BJR

https://doi.org/10.1259/bjr.20180123

Received: 29 January 2018 | Revised: 20 March 2018 | Accepted: 06 April 2018

Cite this article as:
Henderson SA, Muhammad Gowdh N, Purdie CA, Jordan LB, Evans A, Brunton T, et al. Breast cancer: influence of tumour volume estimation method at MRI on prediction of pathological response to neoadjuvant chemotherapy. *Br J Radiol* 2018; **91**: 20180123.

FULL PAPER

Breast cancer: influence of tumour volume estimation method at MRI on prediction of pathological response to neoadjuvant chemotherapy

[1]SHELLEY A HENDERSON, PhD, [2]NAZLEEN MUHAMMAD GOWDH, FRCR, MBChb, [3]COLIN A PURDIE, FRCPath, [3]LEE B JORDAN, FRCPath, [4]ANDREW EVANS, MRCP, FRCR, [5]TRACY BRUNTON, DCR(R), [6]ALASTAIR M THOMPSON, FRCSEd and [4]SARAH VINNICOMBE, MRCP, FRCR

[1]Department of Medical Physics, Ninewells Hospital, Dundee, UK
[2]Department of Radiology, Aberdeen Royal Infirmary, Aberdeen, UK
[3]Department of Pathology, Ninewells Hospital, Dundee, UK
[4]Division of Imaging and Technology, University of Dundee, Dundee, UK
[5]MRI Department, Clinical Research Centre, University of Dundee, Dundee, UK
[6]Department of Breast Surgical Oncology, University of Texas MD Anderson Cancer Center, Houston, TX, USA

Address correspondence to: Dr Shelley A Henderson
E-mail: *shelley.henderson1@nhs.net*

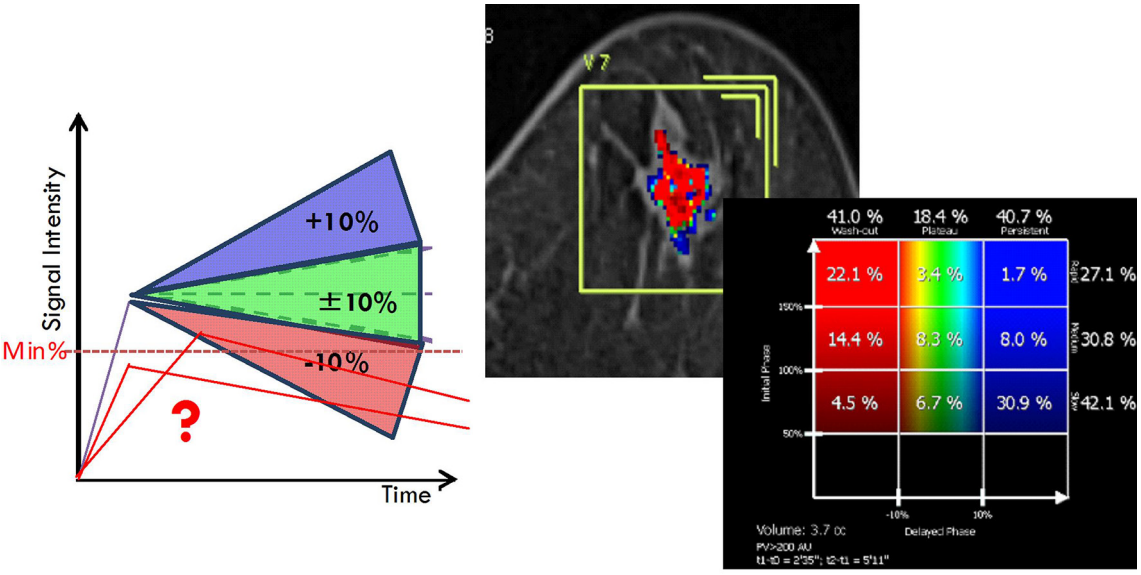
Objective: Does method of tumour volume measurement on MRI influence prediction of treatment outcome in patients with primary breast cancer undergoing neoadjuvant chemotherapy (NAC)?.
Method: The study comprised of 136 women with biopsy-proven breast cancer scheduled for MRI monitoring during NAC treatment. Dynamic contrast-enhanced images were acquired at baseline (pre-NAC) and interim (post three NAC cycles) time points. Functional tumour volumes (FTVs), automatically derived using vendor software and enhancing tumour volumes (ETVs), user-derived using a semi-automated thresholding technique, were calculated at each time point and percentage changes calculated. Response, assessed using residual cancer burden (RCB) score on surgically resected specimens, was compared statistically with volumetric changes and receiver operating characteristic analysis performed.
Results: Mean volumetric differences for each RCB response category were (FTV/ETV): pathological complete response (pCR) 95.5/96.8%, RCB-I 69.8/66.7%, RCB-II 64.0/65.5%, RCB-III 25.4/24.0%. Differences were significant between pCR and RCB-II/RCB-III categories ($p < 0.040$; unpaired t-test) using FTV measures and between pCR and RCB-I/RCB-II/RCB-III categories ($p < 0.006$; unpaired t-test) when ETV was used. Receiver operating characteristic analysis for pCR identification post-NAC yielded area under the curve for FTV/ETV of 0.834/0.920 respectively. Sensitivity and specificity for FTV was 80.0 and 76.8% for FTV and 81.0 and 91.8% for ETV.
Conclusion: ETV changes can identify patients likely to achieve a complete response to NAC. Potentially, this could impact patient management regarding the possible avoidance of post-NAC surgery.
Advances in Knowledge: Interim changes in ETV are more useful than FTV in predicting final pathological response to NAC. ETV differentiates patients who will achieve a complete response from those who will have residual disease.

INTRODUCTION

Neoadjuvant chemotherapy (NAC) is increasingly used in the treatment of primary breast cancer to downstage locally advanced disease and axillary nodal involvement prior to surgery and render breast conserving surgery and sentinel node biopsy feasible.[1,2] It has been demonstrated that the response of the primary tumour to NAC correlates with patient survival[3,4] and the amount of residual tumour left at surgical resection, as quantified by the residual cancer burden (RCB) score, is directly linked with patient outcomes.[5,6] With modern chemotherapy regimens, pathological complete response (pCR) to NAC is increasingly common. If patients who ultimately achieve a pCR could be confidently identified pre-operatively, they might not, in the future, require surgical intervention after NAC completion.[7,8] For this reason, identification of relevant biomarkers of response are essential for improving and personalising patient management.

Several imaging modalities have been used to assess response to NAC, including mammography, ultrasound,

Figure 1. (a) Pixels are colour coded after reaching a minimum enhancement threshold (Min %). They are then coloured according to whether there is persistent enhancement (+10%), plateau enhancement (± 10%) or washout enhancement (–10%). However, the technique does not take into account slower or lower enhancing pixels as shown.as shown. (b) Example image of how pixels are coloured and the corresponding calculated functional tumour volume, considered in terms of washout, plateau and persistent enhancement as shown in (a).



(a) *(b)*

shearwave elastography and MRI. Breast MRI, particularly dynamic contrast-enhanced imaging (DCE-MRI), appears to be more accurate than clinical examination, mammography or ultrasound in assessing final response to NAC.[9,10] More recently, however, attention has focused on the correlation of early changes in breast MRI with pathological response. Various parameters have been investigated, including tumour size changes,[11] diffusion changes,[12,13] contrast uptake and washout kinetics[14,15] as well as more complex pharmacokinetic modelling parameters[16] and textural analysis.[17–19]

However, the ACRIN 6657 I-SPY TRIAL reported that the changes between baseline and interim MRI in a simpler metric, functional tumour volume (FTV), were the most useful in the early prediction of response.[20,21] Furthermore, baseline tumour volume and changes in response to treatment carry important prognostic information, smaller reductions in volume being associated with poorer patient-related outcomes.[11,22] The FTV reflects the number of pixels within a breast tumour that reach a minimum pre-defined threshold of signal enhancement early after contrast administration; all such pixels are summed to produce a volume. The enhancement threshold generally used is 50% but it may be as high at 100%. Pixels that attain the threshold are usually also "coloured" to provide further information as to whether signal intensity reduces, plateaus, or continues to increase steadily after the peak enhancement, reflecting the presence or absence of contrast washout and thus, providing a visual summary of tumour enhancement kinetics (Figure 1). However, FTV measures require either dedicated computer-aided detection (CAD) software or MR vendor-specific post-processing programmes, and there is no clear evidence to indicate which enhancement threshold should be used. Standardised thresholds are likely to result in lesions with lower or slower enhancement being erroneously missed or measured as smaller than they actually are. Furthermore, thresholds are likely to vary from scanner to scanner due to different coil architectures and will be specific to a given set of imaging parameters and contrast agent.

A simpler measure of tumour volume is measurement of the total number of pixels that enhance around 2 min post contrast injection, denoted the enhancing tumour volume (ETV). The ETV can be measured without sophisticated computer software, using either commercial or freely available packages with operator-defined thresholding methods. Such techniques, while potentially less reproducible, allow for user experience and interpretation to be taken into account and might go some way to obviate the issue of erroneous exclusion of viable tumour as a result of changed enhancement patterns in response to NAC.[23]

The aim of this study was to investigate whether changes in ETV or FTV between baseline and interim MRI are useful in predicting a pCR in patients undergoing NAC for primary breast cancer. As ETV is a user-interactive technique, the reproducibility of ETV was also measured.

METHODS AND MATERIALS

Patients

This was a single-institution study performed on females with biopsy-proven primary breast cancer scheduled for NAC and

referred for MRI examinations for treatment monitoring. All females gave written consent for images to be used for research purposes and ethical approval was waived for this anonymised, retrospective study.

From 172 consecutively scanned patients, a total of 36 patients were excluded from the study due to excessive motion (n = 18), poor image quality due to fat saturation failure (n = 9), incorrect timing of MRI examinations relative to NAC treatment (n = 4), failure to complete NAC prior to surgery (n = 2), no final pathology available (n = 2) or non-standard management (n = 1). The remaining 136 patients comprised the study cohort, who underwent MRI examinations at baseline (prior to NAC) and at interim (after 2 or 3 cycles of NAC).

In these 136 patients, 40 cancers were oestrogen receptor positive (ER+, Allred score ≥3) and HER2 negative; 9 were HER2+ (immunohistochemistry 3 or 2 + with fluorescence *in situ* hybridisation amplified) and ER-; and 50 were classified as triple negative breast cancer (Allred score <3 and HER2 0, 1 or 2 + with fluorescence *in situ* hybridisation non-amplified). There was also a hybrid group (both ER and HER2 positive) of 37 patients.

28 patients received six cycles of FEC (fluorouracil, epirubicin and cyclophsphamide); 62 female, three cycles of FEC followed by three cycles of docetaxel. All patients with HER2 positive disease underwent three cycles of FEC and either three cycles of docetaxel and trastuzumab (n = 41) or docetaxel with trastuzumab-emtansine (TDM1) (n = 5).

MRI examination

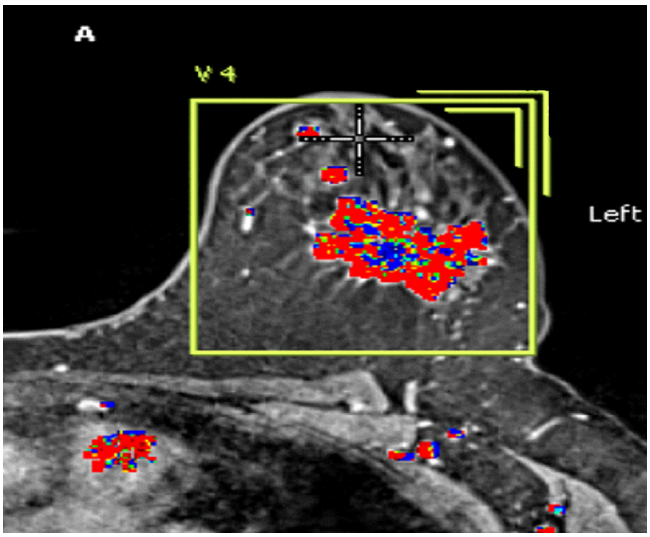
Patients were scanned prior to NAC and after NAC cycle 2 (n = 22) or cycle 3 (n = 114). At interim MRI examination, all patients had received only FEC treatment.

All MRI examinations were performed on a 32-channel 3.0 T Siemens Magnetom Trio or Siemens Prisma^{FIT} (Erlangen, Germany) MRI scanner using a dedicated bilateral breast coil. The imaging protocol consisted of standard anatomical imaging (T_2 weighted turbo spin echo, diffusion-weighted imaging and a high resolution T_1 weighted post-contrast sequence). The DCE-MRI sequence was performed using a three-dimensional spoiled gradient echo volumetric sequence (FLASH) with fat suppression (repetition time/echo time/α = 3.4 ms/1.22 ms/6°, voxel size = 0.8 × 0.8 × 0.8 mm^3, parallel imaging factor × 2). The complete dynamic acquisition consisted of eight volumetric acquisitions (each of 49–58 s), with contrast injected at the end of the second volume acquisition. All patients received a 0.1 mmol kg^{-1} dose of Dotarem (Guerbet , France), injected at 2.0 ml s^{-1}, followed by 20 ml saline flush, using a power injector (Spectris Solaris EP; Medrad, Pittsburgh, PA).

Image analysis

All volumetric measurements were performed blinded to patient outcome, clinical information and final pathology. In patients with multifocal disease, only the largest lesion was analysed for each scan. All patients had only unilateral disease.

Figure 2. Functional tumour volume calculation using Siemens BreVis software. All coloured pixels have reached minimum enhancement percentage of 50% and are assigned a colour based on whether time–intensity curve demonstrates washout (red), plateau (green) or persistent (blue) enhancement. A region of interest placed over the region will count pixels to determine functional tumour volume.

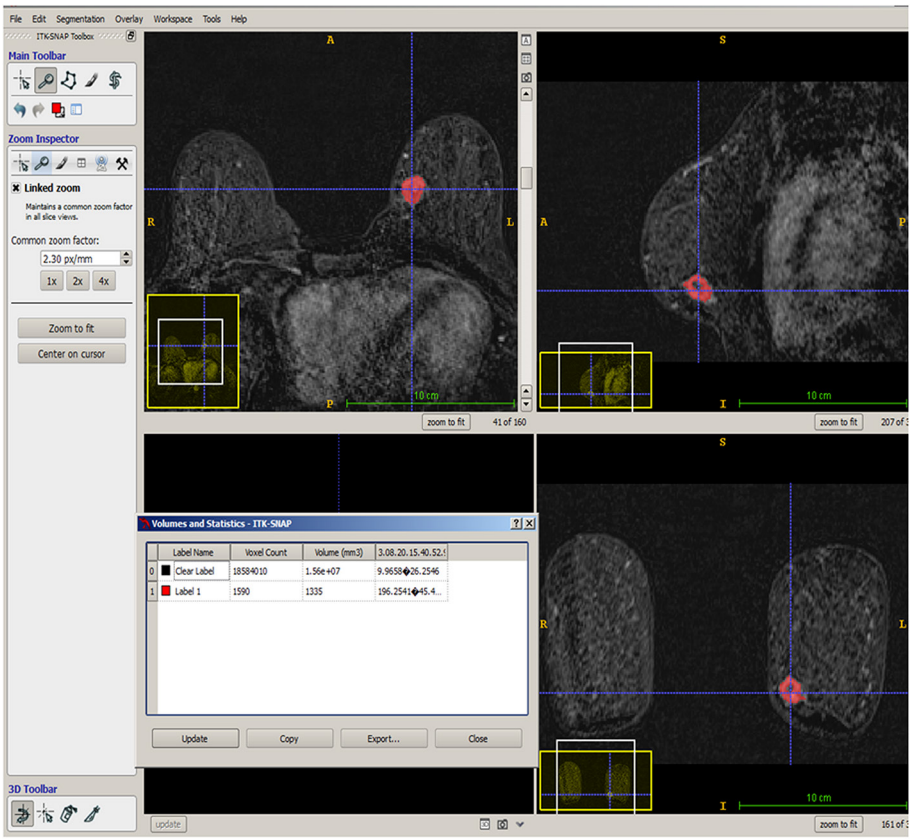


FTV was measured using a fully automated breast CAD package (SyngoVia BreVis; Siemens, Erlangen, Germany) with a standardised threshold value of 50% at 2 min post-contrast injection (as per manufacturers default setting). Pixels were coloured on the basis of whether enhancement was deemed to be washout (decrease in signal intensity of more than 10%), plateau (signal intensity that remained within ± 10%) or persistent (signal intensity increase by more than 10%) relative to the peak enhancement. Volumes of interest were drawn around regions of enhancement for pixel counting of those that met enhancement criteria (Figure 2).

ETV measurements were measured offline using a stand-alone PC. The 2 min post-contrast subtraction series Digital Imaging and Communications in Medicine images were exported onto this workstation and analysis was performed using a semi-automated threshold method in ITK-Snap software.[24] Threshold and smoothing values were user-defined on a case-by-case basis using observer experience and clinical knowledge to match the thresholding mask to the enhancement on the 2 min post-contrast image (Figure 3). No standardised thresholds were used for this technique. Where there was evidence of linear non-mass enhancement extending from sold mass lesions, which was suspicious of ductal carcinoma *in situ*, this was also included within the measured volume. Measuring the ETV at baseline and interim look approximately 7 min per patient.

Due to the interactive nature of the volumetric calculation, intra- and interobserver repeatability was calculated for ETV measures. All data was analysed twice by SAH with a 1-month time interval between analysis sessions, and a subset of 100 patients was also analysed by NMG/SV.

Figure 3. Enhancing tumour volume calculation using ITK-SNAP software. Only enhancing pixels are included in the volumetric measurement. User-thresholding is used to define the minimum intensity to include within the volume of interest.



Assessment of response

Pathological response was assessed by a specialist breast pathologist (CAP) on the resected specimen, based on tumour bed dimensions, cellularity and axillary node burden as outlined by Symmans et al.[5] The RCB was calculated and the variable dichotomised into an index, defining whether patients achieved a pCR, or had minimal (RCB-I), moderate (RCB-II) or marked (RCB-III) residual disease post-treatment. These RCB groups correspond with the risk of distant relapse-free survival.[5]

Statistics

Intra- and interobserver repeatability was assessed using a Bland–Altman plot and the coefficient of repeatability calculated using Equation 1.

$$CoR = 1.96 \times \sqrt{\frac{\sum (m_2 - m_1)^2}{n - 1}} \quad (1)$$

Where $m_{1,2}$ are the first and second measures made and n is the number of patients in the cohort.

As it was the change in volume that we were interested in, percentage volume reductions between baseline and interim MRI examinations were calculated for ETV and FTV measures for every patient, and these were compared with the ultimate pathological response, as measured using the RCB score.

A one-way ANOVA was performed to determine if there were any significant differences across all response groups for the baseline volumes, as measured using either ETV or FTV, to ensure no bias to the final results. Baseline volumes for each immunophenotype were also tested for any relationship with ultimate pathological response, using a Mann–Whitney *U* test. Correlation between the two volumetric assessment methods was assessed using a Pearson's correlation coefficient, to ascertain consistency of measures between each method.

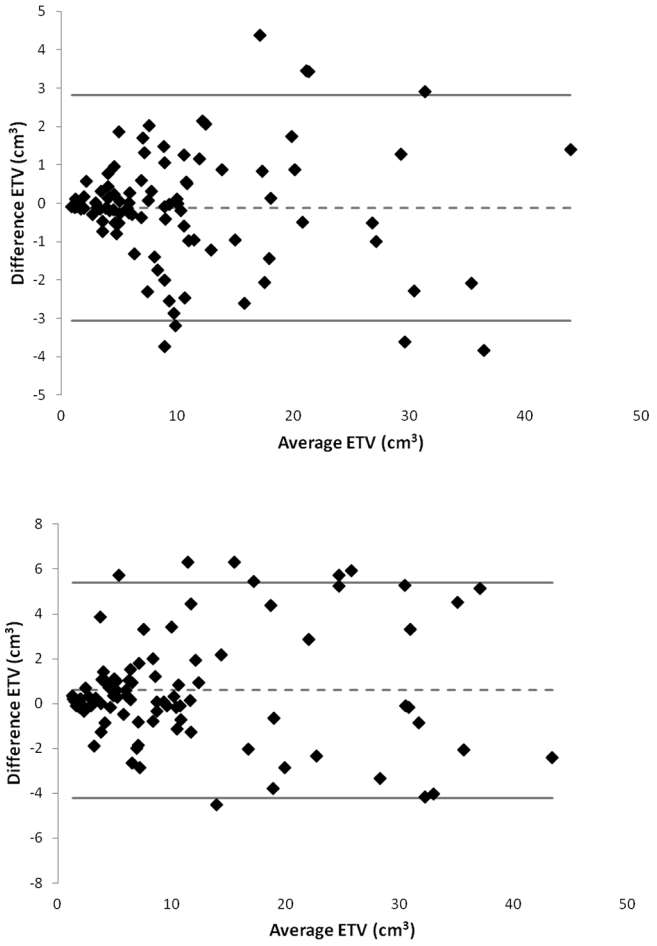
Unpaired *t*-tests were used for comparisons between the individual response categories, and receiver operating characteristic curves (ROC) were generated using R Studio (v. 1.0143, www.R-project.org). The area under the ROC curves was calculated (AUROC) and optimal thresholds to identify pCR post-NAC treatment were derived using Youden's index. Sensitivity, specificity, accuracy and positive and negative predictive values (PPV and NPV) were calculated based on these optimal thresholds.

RESULTS

Patient cohort

Within the cohort of 136 patients, there were a total of 151 lesions identified, however, the volumetric analysis was restricted to the largest, index lesion in each patient. Of these lesions, 109 were masses and there were 27 non-mass lesions. At final pathological examination, 24 patients were categorised as pCR, 20 as RCB-I,

Figure 4. Bland–Altman plots of intraobserver (top) and interobserver (bottom) repeatability. ETV, enhancing tumour volume; FTV, functional tumour volume.



64 as RCB-II and 28 as RCB-III. When broken down into immunohistochemical subtype, there were 40 ER + cancers (pCR: 7, RCB-I: 6, RCB-II: 22, RCB-III: 5), 9 HER2 + cancers (pCR: 2, RCB-I: 1, RCB-II: 2, RCB-III: 4), 37 hybrid cancers (pCR: 8, RCB-I: 5, RCB-II: 13, RCB-III: 11) and 50 triple negative cancers (pCR: 7, RCB-I: 8, RCB-II: 27, RCB-III: 8).

Reproducibility of ETV measurements

Intraobserver repeatability was calculated to be 2.7 cm^3, with the average lesion volume 10.5 cm^3. The interobserver repeatability coefficient was higher at 4.7 cm^3 with average lesion volume size of 11.7 cm^3. Bland–Altman plots are shown in Figure 4.

RCB response

There were no significant differences across all response categories for the baseline (presenting) tumour volume as measured using FTV (p = 0.642, one-way ANOVA) or ETV (p = 0.149, one-way ANOVA). Neither were there any significant differences across response categories when considered in terms of immunophenotypes (p > 0.429, one-way ANOVA).

In terms of immunophenotype, in the ER + lesions there was a significant difference between the baseline (presenting) ETVs for patients who ultimately achieved a pCR compared with those with residual disease (RCB-I, II or III) (p = 0.028; Mann–Whitney U test). There was no significant difference in the FTV (p = 0.191; Mann–Whitney U test), or in any measure of tumour volume for the HER2 + or hybrid lesions (p > 0.85).

As expected, there was a significant correlation between the FTV and the ETV measurements with a Pearson's correlation coefficient of 0.689 (p < 0.001).

Average and standard deviation percentage volume reductions are shown for both FTV and ETV techniques in Table 1 and Figure 5, for each response category. Patients who achieved a pCR had the greatest percentage reduction in tumour volume using both techniques, while those who had extensive residual disease at final pathology had the lowest percentage reduction in tumour volume.

Pairwise statistical comparisons demonstrate that for FTV measurement, there are significant differences between pCR and RCB-II and RCB-III categories (p < 0.040, unpaired t-test), but not between the pCR and RCB-I categories (p = 0.156, unpaired t-test). Significant differences were demonstrated for all RCB-I, RCB-II and RCB-III comparisons (p < 0.017, unpaired t-test).

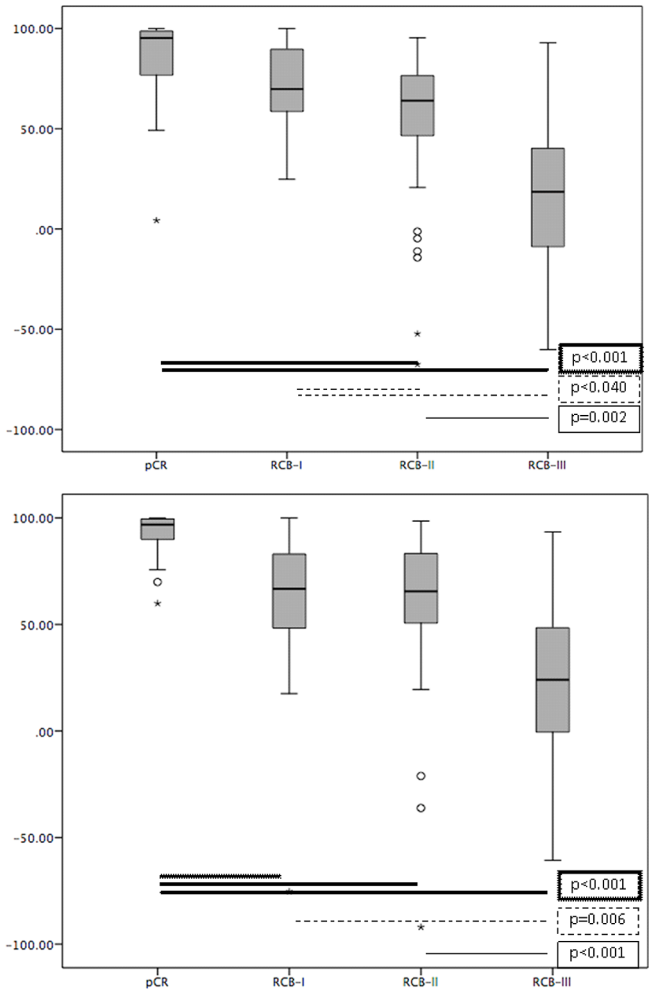
For ETV, there were significant differences between pCR and all other categories (p < 0.006, unpaired t-test). There were significant differences in all other comparisons of categories, with the exception of RCB-I and RCB-II responders (p = 0.829, unpaired t-test). Notably for ETV, pCR demonstrates a significantly distinct volumetric change between baseline and interim MRI relative to all patients who had some form of residual disease.

Table 1. Average reductions in volume for both volumetric measurement methods in each pathological response category

	FTV		ETV	
	Average volume reduction	Standard deviation	Average volume reduction	Standard deviation
pCR	95.2 %	23.8 %	96.8 %	11.1 %
RCB-I	69.8 %	23.6 %	66.7 %	41.9 %
RCB-II	64.0 %	35.2 %	65.5 %	34.3 %
RCB-III	25.4 %	39.4 %	24.0 %	35.7 %

ETV, enhancing tumour volume; FTV, functional tumour volume; pCR, pathological complete response.

Figure 5. Percentage volume reduction for each response category for FTV (upper graph) and ETV (lower graph). Significant differences between response categories are indicated ($p < 0.05$; unpaired *t*-test). ETV, enhancing tumour volume; FTV, functional tumour volume; pCR, pathological complete-response.

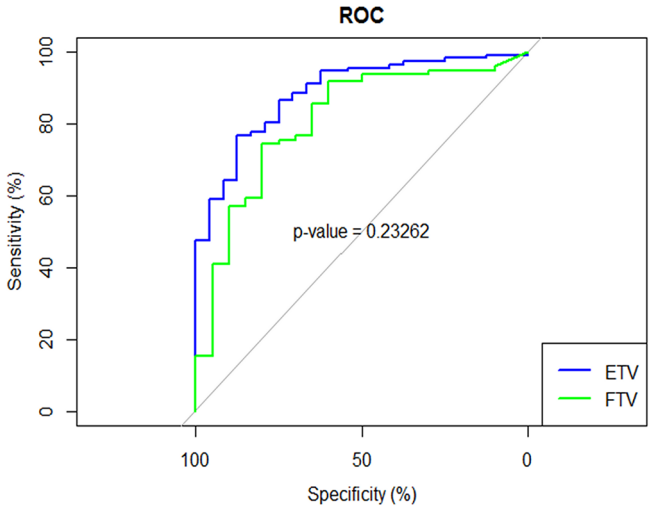


When considered in terms of immunohistochemical subtype, significance between pCR and RCB-I categories were lost for ETV measures ($p > 0.051$, unpaired *t*-test), and for both ETV and FTV, there were no significant differences in any metric between RCB-I and RCB-II ($p > 0.081$, unpaired *t*-test).

ROC analysis was used to compare the prediction of pCR after treatment using both FTV and ETV volumetric changes between baseline and interim MRI. The resulting AUC was good for FTV (AUROC = 0.834) and excellent for ETV (AUROC = 0.920), however, there were no significant differences between the curves ($p = 0.233$), as shown in Figure 6.

Using Youden's index, optimal thresholds for pCR prediction were derived. For FTV, a 75.6% reduction resulted in a sensitivity and specificity of 80.0 and 76.8% respectively, while a reduction of 89.8% in ETV measure gave a sensitivity and specificity of 81.0 and 91.8%. Full diagnostic performance is presented in Table 2, which demonstrates the superior rates for ETV measures.

Figure 6. ROC curves demonstrating sensitivity and specificity in prediction of pCR after treatment based on volumetric changes measured between baseline and interim MRI using FTV and ETV methods. ETV, enhancing tumour volume; FTV, functional tumour volume; ROC. receiver operating characteristic.



DISCUSSION

NAC is increasingly used in breast cancer management, and females with a pCR have improved overall survival.[25–27] Image analysis has the potential to assess response during therapy, prior to surgical resection, however, there is still lack of any consensus as to the best metric that correlates with ultimate pathological outcome.

We have demonstrated that changes in ETV between baseline and interim MRI may provide a more accurate predictive assessment of pCR to NAC compared with FTV. ROC curves demonstrate no statistically significant differences between the two methods, however, there are improved diagnostic performance criteria for the ETV method compared with a comparable assessment of FTV changes.

Table 2. Diagnostic performance criteria in the prediction of pCR post-NAC treatment when using volumetric changes between baseline and interim MRI as measured using FTV or ETV

	FTV	ETV
	Optimal threshold: 75.6%	Optimal threshold: 89.8%
Sensitivity	80.0%	81.0%
Specificity	76.8%	91.8%
Accuracy	77.4%	89.8%
PPV	42.1%	68.0%
NPV	77.4%	95.7%

ETV, enhancing tumour volume; FTV, functionaltumour volume; NPV, positive-predictive value; PPV, negative-predictive value.

Due to the user-thresholding of the ETV method, reproducibility assessment was performed, which resulted in average reproducibility for intraobserver repeatability, but only moderate reproducibility for interobserver measures. It should be noted, however, that while one observer had a number of years of experience in using the software, the other observer had very limited experience; therefore it is likely that the gap in intra- and interobserver variability could be reduced substantially. Although some training in the use of ITK-SNAP is required, it is easy to use, freely available and once familiar with the functions, segmentation can be carried out quickly, typically taking a few minutes per data set. On the other hand, a fully automated method would potentially be more useful in the clinical arena once fully validated. Non-mass enhancement was significantly more subjective to measure, and generally required a greater degree of interpretation and experience in defining the extent of enhancement. While only inclusion of solid mass lesions would likely have increased the reproducibility of the ETV technique, it would be unlikely to reflect true tumour burden.

No standardised thresholds were utilised in ETV measurements, as this would then limit the technique to the drawbacks associated with FTV measurement—*i.e.* slowly or lesser enhancing tumours would be underestimated. The semi-automated technique of utilising ETV means that a degree of observer judgement based on experience can be combined by increasing or decreasing thresholds to accurately include the regions of suspicious enhancement, *e.g.* in inclusion of linear enhancement reflecting ductal carcinoma *in situ* that generally enhances to a lesser degree than mass lesions. It should also be noted that it has previously been reported that different tumour immunophenotypes require different enhancement thresholds due to differences in enhancement patterns,[28] reflecting our findings here that a user-interactive technique is likely to be more accurate in describing tumour volumes.

Reports suggest that FTV can be used to predict not only pCR to NAC, but also recurrence-free survival.[20,21] However, it has also been reported that the thresholds used for FTV calculation have a significant impact on the measured volumes[29–31] and thresholds also are dependent on the breast cancer immunohistochemical subtype.[28] It has been shown that taxanes have an anti-angiogenic effect and hence, may result in diminished enhancement when there is in fact residual disease,[23] meaning that automated methods have the potential to underestimate the amount of disease present. On the other hand, use of an interactive thresholding technique enables an experienced reader to compensate for this, which may explain in part the superior differentiation between ultimate pCR and RCB-I groups. There are also differences in the way that different software platforms implement the FTV measure[31] and such measures are dependent on good quality dynamic data, so poor fat saturation and motion can result in incorrect or even failure of calculation,[31] which is not an issue for ETV measures as they are used on a standalone image. By contrast, with user-defined interactive platforms, there is the possibility of compensating for misregistration and poor fat suppression.

There have been numerous studies that have reported the utility of tumour volume, and the importance of this parameter importance in prediction of NAC outcome.[11,32–34] Tumour volume has been reported to be more indicative of ultimate outcome to treatment than other simpler measures such as tumour diameter,[11,21] diffusion-weighted parameters[33] and more complex pharmacokinetic modelling parameters.[34] In fact, the I-SPY trial concluded that FTV performed better than maximum diameter measurements at predicting ultimate outcome at all time points—early, interim and post-treatment imaging examinations.[21] As FTV measures are readily available on all commercial breast CAD packages, which are recommended for use in reporting breast MRI examinations, it is therefore, a more appropriate metric to use in clinical reporting than tumour diameter measurements. However, to date, there have been no direct comparisons of the two volumetric measures considered in this study—FTV and ETV.

As FTV measures are dependent entirely on enhancement thresholds, in turn this will result in a likely dependence on factors affecting signal to noise ratio such as field strength, coil architecture as well as imaging parameters, which to our knowledge has not been investigated within the literature and requires clarification prior to implementation as a clinical prediction tool. While ETV requires transfer of data offline, it is less likely to be prone to such factors and potentially will provide a more robust measurement tool.

The ability to confidently predict a pCR to NAC on imaging assessment alone could potentially facilitate novel approaches to surgical management, *e.g.* only performing percutaneous sampling of the tumour bed and radiotherapy treatment and dispensing with surgical intervention.[7,8] Such feasibility studies are indeed already underway in the USA and the UK and are a step towards personalised treatment.

When considering subtypes of breast cancer, similar trends were observed, however, these were not significant between pCR and RCB-I responders, most likely due to the smaller numbers in each comparative groups. The ETV pCR *vs* RCB-I comparison for hybrid cancers was approaching significance (p = 0.051, unpaired t-test) with 37 patients, and therefore, warrants further investigation in larger cohorts. Unfortunately, it was not possible to perform response assessment in the ER-HER2 + group as there were only nine patients in this cohort. While the reported rates within the literature for pCR may appear higher, it should be noted that there is a discrepancy in the use of the term "pCR" within the literature, with some groups reporting pCR as no evidence of invasive cancer or *in situ* disease in the breast or nodes, no evidence of invasive cancer in the breast or nodes (irrespective of *in situ* disease) or no evidence of invasive cancer in the breast (irrespective of *in situ* disease or nodal involvement).

We considered only baseline and interim MRI examinations in this study, in line with previous findings from the I-SPY trial that changes in FTV between baseline and interim MRI were better

predictors of response than differences between baseline and post-treatment.[21] There is, however, the possibility that future work could also include final MRI findings to establish whether this would further increase sensitivity in outcome prediction. We were only able to evaluate one manufacturer's software platform in this study, and this was proprietary software. However, it is likely that due to the standardised method associated with FTV measurement, that results would be similar across multiple vendor software platforms; this needs to be confirmed. The method we used for measurement of ETV was only semi-automated method, and further investigation into fully automated methods is required, particularly in the light of the substantial interobserver variability we found. A robust fully automated method established through deep learning could obviate these issues and allow widespread multicentre validation studies, which would be required before adoption of this technique in the clinical arena.

We conclude that there are no significant differences between the overall diagnostic performances in using ETV or FTV to predict outcome to NAC treatment after an interim MRI examination. However, at the interim examination, the use of ETV measures indicates which patients are likely to have a pCR as opposed to minimal residual disease at completion, whereas FTV measures were not able to make this differentiation. Thus, our data indicate that using ETV measures, it may be possible to identify patients who may not require post-therapy surgery. Before more extensive clinical trials can be established to verify this, further work is required with greater patient numbers in order to establish whether such predictions can be applied more reliably within individual immunohistochemical breast cancer subtypes.

ACKNOWLEDGMENTS

The authors would like to thank Breast Cancer Now for partial funding of this research, as well as the patients who took part in the study.

FUNDING

Partial funding provided by Breast Cancer Now.

REFERENCES

1. Makhoul I, Kiwan E. Neoadjuvant systemic treatment of breast cancer. *J Surg Oncol* 2011; **103**: 348–57. doi: https://doi.org/10.1002/jso.21696
2. Liu SV, Melstrom L, Yao K, Russell CA, Sener SF. Neoadjuvant therapy for breast cancer. *J Surg Oncol* 2010; **101**: 283–91. doi: https://doi.org/10.1002/jso.21446
3. Bonadonna G, Valagussa P, Brambilla C, Ferrari L, Moliterni A, Terenziani M, et al. Primary chemotherapy in operable breast cancer: eight-year experience at the Milan Cancer Institute. *J Clin Oncol* 1998; **16**: 93–100. doi: https://doi.org/10.1200/JCO.1998.16.1.93
4. Kuerer HM, Newman LA, Smith TL, Ames FC, Hunt KK, Dhingra K, et al. Clinical course of breast cancer patients with complete pathologic primary tumor and axillary lymph node response to doxorubicin-based neoadjuvant chemotherapy. *J Clin Oncol* 1999; **17**: 460–9. doi: https://doi.org/10.1200/JCO.1999.17.2.460
5. Symmans WF, Peintinger F, Hatzis C, Rajan R, Kuerer H, Valero V, et al. Measurement of residual breast cancer burden to predict survival after neoadjuvant chemotherapy. *J Clin Oncol* 2007; **25**: 4414–22. doi: https://doi.org/10.1200/JCO.2007.10.6823
6. Symmans WF, Wei C, Gould R, Yu X, Zhang Y, Liu M, et al. Long-term prognostic risk after neoadjuvant chemotherapy associated with residual cancer burden and breast cancer subtype. *J Clin Oncol* 2017; **35**: 1049–60. doi: https://doi.org/10.1200/JCO.2015.63.1010
7. Ring A, Webb A, Ashley S, Allum WH, Ebbs S, Gui G, et al. Is surgery necessary after complete clinical remission following neoadjuvant chemotherapy for early breast cancer? *J Clin Oncol* 2003; **21**: 4540–5. doi: https://doi.org/10.1200/JCO.2003.05.208
8. van la Parra RF, Kuerer HM. Selective elimination of breast cancer surgery in exceptional responders: historical perspective and current trials. *Breast Cancer Res* 2016; **18**: 28–35. doi: https://doi.org/10.1186/s13058-016-0684-6
9. Croshaw R, Shapiro-Wright H, Svensson E, Erb K, Julian T. Accuracy of clinical examination, digital mammogram, ultrasound, and MRI in determining postneoadjuvant pathologic tumor response in operable breast cancer patients. *Ann Surg Oncol* 2011; **18**: 3160–3. doi: https://doi.org/10.1245/s10434-011-1919-5
10. Shin HJ, Kim HH, Ahn JH, Kim SB, Jung KH, Gong G, et al. Comparison of mammography, sonography, MRI and clinical examination in patients with locally advanced or inflammatory breast cancer who underwent neoadjuvant chemotherapy. *Br J Radiol* 2011; **84**: 612–20. doi: https://doi.org/10.1259/bjr/74430952
11. Partridge SC, Gibbs JE, Lu Y, Esserman LJ, Tripathy D, Wolverton DS, et al. MRI measurements of breast tumor volume predict response to neoadjuvant chemotherapy and recurrence-free survival. *AJR Am J Roentgenol* 2005; **184**: 1774–81. doi: https://doi.org/10.2214/ajr.184.6.01841774
12. Fangberget A, Nilsen LB, Hole KH, Holmen MM, Engebraaten O, Naume B, et al. Neoadjuvant chemotherapy in breast cancer-response evaluation and prediction of response to treatment using dynamic contrast-enhanced and diffusion-weighted MR imaging. *Eur Radiol* 2011; **21**: 1188–99. doi: https://doi.org/10.1007/s00330-010-2020-3
13. Wu LM, Hu JN, Gu HY, Hua J, Chen J, Xu JR. Can diffusion-weighted MR imaging and contrast-enhanced MR imaging precisely evaluate and predict pathological response to neoadjuvant chemotherapy in patients with breast cancer? *Breast Cancer Res Treat* 2012; **135**: 17–28. doi: https://doi.org/10.1007/s10549-012-2033-5
14. Woolf DK, Padhani AR, Taylor NJ, Gogbashian A, Li SP, Beresford MJ, et al. Assessing response in breast cancer with dynamic contrast-enhanced magnetic resonance imaging: are signal intensity-time curves adequate? *Breast Cancer Res Treat* 2014; **147**: 335–43. doi: https://doi.org/10.1007/s10549-014-3072-x
15. Martincich L, Montemurro F, De Rosa G, Marra V, Ponzone R, Cirillo S, et al. Monitoring response to primary chemotherapy in breast cancer using dynamic contrast-enhanced magnetic resonance imaging. *Breast Cancer Res Treat*

2004; **83**: 67–76. doi: https://doi.org/10.1023/B:BREA.0000010700.11092.f4
16. Pickles MD, Lowry M, Manton DJ, Gibbs P, Turnbull LW. Role of dynamic contrast enhanced MRI in monitoring early response of locally advanced breast cancer to neoadjuvant chemotherapy. *Breast Cancer Res Treat* 2005; **91**: 1–10. doi: https://doi.org/10.1007/s10549-004-5819-2
17. Ahmed A, Gibbs P, Pickles M, Turnbull L. Texture analysis in assessment and prediction of chemotherapy response in breast cancer. *J Magn Reson Imaging* 2013; **38**: 89–101. doi: https://doi.org/10.1002/jmri.23971
18. Henderson S, Purdie C, Michie C, Evans A, Lerski R, Johnston M, 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. *Eur Radiol* 2017; **27**: 4602–11. doi: https://doi.org/10.1007/s00330-017-4850-8
19. Parikh J, Selmi M, Charles-Edwards G, Glendenning J, Ganeshan B, Verma H, et al. Changes in primary breast cancer heterogeneity may augment midtreatment MR imaging assessment of response to neoadjuvant chemotherapy. *Radiology* 2014; **272**: 100–12. doi: https://doi.org/10.1148/radiol.14130569
20. Hylton NM, Blume JD, Bernreuter WK, Pisano ED, Rosen MA, Morris EA, et al. Locally advanced breast cancer: MR imaging for prediction of response to neoadjuvant chemotherapy--results from ACRIN 6657/I-SPY TRIAL. *Radiology* 2012; **263**: 663–72. doi: https://doi.org/10.1148/radiol.12110748
21. Hylton NM, Gatsonis CA, Rosen MA, Lehman CD, Newitt DC, Partridge SC, et al. Neoadjuvant therapy for breast cancer: functional tumor volume by MR imaging predicts recurrence free survival-results from the ACRIN 6657/CALGB 150007 I-SPY 1 TRIAL. *Radiology* 2016; **279**: 44–55. doi: https://doi.org/10.1148/radiol.2015150013
22. Yi A, Cho N, Im SA, Chang JM, Kim SJ, Moon HG, et al. Survival outcomes of breast cancer patients who receive neoadjuvant chemotherapy: association with dynamic contrast-enhanced MR imaging with computer-aided evaluation. *Radiology* 2013; **268**: 662–72. doi: https://doi.org/10.1148/radiol.13121801
23. Schrading S, Kuhl CK. Breast cancer: influence of taxanes on response assessment with dynamic contrast-enhanced MR imaging. *Radiology* 2015; **277**: 687–96. doi: https://doi.org/10.1148/radiol.2015150006
24. Yushkevich PA, Yang Gao, Gerig G. ITK-SNAP: an interactive tool for semi-automatic segmentation of multi-modality biomedical images. *Conf Proc IEEE Eng Med Biol Soc* 2016; **2016**: 3342–5. doi: https://doi.org/10.1109/EMBC.2016.7591443
25. Cortazar P, Zhang L, Untch M, Mehta K, Costantino JP, Wolmark N, et al. Pathological complete response and long-term clinical benefit in breast cancer: the CTNeoBC pooled analysis. *Lancet* 2014; **384**: 164–72. doi: https://doi.org/10.1016/S0140-6736(13)62422-8
26. Mougalian SS, Hernandez M, Lei X, Lynch S, Kuerer HM, Symmans WF, et al. Ten-year outcomes of patients with breast cancer with cytologically confirmed axillary lymph node metastases and pathologic complete response after primary systemic chemotherapy. *JAMA Oncol* 2016; **2**: 508–16. doi: https://doi.org/10.1001/jamaoncol.2015.4935
27. von Minckwitz G, Untch M, Blohmer JU, Costa SD, Eidtmann H, Fasching PA, et al. Definition and impact of pathologic complete response on prognosis after neoadjuvant chemotherapy in various intrinsic breast cancer subtypes. *J Clin Oncol* 2012; **30**: 1796–804. doi: https://doi.org/10.1200/JCO.2011.38.8595
28. Jafri NF, Newitt DC, Kornak J, Esserman LJ, Joe BN, Hylton NM. Optimized breast MRI functional tumor volume as a biomarker of recurrence-free survival following neoadjuvant chemotherapy. *J Magn Reson Imaging* 2014; **40**: 476–82. doi: https://doi.org/10.1002/jmri.24351
29. Keller BM, McCarthy AM, Chen J, Armstrong K, Conant EF, Domchek SM, et al. Associations between breast density and a panel of single nucleotide polymorphisms linked to breast cancer risk: a cohort study with digital mammography. *BMC Cancer* 2015; **15**: 143–54. doi: https://doi.org/10.1186/s12885-015-1159-3
30. Lo WC, Li W, Jones EF, Newitt DC, Kornak J, Wilmes LJ, et al. Effect of imaging parameter thresholds on MRI prediction of neoadjuvant chemotherapy response in breast cancer subtypes. *PLoS One* 2016; **11**: e0142047. doi: https://doi.org/10.1371/journal.pone.0142047
31. Newitt DC, Aliu SO, Witcomb N, Sela G, Kornak J, Esserman L, et al. Real-time measurement of functional tumour volume by MRI to assess treatment response in breast cancer neoadjuvant clinical trials: validation of the Aegis SER software platform. *Transl Oncol* 2014; **7**: 94–100. doi: https://doi.org/10.1593/tlo.13877
32. Heldahl MG, Lundgren S, Jensen LR, Gribbestad IS, Bathen TF. Monitoring neoadjuvant chemotherapy in breast cancer patients: improved MR assessment at 3 T? *J Magn Reson Imaging* 2011; **34**: 547–56. doi: https://doi.org/10.1002/jmri.22642
33. Jensen LR, Garzon B, Heldahl MG, Bathen TF, Lundgren S, Gribbestad IS. Diffusion-weighted and dynamic contrast-enhanced MRI in evaluation of early treatment effects during neoadjuvant chemotherapy in breast cancer patients. *J Magn Reson Imaging* 2011; **34**: 1099–109. doi: https://doi.org/10.1002/jmri.22726
34. O'Flynn EA, Collins D, D'Arcy J, Schmidt M, de Souza NM. Multi-parametric MRI in the early prediction of response to neo-adjuvant chemotherapy in breast cancer: value of non-modelled parameters. *Eur J Radiol* 2016; **85**: 837–42. doi: https://doi.org/10.1016/j.ejrad.2016.02.006