oversampling was not required with this outcome. During each run 80% of the data was randomly selected for training, and the remaining 20% for testing. The model had 100 and 50 hidden nodes for the two hidden layers respectively, with 300 epochs during each run.

5.5.3 Hormone receptor status

Due to the exploratory nature of this work and small numbers in certain subgroups each receptor was considered in isolation. It was not possible to compare derived breast cancer subtypes. Thus, separate models were developed to classify lesions as ER+ vs ER-, PR+ vs PR-, HER2+ vs HER2-.

Due to the small number of ER- and HER2+ samples, oversampling was performed to balance the number of samples to aid the classification. Random sample were generated to balance the class distribution at 2:1 (ER+:ER- and HER2-:HER2+).During each run 80% of the data was randomly selected for training, and

the remaining 20% for testing. Each model had 100 and 50 hidden nodes for the two hidden layers respectively, with 300 epochs during each run.

5.5.4 Response to Neoadjuvant Chemotherapy

A predictive model was developed to classify lesions into those that completely responded to neoadjuvant chemotherapy (pCR) and those that had residual disease (non-pCR), based on textural analysis features of the baseline CESM.
Due to the small number of pCR samples, oversampling was performed to balance the number of samples to aid the classification. Random samples were generated to balance the class distribution at 2:1 (non-pCR:pCR). During each run 80% of the data was randomly selected for training, and the remaining 20% for testing. The model had 50 and 100 hidden nodes for the two hidden layers respectively, with 300 epochs during each run.

5.6 Dynamic tumour enhancement with CESM

Dynamic enhancement characteristics produced by MRI can be useful in predicting malignancy, histological grade of a lesion and response to neoadjuvant chemotherapy. Therefore, I analysed quantitative lesion enhancement on initial and delayed CESM images to investigate whether it may be possible to derive similar information. It was not possible to compare the enhancement characteristics of benign and malignant lesions or to generate cut-off values for wash-out characteristics due to the paucity of benign lesions. Consequently, the cut-off of +/- 10% change previously described for MRI, and more recently CESM, was used to differentiate type 1, 2 and 3 curves.[51 127 131]
The 95%CI, 90%CI and mean enhancement values from histogram analysis were used; lesion freehand ROIs (FH_ROI) and ellipsoid ROIs (ellipsoid_ROI) were considered separately. The following indices were calculated:

5.6.1 Relative lesion enhancement (RLE):

This is a measure of early enhancement, as no pre-enhanced subtracted images are produced by CESM, background enhancement was used as an alternative.

$$RLE = \frac{\text{Initial enhancement-background enhancement}}{} \times 100$$

Background enhancement

5.6.2 Washout Index (WI):

This is a measure of later enhancement characteristics, similar to MRI but with background ROI used instead of pre-enhancement signal intensity.

$$WI = \frac{\text{Initial enhancement-delayed enhancement}}{\text{Background enhancement}} \times 100$$

5.6.3 Relative enhancement difference (RD):

This may be more appropriate for CESM due to the lack of pre-enhanced views, and early promising results have been shown[131]

$$RD = \frac{\text{delayed enhancement-initial enhancement}}{\text{Initial enhancement}} \times 100$$

5.6.4 Subtracted relative enhancement difference (SRD):

First, subtracted values were calculated for both initial and delayed images by subtracted background ROI from lesion ROI, thus producing an indication of the excess enhancement seen in the lesion. These values were then used to calculate the subtracted relative enhancement difference.

$$SRD = \frac{\text{Subtracted delayed enhancement-subtracted initial enhancement}}{\text{Subtracted initial enhancement}} \times 100$$

5.6.5 Enhancement Curve:

Curve types were generated using the SRD. Due to the narrow spectrum of greyscale values seen in CESM, within the extremely large possible range neither the WI nor RD can generate values close the 10% cut-off used in MRI. Type I: SRD < -10%, Type II: SRD -10% – 10%, Type III: SRD > 10%

It was hypothesised that more aggressive lesions would demonstrate rapid wash-in followed by washout characteristics, as seen on MRI. In other words, a higher RLE followed by a negative value for RD and SRD. By contrast more benign lesions would demonstrate a lower RLE followed by a positive RD and SRD. It was

hypothesised that WI would be higher amongst more aggressive lesions. The following relationships were assessed:

- CESM and MRI enhancement characteristics: Due to the lack of benign lesions in our dataset it was not possible to directly assess and compare the enhancement characteristics of benign and malignant lesions. Therefore, the MRI TIC was used as a surrogate marker.
- Qualitative and quantitative CESM enhancement characteristics
- CESM dynamic contrast features and tumour grade: this was only performed on lesions treated with primary surgery. Patients treated with NACT were excluded as chemotherapy can affect tumour grade at surgery.
- CESM dynamic contrast features at pre-treatment and mid-treatment and complete pathological response

Statistical analysis of continuous variables was performed using either a two-tailed Mann Whitney U test for non-parametric data or Kruskal-Wallis test for independent samples, for data with two or three possible outcome categories respectively. Analysis of ordinal data was performed using linear weighted Cohen's Kappa.

Chapter 6: Results (CONDOR)

6.1 Background & Feasibility

Forty-one women were identified as potential participants following the decision to consider NACT at the MDT meeting. The recruitment pathway is illustrated below.

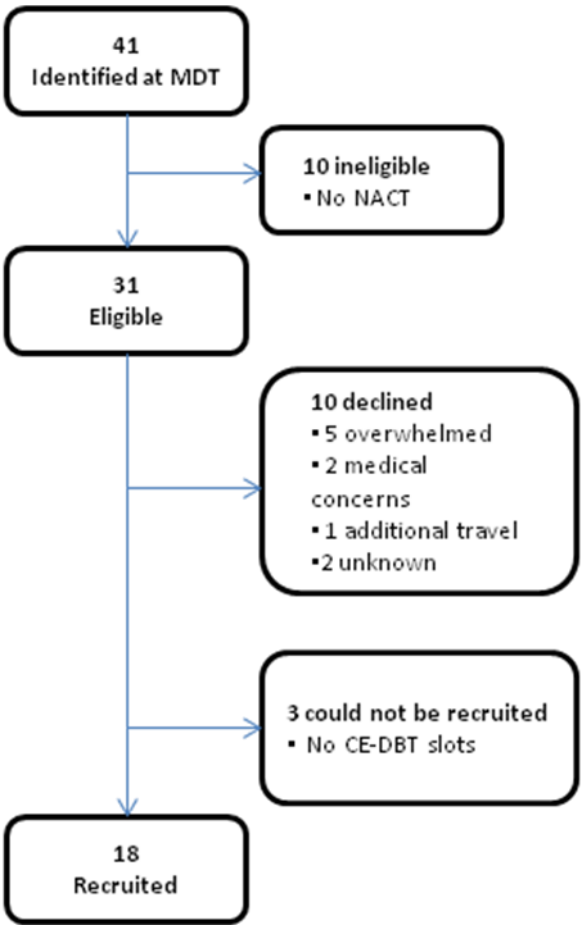


Figure 8: CONDOR recruitment pathway

Recruitment was closed early due to the onset of the COVID pandemic. Ultimately eighteen patients were recruited, with a recruitment rate of 58% of eligible patients. The average age of participants was 52.7 years (range 32-72 years). Of note three patients could not be recruited due to logistical issues; there was insufficient time and availability of equipment to arrange the pre-treatment CE-DBT between the oncology appointment when the decision to treat with NACT was made, and the first chemotherapy cycle.

6.1.1 Tumour features

Thirteen women had unifocal disease, three had unilateral multifocal disease (two tumours), and two had bilateral disease. One of the women with bilateral disease had three distinct tumours. In total, there were 24 invasive cancers in 18 women. Twenty-three (96%) of the cancers were invasive ductal carcinoma no special type (IDC NST). There was a single invasive lobular cancer (ILC), this tumour was occult on all imaging in a patient with bilateral IDC.

Tumour features are illustrated in the figure 9, below.

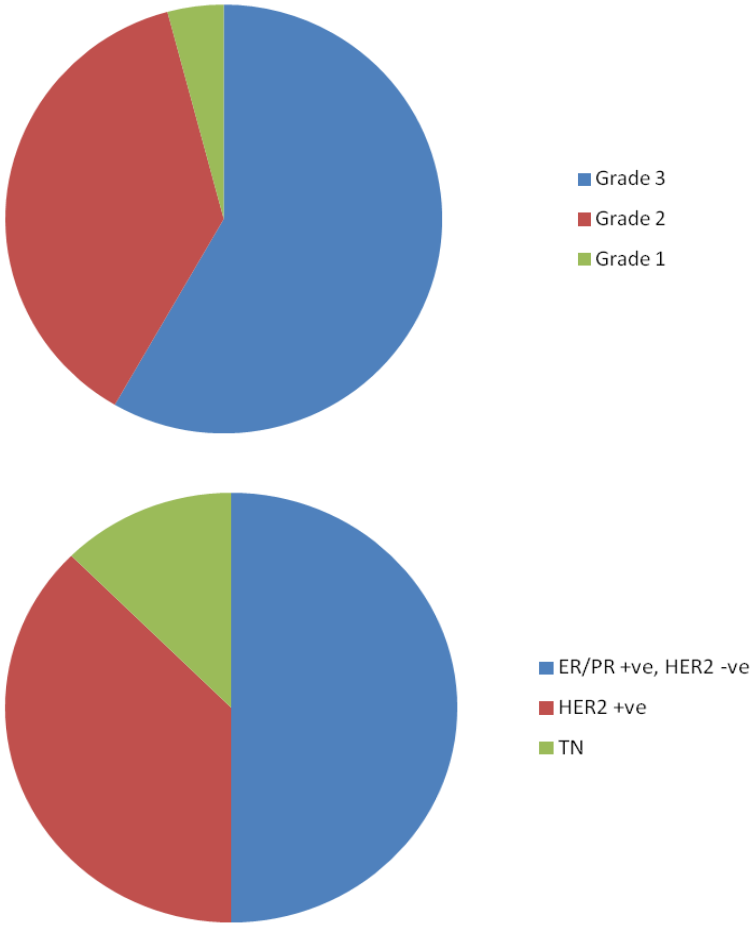


Figure 9: Tumour grade and receptor status

With respect to invasive tumour grade: fourteen (58%) were grade 3, nine (38%) grade 2 and one (45) grade 1. The single grade 1 tumour was an unexpected finding in a woman with contralateral grade 3 IDC NST. Regarding receptor status; twelve (50%) were ER/PR+ve HER-ve, nine (37%) were HER2+ve and three (13%) were triple negative.

6.1.2 Chemotherapy regimes

The chemotherapy regimens received by the 18 participants are detailed below.

n	FEC 75	Taxane	Cyclophosphamide	Reason
12	3	3 (Docetaxal)	0	Local standard of care
1	3	7 (1 Docetaxal, 6 Paclitaxel)	0	Reaction to Docetaxel (rash)
1	3	5 (Docetaxal)	0	Bone metastases, palliative
2	6	0	0	Unknown
2	0	4	4	Frail, comorbidities

Table 10: Chemotherapy regimes

One participant (study ID 15) who developed bone metastases whose treatment became palliative had five cycles of docetaxel prior to surgery. The final two cycles were given following end-of-treatment imaging (MRI alone). This case was therefore excluded from analysis of final imaging and imaging-pathology correlation. The remaining patients had imaging after cycle 3 and at the end of treatment (irrespective of total number of cycles of chemotherapy)

6.1.3 Adverse events and study withdrawal

There were no significant adverse events, including contrast reactions. One patient (study ID 03) withdrew from the study at mid-treatment due to difficult intravenous access and pain at previous injection site.

Participants with missing or incomplete studies are detailed in table 11.

Study ID	Modality	Time point	Reason
03	MRI, CE-DBT	Mid-treatment, End-of-treatment	Withdrew
06	MRI	End-of-treatment	Not standard of care, 4 cycles of chemotherapy
07	MRI	End-of-treatment	Not standard of care, 4 cycles of chemotherapy
04	Delayed CESM	Pre-treatment	Unilateral delayed imaging, bilateral cancer
06	Delayed CESM	Mid-treatment	Equipment failure
15	CE-DBT	End-of-treatment	Developed metastases, palliative, COVID
10	DBT, CESM	Mid-treatment	Human error
17	CE-DBT	Mid-treatment	COVID
18	CE-DBT	Mid-treatment	COVID

Table 11: Details of incomplete studies

As described above; two participants received four, rather than six cycles of chemotherapy, therefore following the local standard-of-care pathway they did not receive an end-of-treatment MRI. Due to the COVID pandemic it was not possible to offer mid-treatment CE-DBT to two participants and end-of-treatment CE-DBT to a third participant especially as she had developed metastases and become palliative. One patient only had a digital mammogram (without contrast) at mid-treatment due to human error. A further patient did not have a delayed image due to equipment failure. Of note, one patient with bilateral cancer only had unilateral delayed images at pre-treatment in accordance with the original protocol. A protocol amendment was subsequently made to allow bilateral delayed CESM in women with bilateral tumours.

Therefore, there were 47 patient episodes with *matched* imaging (i.e. CE-DBT and MRI). 18 at pre-treatment, 15 at mid-treatment and 14 at end-of treatment. Twelve patients had imaging with both techniques at all time-points.

6.1.4 Timing of imaging and surgery

The interval between the CE-DBT and MRI (in any order), and the interval from imaging to surgery is illustrated in the table below.

	Interval between CE-DBT & MRI (days)			Interval between imaging and surgery (days)	
Study ID	Initial	Mid-treatment	End-of-treatment	CE-DBT	MRI
01	7	11	4	27	23
02	18	6	0	33	33
03	18	-	-	-	-
04	13	8	5	15	20
05	15	3	6	14	20
06	6	4	-	14	-
07	0	4	-	7	-
08	31	0	2	22	24
09	7	0	0	28	28
10	6	2	0	24	24
11	7	3	0	31	31
12	0	4	4	34	30
13	11	8	3	26	29
14	7	1	4	42	38
15	11	3	-	-	64
16	14	6	6	13	19
17	1	-	4	15	19
18	0	-	13	36	23
Mean (all)	9.56	4.20	3.64	23.81	28.33
Range (all)	0-31	0-11	0-13	7 - 42	19 - 64
Mean (matched)				25.71	25.79
Range (matched)				13 - 42	19 - 38

Table 12: Interval between imaging techniques, and imaging techniques and surgery

The interval between CESM and MRI at pre-treatment was significantly longer than the interval at both mid-treatment ($p = 0.21$) and end-of treatment ($p = 0.29$). However, there was no significant difference between mid- and end-of-treatment imaging, $p = 0.77$.

There is an outlier for in the pre-treatment imaging (study ID 08). This participant had CE-DBT at diagnosis as part of a different research trial, CONTEST. The decision to give NACT and pre-treatment MRI were delayed as the patient did not attend a number of subsequent appointments, hence there was a gap of 31 days between pre-treatment CE-DBT and MRI. When this outlier is removed from analysis, the difference in imaging interval between pre- and mid-treatment remains significant (*p*

= 0.04) however the difference between pre- and end-of-treatment is no longer significant $p = 0.57$.

Time from final imaging to surgery for all patients was on average 23.81 days (range 7 – 42) and 28.33 days (range 19 – 66) for CE-DBT and MRI respectively. The outlier (study ID 15) had an additional two cycles of chemotherapy following end of treatment MRI, as explained previously. For cases with matched data the time to surgery did not significantly vary 25.71 days, (13 – 42 days); 25.79 days (19 – 38 days) for CE-DBT and MRI respectively, $p = 0.711$.

6.2 Primary Outcomes

6.2.1 Prediction of pCR on post-chemotherapy images

Only patients with matched CE-DBT and MRI end-of-treatment imaging were included in this analysis. Of 16 breasts with cancer in 14 women, 10 demonstrated a complete pathological response (pCR), with a prevalence of 62.5%. Sensitivity was defined as the proportion of lesions demonstrating pCR at surgical excision with a corresponding imaging complete response; and specificity the proportion of lesions with residual invasive disease (non-pCR) with an incomplete response on imaging The diagnostic accuracy of each imaging modality in predicting pCR is illustrated in table 13.

Modality	Imaging response	Pathological response		Sensitivity (95% CI)	Specificity (95% CI)	PPV (95% CI)	NPV (95% CI)	Diagnostic accuracy
		pCR	Non-pCR					
LE mammo	CR	5	3	55.56 21.20 – 86.30	57.14 18.41 – 90.10	62.50 37.17 – 82.44	50.00 27.44 – 72.56	56.25 29.88 – 80.25
	Non-CR	4	4					
DBT	CR	5	2	55.56 21.20 – 86.30	71.43 29.04 – 96.33	71.43 40.31 – 90.25	55.56 34.42 – 74.86	62.50 35.43 – 84.80
	Non-CR	4	5					
CESM (CE)	CR	9	3	100.00 66.37 – 100.00	57.14 18.41 – 90.10	75.00 56.05 – 87.59	100.00	81.25 54.35 – 95.95
	Non-CR	0	4					
CESM (CE+calc)	CR	5	3	62.50 24.49 – 91.48	57.14 18.41 – 90.10	62.50 37.78 – 82.06	57.14 30.72 – 80.04	60.00 32.29 – 83.66
	Non-CR	3	4					
CE-DBT	CR	5	2	55.56 21.20 – 86.30	71.43 29.04 – 96.33	71.43 40.31 – 90.25	55.56 34.42 – 74.86	62.50 35.43 – 84.80
	Non-CR	4	5					
MRI	CR	4	1	44.44 13.70-78.80	85.71 42.13 – 99.64	80.00 36.13 – 96.59	54.55 38.33 – 69.85	62.50 35.43 – 84.80
	Non-CR	5	6					

Table 13: Diagnostic accuracy for predicting pCR according to imaging technique

The greatest diagnostic accuracy, 81.25% was seen with CESM(CE). This technique identified all patients who had a pCR, however three patients with incomplete pathological response were considered to be complete responders due to lack of residual enhancement. In two of these breasts the foci of residual invasive disease

measured 6mm or less, as shown in figure 10. In one case, there were foci of invasive disease over an area on 72mm, however, the pathology in this case is reported as ‘marked and almost complete pathological response to neoadjuvant chemotherapy…there has been more than 90% loss of tumour cellularity.’

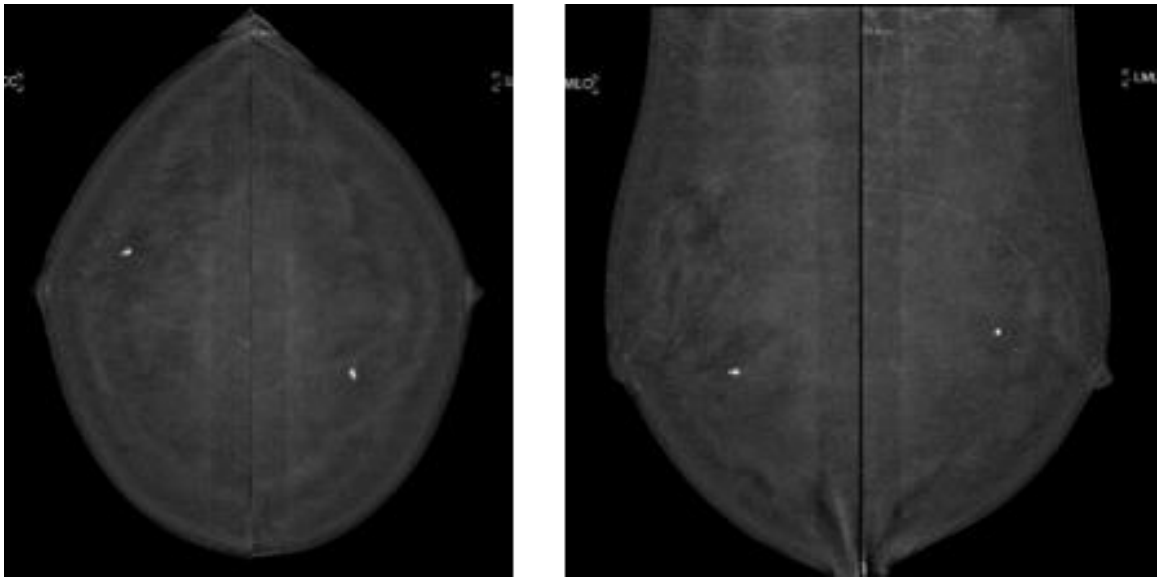


Figure 10: End-of-treatment CESM in a patient with bilateral cancer demonstrating no residual enhancement.
In the right breast 6mm residual grade 1 IDC was present (false negative), in the left breast 1mm residual DCIS, no residual invasive disease (true negative)

LE mammogram had lower sensitivity for pCR with identical specificity, thus the combination of microcalcification and enhancement did not improve accuracy, resulting in a lowered sensitivity with no improvement in specificity and therefore a lower overall diagnostic accuracy.

DBT, CE-DBT and MRI had equivalent accuracy for pCR of 62.50%. These techniques demonstrated lower sensitivity than CESM (CE) but higher specificity. In the case of DBT two cases of residual invasive disease were not identified, one patient with two foci of 0.8mm and 4mm, and the patient with the 72mm area showing almost complete response described above. These three cases were the same cases that CESM(CE) failed to identify. MRI failed to identify a 6mm site of invasive disease. However, DBT and MRI only identified 5 (55.6%) and 4 (44.4%) patients with pCR respectively, resulting in a lower sensitivity. The results of a combined CE-DBT study were identical to that of DBT alone, with an incremental increase in specificity but larger drop in sensitivity.

6.2.2 Prediction of residual WTS on post-chemotherapy imaging

Correlation between whole tumour size (WTS) and the size of residual tumour predicted on each imaging modality are displayed in the figures 11 & 12 and table 14 below.

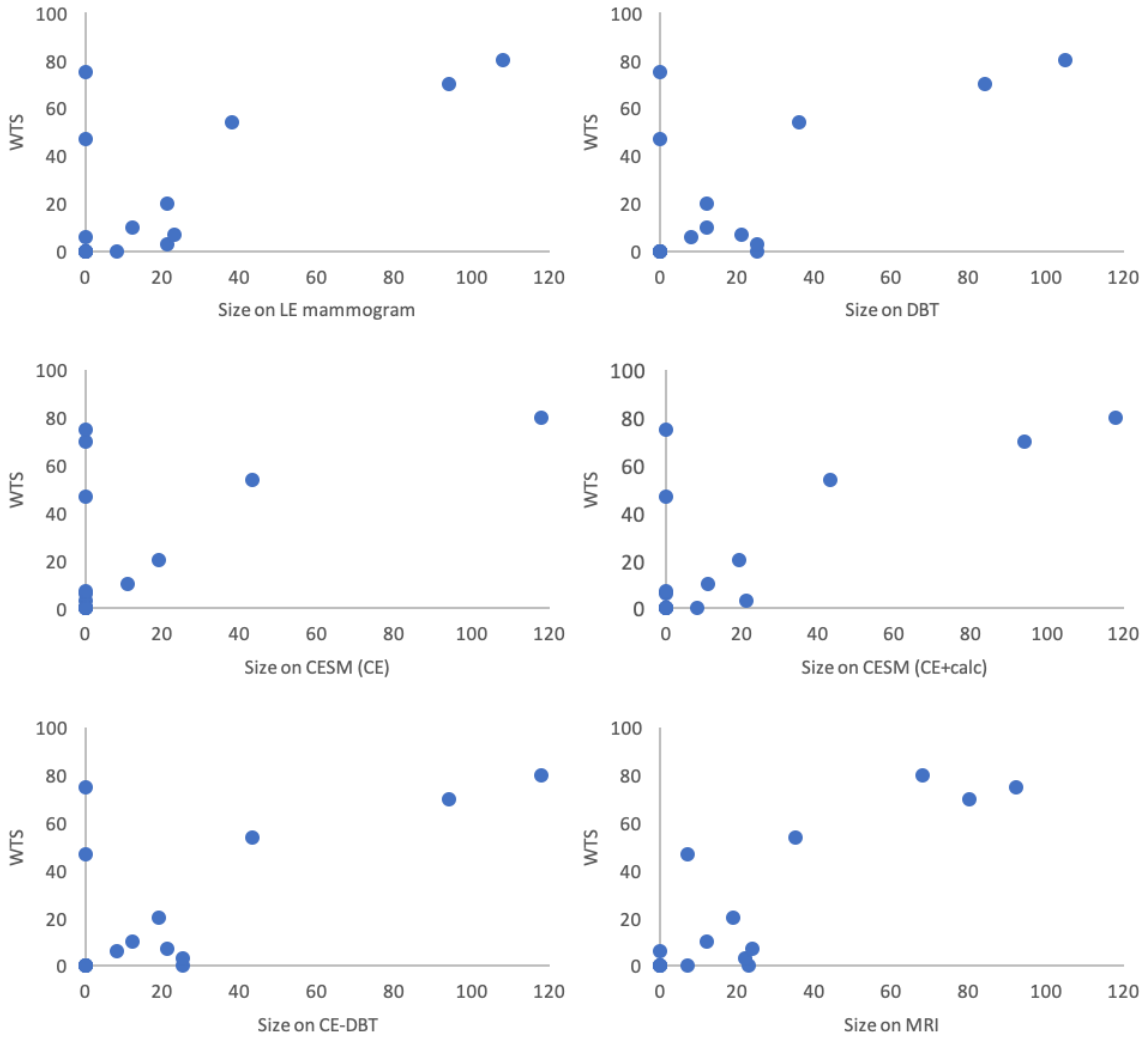


Figure 11: Final imaging size vs whole tumour size

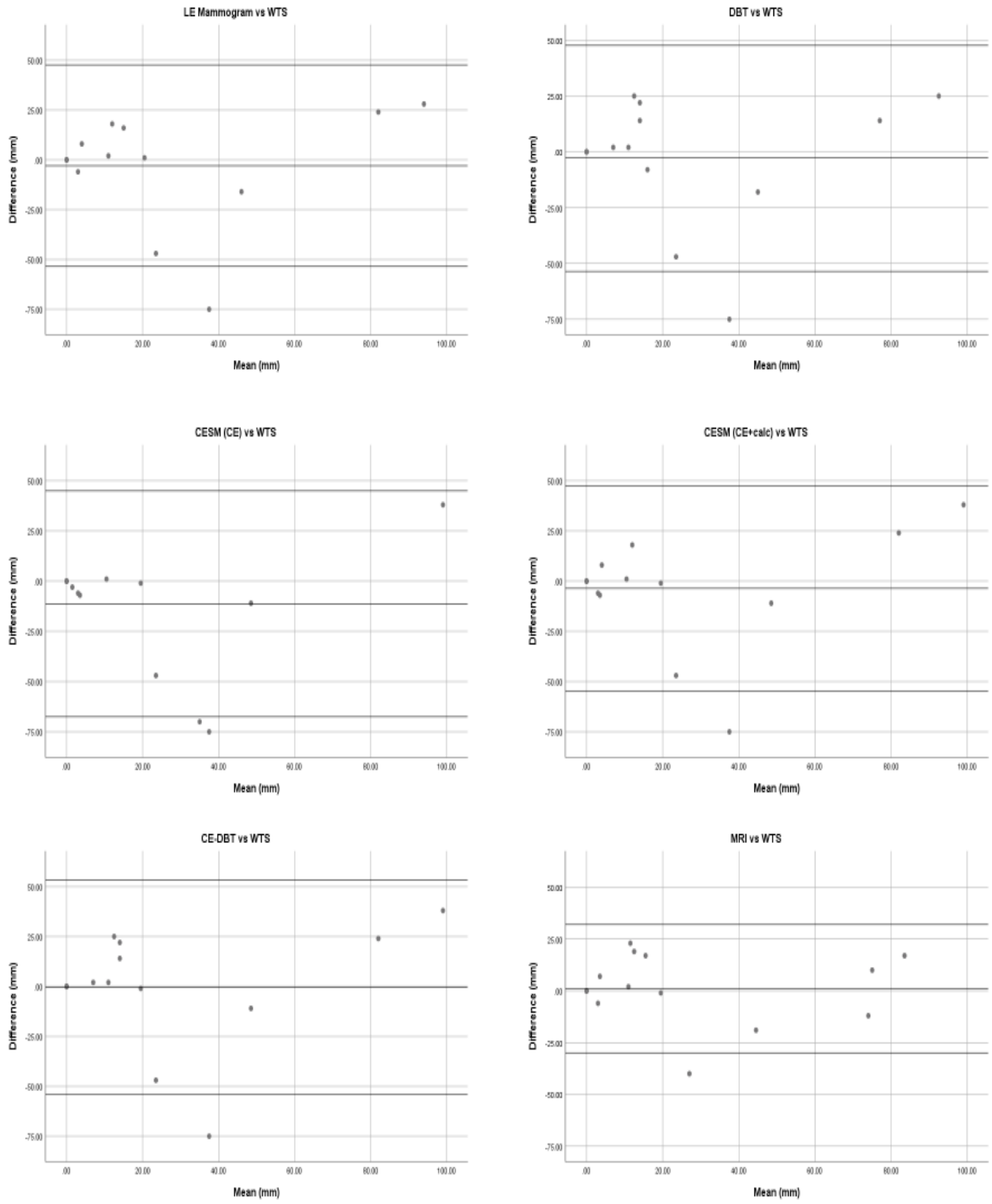


Figure 12: Bland Altman plots illustrating the difference between final imaging size and whole tumour size.

	Mean lesion size (mm)	Imaging size-WTS (mm)	Lin's concordance coefficient	95% Confidence Interval	
				Lower	Upper
WTS	21.9				
LE Mammogram	19.1	-2.8	0.68	0.33	0.86
DBT	19.3	-2.6	0.64	0.27	0.85
CESM (CE)	11.2	-10.7	0.52	0.12	0.78
CESM (CE+calc)	18.5	-3.4	0.69	0.36	0.87
CE-DBT	21.5	-0.4	0.65	0.29	0.85
MRI	22.8	1.0	0.86	0.67	0.95

Table 14: Lin's Concordance co-efficient for predicting whole tumour size according to imaging modality.

When considered separately, the components of CE-DBT confer similar reliability, with concordance coefficients for LE mammography, DBT and CESM(CE) of 0.68, 0.65 and 0.53 respectively. The combined assessment CESM(CE+calc) increases concordance to 0.70. No benefit is seen when combining with DBT, with an overall CE-DBT concordance of 0.67. MRI confers the strongest concordance (0.87).
Due to the small sample size, the confidence intervals are relatively large, however the mean values suggest MRI tends to over-estimate disease, whereas CE-DBT tends to under-estimate.

6.2.3 Prediction of residual ITS on post-chemotherapy imaging

As with WTS, correlation between invasive tumour size (ITS) and the size of residual tumour predicted on each imaging modality are displayed in the figures 13, 14 and table 15 below.

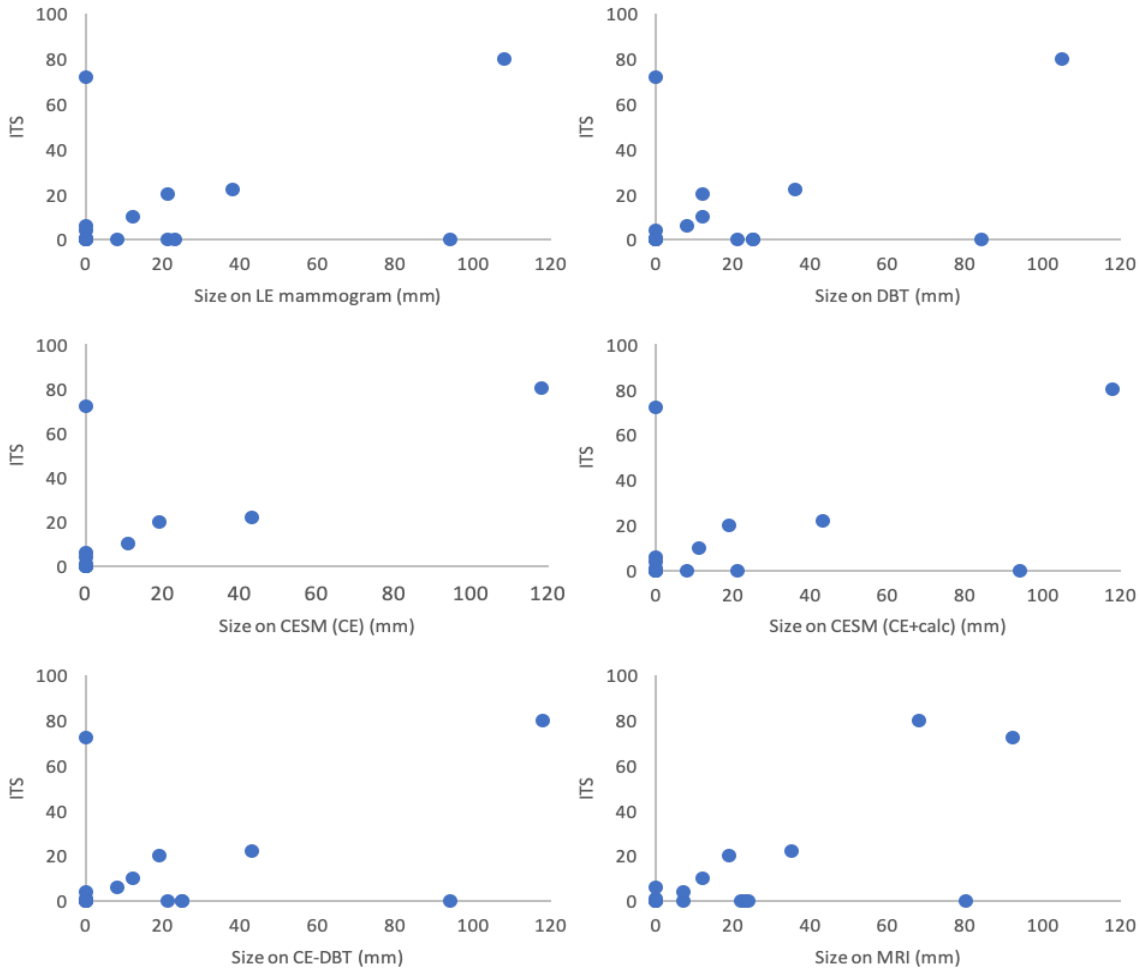


Figure 13: Final imaging size vs invasive tumour size

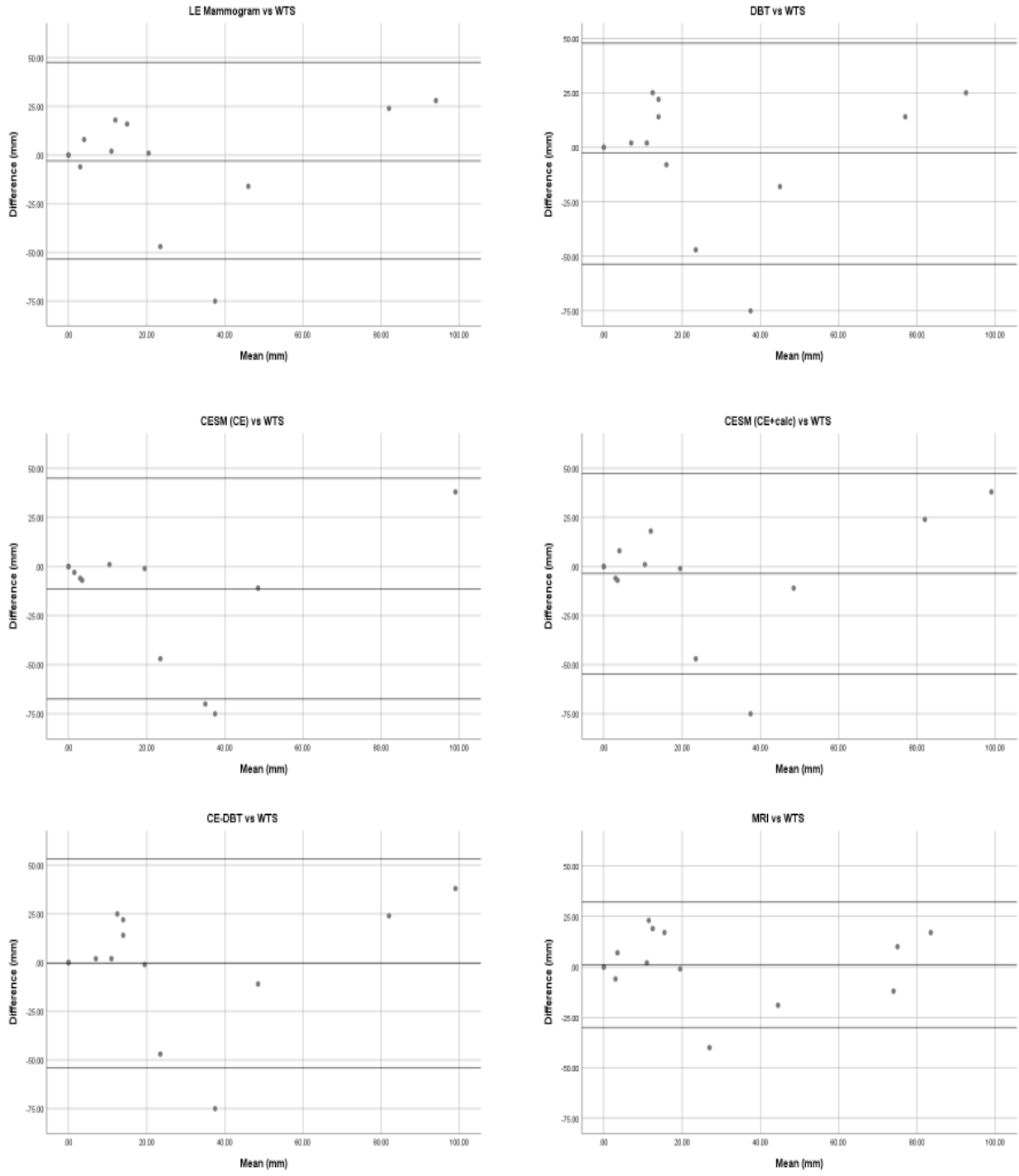


Figure 14: Bland Altman plots illustrating the difference between final imaging size and invasive tumour size

	Mean lesion size (mm)	Imaging size – ITS (mm)	Lin's concordance coefficient	95% Confidence Interval	
				Lower	Upper
ITS	11.9				
LE Mammogram	18.1	6.1	0.43	<0.01	0.73
DBT	18.2	6.3	0.43	<0.01	0.73
CESM(CE)	10.6	-1.3	0.70	0.39	0.88
CESM(CE+calc)	17.4	5.5	0.46	<0.01	0.74
CE-DBT	20.3	8.3	0.43	0.02	0.72
MRI	21.6	9.7	0.66	0.34	0.85

Table 15: Lin's Concordance co-efficient for predicting invasive tumour size according to imaging modality.

Both CESM(CE) and MRI confer similar reliability for predicting ITS; concordance coefficients: 0.7 and 0.66 respectively. The mean lesion size indicates that whilst CESM(CE) tends to slightly under estimate tumour size, MRI tends to over-estimate disease extent. The incremental addition of LE and DBT to CESM(CE), lowers the reliability with LE, DBT, CESM(CE+Calc) and CE-DBT conferring poor reliability.

6.3 Secondary Outcomes

6.3.1 Diagnostic accuracy for detection of multifocality of initial imaging

Five patients had multifocal disease, with six additional sites of disease confirmed on pathology; thus 24 lesions were present in thirty-six breasts, as detailed in table 16 below.

No. of patients	Multifocal	Bilateral	No lesions per participant	Total no. lesions
13	N	N	1	13
3	Y	N	2	6
1	Y	Y	2	2
1	Y	Y	3	3
Total: 18				24

Table 16: Presence of multifocality and bilaterality according to pathology

The resulting accuracy, sensitivity and specificities of the individual modalities are as follows.

	Positive		Negative		Sensitivity (95% CI)	Specificity (95% CI)	PPV	NPV	Accuracy (%)
	TP	FN	TN	FP					
LE	2	4	31	0	33.3 4.3-77.7	100.0 88.8-100.0	100.00	88.6 81.5-93.2	89.2 74.6-97.0
DBT	3	3	31	0	50.0 11.8-88.2	100.0 88.8-100.0	100.0	91.2 82.3-95.8	91.9 78.1-98.3
CESM (CE)	5	1	30	1	83.3 35.9-99.6	96.8 83.3-99.9	83.33 41.3-97.3	96.77 83.4-99.5	94.6 81.8-99.3
CESM (CE+calc)	5	1	30	1	83.3 35.9-99.6	96.8 83.3-99.9	83.33 41.3-97.3	96.77 83.4-99.5	94.6 81.8-99.3
CE-DBT	5	1	30	1	83.3 35.9-99.6	96.8 83.3-99.9	83.33 41.3-97.3	96.77 83.4-99.5	94.6 81.8-99.3
MRI	4	2	29	2	66.7 22.3-95.7	93.6 78.6-99.2	66.7 31.8-89.6	93.6 82.3-98.0	89.2 74.6-97.0

Table 17: Diagnostic accuracy of diagnosing additional lesions (multifocality) according to imaging modality

An example of multifocal disease, visible on all imaging techniques is illustrated in figure 15. The index lesion was a grade 3 ductal of no special type (NST), ER+ve, PR-ve, HER2-ve. The smaller satellite was a triple negative grade 3 ductal NST.

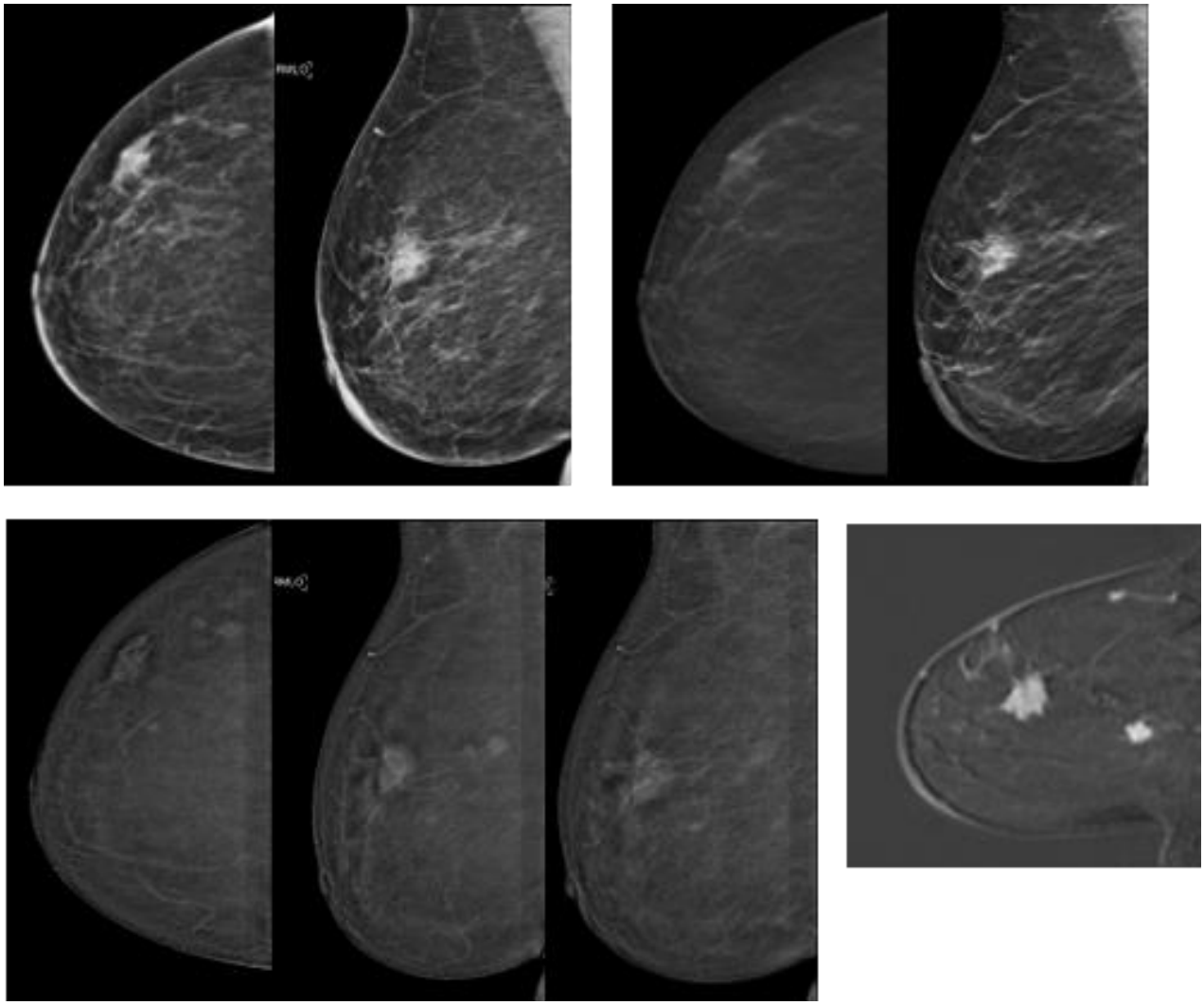


Figure 15: (Clockwise from top left): RCC and RMLO views from CE-DBT study: LE mammogram, selected slice from DBT study, CESM with delayed image showing washout and corresponding sagittal reformat from MRI study.

One site of bilateral disease, an invasive lobular cancer, was occult on all imaging modalities. The contrast-enhanced CESM image identified all other sites of multifocality with one false positive due to residual enhancement in a papilloma excision bed, as shown in figure 16.

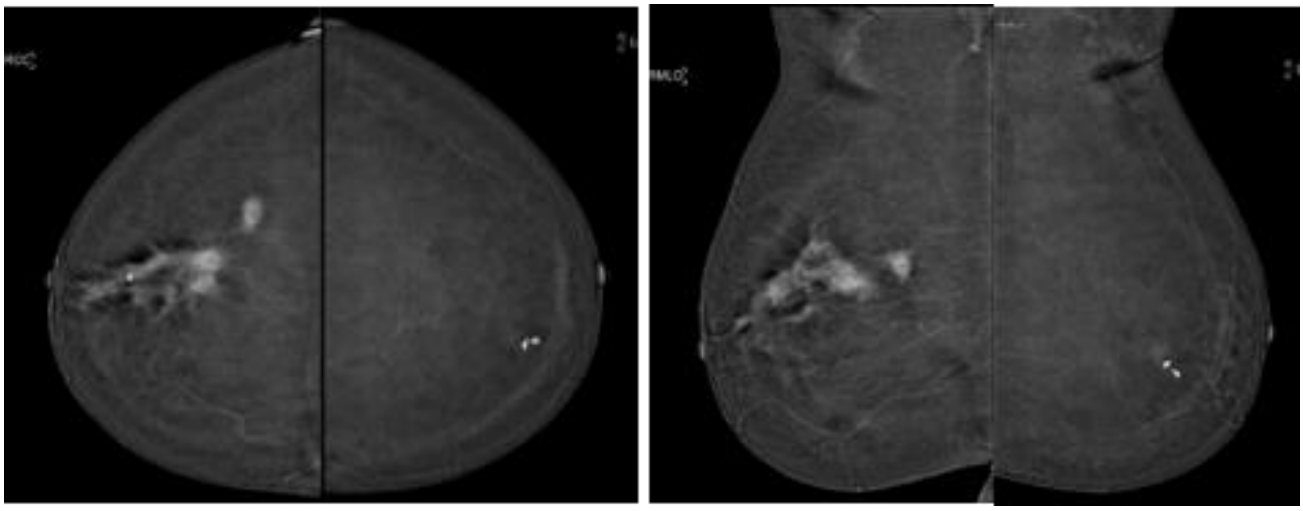


Figure 16: Bilateral CESM images. Left breast: Multifocal grade 2 ductal NST, Right breast: faint enhancement adjacent to marker clip represents papilloma excision bed (false positive)

In addition to the occult ILC, a contralateral grade 1 tumour was not identified on MRI, as illustrated in figure 17. MRI false positive results were seen in two breasts, resulting in three benign biopsies.

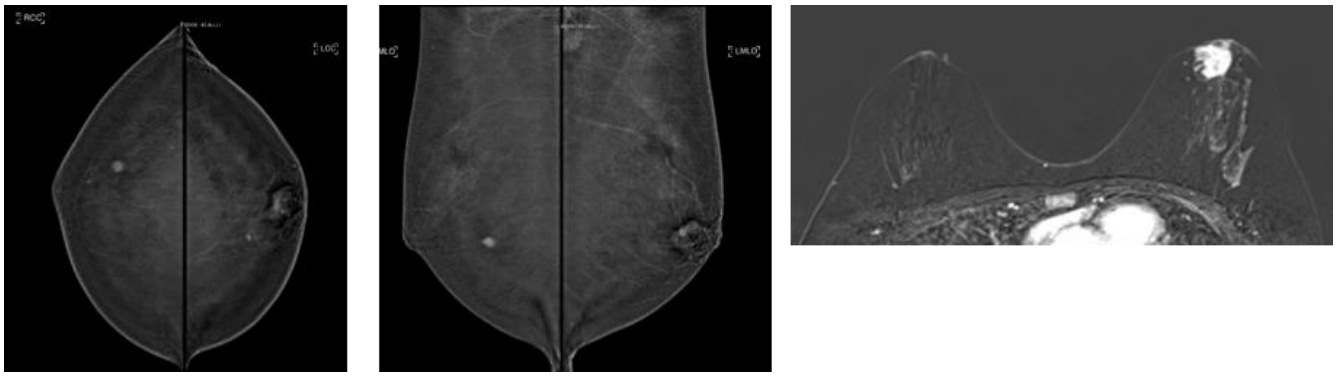


Figure 17: (Left to right). Bilateral CESM images showing a right grade 1 ductal carcinoma and a symptomatic left grade 3 ductal carcinoma. The small left cancer is not seen on MRI.

A greater extent of multifocal disease was occult on both LE mammogram and DBT with only half of additional disease identified on DBT and a third on mammogram. As illustrated in figure 18. Neither modality generated any false positives.

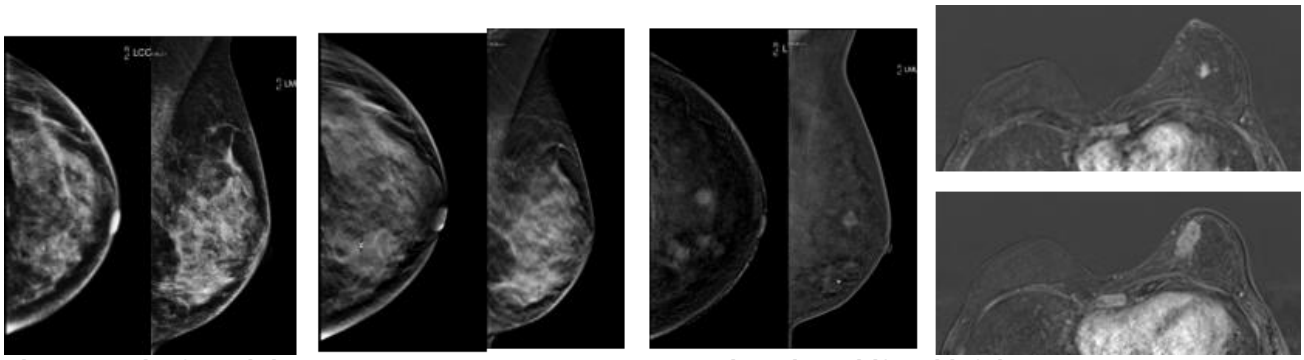


Figure 18: (Left to right: LE mammogram, DBT, CESM and MRI). Multifocal left breast cancer: Index lesion is subtle but visible on LE and DBT. Satellite is only seen on CESM and MRI.

6.3.2 Change in size between initial and interim imaging: predicting response

Four patients did not have a mid-treatment CE-DBT and in one case the MRI total disease extent was unobtainable, therefore data from thirteen patients was included. For one patient with bilateral disease the RCB-class was only available for one breast, thus there were 14 lesions considered in relation to pCR status and 13 lesions in relation to RCB score. Results are shown in the tables 18 & 19 below. As expected the pCR group demonstrated greater average reductions in size than the non-pCR group. However, this did not reach statistical significance. The most notable trend was seen in percentage-size change on CESM which neared statistical significance (median: -34.95 vs -17.56, p=0.098). With regards to predicting RCB-status, a significant difference was seen between the RCB 0-I and RCB II-III groups for percentage-size change on MRI (-31.25 vs -12.67, p=0.045) a similar trend was seen for CESM although this did not reach statistical significance (-37.50 vs -17.56, p=0.085).

Imaging technique			Pathological Complete response		*p*
			Yes (n=6)	No (n=8)	
Change in size					
	LE_mammogram				
		Median	-0.50	-9.00	
		IQR	27.50	47.75	0.746
	DBT				
		Median	-5.50	-6.00	
		IQR	19.75	21.00	0.698
	CESM				
		Median	-16.00	-10.50	
		IQR	23.25	17.25	0.698
	CESM+calc				
		Median	-16.00	-10.50	
		IQR	23.25	17.25	0.698
	CE-DBT				
		Median	-16.00	-10.50	
		IQR	24.25	7.50	0.747
	MRI				
		Median	-8.00	-4.00	
		IQR	10.25	31.25	0.299
Percentage change in size					
	LE_mammogram				
		Median	-1.47	-33.48	
		IQR	91.32	149.68	0.605
	DBT				
		Median	-13.26	-14.39	
		IQR	64.53	66.30	0.897
	CESM				
		Median	-34.95	-17.56	
		IQR	78.52	28.24	0.093
	CESM+calc				
		Median	-34.95	-17.56	
		IQR	78.52	28.24	0.093
	CE-DBT				
		Median	-34.95	-20.49	
		IQR	83.07	19.99	0.196
	MRI				
		Median	-12.67	-30.63	
		IQR	55.14	8.63	0.142

Table 18: Change in imaging lesion size in relation to pCR-status

Imaging technique		RCB class		Mann-Whitney U	Pearsons correlation coefficient	
		I-II (n=7)	III-IV (n=6)	*p*	*r*	*p*
Change in size						
LE_mammogram						
	Median	-1.00	-10.50			
	IQR	20.00	68.00	0.617	0.104	0.735
DBT						
	Median	-1.00	-6.00			
	IQR	24.00	13.00	0.720	0.162	0.598
CESM						
	Median	-24.00	-10.50			
	IQR	21.00	17.75	0.774	0.139	0.651
CESM+calc						
	Median	-24.00	-10.50			
	IQR	21.00	17.75	0.774	0.139	0.651
CE-DBT						
	Median	-12.00	-10.50			
	IQR	21.00	10.25	0.617	0.197	0.519
MRI						
	Median	-8.00	-4.00			
	IQR	27.00	24.50	0.150	0.246	0.419
Percentage change in size						
LE_mammogram						
	Median	-2.94	-38.97			
	IQR	61.68	119.59	0.520	0.061	0.844
DBT						
	Median	-8.33	-14.39			
	IQR	78.10	31.99	0.475	0.140	0.649
CESM						
	Median	-37.50	-17.56			
	IQR	77.42	32.79	0.086	0.415	0.159
CESM+calc						
	Median	-37.50	-17.56			
	IQR	77.42	32.79	0.086	0.415	0.159
CE-DBT						
	Median	-29.41	-17.56			
	IQR	81.25	19.78	0.198	0.387	0.191
MRI						
	Median	-31.25	-12.67			
	IQR	10.34	36.85	0.045	0.497	0.084

Table 19: Change in imaging lesion size in relation to RCB-class and RCB-score

6.4 Results of Patient Preference Questionnaires

Of the 18 patients who had initial imaging, 17 and 14 patients completed questionnaires for CE-DBT and MRI respectively, with 13 patients completing both. At end-of-treatment, 15 completed questionnaires following CE-DBT and 10 completed questionnaires following MRI, with 9 completing both. Thus, there were a total of 22 participant episodes completed by 16 patients, where all questionnaires were completed. These matched questionnaires were used for further comparative statistical analysis. The content of free-text responses was reviewed for all completed questionnaires (CE-DBT n = 32, MRI n = 24). Free-text comments were grouped according to content of the responses. Details of free text responses are shown below:

6.4.1 If you felt any anxiety about the test or during the test, please tell us what this was about:

6.4.1.1 CE-DBT

General anxiety

- Anxious about what will be found.
- A little anxious. Mainly because it's the first time I have been to the breast screening clinic.
- Just a bit nervous coming for test. Staff were friendly so that helped me tremendously.

Breast pain, discomfort or compression

- Due to the left breast being sore was very uncomfortable. Nurses, radiographer understanding and put me to ease
- None apart from being slightly painful on right side breast
- scared about the lump getting compressed

Cannulation / contrast

- After injecting the dye I felt an uncomfortable hot flush throughout my body which I wasn't expecting
- The only anxiety I've had was having an injection (due to bad experience when I was a child)
- Anxious about cannulation as veins have been a problem

- A little anxious when needle was inserted

6.4.1.2 MRI

General anxiety

- Before test worried about unknown, not knowing what it would be like.
- Just a little at the beginning until I calmed my breathing down
- Just a bit nervous
- I felt apprehensive beforehand
- Never done anything like this before- fear of the unknown

Enclosed

- Being enclosed
- Because my face was in an enclosed space I tried not to panic as felt too hot

Noise

- I got a surprise when there was a loud siren type sound at the start. I knew there would be 'banging' but didn't expect the siren and wondered if something was wrong. It was mild and short lived

Controlling breathing / keeping still

- Worried about staying still for so long.
- Difficulty in controlling breathing

Non-breast pain

- anxiety with my left leg, apart from that it was fine

6.4.2 If you noticed any strange feelings anywhere in your body when the dye was going in, please describe what you felt:

6.4.2.1 CE-DBT

Heat / flushing

- Heat
- Feeling hot from head to toe as it passed through body
- Just a slightly warm sensation.
- The dye was injected in my left arm, I felt hot sensation in right arm followed by chest stomach and face
- Hot sensation throughout my body felt the dye going through

- Warm flushing sensation from my groin area up through my body to chest area
- warm sensation all over
- hot feeling for a minute
- hot feeling throughout body
- warm flushing
- Hands burning
- Warm feeling in palms of hands and groin
- Not feeling face and lower body
- Very hot flush briefly, skin at chest/throat area felt hot, warm feeling at bladder & feeling like I was passing urine!
- Hot flush, hot sensation down below, hot metallic sensation coming up my chest & throat
- Body felt hot and tingly most noticeable at back of throat and feeling like going to wet self
- Hot flush. Like acid travelling up my throat & a sensation like I wee myself.
- Fingers, arms body

Numbness

- Numbness in fingers when injection needle went in. I then felt strange taste in mouth, then flushing, then a little like I'd pee'd myself. I also felt a little nauseous for a few seconds but it past

Need to urinate

- yes, felt like I was passing water
- It the start felt needed go toilet pass water
- Felt as if I needed to pee
- Feeling I needed to go bathroom pass water
- needing to urinate, hottness all over, dizziness

Strange taste

- Taste in mouth
- tingling, odd taste. Slightly flushed
- Metallic taste, heat to body, need for toilet sensation
- Metallic taste in my mouth. Below waist (hot feeling)

- A metallic taste in mouth followed by a warm flush from head down chest and arms
- flush of hot throughout felt like I had peed my pants :-)
- I had a funny taste in my mouth, there was a feeling like I was peeing myself and my hands and feet felt hot

6.4.2.2 MRI

Heat / flushing

- Felt a flush feeling. It felt similar to the dye you get from a CT scan.
- Hot flush.

Cold Sensation

- slightly cold feeling
- Coldness to arm. Feeling of need to urinate.
- I felt a cold sensation where the line went around my thumb where I was holding it
- Cold feeling in arm where dye went in, warm feeling in groin.
- Cold sensation in hand

Need to urinate

- I felt a funny taste in my mouth and the feeling like you were peeing
- Felt like I needed a wee

Strange taste

- Metallic taste

6.4.2 If you felt any pain or discomfort in any other parts of the body during the test, please tell us about it:

6.4.3.1 CE-DBT

Leg (Sciatic)

- yes as I have siatic nerve pain in my left leg apart from that mammogram was ok!

6.4.3.2 MRI

Leg (Sciatic)

- Due to mild degenerate arthritis in my leg bulging 5-6-7 doctor thinks sciatic nerve is causing pain in my leg

Shoulders /upper limbs

- Arms & hands but recently had 2 biopsies 1 underarm, and 1 in breast and feel the discomfort was due to this.
- strain on shoulders
- None - shoulders ached- but due to positioning not the scan.
- Shoulders slightly
- My shoulder felt stiff and I felt a pressure pain on my forehead where it was resting on the head rest

Face / forehead

- pressure on face, paper perhaps
- Slightly uncomfortable position of arms above head, generally tense muscles
- My forehead was really sore from pressing on the head rest. I had a pressure mark briefly on my forehead after the test. My shoulders were really sore from not moving especially my right shoulder

6.4.3 Please tell us anything else you think people should know about what it's like having the test:

6.4.3.2 CE-DBT

Wish to reassure other women / share positive experience

- Very little to worry about. Most uncomfortable thing was the cannula in arm.
- Its nothing to worry about. It is relatively quick and not too different from a standard mammogram
- Interesting
- Absolutely fine. No problem.
- Overall it it positive as its much quicker than having MRI and much more pleasant experience
- There is nothing to worry about.
- none I was put very much at ease
- Nothing to it, no worse than a normal mammogram other than having the canula put in.
- Its no problem
- Fairly easy procedure

- It is totally fine. Nothing to be worried about.
- Staff so kind and friendly felt little discomfort.
- Nothing to worry about.
- its very easy and painless and hopefully useful

Sensations associated with contrast

- Be prepare to feel if you are having a wee
- The experience of passing water
- Nothing - it is just like having a mammogram except for the dye when it feels funny when it first goes in
- Staff put you at ease. Taste in mouth as dye goes in
- Apart from the flushing as the dye was injected it wasn't very different from a normal mammogram

Breast pain / discomfort

- It will hurt the breast with the lump. Other breast will just be uncomfortable.

6.4.4.2 MRI

Noise

- Very comfortable & relaxing I think it would be good to know more about the noises before it starts
- Be prepared for the noises
- Too noisy to hear music! Fear of swallowing or breathing even! Point in gaps in scanning
- Noisy
- There is a lot of beeping and banging noises which are normal and nothing to worry about
- The machine seemed much less noisy this time from 1st time (as if it had been oiled!)

Headache / off-balance

- After the test I felt slight headache and a bit off balance

Wish to reassure other women / share positive experience

- It will be fine
- information given was accurate
- Try to stay calm
- It gets easier after the first one

Leaflet / more info

- leaflet other people's (positive) experiences

Non-breast pain

- I found the test ok my pain in my leg was letting me down so apart from that it was fine

6.4.4 Please tell us the reason for your answer above: [preference of technique]

6.4.4.1 CE-DBT

Quicker technique

- Quicker
- No comparison. Time factor 1.5 hours as opposed to 10 mins.
- It is much more comfortable and quicker experience. Even though it is slightly painful having the mammogram that out weighs the anxiety of being in MRI scanner for so long
- Shorter and less intense.
- Shorter / over sooner, less worrying
- Not as long
- Time factor and comfort.
- Less time/ to be in a position of length of time can be difficult if having a MRI.
- Quicker
- CE-DBT is quicker, more comfortable and it is easy for the staff to put you at ease as you can speak throughout. The CE-DBT experience was considerably better
- The CE-DBT is a more pleasant experience. It is over quicker and you don't get discomfort - the MRI is uncomfortable and long. Also you can see the people who are doing the CE-DBT and speak to them which puts you at ease
- I experienced more anxiety during MRI, which was a longer procedure and afterwards I had a sore head.

More comfortable

- More comfortable and wouldn't be so claustrophobic.
- Its was easier and I felt more comfortable with it

- Don't need to lie still for exam
- Uncomfortable - some anxiety unlike the mammogram. 30 minutes seems like a long time.

Less intimidating / more in control

- The CE-DBT was less intimidating
- Mammogram feels like I am more in control. MRI was fine though.

Feeling unwell after MRI

- Feelings on the day only? I fainted today which made me feel uneasy.
- I felt unwell after the MRI - headache, disorientated. The MRI took longer, was more uncomfortable

6.4.4.2 MRI

More comfortable

- Mammogram too uncomfortable.
- I would chose the MRI because it was less painful - although slightly claustrophobic
- Easier
- Less handling

More confidence in technique

- Unsure with this answer, I've gone for scan, maybe more detailed.
- Only because its the usual results that would be sent but really wouldn't mind either if results were the same.
- Because it is known, but if info/results were the same, either would be acceptable.

The categorised responses are shown in table 20, where comments include more than one theme they were counted for each theme separately. For example, the following comment pertaining to strange sensation associated with contrast; ‘Metallic taste, heat to body, need for toilet sensation’ was categorised as strange taste, heat/flushing and need to urinate.

Question	Free-text response (grouped)	True for CE-DBT (n = 32)	True for MRI (n = 24)
If you felt any anxiety about the test or during the test, please tell us what this was about:			
	General anxiety	3 (9%)	5 (21%)
	Breast pain / discomfort / compression	3 (9%)	0
	Cannulation / contrast	4 (13%)	0
	Being enclosed	0	3 (13%)
	Noise	0	1 (4%)
	Controlling breathing / Keeping still	0	2 (8%)
	Non-breast pain	0	1 (4%)
If you noticed any strange feelings anywhere in your body when the dye was going in, please describe what you felt:			
	Heat / flushing	23 (71%)	2 (8%)
	Cold sensation	0	6 (25%)
	Numbness	2 (6%)	0
	Need to urinate	12 (0.38)	3 (0.13)
	Strange taste	7 (0.22)	2 (0.08)
If you felt any pain or discomfort in any other parts of the body during the test, please tell us about it:			
	Leg (sciatic)	1 (3%)	1 (4%)
	Shoulders / upper limbs	0	7 (29%)
	Face / forehead	0	3 (13%)
Please tell us anything else you think people should know about what it's like having the test:			
	Noise	0	5 (20%)
	Headache / off balance	0	2 (8%)
	Wish to reassure other women / share positive experience	14 (44%)	5 (21%)
	Leaflet / more information	0	2 (8%)
	Non-breast pain	0	1 (4%)
	Sensation associated with contrast	5 (6%)	0
	Breast pain / discomfort	1 (3%)	0
Please tell us the reason for your answer above: [preference of technique]			
	Quicker technique	13 (41%)	0
	More comfortable	11 (34%)	4 (17%)
	More confidence in technique	1 (3%)	3 (13%)
	Less intimidating / more in control	3 (9%)	0
	Feeling unwell after MRI	2 (6%)	0

Table 20: Summarised free-text responses from all completed questionnaire

Outcome data for questions answered with four-point categorical response format is shown in table 21 and figure 19; statistically significant results are given in bold.

	Rating	CE-DBT	MRI	*p*
Overall how much anxiety did you feel during the test?				
	None	14	9	
	Mild	6	10	
	Moderate	2	2	
	Severe	0	1	0.052
How much pain did you feel when the needle was put in?				
	None	11	10	
	Mild	11	12	
	Moderate	0	0	
	Severe	0	0	0.655
Overall, how much pain did you feel in your breasts during the test?				
	None	12	20	
	Mild	7	1	
	Moderate	1	1	
	Severe	1	0	**0.021**
Overall, how much discomfort did you feel in your body during the test, not including in your breasts?				
	None	16	11	
	Mild	5	8	
	Moderate	1	2	
	Severe	0	1	**0.046**
How much did the staff put you at ease during the test?				
	Very Much	21	15	
	Moderately	1	5	
	A little	0	2	
	Not at all	0	0	**0.023**
During the test, how confident did you feel that you could say stop if you needed to?				
	Very Much	22	19	
	Moderately	0	1	
	A little	0	1	
	Not at all	0	1	0.109
How unpleasant was the feeling of the dye going in?				
	Not at all	15	18	
	A little	6	3	
	Moderately	1	1	0.257
	Very Much	0	0	
How would you rate your overall experience?				
	Excellent	14	8	
	Good	8	7	
	Fair	0	7	
	Poor	0	0	**0.008**

Table 21:Patient experience from matched questionnaire, categorical data

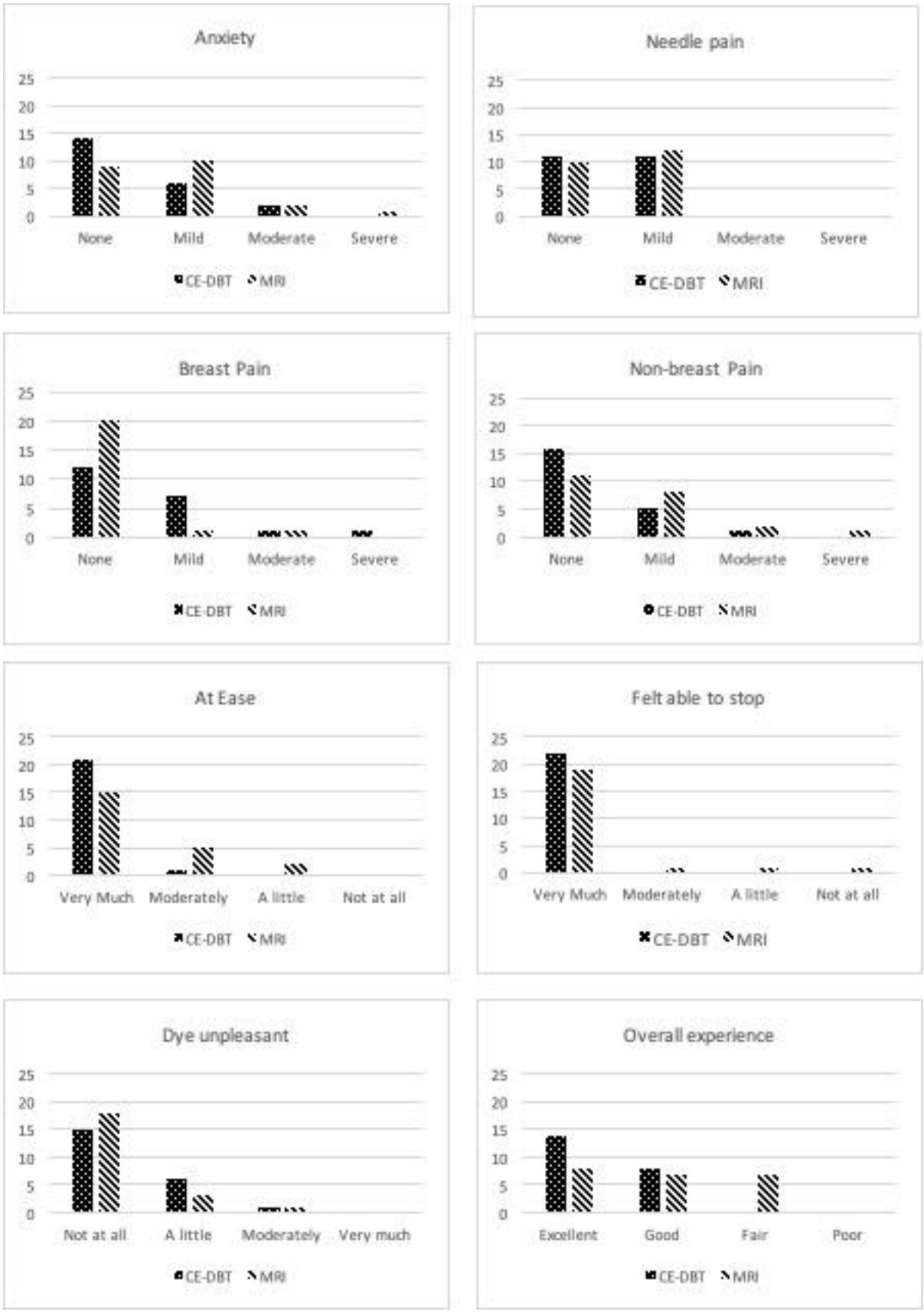


Figure 19: Patient experience from matched questionnaires, categorical data

Answers measured using VAS are displayed in table 22 and figure 20.

	CE-DBT (mean±S.D.)	MRI (mean±S.D.)	*p*
Anxiety	6.45 ± 8.06	16.91 ± 20.77	**0.003**
Breast pain	11.14 ± 18.60	3.86 ± 9.92	**0.011**

Table 22: Patient experience from matched questionnaires, questions answered using VAS

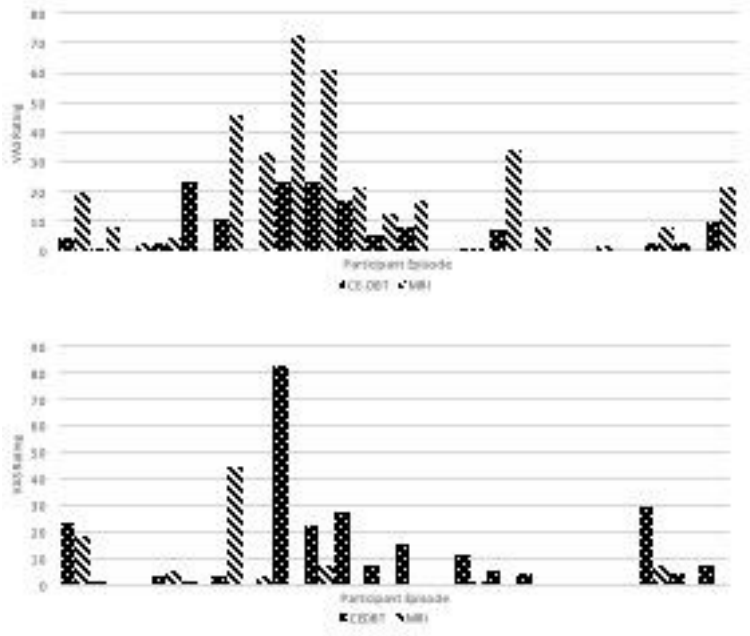


Figure 20:Patient experience, questions answered using VAS. Above: anxiety, beneath: breast pain

Significant differences in favour of CE-DBT were seen for non-breast pain (p = 0.046), being put at ease by staff (p = 0.023) and overall experience (p = 0.008). Anxiety was lower for CE-DBT when measured using VAS (p = 0.003); there was no statistically significant difference when measured with the categorical scale. By contrast, breast pain was significantly higher with CE-DBT when measured with both VAS (p = 0.011) and the categorical scale (p = 0.021). No statistically significant difference was seen between CE-DBT and MRI in patients' confidence that they could stop the test if needed.

6.4.6 Patient preference

Patient preference was recorded after 31 episodes, 16 following initial imaging, 15 following final imaging, one patient selected both CE-DBT and MRI after final

imaging this episode was excluded from further analysis. Eleven patients recorded a preference at initial and final imaging. Results are displayed in table 21.

Time-point	Preference	
	CE-DBT (%)	MRI (%)
Pre-treatment	13 (81%)	3 (19%)
End-of-treatment	10 (71%)	4 (29%)
Total	23 (77%)	7 (23%)

Table 23: Patient preference

Overall, on 77% of occasions patients preferred CE-DBT. Of the eleven patients who responded at both time points there was no significant change in in the proportion preferring CE-DBT; 10 (91%) and 8 (82%) at initial and final imaging respectively, $p = 0.25$.

Chapter 7: Results (TACESM)

7.1 Textural analysis modelling: Benign vs Malignant

There were 238 patients with a total of 269 enhancing mass lesions included. Of these 14 were benign and 255 were malignant. As described in the methods (chapter 5), the imbalance of benign/malignant lesions was mitigated through over-sampling. When considering model accuracy for identifying benign and malignant lesions, the following sub-analyses were performed.

7.1.1 Type of ROI

A total of 542 lesion views were included; all had TA features generated from both a FH_ROI and ellipsoid_ROI. Two models were developed based on the ROI-type. The relative accuracy of the two models is detailed in table 24 below.

ROI type	Accuracy (SD)	F1 score (SD)	AUC (SD)	Sensitivity (SD)	Specificity (SD)	PPV (SD)	NPV (SD)
FH	0.914(0.013)	0.928(0.01)	0.974(0.006)	0.953(0.028)	0.891(0.013)	0.841(0.016)	0.969(0.016)
Ellipsoid	0.947(0.009)	0.955(0.007)	0.986(0.004)	0.998(0.005)	0.916(0.014)	0.878(0.018)	0.998(0.003)

Table 24: ROI_model accuracy

Whilst the accuracy of both models was greater than 0.9 which is deemed to be very good,[149 150] the Ellipsoid_ROI model demonstrated consistently better results across all accuracy measures. Two-tailed independent sample mean test was conducted and showed these differences to be statistically significant (p <0.05), confirming that the model using ellipsoid-ROI had better performance.

7.1.2 Mammographic view

A total of 270 lesions were seen on CC view and 272 on MLO, with 263 seen on both CC and MLO. TA data generated from both the FH_ROI and ellipsoid_ROI was included. Three models were developed based on the ROI-type; one using CC_ROI only, one using MLO_ROI only and a third using lesions which had both CC_ROI and MLO_ROI. The relative accuracy of the three models alongside a comparison of

accuracy calculated using one-way analysis of variance (ANOVA) is detailed in table 25 below.

	Mammographic View			ANOVA	
	MLO	CC	Both	F	*p*
Accuracy (SD)	0.947(0.015)	0.955(0.011)	0.955(0.006)	5.026	**0.009**
F1 score (SD)	0.959(0.012)	0.964(0.009)	0.966(0.004)	4.855	**0.010**
AUC (SD)	0.985(0.009)	0.988(0.007)	0.987(0.004)	1.438	0.243
Sensitivity (SD)	0.954(0.029)	0.942(0.009)	0.954(0.011)	4.142	**0.019**
Specificity (SD)	0.944(0.020)	0.962(0.014)	0.956(0.007)	12.287	**0.000**
PPV (SD)	0.904(0.029)	0.935(0.023)	0.915(0.012)	14.683	**0.000**
NPV (SD)	0.974(0.015)	0.966(0.005)	0.977(0.005)	3.266	**0.043**

Table 25: Mammographic view model accuracy measures (ANOVA)

All three models demonstrate high levels of accuracy (>0.94). Significant differences were demonstrated between all measures other than AUC, which was consistently very high (>0.98). Post-hoc analysis on all other measures was conducted using multiple-comparison tests with Bonferroni correction, as shown in table 26 below

	Difference between mammographic views (*p* value)		
	MLO vs CC	MLO vs Both	CC vs Both
Accuracy	<0.00001	<.00001	0.869
F1 score	<0.00001	<.00001	<.00001
Sensitivity	0.048	0.959	0.036
Specificity	0.020	0.024	0.457
PPV	0.001	0.036	0.025
NPV	0.035	0.089	0.029

Table 26:Post-hoc direct comparison of mammographic view model accuracy measures

Overall, the model using both CC_ROI and MLO_ROI produced the most consistently good results across all measures. The CC_ROI model demonstrated a greater ability to detect benign lesions (higher specificity and NPV) whilst the MLO-ROI demonstrated a greater ability to detect malignant lesions (higher sensitivity and PPV) although overall accuracy and F1 score were lower. Although significant differences were seen, all models produced very strong results.

7.1.3 Vendor

The models designed to compare data acquired from Hologic and GE equipment, as described in the method section, were run. However due to small samples sizes for the Hologic dataset, when we divided the data into training and test set, there was no Hologic sample in the test set. Sampling was repeated 30 times and on the majority

of occasions there were no samples in the test set; sometimes one or two samples were present. This led to inconclusive results which cannot be reported.

7.2 Textural analysis modelling: Tumour Grade

152 lesions (19 Grade 1, 71 grade 2 and 62 grade 3) were included. All lesions views (CC/MLO) and ROI types (FH/Ellipsoid) were included to increase the size of the dataset, as shown in table 27. Due to the very small set of grade 1 lesions these were combined with grade 2 lesions to form a low grade (LG) dataset while grade 3 lesions were considered high grade (HG). Model results are shown in table 28.

	Pathological Grade		
	1(%)	2(%)	3(%)
No. of lesions	19(12.5)	71(46.7)	62(40.8)
No. of lesion views	72(12.3)	274(46.8)	240(41.0)

Table 27: Number of lesions and lesion views according to pathological grade

Tumour grade	Accuracy(SD)	F1 score (SD)	AUC (SD)	Sensitivity (SD)	Specificity(SD)	PPV(SD)	NPV(SD)
LG vs HG	0.630(0.026)	0.556(0.033)	0.653(0.013)	0.738(0.043)	0.504(0.040)	0.635(0.020)	0.624(0.038)

Table 28: Pathological grade model accuracy

The models produced limited accuracy, which may be partly due to small numbers of grade 1 tumours.

7.3 Textural analysis modelling: Hormone receptor status

Hormone receptor status was recorded for malignant lesions as displayed in table 29 below. All available lesions views (CC/MLO) and ROI types (FH/Ellipsoid) were included to increase the size of the data set.

	ER		PR		HER2	
	+ve(%)	-ve(%)	+ve(%)	-ve(%)	+ve(%)	-ve(%)
No. of lesions	193(77.5)	56(22.5)	164(66.1)	84(33.9)	50(20.3)	196(79.7)
No. of lesion views	754(77.7)	217(22.3)	641(66.4)	325(33.6)	191(19.9)	769(80.1)

Table 29: Number of lesions and lesion views according to receptor type

Relative model accuracy for each hormone receptor is shown in table 30

Receptor status	Accuracy (SD)	F1 score (SD)	AUC (SD)	Sensitivity (SD)	Specificity (SD)	PPV (SD)	NPV (SD)
ER	0.767(0.019)	0.820(0.017)	0.819(0.014)	0.674(0.045)	0.816(0.033)	0.662(0.034)	0.826(0.017)
PR	0.667(0.033)	0.763(0.031)	0.640(0.020)	0.453(0.064)	0.75(0.058)	0.419(0.047)	0.779(0.015)
HER	0.791(0.018)	0.697(0.025)	0.836(0.014)	0.833(0.029)	0.711(0.04)	0.85(0.0160)	0.686(0.035)

Table 30: Hormone receptor model accuracy

Moderate accuracy is demonstrated in the model differentiating ER status; accuracy is incrementally lower for HER2 and PR status.

7.4 Textural analysis modelling: Response to neoadjuvant chemotherapy (NACT)

A total of 99 lesions were treated with NACT, of which 28 had a complete pathological response (pCR). All lesion views (CC/MLO) and ROI types (FH/Ellipsoid) were included to increase the size of the data set, resulting in 103 lesion views for pCR and 274 for non-pCR (table 31). Model accuracy is shown below in table 32.

	pCR(%)	Non-pCR(%)
No. of Lesions	28(29.2)	68(70.8)
No. of Lesions views	103(27.3)	274(72.7)

Table 31: Number of lesions and lesion views according to pCR status

Accuracy (SD)	F1 score (SD)	AUC (SD)	Sensitivity (SD)	Specificity (SD)	PPV (SD)	NPV (SD)
0.704(0.041)	0.559(0.051)	0.744(0.027)	0.732(0.05)	0.638(0.059)	0.83(0.026)	0.5(0.057)

Table 32: Model accuracy for predicting pCR

Whilst model accuracy is 0.704, F1 score is lower at 0.559 suggesting poor reliability. This may be due to the small subset of lesions which demonstrated a pCR.

7.5 CESM dynamic tumour enhancement: association with MRI TIC

Dynamic enhancement data for matched CESM and MRI studies was available for 35 patient episodes. Association between CESM dynamic enhancement features

and MRI TIC is shown in tables 33 & 34. Significant differences were seen for all measures of WI, RD and SRD using the ellipsoid ROI according to MRI TIC group. With respect to the freehand ROI, significant differences were seen for the 90th percentile measures of WI, RD and SRD only. No difference with any measure of RLE was seen.

Using a cut-off of 10% change in SRD, as previously described, the association of CESM 'washout curve' and MRI TIC was assessed (see table 35). The strongest relationship, approaching moderate agreement, was seen with the 'curve' generated from the 90^{th} percentile measures of the ellipsoid ROI (kappa: 0.483). Poor agreement as seen with all other measures (kappa: 0.158 – 0.393). Poor agreement was also seen between qualitatively-assessed CESM washout and MRI curve (kappa: 0.302).

	RLE_Perc99		RLE_Perc90		RLE_Mean		WI_Perc99		WI_Perc90		WI_Mean		RD_Perc99		RD_Perc90		RD_Mean		SRD_Perc99		SRD_Perc90		SRD_Mean	
	1-2	3	1-2	3	1-2	3	1-2	3	1-2	3	1-2	3	1-2	3	1-2	3	1-2	3	1-2	3	1-2	3	1-2	3
Mean	1.81	2.14	1.34	1.51	0.74	0.76	-0.32	0.37	-0.28	0.28	-0.06	0.20	0.35	-0.37	0.28	-0.28	0.06	-0.20	33.52	-78.26	44.96	-25.33	94.40	-36.80
Median	2.08	1.85	1.40	1.24	0.53	0.65	-0.43	0.48	-0.38	0.38	-0.17	0.21	0.42	-0.48	0.38	-0.38	0.17	-021	15.48	-21.43	31.23	-22.86	57.25	-12.96
Min	0	0	0	0	0	-1	-2	-1	-1	-1	-1	-1	-2	-1	-1	-1	-1	-1	-89	-1100	-83	-143	-87	-603
Max	3	5	3	5	2	3	2	1	1	1	1	1	2	1	1	1	1	0	167	26	300	100	437	239
IQR	2	2	1	1	1	1	1	1	1	1	0	1	1	1	1	1	0	1	98	47	69	50	146	77
p	0.749		0.906		0.814		**0.003**		**0.001**		**0.037**		**0.003**		**0.001**		**0.037**		**0.002**		**0.003**		**0.01**	

Table 33: Correlation with MRI curve and enhancement characteristics generated from an ellipsoid ROI

	RLE_Perc99		RLE_Perc90		RLE_Mean		WI_Perc99		WI_Perc90		WI_Mean		RD_Perc99		RD_Perc90		RD_Mean		SRD_Perc99		SRD_Perc90		SRD_Mean	
	1-2	3	1-2	3	1-2	3	1-2	3	1-2	3	1-2	3	1-2	3	1-2	3	1-2	3	1-2	3	1-2	3	1-2	3
Mean	2.06	2.20	1.34	1.49	0.57	0.77	-0.29	0.19	-0.19	0.16	-0.03	0.16	0.28	-0.19	0.18	-0.16	0.03	-0.16	25.93	-16.66	24.88	-18.86	-12.24	-31.99
Median	2.01	1.90	1.43	1.24	0.56	0.67	-0.24	0.28	-0.24	0.33	-0.10	0.23	0.23	-0.28	0.24	-0.33	0.10	-0.23	3.66	-17.31	18.42	-20.77	20.93	-15.08
Min	0	1	0	1	0	0	-2	-1	-1	-1	0	-1	-1	-1	-1	-1	-1	-1	-82	-56	-79	-100	-657	-288
Max	3	5	2	4	1	2	1	1	1	1	1	1	2	1	1	1	0	1	267	25	200	27	288	231
IQR	1	1	1	1	1	1	1	1	0	1	0	1	1	1	0	1	0	1	74	46	49	53	57	93
p	0.662		0.449		0.637		0.069		**0.017**		0.074		0.072		**0.015**		0.074		0.096		**0.015**		0.106	

Table 34: Correlation with MRI curve and enhancement characteristics generated from a freehand ROI

CESM curve		MRI Curve			kappa
		1	2	3	
Ellipsoid ROI Per99					
	1	**7**	0	3	
	2	2	**1**	4	0.393
	3	4	0	**14**	(0.114-0.673)
Ellipsoid ROI Per90					
	1	**10**	1	2	
	2	0	**1**	1	0.483
	3	4	5	**12**	(0.230-0.736)
Ellipsoid ROI Mean					
	1	**9**	1	9	
	2	3	**0**	1	0.315
	3	1	0	**11**	(0.063-0.567)
Freehand ROI Per99					
	1	**4**	1	3	
	2	4	**0**	6	0.158
	3	5	0	**11**	(-0.114-0.430)
Freehand ROI Per90					
	1	**7**	1	6	
	2	3	**0**	3	0.277
	3	3	0	**12**	(0.002-0.553)
Freehand ROI Mean					
	1	**7**	1	7	
	2	3	**0**	3	0.230
	3	3	0	**11**	(-0.045-0.505)
Qualitative CESM curve					
	1	**3**	0	2	
	2	6	**1**	2	0.302
	3	4	0	**17**	(0.068-0.537)

Table 35: Correlation between washout curves generated from CESM and MRI images

7.6 CESM dynamic tumour enhancement: Association between qualitative and quantitative CESM enhancement

Matched data for qualitative and quantitative CESM enhancement characteristics was available for 38 patient episodes. Association between quantitative enhancement characteristics and qualitative curve type are shown in tables 36 & 37. No significant difference in RLE was seen between the qualitative curve types generated by either ellipsoid or freehand ROI. Significant differences were observed for all other quantitative measures. On subsequent pairwise comparison, there were

consistently significant differences between the quantitative values in curve types II and III. However, there was no significant difference between curve types I and II for any measure and a significant difference between curve type I and III was only seen in half of the measures.

CESM enhancement characteristic	Qualitative CESM curve type			*p*	Pairwise comparison (*p* value)		
	Type I	Type II	Type III		I – II	I – III	II – III
RLE_Per99							
Median	1.71	2.08	1.99				
IQR	1.01	1.77	1.55	0.625	N/A	N/A	N/A
RLE_Per90							
Median	1.22	1.76	1.33				
IQR	0.71	1.21	1.24	0.831	N/A	N/A	N/A
RLE_Mean							
Median	0.65	0.66	0.65				
IQR	1.05	1.18	1.23	0.970	N/A	N/A	N/A
WI_Per99							
Median	-0.17	-0.61	0.57				
IQR	0.59	0.71	0.43	**0.000**	1.000	0.064	**0.000**
WI_Per90							
Median	-0.19	-0.52	0.38				
IQR	0.42	0.71	0.52	**0.000**	1.000	**0.018**	**0.001**
WI_Mean							
Median	-0.20	-0.32	0.34				
IQR	0.43	0.50	0.52	**0.000**	1.000	**0.025**	**0.002**
RD_Per99							
Median	0.16	0.60	-0.55				
IQR	0.59	0.67	0.43	**0.001**	1.000	0.061	**0.000**
RD_Per90							
Median	0.19	0.52	0.38				
IQR	0.41	0.71	0.51	**0.000**	1.000	**0.018**	**0.001**
RD_Mean							
Median	0.20	0.32	-0.33				
IQR	0.43	0.49	0.52	**0.001**	1.000	**0.027**	**0.002**
SRD_Per99							
Median	-0.85	24.14	-33.33				
IQR	57.86	86.14	44.57	**0.001**	1.000	0.105	**0.002**
SRD_Per90							
Median	14.97	36.36	-33.33				
IQR	59.67	70.71	49.17	**0.003**	1.000	0.051	**0.009**
SRD_Mean							
Median	64.19	42.26	-14.02				
IQR	194.71	170.58	86.62	**0.003**	1.000	0.030	**0.015**

Table 36: Association between quantitative CESM enhancement characteristics derived from an ellipsoid ROI and qualitative CESM enhancement curve

CESM enhancement characteristic		Qualitative CESM curve type			*p*	Association (*p* value)		
		Type I	Type II	Type III		I – II	I – III	II – III
RLE_Per99								
	Median	1.61	2.01	2.22				
	IQR	1.01	1.18	1.30	0.356	N/A	N/A	N/A
RLE_Per90								
	Median	1.14	1.52	1.33				
	IQR	0.70	0.58	0.80	0.628	N/A	N/A	N/A
RLE_Mean								
	Median	0.66	0.53	0.68				
	IQR	1.18	0.75	0.75	0.612	N/A	N/A	N/A
WI_Per99								
	Median	-0.33	-0.88	0.38				
	IQR	0.43	1.34	0.81	**0.002**	1.000	0.083	**0.005**
WI_Per90								
	Median	-0.26	-0.38	0.33				
	IQR	0.25	0.57	0.38	**0.000**	1.000	**0.013**	**0.003**
WI_Mean								
	Median	-0.16	-0.17	0.32				
	IQR	0.39	0.53	1.18	**0.003**	1.000	**0.022**	**0.019**
RD_Per99								
	Median	0.33	0.84	-0.37				
	IQR	0.42	1.31	0.79	**0.002**	1.000	0.070	**0.005**
RD_Per90								
	Median	0.26	0.37	-0.33				
	IQR	0.25	0.56	0.38	**0.001**	1.000	**0.013**	**0.003**
RD_Mean								
	Median	0.16	0.17	-0.31				
	IQR	0.39	0.53	0.64	**0.003**	1.000	**0.022**	**0.019**
SRD_Per99								
	Median	4.29	14.18	-21.43				
	IQR	84.46	72.04	41.46	**0.006**	1.000	0.080	**0.019**
SRD_Per90								
	Median	18.70	24.01	-24.00				
	IQR	65.00	48.37	50.40	**0.000**	1.000	**0.017**	**0.001**
SRD_Mean								
	Median	5.42	35.21	-33.43				
	IQR	254.32	53.76	90.88	**0.013**	0.753	0.732	**0.011**

Table 37: Association between quantitative CESM enhancement characteristics derived from a freehand ROI and qualitative CESM enhancement curve

The agreement between qualitative CESM curves and those generated for SRD values are shown in table 38. Poor correlation is seen for all values, with the best agreement seen for FH Per90 (kappa: 0.429).

		CESM qualitative			kappa
		1	2	3	
Ellipsoid (ROI) Per99					
	1	2	6	2	
	2	1	2	4	0.300
	3	3	1	17	(0.062-0.538)
Ellipsoid ROI Per90					
	1	**3**	6	5	
	2	3	**2**	3	0.347
	3	0	1	**15**	(0.147-0.547)
Ellipsoid ROI Mean					
	1	**5**	8	7	
	2	1	**0**	3	0.320
	3	0	1	**13**	(0.136-0.505)
Freehand (ROI) Per99					
	1	**2**	5	2	
	2	3	**1**	6	0.283
	3	1	2	**15**	(0.065-0.502)
Freehand ROI Per90					
	1	**4**	6	4	
	2	2	**2**	3	0.429
	3	0	1	**16**	(0.226-0.631)
Freehand ROI Mean					
	1	**3**	7	5	
	2	2	**1**	4	0.276
	3	1	1	**14**	(0.07-0.484)

Table 38: Correlation between qualitative and quantitative CESM washout curve

7.7 Association between dynamic contrast and tumour grade

Eighteen lesions were identified which were treated with primary surgery. Of these lesions two were grade 1, nine were grade 2 and eight were grade 3. Ellipsoid_ROI data was missing for one grade 3 lesion. Therefore, eighteen lesions were included in FH_ROI analysis and seventeen in ellipsoid_ROI analysis. Results are displayed in tables 39 & 40; no significant variation between different grade is identified for any of the quantitative measures.

CESM enhancement characteristic		Tumour grade			*p*
		1	2	3	
RLE_Per99					
	Median	3.96	2.46	2.29	
	IQR	-	1.98	1.05	0.405
RLE_Per90					
	Median	3.81	2.18	1.95	
	IQR	-	1.56	1.73	0.248
RLE_Mean					
	Median	3.14	1.61	1.41	
	IQR	-	1.09	1.63	0.149
WI_Per99					
	Median	0.64	0.38	0.59	
	IQR	-	1.42	0.65	0.614
WI_Per90					
	Median	0.74	0.24	0.33	
	IQR	-	0.90	0.45	0.641
WI_Mean					
	Median	0.58	-0.1	0.33	
	IQR	-	0.88	0.46	0.430
RD_Per99					
	Median	-0.61	-0.36	-0.58	
	IQR	-	1.39	0.63	0.736
RD_Per90					
	Median	-0.71	-0.23	-0.32	
	IQR	-	0.89	0.44	0.720
RD_Mean					
	Median	-0.56	0.12	-0.32	
	IQR	-	0.86	0.45	0.430
SRD_Per99					
	Median	-17.36	-19.13	-30.87	
	IQR	-	49.36	27.43	0.252
SRD_Per90					
	Median	-15.82	-19.57	-33.35	
	IQR	-	60.19	36.56	0.739
SRD_Mean					
	Median	-8.21	-15.76	-17.96	
	IQR	-	33.16	52.38	0.487

Table 39: Association between quantitative CESM enhancement characteristics derived from an ellipsoid ROI and invasive tumour grade

CESM enhancement characteristic		Tumour grade			*p*
		1	2	3	
RLE_Per99					
	Median	3.86	2.27	2.27	
	IQR	-	1.54	1.51	0.373
RLE_Per90					
	Median	3.57	1.97	1.55	
	IQR	-	1.27	1.41	0.165
RLE_Mean					
	Median	2.61	1.28	0.94	
	IQR	-	0.92	1.29	0.123
WI_Per99					
	Median	0.64	0.40	0.28	
	IQR	-	1.07	0.46	0.957
WI_Per90					
	Median	0.67	0.29	0.24	
	IQR	-	0.88	0.44	0.839
WI_Mean					
	Median	0.47	0.04	0.12	
	IQR	-	0.64	0.34	0.584
RD_Per99					
	Median	-0.61	-0.39	-0.28	
	IQR	-	1.05	0.45	0.957
RD_Per90					
	Median	-0.64	-0.28	-0.23	
	IQR	-	0.86	0.43	0.839
RD_Mean					
	Median	-0.45	-0.35	-0.11	
	IQR	-	0.63	0.34	0.653
SRD_Per99					
	Median	-17.73	-18.14	-23.11	
	IQR	-	40.88	31.89	0.674
SRD_Per90					
	Median	-15.28	-17.98	-18.64	
	IQR	-	41.40	28.08	0.738
SRD_Mean					
	Median	-5.66	-14.17	-14.02	
	IQR	-	45.55	44.14	0.774

Table 40: Association between quantitative CESM enhancement characteristics derived from a freehand ROI and invasive tumour grade

7.8 Association between CESM dynamic contrast features and pCR

Quantitative dynamic enhancement data were available for 24 lesions at pre-treatment and 12 lesions at mid-treatment. Data for 11 lesions was present at both time-points. Results from pre-treatment imaging are shown in tables 41 & 42, with results from mid-treatment in tables 43 & 44.

At pre-treatment FH_WI_Perc99 showed significant difference between lesions that completely responded and those which did not. Lesions in the pCR group had a significantly higher WI_Perc99 (Mean 0.46 vs -0.14, median 0.38 vs 0.00, $p = 0.049$). No significant differences were observed between the pCR and non-pCR groups at mid-treatment

	RLE_Perc99		RLE_Perc90		RLE_Mean		WI_Perc99		WI_Perc90		WI_Mean		RD_Perc99		RD_Perc90		RD_Mean		SRD_Perc99		SRD_Perc90		SRD_Mean	
	pCR	No pCR	pCR	No pCR	pCR	No pCR	pCR	No pCR	pCR	No pCR	pCR	No pCR	pCR	No pCR	pCR	No pCR	pCR	No pCR	pCR	No pCR	pCR	No pCR	pCR	No pCR
Mean	2.41	2.20	1.92	1.58	1.21	0.79	0.54	0.06	0.43	0.16	0.29	0.19	-0.53	-0.06	-0.42	-0.16	-0.29	-0.19	-31.53	1.07	-39.27	-6.03	-39.97	1.71
Median	2.56	2.03	2.24	1.42	1.31	0.63	0.57	0.33	0.33	0.19	0.27	0.08	-0.56	-0.32	-0.333	-0.19	-0.27	-0.08	-25.00	-7.81	-22.86	-2.44	-12.96	10.65
Min	1	1	0	0	0	0	0	-1	0	-1	0	-1	-1	-1	-1	-1	-1	-1	-105	-53	-143	-81	-603	-290
Max	4	5	3	5	3	3	1	1	1	1	1	1	0	1	0	1	0	0	-3	93	18	78	239	237
IQR	2	1	2	1	2	1	0	1	0	1	1	1	0	1	0	1	1	1	23	81	26	65	42	134
p	0.297		0.235		0.235		0.125		0.235		0.664		0.140		0.235		0.664		0.235		0.235		0.434	

Table 41: Association between pre-treatment quantitative ellipsoid CESM enhancement characteristics and complete pathological response

	RLE_Perc99		RLE_Perc90		RLE_Mean		WI_Perc99		WI_Perc90		WI_Mean		RD_Perc99		RD_Perc90		RD_Mean		SRD_Perc99		SRD_Perc90		SRD_Mean	
	pCR	No pCR	pCR	No pCR	pCR	No pCR	pCR	No pCR	pCR	No pCR	pCR	No pCR	pCR	No pCR	pCR	No pCR	pCR	No pCR	pCR	No pCR	pCR	No pCR	pCR	No pCR
Mean	2.40	2.38	1.77	1.64	1.07	0.72	0.46	-0.14	0.32	0.10	0.31	0.10	-0.44	0.14	-0.31	-0.10	-0.31	-0.10	-21.76	4.63	-22.52	-3.73	-16.13	-9.98
Median	2.69	2.08	1.77	1.38	1.15	0.65	0.38	0.00	0.29	0.05	0.26	0.08	-0.38	0.00	-0.28	-0.05	-0.26	-0.08	-23.81	2.56	-16.22	2.25	-12.22	18.37
Min	1	1	1	1	0	0	0	-2	-1	0	0	-1	-1	-1	-1	-1	-1	-1	-56	-51	-100	-61	-226	-288
Max	4	5	3	4	2	2	1	1	1	1	1	1	0	2	1	0	0	1	7	131	27	50	168	288
IQR	2	1	1	0	1	1	1	1	0	1	1	1	1	1	0	1	1	1	20	52	37	41	37	109
p	0.582		0.385		0.192		**0.049**		0.401		0.235		0.060		0.417		0.235		0.077		0.311		0.434	

Table 42: Association between pre-treatment quantitative freehand CESM enhancement characteristics and complete pathological response

	RLE_Perc99		RLE_Perc90		RLE_Mean		WI_Perc99		WI_Perc90		WI_Mean		RD_Perc99		RD_Perc90		RD_Mean		SRD_Perc99		SRD_Perc90		SRD_Mean	
	pCR	No pCR	pCR	No pCR	pCR	No pCR	pCR	No pCR	pCR	No pCR	pCR	No pCR	pCR	No pCR	pCR	No pCR	pCR	No pCR	pCR	No pCR	pCR	No pCR	pCR	No pCR
Mean	1.58	1.84	1.18	1.33	0.85	0.68	0.27	-0.03	0.06	-0.07	0.17	-0.03	-0.26	0.03	-0.06	0.07	-0.17	0.03	-175.8	2.28	54.83	9.91	33.81	82.32
Median	1.57	1.75	1.12	1.33	0.96	0.49	0.24	-0.17	-0.07	-0.14	0.06	-0.02	-0.24	0.16	0.06	0.14	-0.07	0.02	-29.20	-8.62	28.90	1.92	11.29	15.64
Min	0.05	0.24	-0.38	0.14	-0.77	0.02	-0.66	-0.61	-0.76	-0.57	-0.44	-0.36	-1.52	-0.70	-1.40	-0.56	-1.38	-0.49	-1100	-51.43	-83.33	-52.00	-86.99	-90.78
Max	3.45	3.31	2.76	2.01	1.99	1.78	1.56	0.71	1.43	0.57	1.40	0.50	0.66	0.59	0.75	0.60	0.43	0.36	166.0	100.0	300.0	100.0	266.9	437.3
IQR	2.56	1.42	2.36	1.07	1.97	0.99	1.67	0.94	1.30	0.46	0.82	0.43	1.64	0.92	2.15	0.46	0.81	0.43	402.7	51.10	204.9	53.71	143.8	229.2
p	0.749		0.749		0.522		0.631		0.873		0.749		0.631		0.873		0.749		0.631		0.575		0.873	

Table 43:Association between mid-treatment quantitative ellipsoid CESM enhancement characteristics and complete pathological response

	RLE_Perc99		RLE_Perc90		RLE_Mean		WI_Perc99		WI_Perc90		WI_Mean		RD_Perc99		RD_Perc90		RD_Mean		SRD_Perc99		SRD_Perc90		SRD_Mean	
	pCR	No pCR	pCR	No pCR	pCR	No pCR	pCR	No pCR	pCR	No pCR	pCR	No pCR	pCR	No pCR	pCR	No pCR	pCR	No pCR	pCR	No pCR	pCR	No pCR	pCR	No pCR
Mean	2.00	1.66	1.30	1.16	0.86	0.61	0.45	-0.22	0.10	-0.16	0.15	-0.05	-0.44	0.21	-0.10	0.16	-0.15	0.05	-35.49	37.76	-17.53	33.30	-21.12	-86.85
Median	1.57	1.61	1.19	1.19	0.88	0.38	0.38	-0.33	0.00	-0.24	-0.04	-0.08	-0.37	0.33	0.00	0.24	0.04	0.08	-44.00	1.14	0.00	13.15	-4.72	1.99
Min	1.18	0.14	0.71	0.09	0.40	-0.01	-0.47	-0.85	-0.48	-0.67	-0.15	-0.49	-1.25	-0.37	-0.98	-0.42	-0.70	-0.38	-82.35	-36.67	-79.49	-50.00	-82.27	-657.4
Max	3.36	2.84	1.86	2.06	1.15	1.86	1.28	0.38	1.00	0.43	0.70	0.38	0.47	0.83	0.47	0.66	0.15	0.49	12.90	266.7	20.00	200.0	28.52	231.4
IQR	1.56	1.36	0.93	0.88	0.53	0.95	1.04	0.95	1.17	0.63	0.66	0.54	1.03	0.93	1.15	0.62	0.66	0.53	64.12	92.98	83.28	83.60	87.57	316.5
p	0.837		0.818		0.589		0.109		0.423		0.337		0.078		0.423		0.337		0.100		0.522		0.749	

Table 44:Association between mid-treatment quantitative

Chapter 8: Discussion (CONDOR)

8.1 Feasibility

This study has demonstrated that the use of monitoring response to NACT within clinical care is feasible, although not without challenges. Unfortunately, the intended number of participants could not be recruited, with 18 out of a planned 25 (72%) included prior to the study closing due to the COVID-19 pandemic. Of those patients meeting the inclusion criteria, 58% were successfully recruited. This may have increased to 68% had we been able to offer pre-treatment CE-DBT within an appropriate time-frame to an additional three eligible patients. The logistics of booking the pre-treatment imaging were more complex due to the limited time frame between the decision to treat with NACT and start of NACT. Whilst the pre-treatment MRI scans were booked by the Oncologist as standard-of-care at the oncology appointment, the patient only became eligible for the study having agreed to proceed with NACT at this appointment. Therefore, they had to be subsequently contacted to ascertain whether they wished to participate and only then could the CE-DBT be arranged. Subsequent imaging at mid- and end-of-treatment was planned in advance in accordance with chemotherapy cycles. It was therefore easier to co-ordinate the later CE-DBT and MRI appointments. With regard to a future, multi-centre trial, it is likely that enrolment would be higher in centres routinely using CE-DBT at time of diagnosis as this appears to be the greatest limiting factor. To further improve recruitment, it may be possible to provide patient information leaflets at the results appointment so that discussion and recruitment could occur at the first oncology appointment for those proceeding with NACT. Assistance with transport and co-ordination of appointments to reduce travel difficulties may also help.
There were no adverse outcomes reported during the trial although one patient withdrew at mid-treatment due to pain at the cannulation site. However, there were several incomplete studies. Of note, one patient did not have a CESM(CE) or DBT study due to human error and for a further patient the delayed CESM(CE) was not acquired due to equipment failure. One early recruit with bilateral disease was unable to have bilateral delayed images on initial CESM due to the original protocol

stipulations – this was amended for future patients with bilateral disease. The remainder of the incomplete studies were unavoidable – either a result of COVID, chemotherapy regimens or progressive disease.

8.2 Primary Outcomes

8.2.1 Predicting response on post-chemotherapy images

With moves to reduce and potentially cease operating on patients with a complete pathological response following NACT, the ability to identify such cases pre-operatively is becoming increasingly important.

With regard to identifying pCR, we have demonstrated that CESM(CE) has highest overall accuracy: 81.25%, with excellent sensitivity but considerably lower specificity - 100% and 57.14% respectively. The addition of DBT improved specificity to 71.43% but to the detriment of sensitivity and overall accuracy. Consistent with previous evidence, lower accuracy for detecting pCR was seen with LE mammography [113 151-154] and the accuracy of CESM was not improved by combining with LE images. CE-DBT, DBT and MRI demonstrated equal accuracy of 62.50%. However, MRI demonstrated both the highest specificity (85.71%) and lowest sensitivity (44.44%) of all techniques.

Review of the published literature identified eight papers concerning CESM and the prediction of response to NACT[111-113 151 155-158], one pre-print[152] and a conference abstract[159]. There is noticeable heterogeneity between studies, including in the definition of pCR. In three papers[151 158 159] it was defined as no residual invasive disease yTP0/is, as in this study. In three papers [111 112 156] it was defined as no residual invasive or in situ (yTP0), and one considered both definitions [155].

Prevalence of pCR ranges from 14%-36%, which is considerably lower than in this study (62.5%). Timing of imaging ranged from 4 weeks prior to the final cycle of chemotherapy to less than 10 days prior to surgery. Our mean interval of 25 days is within this broad range. When reported, NACT regimes varied both within and between studies. One study also included patients receiving neoadjuvant endocrine therapy.[112]

A recently published systematic review and meta-analysis[160] of the accuracy of CESM for prediction of pCR included six of these papers[111-113 151 156 157]. However,

although the authors recognised some heterogeneity between the studies, they did not appreciate that the studies calculated diagnostic accuracy for differing objectives. Whilst the Forest plot included in the meta-analysis purports to show 'CESM sensitivity and specificity to predict pCR' only two papers actually calculated this;[111] [112] two calculated the opposite – the diagnostic accuracy for detecting *residual disease*[113] [151] and two papers reported the accuracy for identifying 'responders' either defined as Miller-Payne grade 3-5, i.e. ≥ 30-90% reduction in tumour cells[157], or as 'all tumour responses'.[156]

Therefore, I sought to calculate the true CESM sensitivity and specificity for prediction of pCR from the raw data generated in each study. It is possible to derive the raw results for five of the six papers in the meta-analysis and two additional studies. Three studies were excluded from further analysis due to lack of raw data[157] [159] or patient overlap with other studies.[155] Diagnostic accuracy results are displayed alongside the results of my study in Table 45 and the Forest Plot in figure 21. My results are broadly similar to previous studies, but wide variation is demonstrated.

Study	CESM	Pathology		Sensitivity (95% CI)	Specificity (95% CI)	PPV (95% CI)	NPV (95% CI)	Diagnostic accuracy (95% CI)
		pCR	Non-pCR					
Barra, 2017	CR	2	1	100.00 (15.81-100)	83.33 (35.88-99.58)	66.67 (25.05-92.29)	100.00	87.50 (47.35-99.68)
	Non-CR	0	5					
Barra, 2018	CR	7	6	87.50 (47.35-99.68)	76.00 (54.87-90.64)	53.85 (35.64-71.08)	95.00 (74.99-99.18)	78.79 (61.09-91.02)
	Non-CR	1	19					
ElSaid, 2017	CR	5	0	83.33 (35.88-99.58)	100.00 (78.20-100.00)	100.00	93.75 (71.48-98.90)	95.24 (76.18-99.88)
	Non-CR	1	15					
Iotti, 2017	CR	8	6	100.00 (68.75-93.98)	84.21 (68.75-93.98)	57.14 (39.02-73.54)	100.00	86.96 (73.74-95.06)
	Non-CR	0	32					
Kamal, 2020	CR	20	1	95.24 (76.18-99.88)	98.33 (91.06-99.96)	95.24 (74.07-99.29)	98.33 89.70-99.75)	97.53 (91.36-99.70)
	Non-CR	1	59					
Patel, 2018	CR	19	15	95.00 (75.13-99.87)	66.67 (51.05-80.00)	55.88 (45.29-65.96)	96.77 (81.45-99.51)	75.38 (63.13-85.23)
	Non-CR	1	30					
Savaridas	CR	9	3	100.00 (67.37-100.00)	57.14 (18.41-90.10)	75.00 (56.05-87.59)	100.00	81.25 (54.35-95.95)
	Non-CR	0	4					
Steinhof-Radwanska, 2021	CR	18	12	85.71 (63.66-96.95)	71.43 (55.42-84.28)	60.00 (47.41-71.39)	90.91 (77.51-96.67)	76.19 (63.79-86.02)
	Non-CR	3	30					

Table 45: Summary of papers reporting diagnostic accuracy of CESM. CR: complete response

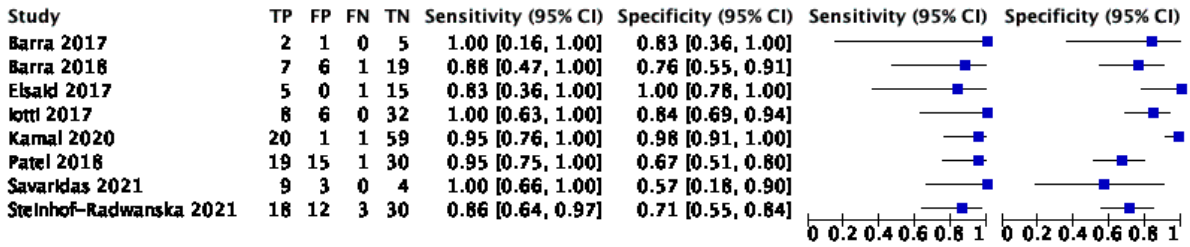


Figure 21: Forest Plot for the sensitivity and specificity of CESM for detecting breast cancer

With regard to MRI, I demonstrated a sensitivity and specificity of 44% and 85% respectively. Accuracy of MRI for predicting pCR varies considerably in the literature; a meta-analysis published in 2013 reported a median sensitivity of 92% (IQR: 85-97%) and specificity of 60% (IQR: 39-96%), AUC 0.88. Variation according to precise pCR definition was noted with lowest accuracy for studies that include residual DCIS in the definition of pCR (AUC 0.83).[154] By contrast, a more recent meta-analysis (2017) reported a pooled sensitivity of 64% (95% CI: 56–70%) and specificity of 92% (95% CI: 89–94%).[161] When Lobbes *et al* conducted a systematic review they declined to perform meta-analysis citing wide variations in study design. The reported ranges were: sensitivity 25-100% and specificity 50-97%.[82]

Three published studies directly compared the diagnostic accuracy of CESM with MRI for predicting pCR,[111 112 151].The results of these studies are shown in the table below, alongside my results.

Study	MRI	Pathology		Sensitivity (95% CI)	Specificity (95% CI)	PPV (95% CI)	NPV (95% CI)	Diagnostic accuracy (95% CI)
		pCR	Non-pCR					
Patel, 2018	CR	19	14	95.00 (75.13-99.87)	68.89 (53.35-81.83)	57.58 (46.48-67.95)	96.88 (81.96 – 99.53)	76.92 (64.81-86.47)
	Non-CR	1	31					
Barra, 2018	CR	6	2	75.00 (34.91-96.81)	92.00 (73.97-99.02)	75.00 (42.81-92.32)	92.00 (77.50 – 97.46)	87.88 (71.80-96.60)
	Non-CR	2	23					
Iotti, 2017	CR	7	15	87.50 (47.35-99.68)	60.53 (43.39-75.96)	31.82 (22.53-42.82)	95.83 (78.32 – 99.32)	65.22 (49.75-78.65)
	Non-CR	1	23					
Savaridas	CR	4	1	44.44 (13.70-78.80)	85.71 (42.13-99.64)	80.00 (36.13-96.59)	54.55 (38.33 – 69.85)	62.50 (35.43-84.80)
	Non-CR	5	6					

Table 46: Summary of comparative papers reporting diagnostic accuracy of MRI

CONDOR demonstrates a comparable specificity and diagnostic accuracy but noticeably lower sensitivity. This may be partly related to pCR definition. In both my

study and that of Barra *et al*, pCR is defined as the absence of invasive disease. The relatively low sensitivity in both cases may be attributed to residual *in situ* disease identified on MRI. For example, in 3 of 5 cases of MRI false negative in my study there was residual in-situ disease, thus if pCR was defined as ypT0 they would have been classified was true positives. It is also important to note that whilst it is useful to identify a complete response to avoid unnecessary surgery, it is critical that residual disease is not left within the breast due to falsely reassuring imaging. For this reason, the need for a high specificity outweighs the need for high sensitivity.
There is a dearth of published data for the use of DBT in the context of NACT with no studies directly comparing diagnostic accuracy for pCR in relation to, or combined with, CESM. My results are similar to two studies which compared DBT to mammography and ultrasound, reporting a sensitivity of 44.7-50% and specificity of 91-97.6%.[110 162] The accuracy of DBT was lower than CESM and I did not demonstrate any additive value in combining DBT with CESM; therefore my findings do not support use of CE- DBT for detection of pCR.

In addition to being able to identify complete response pre-operatively, it is important to quantify the size of residual disease to guide surgical decision making – whether breast conserving surgery is feasible – and also to improve surgical margins and reduce surgical re-excision rates. Whilst presence of residual *in situ* disease in the absence of invasive disease does not affect survival or local recurrence rate,[87] it is important for surgical decision making as it needs to be excised. Therefore, I considered both residual whole tumour size (WTS) – for surgical decision making, and invasive tumour size (ITS) – for prognostication.
In my data set, MRI had the greatest accuracy for predicting residual WTS; concordance co-efficient 0.86 (CI:0.67-0.95). Promising results were seen for CESM, especially when the presence of micro-calcification was considered in addition to residual enhancement, concordance co-efficient 0.69 (CI: 0.36-0.87). By contrast, with respect to residual ITS, the greatest concordance is seen with CESM(CE), 0.70 (CI: 0.39-0.88) followed by MRI, 0.66 (CI: 0.34-0.85). No benefit was demonstrated by incorporating DBT to produce a full CE-DBT score for predicting either WTS or ITS.
My results are consistent with published data on CESM for the prediction of residual disease, with concordance coefficients which range from moderate to good, 0.7-

0.81.[111 112 151] Of these three papers, Iotti *et al* considered a combined measure using both the recombined image and micro-calcification on the LE energy and specify that residual *in situ* disease is included in the pathological measurement.[111] By contrast Barra *et al* and Patel *et al* report only residual enhancement on the recombined image alone and the exact definition of residual malignancy remains unclear.[112 151] Variation between CESM and pathology size range greatly in the published literature, from 85mm to 2.75mm,[111 112 151 152 158] this is likely related to the varied definitions of pathological and CESM size as described above. I report a mean underestimation of WTS by 10.7mm for CESM(CE), improving to 3.4mm for CESM(CE+calc). With respect to ITS, CESM(CE) produces mean underestimation of 1.3mm whereas CESM(CE+calc) results in a mean overestimation of 5.5mm.

It is widely accepted that the presence of residual mammographic microcalcifications is not consistently related to residual disease. However, even in the context of loss of MRI enhancement it is not possible to predict absence of residual disease with sufficient accuracy to avoid complete excision of tumour bed calcifications.[163 164] My results suggest that this finding is also true for persistent microcalcifications in the absence of CESM enhancement. This is further supported by Iotti *et al* who report that whilst the combination of CE+calc increases sensitivity in the detection and accurate measurement of residual disease, it also increases the false positive rate.[155]

Concordance for MRI varied greatly across the three papers that directly compared CESM and MRI, from poor (0.4) to good (0.79), with the average size differential between pathology and MRI ranging from 18mm-5.4mm.[111 112 151] - all demonstrated poorer results than this study. However, a systematic review of 35 MRI studies reported correlation coefficients ranging from 0.21-0.982 with a median value of 0.698[82] and whilst these cannot be directly compared to concordance coefficients it suggests our results are reflective of previous larger studies.

Average mammogram-pathology size variation is reported by two comparative studies and varies dramatically, 7mm- 63mm. [151 152] This is likely in-part due to the fact that the imaging for Steinhof-Radwanska was performed prior to the final cycle of chemotherapy.[152] A further study of 104 patients, which compared FFDM with US, reported that mammographic size was accurate within 10mm in less than a third of cases, and could not be measured in almost half.[165] By comparison, I found an average difference of 2.8mm between LE mammography and pathology.

Two papers consider DBT for assessing residual disease after NACT. Park *et al* reported an intraclass correlation coefficient (ICC) of 0.63, with mean difference between DBT and pathology of 16.6mm [110]. Skarping *et al* reported that DBT was accurate at predicting invasive size on pathology to within 2mm in 27% of cases[162]. Our results appear to be at least as good as those.

In summary, I have demonstrated promising results for the use of CESM in predicting response to neoadjuvant chemotherapy on post-treatment imaging. My results suggest CESM may have greatest accuracy for predicting pCR and residual invasive tumour size. I recommend that the RC images are reported in parallel with residual microcalcifications on the LE mammograms to improve accuracy of predicting residual *in situ* disease. My findings do not support the addition of DBT to the CESM study.

8.2 Secondary Outcomes

8.2.1 Diagnostic accuracy for detection of multifocality of initial imaging

Within my dataset, CESM demonstrated the greatest accuracy 94.6%, sensitivity 83.3%, specificity 96.8%, PPV 100%, NPV 88.6%; compared to accuracy 89.2%, sensitivity 66.7%, specificity 93.6%, PPV 66.7% and NPV 93.6% of MRI for identifying additional sites of disease.

It is essential to recognise that whilst high levels of specificity are desirable, to avoid unnecessary procedures (biopsies) and/or extensive surgery, this is out-weighed by the need for high levels of sensitivity. It is important to identify all sites of disease pre-treatment to allow marker clips to be inserted so the tumour bed can be subsequently localised.

Only one malignant lesion was missed on CESM, an invasive lobular cancer which was occult on all imaging, possibly due to its proximity to a known IDC giving the radiological appearance of one lesion. By contrast an additional contralateral grade one cancer was missed on MRI. This could have substantial clinical significance as grade one cancers tend not to respond to systemic treatment and there is no possibility it would have been inadvertently excised or treated with radiotherapy alongside the contralateral index lesion.

There was one CESM false positive, a papilloma excision bed. In clinical practice the reader would have had access to this clinical information which modifies interpretation of images. By contrast MRI false positives generated three additional benign biopsies. Although LE and DBT did not identify any false positives both missed at least half the sites of multifocal disease. Therefore, I identified no added value in combining the components of CESM (CE+calc) or CE-DBT.

Three previous studies report the accuracy of CESM and MRI for identifying additional lesions in patients with known breast cancer.[166-168] Consistent with my findings MRI generated a higher number of false-positives than CESM across all three studies. In one study of 52 patients MRI generated 45 false positive results as opposed to just five for CESM.[168] With respect to true positives Li *et al* reported that all lesions were accurately identified by both CESM and MRI, and Lee-Felker *et al* reported that all malignant lesions were identified on CESM with 1/11 occult on MRI. [167 168] By contrast, in Jochelson's dataset only 14/52 lesions were identified on CESM as opposed to 22/25 on MRI; one contralateral lesion was occult on all imaging and only identified following prophylactic mastectomy.

There are two published meta-analyses which consider the overall diagnostic accuracy of CESM. Tagliafico *et al* included 8 studies with a pooled sensitivity of 0.96 (0.96, 1.00) and specificity of 0.58 (0.38, 0.77).[169] However they did not compare with MRI and acknowledge that this was a highly selected cohort with a high prevalence of breast cancer. Indeed, two studies only included patients with known cancer.[68 166] in the remainder CESM was performed due to suspicion of malignancy.[66 67 119 123 170 171]

A subsequent meta-analysis compared the respective diagnostic accuracy of CESM and MRI in thirteen studies. The authors reported pooled sensitivity and specificity of 0.97 (CI:0.95-0.98) and 0.66 (0.59-0.71), compared with 0.97 (0.95-0.98) and 0.52 (0.46-0.58) for CESM and MRI respectively[172]. However, I suggest that the results of this analysis should be treated with a degree of caution. Two included articles were conference abstracts[173 174] and significant patient overlap is suspected between one abstract and a subsequent publication;[167 174] this was not acknowledged in the meta-analysis. Two original articles could not be identified or obtained, despite contacting the author. Whilst they report 'low risk of bias of patient selection' they do not consider the criteria for patient selection nor the prevalence of cancer within the study populations. Of the accessible and verifiable articles, six included patients with

known cancer [68 166-168 173 174] and two included patients with clinical or radiological suspicious disease [67 175 176]. In two instances only the abstract was in English and the methodology was unclear.[177 178] Whilst it is presumed that these factors would affect both imaging modalities equally, so may not be significant for comparative purposes, the overall sensitivity of test will inevitably be affected if 100% of the participants have known disease at enrolment.

In summary, I have demonstrated that the sensitivity for detecting multifocal disease using CESM is similar to MRI. MRI tends to generate a larger number of false positives and therefore lower specificity. Although numbers are small, this suggests that CESM is at least as reliable and may be preferable to MRI. No added benefit was seen when combining CESM with DBT. These findings are supported by the limited existing evidence.

8.2.2 Change in size between initial and interim imaging: predicting response

As breast cancer is a highly heterogeneous disease with variable response to NACT it is very important to accurately identify who will benefit from chemotherapy. In addition to stratifying patients at diagnosis, it is essential to monitor how tumours respond so that chemotherapy can be appropriately tailored, or if necessary stopped, to allow timely surgery.

I considered the relationship between tumour response, assessed with RCB-class and pCR-status, and both absolute and percentage change in tumour size between pre-treatment and mid-treatment (after 3-cycles of chemotherapy). Previous studies have demonstrated that change in MRI lesion size can be a useful early indicator of response and survival outcomes.[117 179-181] Although one previous study included CESM-imaging at mid-treatment,[111] this is the first study to consider change in lesion size on CESM or CE-DBT for early response prediction.

As anticipated, for CESM, CESM(CE+calc), CE-DBT and MRI, greater average size reductions were seen in the pCR and RCB 0-I groups as opposed to non-pCR and RCB II-III. These trends were not seen for LE mammo or DBT, most likely a result of multifocality that was not appreciated on unenhanced initial imaging.

The only trend which reached statistical significance in relation to predicting RCB-status was percentage-size change on MRI, median: -31.25 vs -12.67, p=0.045 for a

RCB 0-I and RCB II-III respectively. There was insufficient power for a significant correlation, Pearson r=0.497, p = 0.084. A similar trend which neared statistically significance was seen for CESM, median: -37.50 vs -17.56, p=0.08; the trend was less clear for CE-DBT, median: -29.41 vs -17.56, p=0.198. With respect to pCR, the most notable trend was seen in percentage-size change on CESM which neared statistical significance, median: -34.95 vs -17.56, p=0.098.

I have demonstrated that percentage-change is a more reliable indicator than absolute change and results suggest that CESM and MRI are more reliable than unenhanced imaging modalities.

The main limitation for this analysis is the small cohort, 14 lesions for pCR status and 13 lesions for RCB score. One patient developed a satellite at mid-treatment yet went on to have a pathological complete response and in such a small cohort an unusual occurrence such as this is likely to skew results. The variation in chemotherapy regimens between patients may also confound results.

At present chemotherapy regimens at our centre are not routinely tailored or modified according to response at mid-point. Currently the chemotherapy plan is agreed pre-treatment and in most cases, is three cycles of FEC followed by 3 cycles of taxane. To maximise potential benefits from early response prediction it would be worth exploring earlier monitoring, perhaps after two cycles of chemotherapy.

8.2.3 Patient Experience

To my knowledge, this is the first study to assess patient experience of CE-DBT; furthermore, it is the first study to assess patient preference for any form of contrast-enhanced mammographic technique when used for assessing response to NACT. This study has demonstrated that, assuming CE-DBT and MRI provided equivalent diagnostic information, the majority of patients prefer CE-DBT. This did not vary between pre-treatment and end-of-treatment, suggesting that previous experience of the techniques did not influence attitude. Overall experience was also significantly more positive for CE-DBT, with 64% reporting it as excellent, as opposed to only 36% reporting an excellent MRI experience. These findings are supported by previous studies that report a patient preference for CESM over MRI both in the setting of local staging and high risk screening. Similar to previous findings of the most common reasons for preference of CE-DBT or CESM were faster time and

greater comfort.[69 182] Unlike the study by Hobbs *et al,* noise level was not cited as a reason for preference in our cohort, although five patients mentioned MRI-associated noise in free-text.

Consistent with a previous study,[69] anxiety was significantly higher in the MRI group when measured using a VAS ($p = 0.003$), and descriptively higher when measured using the categorical rating scale ($p = 0.052$). Free-text reasons for specific anxiety related to MRI concerned the enclosed space, lying still for a prolonged period and noise. Conversely, significantly more positive responses for CE-DBT were seen in relation to being put at ease by staff. We suggest that the close proximity of staff during CE-DBT, enabling them to reassure patients, reduced the anxiety patients experienced.

Anxieties relating to cannulation and/or contrast administration were only recorded in the free text in relation to CE-DBT not MRI. However, unlike the findings of Hobbs *et al,* no significant difference was demonstrated between modalities either in pain on cannulation or unpleasant sensations associated with contrast injection.[69] Sensations described varied between the two techniques, iodinated CE-DBT contrast more commonly associated with heat or flushing, the sensation of passing urine and odd taste, and gadolinium more commonly associated with a cold sensation.

Breast pain relating to mammographic compression is a widely reported patient concern, and has been shown to be associated with non-re-attendance for mammographic screening.[183] Consistent with Hobbs *et al,* it is therefore perhaps not surprising that significantly more women experienced breast pain associated with CE-DBT than MRI (categorical $p = 0.021$, VAS $p=0.011$).[69] However, it is reassuring that despite the increased compression time necessary to allow DBT acquisition in addition to CESM, overall patient preference and experience remains in favour of CE-DBT. Unlike in previous studies, patients were also asked to report on pain in the rest of the body experienced during both techniques. Significantly more women experienced non-breast pain with MRI ($p = 0.046$). Pain was predominantly related to upper limb and pressure on the face / forehead experienced during MRI. This finding has not been previously reported and may offset the increased breast pain experienced with mammographic techniques. These findings may be useful for clinical guidelines and planning.

One limitation of the study is the small sample size of 18 patients. Because of this small sample size, responses for pre-treatment and end-of-treatment were pooled for statistical analysis. The majority of patients (ten), were only included at one time-point. However, the responses of six patients were included at both time-points. Whilst this could potentially bias results, our findings are consistent with previous slightly larger studies.

This study compared patient experience of CE-DBT to that of a 3T MRI scanner, as opposed to a 1.5T MRI scanner. It is possible that the negative experience of some patients associated with MRI may have been compounded by the higher field strength and the narrower bore. However, whilst studies have demonstrated that patients experience symptoms such as vertigo/dizziness, headache and spinal pain more frequently with 3T MRI there is no evidence that these symptoms were common in our cohort.[184] In addition to comfort, the primary reason for CE-DBT preference was cited as the shorter study time. This factor would remain true irrespective of magnetic field strength. Therefore, whilst it is possible that the preference of CE-DBT was magnified by the higher field strength it is unlikely that it would alter overall preference.

Chapter 9: Discussion (TACESM)

As use of CESM becomes more widespread, the need to improve diagnostic accuracy increases. Evidence suggests that the diagnostic accuracy of CESM is non-inferior to MRI,[67 68] and significantly higher than that of FFDM.[66] However, as the use of CESM becomes more widespread it is possible that this will result in increased false positive rates as occurs with MRI.[61] Whilst detecting cancer that will affect treatment and prognosis is key to all imaging techniques, the added burden of false positives to both the patient and health-service is not insignificant, especially if a benign biopsy is performed.[185 186]

The primary aim of this exploratory work was to establish whether CESM textural analysis features could be used to build a model to differentiate benign from malignant lesions. I anticipate the clinical application for this through the use of a decision support system.[187] This would provide additional quantitative information to the reporting radiologist, with the goal of increasing specificity and mitigating the risk of increased false positives. No published data pertaining to textural analysis applied to CESM images was available when the study protocol was prepared in 2017. Therefore this model was proposed based on textural analysis work conducted on breast MRI images, which demonstrated success using models based on GLCM textural data.[92] Subsequent to the inception of this project there has been a flurry of publications pertaining to textural analysis of CESM images.[133-140 188 189]

9.1 TA modelling: Benign vs malignant lesions

I have demonstrated that it is possible to create a highly accurate model to differentiate benign and malignant lesions. The database of lesions on which these models are built is larger than all currently published studies, which may partly explain why this work demonstrates higher levels of accuracy. The most accurate model developed in this study, which included both CC and MLO views, and ellipsoid_ROI and FH_ROI, demonstrates accuracy measures of 0.955, 0.988, 0.954, 0.962, 0.878, 0.998 for accuracy, AUC, sensitivity, specificity, PPV and NPV

respectively. By comparison, ranges of accuracy measures within published studies are consistently lower, as illustrated in table 47 below[135-140 189].

Paper	Model	Accuracy	AUC	Sensitivity	Specificity	PPV	NPV
Fanizzi*	-	0.875	0.931	0.875	0.917	NR	NR
Losurdo*	Embedded STAT	0.807	NR	0.864	0.750	NR	NR
	Wrapper STAT	0.809	NR	0.903	0.716	NR	NR
Patel‡	-	0.90	0.95	0.88	0.92	NR	NR
Gao‡		0.85	0.84	0.89	0.80	NR	NR
Danala‡	RC	0.685	0.737	NR	NR	0.875	0.615
Perek	FT AlexNet	NR	0.843	NR	NR	NR	NR
	RawNet	NR	0.824	NR	NR	NR	NR
Lin	Rad-Score	NR	0.868	0.700	0.800	NR	NR

Table 47: Diagnostic accuracy measures of published studies.*‡Case overlap. NR: not reported

Five of these publications come from two groups; the Mayo Clinic, USA [136 137 189] and Istituto Tumori “Giovanni Paolo II”, Italy [139 140]. Interestingly, despite overlap of the CESM images from which the textural features were derived, substantial variation is seen in model accuracy measures. This is may be related to different modelling techniques and emphasises the need for establishing the basic technique and consistency.

For a decision support tool to be clinically practical, it is important to minimise the additional time required to acquire the necessary information. With regard to a radiomics model this pertains both to the time taken for the images to be segmented and the time for the model to run. Therefore, unlike other studies that generated 100’s or even 1000’s of textural features, I built on existing work which demonstrated that GLCM features of heterogeneity / entropy on breast MRI are important for differentiated benign from malignant lesions.[100 190-192] This is due to the greater intratumoural heterogeneity seen in malignant lesions with imaging textural features shown to correlate with the underlying pathological heterogeneity.[102]

Furthermore, previous studies - with the possible exception of Perek *et al*[135] - include textural feature data from both the low energy (LE) and subtracted recombined image (RC). By contrast I only included feature data from RC images. This reduces workload for radiologists – fewer images to segment – in addition to reducing the required computing power with no loss in diagnostic accuracy.

I have sought to establish the best method of segmentation, both by comparing results of models built with data from freehand or ellipsoid ROIs. To my knowledge this is the first study to address this question. Existing studies use a range of

segmentation methods; including the whole tumour outline (equivalent of FH_ROI)[189]; consistently sized rectangular_ROI either including the whole tumour and some surrounding tissue[139] or contained within the tumour – similar to ellipsoid_ROI[137]; or sample patches around randomly selected pixels.[135] Whilst both ROI models in this study demonstrated high levels of accuracy, the ellipsoid_ROI demonstrated significantly better performance across all measures. This is especially useful for future modelling as segmenting using an ellipsoid ROI is substantially quicker than delineating the precise boundary of the lesion using a freehand technique.

A second novel aspect of this study is that I have investigated whether it is necessary to use both mammographic views, the MLO and the CC. The majority of previous studies have used data from both views,[135-137 189] two studies used whichever view the lesion was better seen on[139 140] and one used CC alone.[138] I have shown that whilst overall the model using ROIs on both imaging views produces a more accurate model; accuracy remains high for models designed with single views alone, with no significant difference in AUC values across all three models. Of the models looking at single views alone, the CC_ROI model demonstrated higher accuracy and F1 score, and was more accurate at identifying benign lesions with a higher specificity and PPV. The MLO_ROI model was better at detecting malignant lesions with a higher NPV and sensitivity. Although the differences between the single-view models and two-view model are statistically significant, the additional time to draw the second ROI and compute the textural analysis figures needs to be taken into account. I would suggest that the marginal gains, for example accuracy of 0.947 vs 0.955 or F1 score of 0.959 vs 0.99 does not justify the additional time required for the two-view model. Furthermore, whilst the results suggest that the MLO model is preferable, if a lesion is only visible on CC I would propose that this should also be included. Future work will include building a model using *either* CC or MLO views. Interestingly the published models that have taken this approach do appear to have higher accuracy than those using both views or CC alone, [138-140] but due to wide variation in other methodological aspects of the studies, direct comparison is not feasible.

Furthermore, this is the first multi-vendor study to look at textural analysis of CESM images. All the models described above derived radiomic data from images acquired on both GE and Hologic equipment. Although it was not possible to test the relative

model accuracy on separate vendor databases, due to small sub-set numbers, this is a very promising finding. Ultimately, for optimal clinical use a radiomics package would be available on PACS (Picture Archiving and Communication System) workstations to allow quantifiable data to be extracted and used in real-time reporting, irrespective of the machine on which the images were acquired.

9.2 TA modelling: Tumour grade

Breast cancers are known to have heterogeneous architecture yet initial grading has to be based on a small sample of tissue derived from core biopsies. Histological grade on core biopsy and surgical specimen are only moderately correlated, with core biopsy tending to under-estimate, simply because the most aggressive part of the tumour has not been sampled. This has the potential to exclude patients who would benefit from neoadjuvant chemotherapy.[193] This paper explores the possibility radiomics may offer as a non-invasive alternative assessment of the whole tumour. Due to the small number of grade 1 lesions, it was not possible to classify lesions into grade 1, grade 2 and grade 3. Therefore, a model was designed comparing lower grades 1+2 (LG) with high grade 3 (HG) cancer. The model included 90 lesions classified as LG, and 62 classified as HG. Accuracy was fairly low at 0.630, with an AUC of 0.653; this is likely to be related to the small sub-set numbers.
This question has been considered by two other groups Istituto Tumori “Giovanni Paolo II”, Italy [188] and Memorial Sloan Kettering Cancer Centre, USA.[133 134] Interestingly, whilst both the Italian group and the later work published by Marino *et al* compared LG (grade 1+ grade 2) vs HG (grade 3) in a similar manner to this study, they also included *in situ* disease in the models.[133 188] This implies that low grade DCIS is synonymous with grade 2 IDC or high grade DCIS with grade 3 IDC, which on a clinical level is simply inaccurate. For example, whilst grade 3 IDC is likely to benefit from neoadjuvant chemotherapy, this would not be routinely offered to a patient with high grade DCIS. Conversely, the initial work from Marino *et al* considers invasive tumours separately but differentiates grade 1 invasive tumours from grades 2 and 3 invasive tumours.[134] The model built by the Italian group utilises seven textural features (from first order histogram and GLCM) similar to this study, however they were derived from both LE and RC images. By contrast, and

consistent with this study, the American group derived textural information from RC images alone, but included multiple feature classes (first-order histogram, GLCM, RLM, GRA, WAV, GEO). There is presumed overlap in the case set included in papers published by Marino *et al,* with the latter also comparing results to MRI radiomics. Despite none of the papers providing detailed measures of accuracy, the results reported are promising. For the models differentiating grade 1+2 from grade 3 tumours, La Forgia *et al* report an AUC of 0.7985, and Marino *et al* reports an accuracy of 77.8% using RUN features.[133 188] The greatest accuracy for differentiating invasive cancers (grade 1 vs grade 2+3) is reported as 90% based on COM features.[134]

Whilst the evidence for using radiomics to support the grading of invasive tumours remains limited, it is showing signs of promise. It is particularly useful to note that two of the published studies report greatest accuracy using models built on COM +/- first order histogram features, as we have done.[133 188]

9.3 TA modelling: Receptor Status

I have demonstrated models for classifying ER, PR and HER2 status, with accuracies greater for HER2+ status (accuracy 0.791, AUC 0.836) and ER status (Accuracy 0.767, AUC 0.819). Lower accuracy was seen for PR status (accuracy 0.667, AUC 0.640). These findings are consistent with the previous studies. La Forgia *et al* reports AUCs of 90.87%, 83.79% and 75.50% for discriminating HER2+/HER2-, ER+/ER- and PR+/PR-respectively.[188] The approach of Marino *et al* is slightly different as they do not discriminate between ER and PR, instead building a model for hormone receptor (HR) status. The earlier paper, with the larger dataset, reports the greatest accuracy for differentiating HR+/HR- lesions as 78.4% using COM+HIS.[134] Interestingly the latter work by this group, reports the greatest accuracy for differentiating HR+/HR- as 95.6% using RUN features instead. This later paper also considers HER2+/HR- vs HER2-/HR+, reporting an accuracy of 97.2% using RLM+WAV features.[133]

It should be noted that although this work is interesting, evidence suggests that despite the heterogeneity of tumours; ER and HER2 status can be accurately

assessed on core biopsy, although weaker concordance between core biopsy and surgical specimens for PR status may be seen.[14 16]

9.5 TA modelling: Predicting response to NACT

I suggest that the greatest utility will be as a decision support tool to identify tumours that will benefit from NACT, but would not be eligible based on core biopsy alone. I explored the possibility of developing a model that differentiated lesions that proceeded to have a pCR and those that did not, based on the radiomic features acquired on the pre-treatment CESM. The ability to create an accurate model was hampered by sample-size, especially the pCR subset which only included 28 lesions. However, results are promising with accuracy of 0.704 and AUC of 0.744. However, the lower F1 score of 0.559 suggests that the reliability is fairly poor and emphases the need to develop this work further on a larger dataset. To my knowledge, only one previous study considered pre-treatment radiomic assessment of CESM images to predict response to NACT.[194] Unlike my work this study uses RECIST criteria to assess response with 'NAC-ineffective' defined as tumours with <30% dimensional decrease. 117 patients were included with a training set of 97 (75 NAC-effective, 22 NAC-ineffective) and a validation set of 20 (10 NAC-effective, 10 ineffective). 792 radiomics features were extracted from the LE and RC images of the CC view. Feature reduction was applied and identified 11 optimal features, of which 9 were derived from the RC images. They report an AUC of 0.861 for the training set and 0.81 for the validation set.[194] However despite the higher AUC values, I would suggest that pCR is a preferential outcome measure as it is a recognised predictor of disease-free and overall survival.[87]

9.6 Dynamic tumour enhancement on CESM

This exploratory work was undertaken to investigate whether it is possible to derive dynamic enhancement characteristics from CESM; which features are the most useful and whether these features may be predictive of histological grade and response to neoadjuvant chemotherapy. In addition, associations with MRI dynamic curves were considered.

Due to the lack of benign lesions with dynamic enhancement within our dataset it was not possible to investigate whether dynamic features generated from CESM could be useful adjuncts for predicting malignancy.

No significant differences in dynamic enhancement features were identified with respect to invasive tumour grade (LG vs HG). Associations between MRI dynamics and tumour grade have been demonstrated,[50 55] but to our knowledge this is the first research to investigate associations between tumour grade and CESM enhancement. Two studies have investigated associations between CESM-enhancement and invasivity of disease with inconsistent results; Rudnicki *et al* reported significant differences between invasive and *in situ* $p < 0.008$, but no significant difference between benign and *in situ* lesions[128]; Liu *et al* described a significant difference between benign lesions and *in situ* disease $p = 0.001$, but no difference between *in situ* and infiltrating disease $p = 1.000$ [131]. It is perhaps therefore not surprising that we were unable to identify significant differences with our small subset groups.

With regard to pCR-status the only significant difference was seen at pre-treatment for FH_WI_Perc99, and this was lost at mid-treatment. However, not only were subset numbers small, there was only overlap between pre- and mid-treatment dynamic features for 11 lesions. Therefore, this finding should be treated with caution and further research with a larger dataset is required to verify the result.

The greatest clinical application of quantitative assessment of lesions on CESM may be in the non-invasive assessment of the whole tumour for prediction of tumour immunohistology (including tumour grade) and prediction of response to NACT. At present initial treatment decisions are based on core-biopsy, which only provides a small proportion of tumour for pathological assessment. Breast cancers are known to be heterogeneous, both in grade and hormone-receptor status. If the least aggressive part of a tumour is biopsied the treatment offered may not be appropriate. Therefore, I suggest that although the results based on this small dataset are limited, further investigation with a larger dataset is justified.

Previous work has demonstrated associations between MRI dynamics and both tumour grade[50 55] and tumour response to NACT.[56 195] Therefore associations between CESM-dynamic features and MRI curve-type (type 1-2 vs type 3) were undertaken as a surrogate marker, revealing promising results.

With regard to ellipsoid-ROI measures, all 'wash-out' characteristics - WI, RD and SRD, demonstrated a significant difference between type 1-2 and type 3 curves, whether using the 99 % value, the 90% value or mean value. With the freehand-ROI, only 90% values of WI, RD and SRD demonstrated significant correlations. As expected, RD and SRD demonstrated lower (more negative) figures in the MRI type 3 group when compared to the MRI type 1-2 group. Washout index (WI), based on previous work with MRI demonstrated the expected inverse relationship, increasing with a greater degree of washout pixels.

There is no published research which compares the 'curves' generated by CESM and MRI. Interestingly, whilst SRD values demonstrated significant differences between the different MRI-curve groups, this did not translate into a strong correlation between CESM-curve and MRI-curve. This suggests that focused work is required to establish whether applying the same cut-off to CESM images as MRI images is appropriate. If 10% change in enhancement was an appropriate cut-off for our data, a correlation with MRI-curves would have been expected.

Two previous studies have compared CESM washout characteristics in benign and malignant lesions using the 10% cut-off. Unlike in this study, they did not use a formal delayed image, instead the routinely acquired MLO and CC views were compared.[127 131] Therefore the time interval between views was considerably shorter than in our study - 102 and 104 seconds versus the minimum six-minute (360 second) delay for our protocol. Significantly more malignant lesions demonstrated decreasing or washout characteristics (a decrease of ≥10%) (67-71%), whereas the majority of benign lesions demonstrated increasing or wash-in characteristics (48-58%), $p < 0.05$. Interpretation and translation of these results is hampered by the that there is no standardised method for calculating dynamic CESM enhancement and no recognised unit of enhancement. Further research is required to establish what an appropriate cut-off for CESM enhancement would be and to clarify how best to calculate this.

It is widely accepted that malignant lesions are significantly more likely to demonstrate intense enhancement on CESM images, as measured both qualitatively and quantitatively, than those which are benign.[120 121 123 125 127 128 130 131] With respect to MRI, early phase enhancement of ≥80% has a 91% sensitivity for prediction of malignancy, and is a component of type 3 MRI-curves.[49 196] Therefore, as a measure of initial enhancement, CESM RLE would be expected to be higher in malignant

lesions. It was also postulated that higher values may be seen more frequently with MRI-type 3 curves. Unfortunately, as previously discussed it is not possible to analyse for associations with malignancy. No significant difference between MRI curve types was identified. This is most likely due to the lack of type-1 curves resulting in the necessary grouping of type 1 and 2. Although less marked, type 2 curves also demonstrate rapid initial enhancement.[196] Other potential factors include the lack of true baseline (unenhanced measure) for CESM meaning that background enhancement has to be used instead so although similar, the CESM-RLE could not be calculated using the same equation as MRI-RLE. Furthermore, differences in the timing of the first post-contrast image may be significant. Peak enhancement on MRI is seen at 2 minutes [196], while initial CESM-images were performed at least 3 minutes after contrast administration. Future work to investigate possible associations between RLE and malignancy is required, both to establish whether there is an association and subsequently whether there is a benefit in calculating RLE in addition to simply measuring the maximum enhancement as previously reported.

Similarly, when matched data for qualitative and quantitative CESM enhancement characteristics was analysed, significant differences were observed for all quantitative measures with the exception of RLE. Interestingly, on subsequent pairwise comparison, whilst significant differences between the quantitative values in curve types II and III were consistently demonstrated, a significant difference between curve type I and III was only seen in half of the measures. This is surprising as the greatest difference would be anticipated between these groups. It is likely that this lack of difference between type I and III curves is related to the small sub-set of type I curves rather than suggesting that these curve types are similar. Whilst there is no published work assessing correlation between dynamic enhancement characteristics measured qualitatively and quantitatively, Rundnicki *et al* concluded no statistically significant difference between qualitative assessment and quantitative assessment of initial enhancement.[132]

Unlike MRI, the CESM images are not generated in a sequential order whilst the patient remains static and due to the method of creating the reconstructed (subtracted) CESM image there is no true pre-enhanced image. Therefore, it is not possible to derive dynamic enhancement features for selected pixels or voxels; rather the entire ROI is analysed each time. The reason for greater accuracy with the

ellipsoid-ROI may be related to reproducibility. Whilst it is not possible for the breast to be positioned identically between the initial and delayed imaging, the ellipsoid-ROI was unchanged. Whilst the intention was for the ellipsoid ROI to encompass the largest area of the enhancing lesion it was not possible to cover the full extent, the discrepancy dependent on the shape of the lesion. The ellipsoid ROI was identical between the initial and delayed image and was placed over the same area of the lesion, thus the same subset of pixels was analysed. By contrast the freehand-ROI was drawn individually on each image to encompass the entire enhancing area at that time point. Therefore, there may be greater variability in the lesion-ROI when using the freehand method. It is also worth noting that the ellipsoid-ROI was quick to draw manually, and is more readily transferable to an automated or semi-automated system. In addition, using the 90th percentile value has demonstrated the most consistently significant results. This may be due to the fact that it is not possible to assess enhancement change in discrete pixel groups; using the 99th percentile will identify the greatest enhancement on the initial image but may miss the washout as the specific pixel will drop below the 99th percentile. The mean will account for this but may underemphasise the initial enhancement; thus I suggest that the 90th percentile offers a compromise. Promising results have been demonstrated with all measures of 'washout' (WI, RD, SRD) but less so for initial lesion enhancement (RLE), although this is likely due to the fact that prediction of malignancy could not be performed.

I recommend that further research concentrates on segmentation with an ellipsoid ROI using values derived from the 90th percentile enhancement level. Moving forward it will be important to establish a recognised and standardised measure of enhancement to allow calibration of equipment. Research can then be extended to clarify the most accurate measure of dynamic enhancement characteristics and cut-off values to permit categorisation as is seen with MRI curves.

9.7 Limitations

Although larger or equivalent to similar publications, the sample sizes in this work remain modest which limits the validity of the modelling. All segmentation was conducted by the same reader, whilst this prevents confounding due to inter-reader variability, it reduces the generalizability of the data modelling. It will be necessary to

demonstrate reproducibility of segmentation with consideration of both inter- and intra-reader variability in future work on larger datasets.

Chapter 10: Conclusion

In this work, I have taken a multifaceted approach to investigate the accuracy of CESM and CE-DBT in comparison to MRI, with particular emphasis on the neoadjuvant chemotherapy setting and application of radiomics.

Consistent with previous studies, I have shown that CESM is a promising technique with similar accuracy to MRI for predicting response to neoadjuvant chemotherapy. However, unlike other studies I have also investigated the incremental benefit of the LE image, RC image and DBT in this setting. I have demonstrated that the RC image alone is the most accurate for identifying pCR (absence of invasive disease) and predicting the size of residual invasive disease. Combining the presence and size of suspicious microcalcification on LE image with RC lesion size improved the accuracy for predicting WTS, i.e. invasive and *in situ* disease, although this remained lower than the accuracy of MRI. The addition of DBT did not improve accuracy for any measure. This is particularly important to note due to the additional radiation dose associated with DBT.

With respect to pre-treatment imaging, it is important to maximise specificity without compromising sensitivity at pre-treatment, to ensure that all malignant foci are detected whilst limiting the number of false positives as this can lead to delays in treatment and distress for the patient whilst they are investigated. CESM demonstrated higher sensitivity and specificity than MRI for detecting multifocality. This accuracy may be further enhanced with the use of radiomics.

I have shown that textural analysis can be applied to CESM images with the ultimate aim to build a radiomics decision support tool. For such a tool to be clinically applicable it needs to be compatible with multi-vendor images, and it is important to maximise the efficiency of feature extraction and analysis. To my knowledge this is the first study to build radiomic models using multivendor CESM images. These models demonstrate consistently high accuracy for discriminating benign from malignant lesions. In the context of NACT, this is especially important at pre-treatment imaging for the reasons described above.

In addition, this is the first study to investigate optimal segmentation techniques. I have established that segmentation with an ellipsoid_ROI is not only quicker than

using a manually delineated FH_ROI, but also more accurate. A model that included both views was only marginally better than a single view model and all models were highly accurate. Furthermore, my models were based on features extracted from the RC only, and showed no loss in accuracy compared to published studies using a combination of features from both LE and RC images. In summary, I have shown that it is only necessary to segment single view RC images using an ellipsoid_ROI to build a highly accurate model to discriminate benign and malignant lesions. The results of this study also indicate that ellipsoid_ROI is preferable for characterising dynamic enhancement characteristics between initial and delayed imaging.

A further potentially important role for CESM radiomics is as a non-invasive way to assess the entire tumour to predict benefit from NACT. At present, this decision is based primarily on the immunohistopathology of core biopsies of the tumour, which represent only a small sample of a potentially heterogeneous tumour. The model designed to predict response to neoadjuvant chemotherapy (pCR) based on pre-treatment radiomic features had limited accuracy, likely due to small subset numbers. However, models with slightly larger datasets to discriminate tumour grade, hormone receptor and HER2 status were more successful. This suggests that there is merit in extending the pCR prediction modelling work with larger datasets.

To allow development of CESM radiomics, harmonisation of methods of image acquisition, post-processing, segmentation and feature calculation will be critical. This will allow sufficiently large and representative datasets on which to develop and validate predictive models and classifiers.

Finally, critical to any change in practice is patient acceptability. I have shown that the majority of patients prefer CE-DBT to MRI, in the context on NACT monitoring. The main disadvantage of CE-DBT reported by patients was breast pain. It is possible that if the DBT element was removed and thus the compression time reduced, that patient acceptance would increase further.

In summary; I have demonstrated that CESM has similar accuracy to MRI, and is preferred by patients for monitoring response to NACT. The combination of LE and RC images improves the accuracy of size estimation of residual *in situ* disease. Radiomic models built using ellipisoid_ROIs from single view RC images are highly accurate for discriminating benign from malignant lesions and may be useful as decision support tools in the future. Further models for predicting tumour

immunohistopathology and response to NACT also show promise but require development with larger datasets.

References

1. Scotland I. Breast Cancer. Cancer Statistics: ISD Scotland, 2017.
2. Pathology reporting of breast disease in surgical excision specimens incorporating the dataset for histological reporting of breast cancer. 2016
3. Gorringe KL, Fox SB. Ductal Carcinoma In Situ Biology, Biomarkers, and Diagnosis. *Frontiers in Oncology* 2017;7 doi: 10.3389/fonc.2017.00248
4. Kneeshaw PJ, Turnbull LW, Smith A, et al. Dynamic contrast enhanced magnetic resonance imaging aids the surgical management of invasive lobular breast cancer. *European Journal of Surgical Oncology (EJSO)* 2003;29(1):32-37. doi: 10.1053/ejso.2002.1391
5. Ciriello G, Gatza Michael L, Beck Andrew H, et al. Comprehensive Molecular Portraits of Invasive Lobular Breast Cancer. *Cell* 2015;163(2):506-19. doi: 10.1016/j.cell.2015.09.033
6. Makki J. Diversity of Breast Carcinoma: Histological Subtypes and Clinical Relevance. *Clinical Medicine Insights: Pathology* 2015;8:CPath.S31563. doi: 10.4137/CPath.S31563
7. Bloom HJG, Richardson WW. Histological grading and prognosis in breast cancer. *Br J Cancer*:359-77.
8. Rakha EA, El-Sayed ME, Lee AHS, et al. Prognostic Significance of Nottingham Histologic Grade in Invasive Breast Carcinoma. *Journal of Clinical Oncology* 2008;26(19):3153-58. doi: 10.1200/jco.2007.15.5986
9. Knuttel FM, Menezes GLG, van Diest PJ, et al. Meta-analysis of the concordance of histological grade of breast cancer between core needle biopsy and surgical excision specimen. *British Journal of Surgery* 2016;103(6):644-55. doi: 10.1002/bjs.10128
10. Carroll JS. EJE PRIZE 2016: Mechanisms of oestrogen receptor (ER) gene regulation in breast cancer. *European Journal of Endocrinology* 2016;175(1):R41-R49. doi: 10.1530/eje-16-0124
11. Relevance of breast cancer hormone receptors and other factors to the efficacy of adjuvant tamoxifen: patient-level meta-analysis of randomised trials. *The Lancet* 2011;378(9793):771-84. doi: 10.1016/s0140-6736(11)60993-8
12. Hammond MEH, Hayes DF, Dowsett M, et al. American Society of Clinical Oncology / College of American Pathologists Guideline for Immunohistochemical Testing of Estrogen and Progesterone Receptors in Breast Cancer (Unabridged Version). *Arch Pathol Lab Med* 2010;134:e48-e72.
13. Bardou V-J, Arpino G, Elledge RM, et al. Progesterone Receptor Status Significantly Improves Outcome Prediction Over Estrogen Receptor Status Alone for Adjuvant Endocrine Therapy in Two Large Breast Cancer Databases. *Journal of Clinical Oncology* 2003;21(10):1973-79. doi: 10.1200/jco.2003.09.099
14. Asogan AB, Hong GS, Arni Prabhakaran SK. Concordance between core needle biopsy and surgical specimen for oestrogen receptor, progesterone receptor and human epidermal growth factor receptor 2 status in breast cancer. *Singapore Medical Journal* 2017;58(3):145-49. doi: 10.11622/smedj.2016062

15. Chen X, Yuan Y, Gu Z, et al. Accuracy of estrogen receptor, progesterone receptor, and HER2 status between core needle and open excision biopsy in breast cancer: a meta-analysis. *Breast Cancer Research and Treatment* 2012;134(3):957-67. doi: 10.1007/s10549-012-1990-z
16. Arnedos M, Nerurkar A, Osin P, et al. Discordance between core needle biopsy (CNB) and excisional biopsy (EB) for estrogen receptor (ER), progesterone receptor (PgR) and HER2 status in early breast cancer (EBC). *Annals of Oncology* 2009;20(12):1948-52. doi: 10.1093/annonc/mdp234
17. Ross JS, Fletcher JA, Linette GP, et al. The HER-2/neu Gene and Protein in Breast Cancer 2003: Biomarker and Target of Therapy. *The oncologist* 2003;8:307-25.
18. Dowsett M, Cooke T, Ellis I, et al. Assessment of HER2 status in breast cancer: why, whne and how? *Eur J Cancer* 2000;36:170-76.
19. Untch M, Rezai M, Loibl S, et al. Neoadjuvant Treatment With Trastuzumab in HER2-Positive Breast Cancer: Results From the GeparQuattro Study. *Journal of Clinical Oncology* 2010;28(12):2024-31. doi: 10.1200/jco.2009.23.8451
20. Baselga J, Bradbury I, Eidtmann H, et al. Lapatinib with trastuzumab for HER2-positive early breast cancer (NeoALTTO): a randomised, open-label, multicentre, phase 3 trial. *The Lancet* 2012;379(9816):633-40. doi: 10.1016/s0140-6736(11)61847-3
21. Di Leo A, Gomez HL, Aziz Z, et al. Phase III, Double-Blind, Randomized Study Comparing Lapatinib Plus Paclitaxel With Placebo Plus Paclitaxel As First-Line Treatment for Metastatic Breast Cancer. *Journal of Clinical Oncology* 2008;26(34):5544-52. doi: 10.1200/jco.2008.16.2578
22. Perou CM, Sørlie T, Eisen MB, et al. Molecular portraits of human breast tumours. *Nature* 2000;406(6797):747-52. doi: 10.1038/35021093
23. Sorlie T, Perou CM, Tibshirani R, et al. Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications. *Proceedings of the National Academy of Sciences* 2001;98(19):10869-74. doi: 10.1073/pnas.191367098
24. Fragomeni SM, Sciallis A, Jeruss JS. Molecular Subtypes and Local-Regional Control of Breast Cancer. *Surgical Oncology Clinics of North America* 2018;27(1):95-120. doi: 10.1016/j.soc.2017.08.005
25. Pilewskie M, Morrow M. Axillary Nodal Management Following Neoadjuvant Chemotherapy. *JAMA Oncology* 2017;3(4):549. doi: 10.1001/jamaoncol.2016.4163
26. Koh J, Kim MJ. Introduction of a New Staging System of Breast Cancer for Radiologists: An Emphasis on the Prognostic Stage. *Korean Journal of Radiology* 2019;20(1):69. doi: 10.3348/kjr.2018.0231
27. RCR guidance on screening and symptomatic breast imaging 2019
28. Tse GMK, Law BKB, F MTK, et al. Hamartoma of the breast: a clinicopathological review. *J Clin Path* 2002;55:951-54.
29. Vasei N, Shishegar A, Ghalkhani F, et al. Fat necrosis in the Breast: A systematic review of clinical. *Lipids in Health and Disease* 2019;18(1) doi: 10.1186/s12944-019-1078-4
30. Wald NJ, Murphy P, Major P, et al. UKCCCR multicentre randomised controlled trial of one and two view mammography in breast cancer screening. *Bmj* 1995;311(7014):1189-93. doi: 10.1136/bmj.311.7014.1189
31. Pisano ED, Gatsonis C, Hendrick E, et al. Diagnostic perfromance of Digital versus Film Mammography for Breast-Cancer Screening. *NEJM* 2005;353(17):1773-83.
32. Jochelson M. Advanced Imaging Tachniques for the detection of Breast Cancer. ASCO Educational Book: Amercian Society of Clinical Oncology:65-69.

33. Mandelson MT. Breast density as a predictor of mammographic detection: Comparison of interval and screen detected cancers.
34. Berg WA, Sechtin AG, Marques H, et al. Cystic Breast Masses and the ACRIN 6666 Experience. *Radiologic Clinics of North America* 2010;48(5):931-87. doi: 10.1016/j.rcl.2010.06.007
35. Stavros AT, Thickman D, Rapp CL, et al. Solid breast nodules: Use of sonography to distinguish between banign and malignant lesions. *Radiology* 1995;196:123-34.
36. Mehta TS, Raza S, Baum JK. Use of Doppler Ultraound in the Evaluation of Breast Carcinoma. *Semin Ultrasound CT MR* 2000;21(4):297-307. doi: 10.1053/sult.2000.8923
37. Park AY, Seo BK. Up-to-date Doppler techniques for breast tumor vascularity: superb microvascular imaging and contrast-enhanced ultrasound. *Ultrasonography* 2018;37(2):98-106. doi: 10.14366/usg.17043
38. Polan RL, Klein BD, Richman RH. Scintimammography in patients with minimal mammographic or clinical findings. *RadioGraphics* 2001;21(3):641-53.
39. Dibble EH, Hunt KN, Ehman EC, et al. Molecular Breast Imaging in Clinical Practice. *American Journal of Roentgenology* 2020;215(2):277-84. doi: 10.2214/ajr.19.22622
40. Huppe AI, Mehta AK, Brem RF. Molecular Breast Imaging: A Comprehensive Review. *Seminars in Ultrasound, CT and MRI* 2018;39(1):60-69. doi: 10.1053/j.sult.2017.10.001
41. Rosen EL, Eubank WB, Mankoff DA. FDG PET, PET/CT, and Breast Cancer Imaging. *RadioGraphics* 2007;27:S215-S30.
42. Ulaner GA. PET/CT for Patients With Breast Cancer: Where Is the Clinical Impact? *American Journal of Roentgenology* 2019;213(2):254-65. doi: 10.2214/ajr.19.21177
43. Balleyguier C, Opolon P, Mathieu MC, et al. New potential and applications of contrast-enhanced ultrasound of the breast: Own investigations and review of the literature. *European Journal of Radiology* 2009;69(1):14-23. doi: 10.1016/j.ejrad.2008.07.037
44. Evans A, Sim YT, Thomson K, et al. Shear wave elastography of breast cancer: Sensitivity according to histological type in a large cohort. *The Breast* 2016;26:115-18. doi: 10.1016/j.breast.2016.01.009
45. Evans A, Whelehan P, Thompson A, et al. Prediction of Pathological Complete Response to Neoadjuvant Chemotherapy for Primary Breast Cancer Comparing Interim Ultrasound, Shear Wave Elastography and MRI. *Ultraschall in der Medizin - European Journal of Ultrasound* 2017;39(04):422-31. doi: 10.1055/s-0043-111589
46. Berg WA, Cosgrove DO, Dore CJ, et al. Shear-wave Elastography Improves the Specificity of Breast US: The BE1 Multinational Study of 939 Masses. *Radiology* 2012;262(2):435-49. doi: 10.1148/radiol.11110640/-/DC1
47. Pooley RA. <rg.254055027.pdf>.
48. Moore MAS. Putting the neo into neoangiogenesis. *Journal of Clinical Investigation* 2002;109(3):313-15. doi: 10.1172/jci0214940
49. Kuhl CK, Mielcareck P, Klaschik S, et al. Dynamic Breast MR Imaging: Are Signal Intensity Time Course Data Useful for Differential Diagnosis of Enhancing Lesions? *Radiology* 1999;211(1):101-10.
50. Mussurakis S, Buckley DL, Horsman A. Dynamic MR imaging of invasive breast cancer: correlation with tumour grade and other histological features. *BJR* 1997;70:446-51.

51. Wang LC, DeMartini WB, Partridge SC, et al. MRI-Detected Suspicious Breast Lesions: Predictive Values of Kinetic Features Measured by Computer-Aided Evaluation. *American Journal of Roentgenology* 2009;193(3):826-31. doi: 10.2214/ajr.08.1335
52. Gutierrez RL, Strigel RM, Partridge SC, et al. Dynamic Breast MRI: Does Lower Temporal Resolution Negatively Affect Clinical Kinetic Analysis? *American Journal of Roentgenology* 2012;199(3):703-08. doi: 10.2214/ajr.11.7836
53. Partridge SC, Stone KM, Strigel RM, et al. Breast DCE-MRI. *Academic Radiology* 2014;21(9):1195-203. doi: 10.1016/j.acra.2014.04.013
54. Williams TC, DeMartini WB, Partridge SC, et al. Breast MRI Imaging: Computer-aided Evaluation Program for Discriminating Benign from Maligant Lesions. *Radiology* 2007;244(1):94-103.
55. Koo HR, Cho N, Song IC, et al. Correlation of perfusion parameters on dynamic contrast-enhanced MRI with prognostic factors and subtypes of breast cancers. *Journal of Magnetic Resonance Imaging* 2012;36(1):145-51. doi: 10.1002/jmri.23635
56. Lv G, Zhou Y, Zheng D, et al. Comparison of Model-Free and Model-Based Dynamic Contrast-Enhanced Magnetic Resonance Imaging Pharmacokinetic Parameters for Predicting Breast Cancers' Response to Neoadjuvant Chemotherapy. *Journal of Computer Assisted Tomography* 2020;44(2):269-74. doi: 10.1097/rct.0000000000001001
57. Amornsiripanitch N, Bickelhaupt S, Shin HJ, et al. Diffusion-weighted MRI for Unenhanced Breast Cancer Screening. *Radiology* 2019;293(3):504-20. doi: 10.1148/radiol.2019182789
58. Park SH, Moon WK, Cho N, et al. Comparison of diffusion-weighted MR imaging and FDG PET/CT to predict pathological complete response to neoadjuvant chemotherapy in patients with breast cancer. *European Radiology* 2011;22(1):18-25. doi: 10.1007/s00330-011-2236-x
59. Zhang L, Tang M, Min Z, et al. Accuracy of combined dynamic contrast-enhanced magnetic resonance imaging and diffusion-weighted imaging for breast cancer detection: a meta-analysis. *Acta Radiologica* 2015;57(6):651-60. doi: 10.1177/0284185115597265
60. Hata T, Takahashi H, Watanabe K, et al. Magnetic resonance imaging for preoperative evaluation of breast cancer: a comparative study with mammography and ultrasonography. *Journal of the American College of Surgeons* 2004;198(2):190-97. doi: 10.1016/j.jamcollsurg.2003.10.008
61. Plana MN, Carreira C, Muriel A, et al. Magnetic resonance imaging in the preoperative assessment of patients with primary breast cancer: systematic review of diagnostic accuracy and meta-analysis. *European Radiology* 2011;22(1):26-38. doi: 10.1007/s00330-011-2238-8
62. Guo BJ, Yang ZL, Zhang LJ. Gadolinium Deposition in Brain: Current Scientific Evidence and Future Perspectives. *Frontiers in Molecular Neuroscience* 2018;11 doi: 10.3389/fnmol.2018.00335
63. Diekmann F. Digital Mammography Using Iodine-Based Contrast. Initial clinical experience with dynamic contrast medium enhancement. *Investigative Radiology*;40(7):397-404.
64. Dromain C, Balleyguier C, Muller S, et al. Evaluation of tumor angiogenesis of breast carcinoma using contrast-enhanced digital mammography. *AJR Am J Roentgenol* 2006;187(5):W528-37. doi: 10.2214/AJR.05.1944

65. Lewin JM, Isaacs PK, Vance V, et al. Dual-energy contrast-enhanced digital subtraction mammography: feasibility. *Radiology* 2003;229(1):261-8.
66. Luczyńska E, Heinze-Paluchowska S, Dyczek S, et al. Contrast-Enhanced Spectral Mammography: Comparison with Conventional Mammography and Histopathology in 152 Women. *Korean Journal of Radiology* 2014;15(6):689. doi: 10.3348/kjr.2014.15.6.689
67. Luczynska E, Heinze-Paluchowska S, Hendrick E, et al. Comparison between breast MRI and contrast-enhanced spectral mammography. *Med Sci Monit* 2015;21:1358-67. doi: 10.12659/MSM.893018
68. Fallenberg EM, Dromain C, Diekmann F, et al. Contrast-enhanced spectral mammography versus MRI: Initial results in the detection of breast cancer and assessment of tumour size. *Eur Radiol* 2014;24(1):256-64. doi: 10.1007/s00330-013-3007-7
69. Hobbs MM, Taylor DB, Buzynski S, et al. Contrast-enhanced spectral mammography (CESM) and contrast enhanced MRI (CEMRI): Patient preferences and tolerance. *J Med Imaging Radiat Oncol* 2015;59(3):300-5. doi: 10.1111/1754-9485.12296
70. Michell MJ, Batohi B. Role of tomosynthesis in breast imaging going forward. *Clinical Radiology* 2018;73(4):358-71. doi: 10.1016/j.crad.2018.01.001
71. Gilbert FJ, Pinker-Domenig K. Diagnosis and Staging of Breast Cancer: When and How to Use Mammography, Tomosynthesis, Ultrasound, Contrast-Enhanced Mammography, and Magnetic Resonance Imaging. 2019:155-66. doi: 10.1007/978-3-030-11149-6_13
72. James JR, Pavlicek W, Hanson JA, et al. Breast Radiation Dose With CESM Compared With 2D FFDM and 3D Tomosynthesis Mammography. *American Journal of Roentgenology* 2017;208(2):362-72. doi: 10.2214/ajr.16.16743
73. Caumo F, Zorzi M, Brunelli S, et al. Digital Breast Tomosynthesis with Synthesized Two-Dimensional Images versus Full-Field Digital Mammography for Population Screening: Outcomes from the Verona Screening Program. *Radiology* 2018;287(1):37-46. doi: 10.1148/radiol.2017170745
74. Sanmugasiva VV, Ramli Hamid MT, Fadzli F, et al. Diagnostic accuracy of digital breast tomosynthesis in combination with 2D mammography for the characterisation of mammographic abnormalities. *Scientific Reports* 2020;10(1) doi: 10.1038/s41598-020-77456-6
75. Chamming's F, Kao E, Aldis A, et al. Imaging features and conspicuity of invasive lobular carcinomas on digital breast tomosynthesis. *The British Journal of Radiology* 2017;90(1073):20170128. doi: 10.1259/bjr.20170128
76. Wasan R, Morel J, Iqbal A. Digital breast tomosynthesis improves the accuracy of the diagnosis of circumscribed lesions because of margin visibility. *Breast Cancer Res*;16(06) doi: https://doi.org/10.1186/bcr3701
77. Expert Panel on Breast I, Slanetz PJ, Moy L, et al. ACR Appropriateness Criteria((R)) Monitoring Response to Neoadjuvant Systemic Therapy for Breast Cancer. *J Am Coll Radiol* 2017;14(11S):S462-S75. doi: 10.1016/j.jacr.2017.08.037
78. Bear HD, Anderson S, Smith RE, et al. Sequential Preoperative or Postoperative Docetaxel Added to Preoperative Doxorubicin Plus Cyclophosphamide for Operable Breast Cancer: National Surgical Adjuvant Breast and Bowel Project Protocol B-27. *Journal of Clinical Oncology* 2006;24(13):2019-27. doi: 10.1200/jco.2005.04.1665
79. Buzdar AU, Ibrahim NK, Francis D, et al. Significantly Higher Pathologic Complete Remission Rate After Neoadjuvant Therapy With Trastuzumab, Paclitaxel, and

Epirubicin Chemotherapy: Results of a Randomized Trial in Human Epidermal Growth Factor Receptor 2–Positive Operable Breast Cancer. *Journal of Clinical Oncology* 2005;23(16):3676-85. doi: 10.1200/jco.2005.07.032
80. Díaz-Redondo T, Lavado-Valenzuela R, Jimenez B, et al. Different Pathological Complete Response Rates According to PAM50 Subtype in HER2+ Breast Cancer Patients Treated With Neoadjuvant Pertuzumab/Trastuzumab vs. Trastuzumab Plus Standard Chemotherapy: An Analysis of Real-World Data. *Frontiers in Oncology* 2019;9 doi: 10.3389/fonc.2019.01178
81. Recommendations for cross-sectional imaging in cancer management, second edition. *Clin Radiol*
82. Lobbes MBI, Prevos R, Smidt M, et al. The role of magnetic resonance imaging in assessing residual disease and pathologic complete response in breast cancer patients receiving neoadjuvant chemotherapy: a systematic review. *Insights into Imaging* 2013;4(2):163-75. doi: 10.1007/s13244-013-0219-y
83. Munn Z, Moola S, Lisy K, et al. Claustrophobia in magnetic resonance imaging: A systematic review and meta-analysis. *Radiography* 2015;21(2):e59-e63. doi: 10.1016/j.radi.2014.12.004
84. Fowler AM, Mankoff DA, Joe BN. Imaging Neoadjuvant Therapy Response in Breast Cancer. *Radiology* 2017;285(2):358-75. doi: 10.1148/radiol.2017170180
85. Eisenhauer EA, Therasse P, Bogaerts J, et al. New response evaluation criteria in solid tumours: Revised RECIST guideline (version 1.1). *European Journal of Cancer* 2009;45(2):228-47. doi: 10.1016/j.ejca.2008.10.026
86. von Minckwitz G, Untch M, Blohmer J-U, et al. Definition and Impact of Pathologic Complete Response on Prognosis After Neoadjuvant Chemotherapy in Various Intrinsic Breast Cancer Subtypes. *Journal of Clinical Oncology* 2012;30(15):1796-804. doi: 10.1200/jco.2011.38.8595
87. Mazouni C, Peintinger F, Wan-Kau S, et al. Residual Ductal Carcinoma In Situ in Patients With Complete Eradication of Invasive Breast Cancer After Neoadjuvant Chemotherapy Does Not Adversely Affect Patient Outcome. *Journal of Clinical Oncology* 2007;25(19):2650-55. doi: 10.1200/jco.2006.08.2271
88. Symmans WF, Peintinger F, Hatzis C, et al. Measurement of Residual Breast Cancer Burden to Predict Survival After Neoadjuvant Chemotherapy. *Journal of Clinical Oncology* 2007;25(28):4414-22. doi: 10.1200/jco.2007.10.6823
89. Symmans WF, Wei C, Gould R, et al. Long-Term Prognostic Risk After Neoadjuvant Chemotherapy Associated With Residual Cancer Burden and Breast Cancer Subtype. *Journal of Clinical Oncology* 2017;35(10):1049-60. doi: 10.1200/jco.2015.63.1010
90. Yau C, Van Der Noordaa M, Wei J, et al. Abstract GS5-01: Residual cancer burden after neoadjuvant therapy and long-term survival outcomes in breast cancer: A multi-centre pooled analysis. *Cancer Research* 2020;80(4 Suppliment) doi: 10.1158/1538-7445.SABCS19-GS5-01
91. Hieken T, Hamy A-S, Darrigues L, et al. Prognostic value of the Residual Cancer Burden index according to breast cancer subtype: Validation on a cohort of BC patients treated by neoadjuvant chemotherapy. *Plos One* 2020;15(6):e0234191. doi: 10.1371/journal.pone.0234191
92. Waugh SA, Purdie CA, Jordan LB, et al. Magnetic resonance imaging texture analysis classification of primary breast cancer. *European Radiology* 2015;26(2):322-30. doi: 10.1007/s00330-015-3845-6

93. Henderson S, Purdie C, Michie C, et al. Interim heterogeneity changes measured using entropy texture features on T2-weighted MRI at 3.0 T are associated with pathological response to neoadjuvant chemotherapy in primary breast cancer. *European Radiology* 2017;27(11):4602-11. doi: 10.1007/s00330-017-4850-8
94. Molina-García D, García-Vicente AM, Pérez-Beteta J, et al. Intratumoral heterogeneity in 18F-FDG PET/CT by textural analysis in breast cancer as a predictive and prognostic subrogate. *Annals of Nuclear Medicine* 2018;32(6):379-88. doi: 10.1007/s12149-018-1253-0
95. Fan M, Yuan W, Zhao W, et al. Joint Prediction of Breast Cancer Histological Grade and Ki-67 Expression Level Based on DCE-MRI and DWI Radiomics. *IEEE Journal of Biomedical and Health Informatics* 2020;24(6):1632-42. doi: 10.1109/jbhi.2019.2956351
96. Szczypinski PM, Strzelecki M, Materka A. Mazda - a software for texture analysis. 2007:245-49. doi: 10.1109/isitc.2007.15
97. Szczypiński PM, Strzelecki M, Materka A, et al. MaZda—A software package for image texture analysis. *Computer Methods and Programs in Biomedicine* 2009;94(1):66-76. doi: 10.1016/j.cmpb.2008.08.005
98. Strzelecki M, Szczypinski P, Materka A, et al. A software tool for automatic classification and segmentation of 2D/3D medical images. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 2013;702:137-40. doi: 10.1016/j.nima.2012.09.006
99. Castellano G, Bonilha L, Li LM, et al. Texture analysis of medical images. *Clin Radiol* 2004;59(12):1061-9. doi: 10.1016/j.crad.2004.07.008
100. Gibbs P, Turnbull LW. Textural analysis of contrast-enhanced MR images of the breast. *Magn Reson Med* 2003;50(1):92-8. doi: 10.1002/mrm.10496
101. Parikh J, Selmi M, Charles-Edwards G, et al. Changes in Primary Breast Cancer Heterogeneity May Augment Midtreatment MR Imaging Assessment of Response to Neoadjuvant Chemotherapy. *Radiology* 2014;272(1):100-12.
102. Davnall F, Yip CS, Ljungqvist G, et al. Assessment of tumor heterogeneity: an emerging imaging tool for clinical practice? *Insights Imaging* 2012;3(6):573-89. doi: 10.1007/s13244-012-0196-6
103. Ahmed A, Gibbs P, Pickles M, et al. Texture analysis in assessment and prediction of chemotherapy response in breast cancer. *J Magn Reson Imaging* 2013;38(1):89-101. doi: 10.1002/jmri.23971
104. Michoux N, Van den Broeck S, Lacoste L, et al. Texture analysis on MR images helps predicting non-response to NAC in breast cancer. *BMC Cancer* 2015;15:574. doi: 10.1186/s12885-015-1563-8
105. Teruel JR, Heldahl MG, Goa PE, et al. Dynamic contrast-enhanced MRI texture analysis for pretreatment prediction of clinical and pathological response to neoadjuvant chemotherapy in patients with locally advanced breast cancer. *NMR Biomed* 2014;27(8):887-96. doi: 10.1002/nbm.3132
106. Chou CP, Lewin JM, Chiang CL, et al. Clinical evaluation of contrast-enhanced digital mammography and contrast enhanced tomosynthesis--Comparison to contrast-enhanced breast MRI. *Eur J Radiol* 2015;84(12):2501-8. doi: 10.1016/j.ejrad.2015.09.019
107. Chou CP, Lewin J, Pan HB. Contrast-enhanced tomosynthesis: The best of both worlds or more of the same? *Eur J Radiol* 2016;85(2):509. doi: 10.1016/j.ejrad.2015.11.033

108. Lobbes MB, Houben I. Contrast-enhanced tomosynthesis: The best of both worlds or more of the same? *Eur J Radiol* 2016;85(2):507-8. doi: 10.1016/j.ejrad.2015.11.024
109. Helal MH, Mansour SM, Zaglol M, et al. Staging of breast cancer and the advanced applications of digital mammogram: what the physician needs to know? *The British Journal of Radiology* 2017;90(1071):20160717. doi: 10.1259/bjr.20160717
110. Park J, Chae EY, Cha JH, et al. Comparison of mammography, digital breast tomosynthesis, automated breast ultrasound, magnetic resonance imaging in evaluation of residual tumor after neoadjuvant chemotherapy. *Eur J Radiol* 2018;108:261-68. doi: 10.1016/j.ejrad.2018.09.032
111. Iotti V, Ravaioli S, Vacondio R, et al. Contrast-enhanced spectral mammography in neoadjuvant chemotherapy monitoring: a comparison with breast magnetic resonance imaging. *Breast Cancer Res* 2017;19(1):106. doi: 10.1186/s13058-017-0899-1
112. Patel BK, Hilal T, Covington M, et al. Contrast-Enhanced Spectral Mammography is Comparable to MRI in the Assessment of Residual Breast Cancer Following Neoadjuvant Systemic Therapy. *Ann Surg Oncol* 2018;25(5):1350-56. doi: 10.1245/s10434-018-6413-x
113. Barra FR, de Souza FF, Camelo R, et al. Accuracy of contrast-enhanced spectral mammography for estimating residual tumor size after neoadjuvant chemotherapy in patients with breast cancer: a feasibility study. *Radiol Bras* 2017;50(4):224-30. doi: 10.1590/0100-3984.2016-0029
114. Macura KJ, Ouwerkerk R, Jacobs MA, et al. Patterns of Enhancement on Breast MR Images: Interpretation and Imaging Pitfalls. *RadioGraphics* 2006;26(6):1719-34.
115. Pickles M, Lowry M, Manton DJ, et al. Role of dynamic contrast enhanced MRI in monitoring early response of locally advanced breast cancer to neoadjuvant chemotherapy. *Breast Cancer Research and Treatment* 2005;91:1-10.
116. Golden DI, Lipson JA, Telli ML, et al. Dynamic contrast-enhanced MRI-based biomarkers of therapeutic response in triple-negative breast cancer. *Journal of the American Medical Informatics Association* 2013;20(6):1059-66. doi: 10.1136/amiajnl-2012-001460
117. Thibault G, Tudorica A, Afzal A, et al. DCE-MRI Texture Features for Early Prediction of Breast Cancer Therapy Response. *Tomography* 2017;3(1):23-32. doi: 10.18383/j.tom.2016.00241
118. Wu J, Gong G, Cui Y, et al. Intratumor partitioning and texture analysis of dynamic contrast-enhanced (DCE)-MRI identifies relevant tumor subregions to predict pathological response of breast cancer to neoadjuvant chemotherapy. *J Magn Reson Imaging* 2016;44(5):1107-15. doi: 10.1002/jmri.25279
119. Luczynska E, Niemiec J, Ambicka A, et al. Correlation between blood and lymphatic vessel density and results of contrast-enhanced spectral mammography. *Polish Journal of Pathology* 2015;3:310-22. doi: 10.5114/pjp.2015.54965
120. Goh Y, Chan CW, Pillay P, et al. Architecture distortion score (ADS) in malignancy risk stratification of architectural distortion on contrast-enhanced digital mammography. *European Radiology* 2021;31:2657-66. doi: 10.1007/s00330-020-07395-3
121. Travieso-Aja MM, Maldonado-Saluzzi D, Naranjo-Santana P, et al. Evaluation of the applicability of BI-RADS® MRI for the interpretation of contrast-enhanced digital mammography. *Radiología (English Edition)* 2019;61(6):477-88. doi: 10.1016/j.rxeng.2019.07.003

122. Chi X, Zhang L, Xing D, et al. Diagnostic value of the enhancement intensity and enhancement pattern of CESM to benign and malignant breast lesions. *Medicine* 2020;99(37):e22097. doi: 10.1097/md.0000000000022097
123. Mohamed Kamal R, Hussien Helal M, Wessam R, et al. Contrast-enhanced spectral mammography: Impact of the qualitative morphology descriptors on the diagnosis of breast lesions. *European Journal of Radiology* 2015;84(6):1049-55. doi: 10.1016/j.ejrad.2015.03.005
124. van Nijnatten TJA, Jochelson MS, Pinker K, et al. Differences in degree of lesion enhancement on CEM between ILC and IDC. *BJR*|*Open* 2019;1(1):20180046. doi: 10.1259/bjro.20180046
125. Tsigginou A, Gkali C, Chalazonitis A, et al. Adding the power of iodinated contrast media to the credibility of mammography in breast cancer diagnosis. *Br J Radiol* 2016;89(1067):20160397. doi: 10.1259/bjr.20160397
126. Luczynska E, Niemiec J, Heinze S, et al. Intensity and Pattern of Enhancement on CESM: Prognostic Significance and its Relation to Expression of Podoplanin in Tumor Stroma – A Preliminary Report. *Anticancer Research* 2018;38(2) doi: 10.21873/anticanres.12327
127. Deng C-Y, Juan Y-H, Cheung Y-C, et al. Quantitative analysis of enhanced malignant and benign lesions on contrast-enhanced spectral mammography. *The British Journal of Radiology* 2018:20170605. doi: 10.1259/bjr.20170605
128. Rudnicki W, Heinze S, Niemiec J, et al. Correlation between quantitative assessment of contrast enhancement in contrast-enhanced spectral mammography (CESM) and histopathology—preliminary results. *European Radiology* 2019;29(11):6220-26. doi: 10.1007/s00330-019-06232-6
129. Lobbes MBI, Mulder HKP, Rousch M, et al. Quantification of enhancement in contrast-enhanced spectral mammography using a custom-made quantifier tool (I-STRIP): A proof-of-concept study. *European Journal of Radiology* 2018;106:114-21. doi: 10.1016/j.ejrad.2018.07.021
130. Lv Y, Chi X, Sun B, et al. Diagnostic Value of Quantitative Gray-Scale Analysis of Contrast-Enhanced Spectral Mammography for Benign and Malignant Breast Lesions. *Journal of Computer Assisted Tomography* 2020;44(3):405-12. doi: 10.1097/rct.0000000000001019
131. Liu Y, Zhao S, Huang J, et al. Quantitative Analysis of Enhancement Intensity and Patterns on Contrast-enhanced Spectral Mammography. *Scientific Reports* 2020;10(1) doi: 10.1038/s41598-020-66501-z
132. Rudnicki W, Heinze S, Popiela T, et al. Quantitative Assessment of Contrast Enhancement on Contrast Enhancement Spectral Mammography (CESM) and Comparison With Qualitative Assessment. *Anticancer Research* 2020;40(5):2925-32. doi: 10.21873/anticanres.14270
133. Marino MA, Leithner D, Sung J, et al. Radiomics for Tumor Characterization in Breast Cancer Patients: A Feasibility Study Comparing Contrast-Enhanced Mammography and Magnetic Resonance Imaging. *Diagnostics* 2020;10(7):492. doi: 10.3390/diagnostics10070492
134. Marino MA, Pinker K, Leithner D, et al. Contrast-Enhanced Mammography and Radiomics Analysis for Noninvasive Breast Cancer Characterization: Initial Results. *Molecular Imaging and Biology* 2019;22(3):780-87. doi: 10.1007/s11307-019-01423-5

135. Perek S, Kiryati N, Zimmerman-Moreno G, et al. Classification of contrast-enhanced spectral mammography (CESM) images. *International Journal of Computer Assisted Radiology and Surgery* 2018;14(2):249-57. doi: 10.1007/s11548-018-1876-6
136. Danala G, Patel B, Aghaei F, et al. Classification of Breast Masses Using a Computer-Aided Diagnosis Scheme of Contrast Enhanced Digital Mammograms. *Annals of Biomedical Engineering* 2018;46(9):1419-31. doi: 10.1007/s10439-018-2044-4
137. Patel BK, Ranjbar S, Wu T, et al. Computer-aided diagnosis of contrast-enhanced spectral mammography: A feasibility study. *Eur J Radiol* 2018;98:207-13. doi: 10.1016/j.ejrad.2017.11.024
138. Lin F, Wang Z, Zhang K, et al. Contrast-Enhanced Spectral Mammography-Based Radiomics Nomogram for Identifying Benign and Malignant Breast Lesions of Sub-1 cm. *Frontiers in Oncology* 2020;10 doi: 10.3389/fonc.2020.573630
139. Fanizzi A, Losurdo L, Basile TMA, et al. Fully Automated Support System for Diagnosis of Breast Cancer in Contrast-Enhanced Spectral Mammography Images. *Journal of Clinical Medicine* 2019;8(6):891. doi: 10.3390/jcm8060891
140. Losurdo L, Fanizzi A, Basile TMA, et al. Radiomics Analysis on Contrast-Enhanced Spectral Mammography Images for Breast Cancer Diagnosis: A Pilot Study. *Entropy* 2019;21(11):1110. doi: 10.3390/e21111110
141. Kinkel K, Hylton N. Challenges to Intepretation of Breast MRI. *J Magn Reson Imaging* 2001;13:821-29.
142. Radiologists TRCo. Guidance on screening and symptomatic breast imaging, fourth edition. London: Royal College of Radiologists, 2019.
143. Lin LI-K. A concordance correlaton coefficient to evaluate reproducibility. *Biometrics* 1989;45:255-68.
144. Heller GZ, Manuguerra M, Chow R. How to analyze the Visual Analogue Scale: Myths, truths and clinical relevance. *Scandinavian Journal of Pain* 2016;13(1):67-75. doi: 10.1016/j.sjpain.2016.06.012
145. Tennant SL, James JJ, Cornford EJ, et al. Contrast-enhanced spectral mammography improves diagnostic accuracy in the symptomatic setting. *Clinical Radiology* 2016;71(11):1148-55. doi: 10.1016/j.crad.2016.05.009
146. Szczypinski PM, Strzelecki M, Materka A, et al. MaZda--a software package for image texture analysis. *Comput Methods Programs Biomed* 2009;94(1):66-76. doi: 10.1016/j.cmpb.2008.08.005
147. Agrawal U, Soria D, Wagner C, et al. Combining clustering and classification ensembles: A novel pipeline to identify breast cancer profiles. *Artificial Intelligence in Medicine* 2019;97:27-37. doi: 10.1016/j.artmed.2019.05.002
148. Desai M, Shah M. An anatomization on breast cancer detection and diagnosis employing multi-layer perceptron neural network (MLP) and Convolutional neural network (CNN). *Clinical eHealth* 2021;4:1-11. doi: 10.1016/j.ceh.2020.11.002
149. Tumuluru P, Lakshmi CP, Sahja T, et al. A Review of Machine Learning Techniques for Breast Cancer Diagnosis in Medical Applications. *Third International conference on I-SMAC* 2019:618-23. doi: 10.1109/I-SMAC47947.2019.9032427
150. Yue W, Wang Z, Chen H, et al. Machine Learning with Applications in Breast Cancer Diagnosis and Prognosis. *Designs* 2018;2(2):13. doi: 10.3390/designs2020013
151. Barra FR, Sobrinho AB, Barra RR, et al. Contrast-Enhanced Mammography (CEM) for Detecting Residual Disease after Neoadjuvant Chemotherapy: A Comparison with

Breast Magnetic Resonance Imaging (MRI). *BioMed Research International* 2018;2018:1-9. doi: 10.1155/2018/8531916
152. Steinhof-Radwańska K, Grazynska A, Lorek A, et al. Radiological assessment of response to neoadjuvant chemotherapy in breast cancer females, using standard and spectral mammography. *ResearchSquare* 2021 doi: 10.21203/rs.3.rs-217660/v1
153. Croshaw R, Shapiro-Wright H, Svensson E, et al. Accuracy of Clinical Examination, Digital Mammogram, Ultrasound, and MRI in Determining Postneoadjuvant Pathologic Tumor Response in Operable Breast Cancer Patients. *Annals of Surgical Oncology* 2011;18(11):3160-63. doi: 10.1245/s10434-011-1919-5
154. Marinovich ML, Houssami N, Macaskill P, et al. Meta-Analysis of Magnetic Resonance Imaging in Detecting Residual Breast Cancer After Neoadjuvant Therapy. *JNCI: Journal of the National Cancer Institute* 2013;105(5):321-33. doi: 10.1093/jnci/djs528
155. Iotti V, Ragazzi M, Besutti G, et al. Accuracy and Reproducibility of Contrast-Enhanced Mammography in the Assessment of Response to Neoadjuvant Chemotherapy in Breast Cancer Patients with Calcifications in the Tumor Bed. *Diagnostics* 2021;11(3):435. doi: 10.3390/diagnostics11030435
156. ElSaid NAE, Mahmoud HGM, Salama A, et al. Role of contrast enhanced spectral mammography in predicting pathological response of locally advanced breast cancer post neo-adjuvant chemotherapy. *The Egyptian Journal of Radiology and Nuclear Medicine* 2017;48(2):519-27. doi: 10.1016/j.ejrnm.2017.03.022
157. Moustafa AFI, Kamal RM, Gomaa MMM, et al. Quantitative mathematical objective evaluation of contrast-enhanced spectral mammogram in the assessment of response to neoadjuvant chemotherapy and prediction of residual disease in breast cancer. *Egyptian Journal of Radiology and Nuclear Medicine* 2019;50(1) doi: 10.1186/s43055-019-0041-8
158. Kamal RM, Saad SM, Moustafa AFI, et al. Predicting response to neo-adjuvant chemotherapy and assessment of residual disease in breast cancer using contrast-enhanced spectral mammography: a combined qualitative and quantitative approach. *Egyptian Journal of Radiology and Nuclear Medicine* 2020;51(1) doi: 10.1186/s43055-020-00275-2
159. Valenti EL, Lopez JCA, Solis MC, et al. Correlation of pathological complete response with Contrast-enhanced mammography in breast cancer patients after neoadjuvant chemotherapy. 2020 San Antinio Breast Cancer Virtual Symposium. San Antonio, Texas, 2021.
160. Tang S, Xiang C, Yang Q. The diagnostic performance of CESM and CE-MRI in evaluating the pathological response to neoadjuvant therapy in breast cancer: a systematic review and meta-analysis. *The British Journal of Radiology* 2020;93(1112):20200301. doi: 10.1259/bjr.20200301
161. Gu Y-L, Pan S-M, Ren J, et al. Role of Magnetic Resonance Imaging in Detection of Pathologic Complete Remission in Breast Cancer Patients Treated With Neoadjuvant Chemotherapy: A Meta-analysis. *Clinical Breast Cancer* 2017;17(4):245-55. doi: 10.1016/j.clbc.2016.12.010
162. Skarping I, Förnvik D, Heide-Jørgensen U, et al. Neoadjuvant breast cancer treatment response; tumor size evaluation through different conventional imaging modalities in the NeoDense study. *Acta Oncologica* 2020:1-10. doi: 10.1080/0284186x.2020.1830167

163. Feliciano Y, Mamtani A, Morrow M, et al. Do Calcifications Seen on Mammography After Neoadjuvant Chemotherapy for Breast Cancer Always Need to Be Excised? *Annals of Surgical Oncology* 2017;24(6):1492-98. doi: 10.1245/s10434-016-5741-y
164. An YY, Kim SH, Kang BJ. Residual microcalcifications after neoadjuvant chemotherapy for locally advanced breast cancer: comparison of the accuracies of mammography and MRI in predicting pathological residual tumor. *World Journal of Surgical Oncology* 2017;15(1) doi: 10.1186/s12957-017-1263-8
165. Keune JD, Jeffe DB, Schootman M, et al. Accuracy of ultrasonography and mammography in predicting pathologic response after neoadjuvant chemotherapy for breast cancer. *The American Journal of Surgery* 2010;199(4):477-84. doi: 10.1016/j.amjsurg.2009.03.012
166. Jochelson M, Dershaw DD, Sung J, et al. Bilateral Contrast-enhanced Dual-Energy Digital Mammography: Feasibility and Comparison with conventional Digital Mammography and MR Imaging in Women with known Breast Carcinoma. *Radiology* 2013;266(3):743-51.
167. Li L, Roth R, Germaine P, et al. Contrast-enhanced spectral mammography (CESM) versus breast magnetic resonance imaging (MRI): A retrospective comparison in 66 breast lesions. *Diagnostic and Interventional Imaging* 2017;98(2):113-23. doi: 10.1016/j.diii.2016.08.013
168. Lee-Felker SA, Tekchandani L, Thomas M, et al. Newly Diagnosed Breast Cancer: Comparison of Contrast-enhanced Spectral Mammography and Breast MR Imaging in the Evaluation of Extent of Disease. *Radiology* 2017;285(2):389-400. doi: 10.1148/radiol.2017161592
169. Tagliafico AS, Bignotti B, Rossi F, et al. Diagnostic performance of contrast-enhanced spectral mammography: Systematic review and meta-analysis. *The Breast* 2016;28:13-19. doi: 10.1016/j.breast.2016.04.008
170. Lobbes MBI, Lalji U, Houwers J, et al. Contrast-enhanced spectral mammography in patients referred from the breast cancer screening programme. *European Radiology* 2014 doi: 10.1007/s00330-014-3154-5
171. Cheung Y-C, Tsai H-P, Lo Y-F, et al. Clinical utility of dual-energy contrast-enhanced spectral mammography for breast microcalcifications without associated mass: a preliminary analysis. *European Radiology* 2015;26(4):1082-89. doi: 10.1007/s00330-015-3904-z
172. Xiang W, Rao H, Zhou L. A meta-analysis of contrast-enhanced spectral mammography versus MRI in the diagnosis of breast cancer. *Thoracic Cancer* 2020;11(6):1423-32. doi: 10.1111/1759-7714.13400
173. Dromain C, Canale S, Bidault F, et al. Value of Contrast-enhanced Spectral Mammography in Women with Newly Diagnosed Breast Cancers Compared to MRI: Preliminary results. Radiological Society of North America 2011 Scientific Assembly and Annual Meeting. Chicago, IL, 2011.
174. Li L, Liao L, Germaine P, et al. Abstract P1-02-06: Retrospective comparison of sensitivity and positive predictive value (PPV) of contrast enhanced spectral mammography (CESM) to contrast enhanced breast MRI (BMRI) in 50 malignant breasts. Thirty-Seventh Annual CTRC-AACR San Antonio Breast Cancer Symposium. San Antonio, TX. Philadelphia (PA): Cancer Res, 2015.

175. Zhang C, Wang Q, Wang J, et al. Feasibility of contrast-enhanced spectral mammography in the diagnosis of breast cancer. *Radiologic Practice* 2014(12):1420-23.
176. Wang Q, Li K, Wang L, et al. Preclinical study of diagnostic performances of contrast-enhanced spectral mammography versus MRI for breast diseases in China. *SpringerPlus* 2016;5(1) doi: 10.1186/s40064-016-2385-0
177. Yu M, Li J. Comparative Study of Contrast-Enhanced Spectral Mammography and Dynamic Contrast-Enhanced Magnetic Resonance Imaging in Diagnosis of Breast Cancer. *China Medical Equipment* 2017;32(11):74-77.
178. Zou M, Wang Y-j, Jin B, et al. Comparison of Diagnostic Efficacy between CESM and CE-MRI in Breast Diseases. *Chin Comput Med Imag* 2018;24(3):211-14. doi: 10.19627/j.cnki.cn31-1700/th.2018.03.006
179. Tudorica A, Oh KY, Chui SYC, et al. Early Prediction and Evaluation of Breast Cancer Response to Neoadjuvant Chemotherapy Using Quantitative DCE-MRI. *Translational Oncology* 2016;9(1):8-17. doi: 10.1016/j.tranon.2015.11.016
180. Romeo V, Accardo G, Perillo T, et al. Assessment and Prediction of Response to Neoadjuvant Chemotherapy in Breast Cancer: A Comparison of Imaging Modalities and Future Perspectives. *Cancers* 2021;13(14):3521. doi: 10.3390/cancers13143521
181. Tahmassebi A, Wengert GJ, Helbich TH, et al. Impact of Machine Learning With Multiparametric Magnetic Resonance Imaging of the Breast for Early Prediction of Response to Neoadjuvant Chemotherapy and Survival Outcomes in Breast Cancer Patients. *Investigative Radiology* 2019;54(2):110-17. doi: 10.1097/rli.0000000000000518
182. Phillips J, Miller MM, Mehta TS, et al. Contrast-enhanced spectral mammography (CESM) versus MRI in the high-risk screening setting: patient preferences and attitudes. *Clinical Imaging* 2017;42:193-97. doi: 10.1016/j.clinimag.2016.12.011
183. Whelehan P, Evans A, Wells M, et al. The effect of mammography pain on repeat participation in breast cancer screening: A systematic review. *The Breast* 2013;22(4):389-94. doi: 10.1016/j.breast.2013.03.003
184. Weintraub MI, Khoury A, Cole SP. Biologic Effects of 3 Tesla (T) MR Imaging Comparing Traditional 1.5 T and 0.6 T in 1023 Consecutive Outpatients. *Journal of Neuroimaging* 2007;17(3):241-45. doi: 10.1111/j.1552-6569.2007.00118.x
185. Schou Bredal I, Kåresen R, Skaane P, et al. Recall mammography and psychological distress. *European Journal of Cancer* 2013;49(4):805-11. doi: 10.1016/j.ejca.2012.09.001
186. Hafslund B, Espehaug B, Nortvedt MW. Effects of False-Positive Results in a Breast Screening Program on Anxiety, Depression and Health-Related Quality of Life. *Cancer Nursing* 2012;35(5):E26-E34. doi: 10.1097/NCC.0b013e3182341ddb
187. Stivaros SM, Gledson A, Nenadic G, et al. Decision support systems for clinical radiological practice — towards the next generation. *BJR* 2010;83:904-14.
188. La Forgia D, Fanizzi A, Campobasso F, et al. Radiomic Analysis in Contrast-Enhanced Spectral Mammography for Predicting Breast Cancer Histological Outcome. *Diagnostics* 2020;10(9):708. doi: 10.3390/diagnostics10090708
189. Gao F, Wu T, Li J, et al. SD-CNN: A shallow-deep CNN for imporving breast cancer diagnosis. *Computerized Medical Imaging and Graphics* 2018;70:53-62. doi: 10.1016/j.compmedimag.2018.09.004

190. Wang Y, Liao X, Xiao F, et al. Magnetic Resonance Imaging Texture Analysis in Differentiating Benign and Malignant Breast Lesions of Breast Imaging Reporting and Data System 4. *Journal of Computer Assisted Tomography* 2020;44(1):83-89. doi: 10.1097/rct.0000000000000969
191. Karahaliou A, Vassiou K, Arikidis NS, et al. Assessing heterogeneity of lesion enhancement kinetics in dynamic contrast-enhanced MRI for breast cancer diagnosis. *Br J Radiol* 2010;83(988):296-309. doi: 10.1259/bjr/50743919
192. Fan M, Xia P, Clarke R, et al. Radiogenomic signatures reveal multiscale intratumour heterogeneity associated with biological functions and survival in breast cancer. *Nature Communications* 2020;11(1) doi: 10.1038/s41467-020-18703-2
193. Kwok TnC, Rakha EA, Lee AHS, et al. Histological grading of breast cancer on needle core biopsy: the role of immunohistochemical assessment of proliferation. *Histopathology* 2010;57(2):212-19. doi: 10.1111/j.1365-2559.2010.03620.x
194. Wang Z, Lin F, Ma H, et al. Contrast-Enhanced Spectral Mammography-Based Radiomics Nomogram for the Prediction of Neoadjuvant Chemotherapy-Insensitive Breast Cancers. *Frontiers in Oncology* 2021;11 doi: 10.3389/fonc.2021.605230
195. Yamaguchi A, Honda M, Ishiguro H, et al. Kinetic information from dynamic contrast-enhanced MRI enables prediction of residual cancer burden and prognosis in triple-negative breast cancer: a retrospective study. *Scientific Reports* 2021;11(1) doi: 10.1038/s41598-021-89380-4
196. Mann RM, Kuhl CK, Kinkel K, et al. Breast MRI: guidelines from the European Society of Breast Imaging. *European Radiology* 2008;18(7):1307-18. doi: 10.1007/s00330-008-0863-7

Appendix

University of Dundee

Pt Initials	Pt Study Number

NHS Tayside

CONDOR Study: confidential

Patient Experience Questionnaire 2 – after the MRI scan

1. If you felt any anxiety about the test or during the test, please tell us what this was about:

2. Overall, how much anxiety did you feel **during** the test?
(Please tick one box.)

☐ None ☐ Mild ☐ Moderate ☐ Severe

3. How much did the staff put you at ease during the test?

☐ Not at all ☐ A little ☐ Moderately ☐ Very much

4. During the test, how confident did you feel that you could say stop if you needed to?

☐ Not at all ☐ A little ☐ Moderately ☐ Very

5. How much pain did you feel when the injection needle was put in?

☐ None ☐ Mild ☐ Moderate ☐ Severe

6. If you noticed any strange feelings anywhere in your body when the dye was going in, please describe what you felt:

7. How unpleasant was the feeling of the dye going in?

☐ Not at all ☐ A little ☐ Moderately ☐ Very

8. Overall, how much pain did you feel in your breasts during the test?

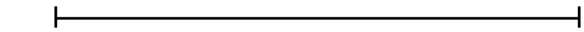
☐ None ☐ Mild ☐ Moderate ☐ Severe

9. If you felt any pain or discomfort in any other parts of your body during the test, please tell us about it:

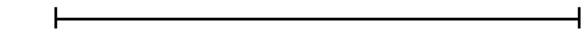
10. Overall, how much discomfort did you feel in your body during the test, not including in your breasts?

☐ None ☐ Mild ☐ Moderate ☐ Severe

11. Overall, how much anxiety did you feel during the test?
(Please put a mark on the line to show your anxiety level.)



12. Overall, how much pain did you feel in your breasts during the test?
(Please put a mark on the line to show your pain level.)



13. How would you rate your **overall** experience during the test?

☐ Poor ☐ Fair ☐ Good ☐ Excellent

14. Please tell us anything else you think people should know about what it's like having the test:

15. If both study tests(the CE-DBT mammogram which you had in the breast clinic and the MRI scan which you had today) gave the same information to your doctor, which would you prefer to have?
(Please tick one box.)

☐ CE-DBT mammogram ☐ MRI scan

16. Please tell us the reasons for your answer above:

Thank you very much for your answers.

For office use only		
Data entry complete	Initials:	
	Date:	