



PET/CT radiomics in breast cancer: promising tool for prediction of pathological response to neoadjuvant chemotherapy

Lidija Antunovic¹ · Rita De Sanctis² · Luca Cozzi³ · Margarita Kirienko³ · Andrea Sagona⁴ · Rosalba Torrissi² · Corrado Tinterri⁴ · Armando Santoro² · Arturo Chiti^{1,3} · Renata Zelic⁵ · Martina Sollini³

Received: 29 November 2018 / Accepted: 12 March 2019 / Published online: 26 March 2019
© Springer-Verlag GmbH Germany, part of Springer Nature 2019

Abstract

Purpose To assess the role of radiomics parameters in predicting pathological complete response (pCR) to neoadjuvant chemotherapy (NAC) in patients with locally advanced breast cancer.

Methods Seventy-nine patients who had undergone pretreatment staging ¹⁸F-FDG PET/CT and treatment with NAC between January 2010 and January 2018 were included in the study. Primary lesions on PET images were delineated, and extraction of first-, second-, and higher-order imaging features was performed using LIFEx software. The relationship between these parameters and pCR to NAC was analyzed by multiple logistic regression models.

Results Nineteen patients (24%) had pCR to NAC. Different models were generated on complete information and imputed datasets, using univariable and multivariable logistic regression and least absolute shrinkage and selection operator (lasso) regression. All models could predict pCR to NAC, with area under the curve values ranging from 0.70 to 0.73. All models agreed that tumor molecular subtype is the primary predictor of the primary endpoint.

Conclusions Our models predicted that patients with subtype 2 and subtype 3 (HER2+ and triple negative, respectively) are more likely to have a pCR to NAC than those with subtype 1 (luminal). The association between PET imaging features and pCR suggested that PET imaging features could be considered as potential predictors of pCR in locally advanced breast cancer patients.

Keywords Breast cancer · Neoadjuvant chemotherapy · Radiomics · Advanced features · ¹⁸F-FDG PET/CT · Treatment response prediction

Electronic supplementary material The online version of this article (<https://doi.org/10.1007/s00259-019-04313-8>) contains supplementary material, which is available to authorized users.

✉ Lidija Antunovic
lidija.antunovic@humanitas.it

¹ Department of Nuclear Medicine, Humanitas Clinical and Research Center- IRCCS, via Manzoni 56, 20089 Rozzano, Milan, Italy

² Department of Medical Oncology and Hematology, Humanitas Clinical and Research Center- IRCCS, Rozzano, Milan, Italy

³ Department of Biomedical Sciences, Humanitas University, Pieve Emanuele, Milan, Italy

⁴ Breast Surgery Department, Humanitas Clinical and Research Center- IRCCS, Rozzano, Milan, Italy

⁵ Clinical Epidemiology Unit, Department of Medicine Solna, Karolinska Institutet, Stockholm, Sweden

Background

Breast cancer, the leading cause of cancer death among women (<http://gco.iarc.fr/today/online>), is a heterogeneous disease that comprises different molecular subtypes characterized by diverse histological characteristics, aggressiveness, therapeutic strategies, and prognostic implications [1]. Growing evidence supports the application of personalized therapies to cancer patients. However, a reliable method for patient stratification in terms of prediction of therapeutic success is still a major issue, including in breast cancer.

Neoadjuvant chemotherapy (NAC) has become a standard approach for patients with locally advanced breast neoplasms [2, 3]. The backbone of NAC is the combination of anthracycline and taxane, with addition of trastuzumab in cases of human epidermal growth factor receptor 2-positive (HER2+) breast cancer [3]. NAC aims at expanding the surgical indications and facilitating breast-conserving surgery.

Furthermore, pathological complete response (pCR) to NAC, defined as the absence of remaining invasive cancer in the breast and axillary nodes on pathological examination of the post-treatment surgical excision specimens (ypT0/TisN0) [4], has been shown to be a surrogate endpoint of long-term event-free and overall survival in breast cancer [5, 6]. The strongest association between pCR and long-term outcome has been reported in the most aggressive subtypes (triple negative, luminal high-grade, and HER2+ hormone-receptor-negative) [6]. On the other hand, patients with pathological residual disease (non-pCR) after NAC are at higher risk of relapse. However, NAC may be limited by a number of toxic side-effects (e.g., hematological ones); consequently, only 73–87% of regimens are completed in the preoperative setting, and impairments affecting daily life may persist even 3 years after treatment itself [7, 8]. Therefore, a staging procedure capable of predicting response to NAC could avoid eventual toxicities in selected patients.

The role of ^{18}F -fluorodeoxyglucose (^{18}F -FDG) positron emission tomography/computed tomography (PET/CT) in breast cancer staging is routinely limited to advanced disease [3, 9]. PET/CT allows imaging of both intratumor heterogeneity and treatment-induced changes. Standard semiquantitative imaging parameters obtained from ^{18}F -FDG PET/CT have been shown to correlate with tumor aggressiveness in breast cancer, and are therefore likely also to correlate with patient outcome [10–16]. Radiomics, an approach able to quantify lesion heterogeneity using medical imaging, is an emerging translational research topic in breast cancer [17]. Recent literature data indicate that features obtained from ^{18}F -FDG PET/CT correlate with tumor histological characteristics and molecular subtypes [18–22], but there is still limited evidence relating to their role as predictive parameters [21–25].

The main objective of this study was to evaluate the role of advanced imaging features obtained from baseline ^{18}F -FDG PET/CT in combination with clinical and pathological parameters in the prediction of tumor response to NAC in breast cancer patients.

Materials and methods

Study population

Data for patients with a diagnosis of locally advanced breast cancer ($n = 291$), treated by NAC and surgery at Humanitas Clinical and Research Center between January 2010 and January 2018, were retrospectively collected. Patient inclusion criteria were as follows: (i) histological diagnosis of primary breast cancer, (ii) performance of ^{18}F -FDG PET/CT for staging purposes before any treatment, and (iii) NAC as primary treatment followed by surgery. Ninety-two patients were selected. Patients who had already been treated for breast cancer ($n = 2$) and those who had suspicion of distant metastasis

on the staging ^{18}F -FDG PET/CT were excluded ($n = 11$). Finally, 79 patients fulfilling the above-mentioned inclusion and exclusion criteria were included in the present analysis. The Institutional Ethics Committee reviewed and approved this retrospective study. Specific informed consent was not required according to local ethics committee rules for retrospective studies.

NAC treatment

A standard NAC protocol with anthracycline and taxane, plus trastuzumab in patients with HER2 amplification [3], was performed in 68/79 patients. In the remaining 11 patients, a different therapeutic regimen was performed in accordance with the oncologist's indication (detailed in Table 1 of the electronic supplementary material). Surgery was performed after completion of NAC in 77/79 patients. In two patients, the standard NAC protocol was discontinued, and surgery was performed irrespective of failure to complete NAC (these patients had no evidence of response to systemic treatment and severe infection, respectively).

Data collection

Data on programmed clinical examination, tumor marker measurement (cancer antigen 15.3), and radiological follow-up were collected from the medical records ($n = 76$). Three patients were referred to other hospitals for further oncological management; therefore, their radiological follow-up was not available. Data on survival status were available for all patients.

Histological characteristics of the primary tumor (expression of hormone receptors, HER2 status, Ki67 value, tumor dimensions and grade, histological tumor type, molecular subtype) and clinical characteristics of each patient (age at diagnosis, menopausal state, evidence of disease relapse or death, type of chemotherapy) were retrieved from the institutional database.

Image acquisition

The imaging acquisition protocol has been previously described [21]. Details according to the Image Biomarker Standardisation Initiative (IBSI) reporting guidelines [26] are provided in Table 2 of the electronic supplementary material. Briefly, ^{18}F -FDG PET/CT images were acquired 60 ± 5 min after ^{18}F -FDG injection in accordance with the European Association of Nuclear Medicine (EANM) guidelines, version 1.0 and, from February 2015, version 2.0 [27] using an integrated PET/CT scanner, either a Siemens Biograph 6 LSO (Siemens, Erlangen, Germany) or a General Electric Discovery 690 (General Electric Healthcare, Waukesha, WI, USA). CT images without contrast enhancement or respiratory gating (free-breathing images) were acquired. All PET images were corrected for attenuation using the acquired CT data.

Image analysis

^{18}F -FDG PET/CT images were retrieved from the Picture Archiving and Communication System (PACS). A board-certified nuclear medicine physician with more than 10 years' experience in PET/CT segmented breast tumor lesions on ^{18}F -FDG PET/CT images using a semiautomatic approach. The volume of interest (VOI) of the breast lesion was defined on PET images with a threshold of 40% of the maximum standardized uptake value (SUV_{max}) using commercial software (PET VCAR; GE Healthcare). Only the primary tumor was segmented. Textural features (i.e., first-, second-, and higher-order imaging parameters) were extracted using dedicated software for radiomics (LIFEx) (<https://www.lifexsoft.org/index.php>). LIFEx, for technical reasons, extracted second- and higher-order imaging parameters only for lesions greater than 64 voxels.

Primary endpoint

The primary endpoint of this study was the prediction of pCR to NAC, defined as the absence of remaining tumor on histopathological examination of the post-treatment surgical excision specimens [4].

Statistical analysis

Clinical data, biological characteristics of the tumor, and imaging-derived features were tested as possible predictors of pCR to NAC. Two different approaches (complete case and sensitivity) were used for model building.

The complete-case approach was applied for the main analysis. Accordingly, only subjects with complete variable information were included in the main analysis. Two models were built within the complete-case approach to predict pCR to NAC. In the first model, clinical data, biological tumor characteristics, and first-order imaging features were included. Univariable logistic regression was used to estimate odds ratios (ORs) and 95% confidence intervals (CIs) for the association of each variable and pCR to NAC. Variables satisfying an a-priori set criterion of $p < 0.2$ were included in the multivariable model. SUV_{max} and total lesion glycolysis (TLG), identified in the literature as potential predictors of outcome [28, 29], were also included in the model (hereafter referred to as model 1). We further tested whether any clinically meaningful interactions would improve the model fit.

In the second model, second- and higher-order imaging features were also evaluated as possible predictors of pCR to NAC. Features identified at univariable logistic regression as additional predictors of the outcome, along with the variables identified in the first model, were included in the model (hereafter referred to as model 2).

Model performance was evaluated through discrimination and calibration. The discriminatory power of the two models was assessed using the area under the receiver operating characteristic (ROC) curve. Area under the ROC curve (AUC) is the probability that the model we developed will assign a higher probability of the outcome to a patient with the outcome than to a randomly chosen patient without the outcome [30]. Since the dataset used for model development was also used to assess model discrimination, the actual AUC estimates could be over-optimistic [31]. Thus, an internal validation with 100 iterations of tenfold cross-validation was used for AUC evaluation. Calibration of the two models was evaluated by plotting predicted and observed probabilities of outcome [31].

As a second approach, a sensitivity analysis was performed. Multiple imputation with chained equation (MICE) was used to impute the missing data [32, 33]. Two models using two thresholds (i.e., 0.5 for model 3 and 0.4 for model 4) were used for predictor selection. Details for imputation analysis and predictor selection for models 3 and 4 are provided in the [supplementary material](#). Model performance was then evaluated as described in the main analysis.

All analyses were performed using STATA (version 15, StataCorp, College Station, TX, USA) and R Statistical Software (Foundation for Statistical Computing, Vienna, Austria).

Results

Baseline characteristics of all patients included in the study are presented in Table 1. Missing values were observed only for tumor dimension ($n = 13$) and information on Ki67 expression ($n = 2$). The mean age of our study population was 50.7 years, and 48% of included patients were postmenopausal women. Thirty-nine percent of patients had HER2+ breast cancer, 32% had triple-negative neoplasms, and the remaining 29% had luminal disease. The mean tumor diameter was 2.95 cm.

Nineteen out of seventy-nine (24%) patients had pCR to treatment, while the remaining 60/79 (76%) had residual disease on surgical specimens.

During follow-up, disease relapse occurred in 16 cases (local recurrence $n = 9$, distant metastases $n = 7$), while two patients died.

In the patient with a multifocal tumor, two areas of focal uptake on PET images were separately segmented and considered as two different lesions. Results in respect of the features extracted are provided in supplementary Table 3. Collectively, all textural features (i.e., $n = 43$) were available in 71/79 cases. In 8/79 patients, lesions were smaller than 64 voxels; hence, only first-order parameters were extracted. Figure 1 provides images of two patients as representative examples.

Table 1 Baseline characteristics of the study population

	Population (<i>n</i> = 79)	
	<i>N</i>	%
Age, years		
Mean (SD)	50.67 (11.93)	
Dimension, cm		
Mean (SD)	2.95 (1.39)	
Median (range)	2.60 (1.00–8.00)	
Estrogen receptors, %		
Mean (SD)	38.62 (43.60)	
Progesterone receptors, %		
Mean (SD)	20 (34)	
Ki67, % (mean, SD)		
Mean (SD)	38.96 (22.50)	
Median (range)	35.00 (5.00–90.00)	
Menopausal state		
No	41	51.90
Yes	38	48.10
Primary tumor		
Unifocal	78	99
Multifocal	1	1
Molecular subtype		
Luminal	23	29.11
HER2+	31	39.24
TN	25	31.65
Standard NAC protocol		
Yes	68 ^a	83.54
No	11	16.46
Response to NAC		
Complete	19	24.05
Incomplete	60	75.95
Follow-up, months		
Median (range)	21.23 (5.16–65.33)	
Disease relapse		
Yes	16	20.25
No	61	79.75
Not available	2	
Death		
Yes	2	2.53
No	77	97.47

SD standard deviation, *HER2+* human epidermal growth factor receptor 2 amplified, *TN* triple negative, *NAC* neoadjuvant chemotherapy

^aIncluding two patients in whom NAC was not completed

Model selection and model performance

Model 1

Results from the univariable and multivariable models used for selection of the variables in the simple prediction model

containing clinical data, biological characteristics of the tumor, and the first-order imaging features are presented in Table 2. Age, molecular subtype, and HER2 status satisfied the inclusion threshold for the univariable analysis. Since the molecular subtype is defined based on the expression of selected receptors, including HER2, the HER2 status was removed from the model due to possible collinearity issues. SUV_{max} and TLG did not satisfy the inclusion threshold in our dataset ($p > 0.2$). Nevertheless, we evaluated whether the performances of the model would be improved by adding SUV_{max} and TLG (Table 2). If the discriminatory performance of the final model was evaluated on the same dataset as was used for the model development, the addition of SUV_{max} and TLG slightly improved the discrimination (AUC = 0.73 vs 0.71). However, the mean AUC over 100 iterations of tenfold cross-validation was lower for the model with SUV_{max} and TLG (mean AUC = 0.59, Fig. 2a) than for the model containing only the age and molecular subtype (mean AUC = 0.61, Fig. 2b). Thus, the model with age and molecular subtype was selected as the final model 1. Calibration for model 1 is presented in Fig. 3.

Model 2

One second-order advanced imaging feature (correlation) and two higher-order imaging features ($coarseness_{NGLDM}$ and $GLNU_{GLZLM}$) satisfied the inclusion threshold for the univariable analysis (Table 3). Results from the univariable and multivariable logistic regression using the variables selected in model 2 are presented in Table 3. The actual AUC for model 2 was 0.72 (Table 3). Calibration for model 2 is presented in Fig. 4.

Models 3 and 4

Using the threshold of 0.5 (conservative criterion, model 3), age, molecular subtype, HER2 status, correlation, and $coarseness_{NGLDM}$ were selected for the final model. Age, molecular subtype, estrogen receptors, HER2 status, type of NAC, Ki67, correlation, $coarseness_{NGLDM}$, and $GLNU_{GLZLM}$ were selected when 0.4 was used as the threshold (model 4). HER2 status was excluded from both models due to the possible collinearity. The actual AUCs for the two models were 0.70 and 0.73, respectively (Table 4). Calibration of the two models can be seen in Fig. 5.

Discussion

The primary aim of the study was to investigate the relationship between pCR to NAC in breast cancer patients and PET-derived features. In particular, we aimed to assess the predictive power of first-, second-, and higher-order imaging

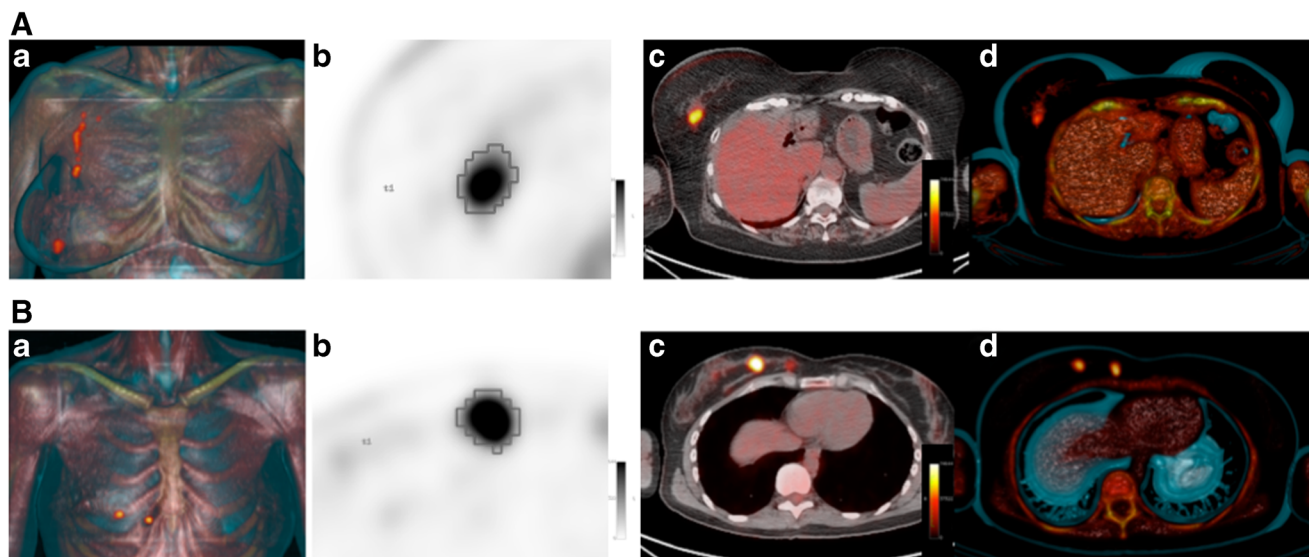


Fig. 1 Representative PET/CT images of a patient with pCR (**a**) and a patient with pathological residual disease (**b**) after NAC. **a** A three-dimensional reconstruction of PET/CT images, axial PET — zoom on the primary breast lesion, fused axial PET/CT, and a transverse view of a three-dimensional reconstruction of PET/CT images (panels *a*, *b*, *c*, and *d* respectively) in an 81-year-old patient with HER2+ breast cancer, estrogen receptor expression 0%, progesterone receptor expression 0%, Ki67 40%, and HER2/neu +++, with no evidence of residual disease at postsurgical histological analysis after neoadjuvant treatment (12 cycles of paclitaxel/trastuzumab). PET parameters (↑ and ↓: above and below the median value): SUV_{maximum} 14.3↑, total lesion glycolysis (TLG) 33.5↓, $Correlation_{\text{GLCM}}$ 0.52↑, $Coarseness_{\text{NGLDM}}$ 0.035↑, $GLNU_{\text{GLZLM}}$ 4.056↓.

b A three-dimensional reconstruction of PET/CT images, axial PET — zoom on the primary breast lesion, fused axial PET/CT, and a transverse view of a three-dimensional reconstruction of PET/CT images (panels *a*, *b*, *c*, and *d* respectively) in a 37-year-old patient with HER2+ breast cancer, estrogen receptor expression 95%, progesterone receptor expression 90%, Ki67 50%, and HER2/neu +++, with persistence of disease at postsurgical histological analysis after neoadjuvant treatment (four cycles of doxorubicin/cyclophosphamide followed by four cycles of docetaxel/trastuzumab). PET parameters: SUV_{maximum} 19.2↑, TLG 17.8↓, $Correlation_{\text{GLCM}}$ 0.295↓, $Coarseness_{\text{NGLDM}}$ 0.041↑, $GLNU_{\text{GLZLM}}$ 2.139↓.

features. The strength of the present study relies on the methodology used for the analysis, determined by the sample size, the retrospective study design, and the number of events (i.e., pCR versus non-pCR). Accordingly, a rigorous statistical plan comparing the complete-case approach (i.e., complete-case dataset) and the sensitivity analysis (i.e., dataset where

missing data were imputed), was applied for building models, management of missing data issues, and selection of variables, in order to obtain reliable results not affected by chance. Two models within the complete-case approach for prediction of pCR to NAC were built; a simple model containing clinical data, biological characteristics of the tumor, and the first-order

Table 2 Odds ratios and 95% CIs from the univariable and multivariable logistic regression models predicting pCR to NAC using the variables selected in model 1

	Univariable analysis		Multivariable analysis			
	OR	95% CI	OR	95% CI	OR	95% CI
Age	0.96	0.92, 1.01	0.97	0.91, 1.02	0.96	0.92, 1.01
Molecular subtype						
Luminal	1.00	—	1.00	—	1.00	—
HER2+	5.78	1.14, 29.37	6.59	1.22, 35.47	5.82	1.11, 30.37
TN	3.32	0.60, 18.45	4.25	0.60, 30.09	3.37	0.59, 19.22
$SUV_{\text{max}}^{\text{a}}$	1.00	0.93, 1.06	0.99	0.89, 1.10	—	—
TLG ^a	1.00	0.99, 1.00	1.00	0.99, 1.00	—	—
Actual AUC			0.73		0.71	
Mean cross-validated AUC			0.59		0.62	

OR odds ratio, CI confidence interval, HER2+ human epidermal growth factor receptor 2 amplified, TN triple negative, SUV_{max} maximum standardized uptake value, TLG total lesion glycolysis, AUC area under the ROC curve

^a Imaging parameters included in the model based on previous knowledge ($p > 0.2$)

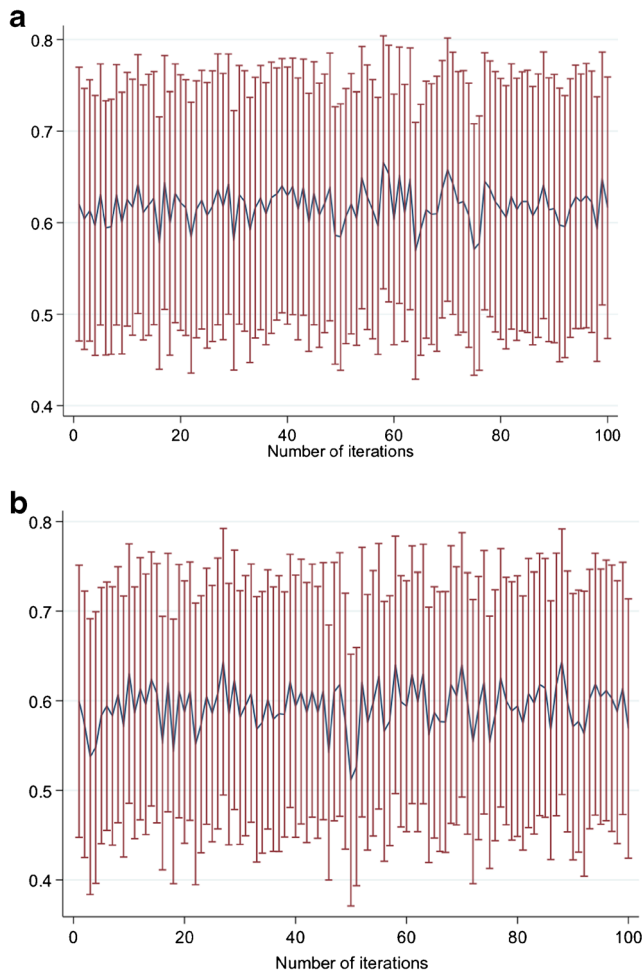


Fig. 2 Variation in the area under the ROC curve (AUC) estimates with the 95% confidence intervals over the 100 iterations of tenfold cross-validation for (a) the model including only age and the molecular subtype and (b) the model including SUV_{max} and TLG in addition to age and molecular subtype

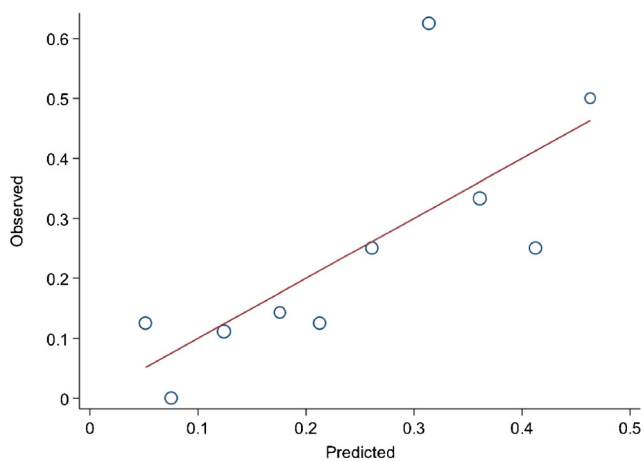


Fig. 3 Calibration of the logistic regression model including only age and molecular subtype (model 1). The size of the circles is proportional to the number of events in the deciles of predicted risk

Table 3 Odds ratios and 95% CIs from the univariable and multivariable logistic regression models predicting pCR to NAC using the variables selected in model 2

	Univariable analysis		Multivariable analysis	
	OR	95% CI	OR	95% CI
Age	0.96	0.92, 1.01	0.96	0.91, 1.02
Molecular subtype				
Luminal	1.00	–	1.00	–
HER2+	5.78	1.14, 29.37	4.24	0.76, 23.60
TN	3.32	0.60, 18.45	1.90	0.30, 12.15
Correlation	0.09	0.01, 1.51	0.33	0.01, 12.16
Coarseness _{NGLDM} ^a	1.02	0.99, 1.05	1.01	0.96, 1.05
GLNU _{GLZLM}	0.93	0.84, 1.04	0.97	0.84, 12.15
Actual AUC			0.72	
Mean cross-validated AUC			0.58	

OR odds ratio, CI confidence interval, HER2+ human epidermal growth factor receptor 2 amplified, TN triple negative, NGLDM neighborhood gray-level different matrix, GLNU_{GLZLM} gray-level non-uniformity gray-level zone length matrix

^a Coarseness_{NGLDM} ranges from 0.00068 to 0.07237, and we thus report the ORs as the change in odds for 0.001 unit increase

imaging features (model 1), and a more complex model in which second- and higher-order features were added (model 2). Age and molecular subtype proved to be the sole, albeit weak, predictors of outcome in the simple model, with slight improvement in the model’s predictive power upon addition of the textural PET-derived features.

¹⁸F-FDG PET imaging parameters have already been proposed as predictors of response to NAC in patients with other tumor types, for example esophageal cancer [34], but information on their predictive role in breast cancer is scarce. Only

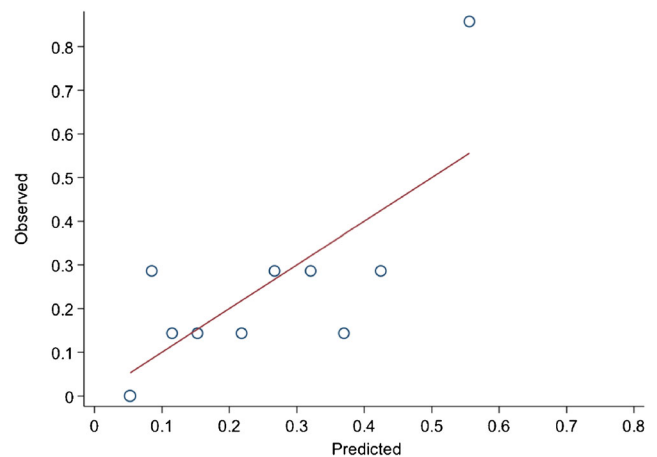


Fig. 4 Calibration of the logistic regression model including only age, molecular subtype, and second- and higher-order imaging features (model 2). The size of the circles is proportional to the number of events in the deciles of predicted risk

Table 4 Odds ratios and 95% CIs from the multivariable logistic regression models predicting pCR to NAC using the variables selected over 50 imputed datasets using thresholds set at 0.5 (model 3) and 0.4 (model 4)

	0.5 threshold		0.4 threshold	
	OR	95% CI	OR	95% CI
Age	0.96	0.91, 1.01	0.96	0.91, 1.01
Molecular subtype ^a	1.52	0.73, 3.18	1.00	0.26, 3.84
Estrogen	–	–	0.99	0.97, 1.02
Type of NAC	–	–	0.56	0.12, 2.55
Ki67	–	–	1.01	0.98, 1.04
Correlation	0.43	0.02, 10.83	0.49	0.02, 15.07
Coarseness _{NGLDM} ^b	1.02	0.98, 1.05	1.01	0.97, 1.05
GLNU _{GLZLM}	–	–	0.97	0.85, 1.11
Actual AUC	0.70		0.73	
Mean cross-validated AUC	0.61		0.56	

OR odds ratio, CI confidence interval, NAC neoadjuvant chemotherapy, NGLDM neighborhood gray-level different matrix, GLNU_{GLZLM} gray-level non-uniformity for gray-level zone length matrix

^a Molecular subtype was treated as a continuous variable

^b Coarseness_{NGLDM} ranges from 0.00068 to 0.07237, and we thus report the ORs as the change in odds for 0.001 unit increase

standard PET parameters (e.g., SUV_{max}, TLG) have been correlated to response to NAC in breast cancer patients [28, 29]. In our study, however, SUV_{max} and TLG did not appear to be predictors of pCR to NAC, and adding them to a simple model containing only age and molecular subtype did not lead to an improvement in the model's predictive power. Very recently, Groheux et al. [35] showed change in ¹⁸F-FDG uptake (Δ SUV_{max}) between PET1 and PET2 (performed before and after two cycles of NAC, respectively) in triple-negative breast cancer to be highly associated with pCR ($p=0.0001$), which was not the case for the absolute value of baseline SUV_{max} ($p=0.11$). The AUC of pCR prediction improved (from 0.63 to 0.76) when baseline SUV_{max} was combined with the genomic grade index ($p=0.016$).

In our study, model 1, which included clinical data only (i.e., age and molecular subtype), and model 2, which also incorporated second- and higher-order imaging features (i.e., correlation, coarseness_{NGLDM}, and GLNU_{GLZLM}), had a similar discrimination performance. Thus, in our hands the inclusion of second- and higher-order imaging features did not improve the discriminatory power of the model. This is somewhat inconsistent with recent evidence of a potential role of high-order PET imaging features as predictors of complete response to NAC in breast cancer [22, 23]. Molina-Garcia et al. [22] found an association between higher-order imaging features, molecular phenotypes, and NAC response, supporting a relationship between local tumor aggressiveness and tumor complexity. Similarly, Yoon et al. [23] found that

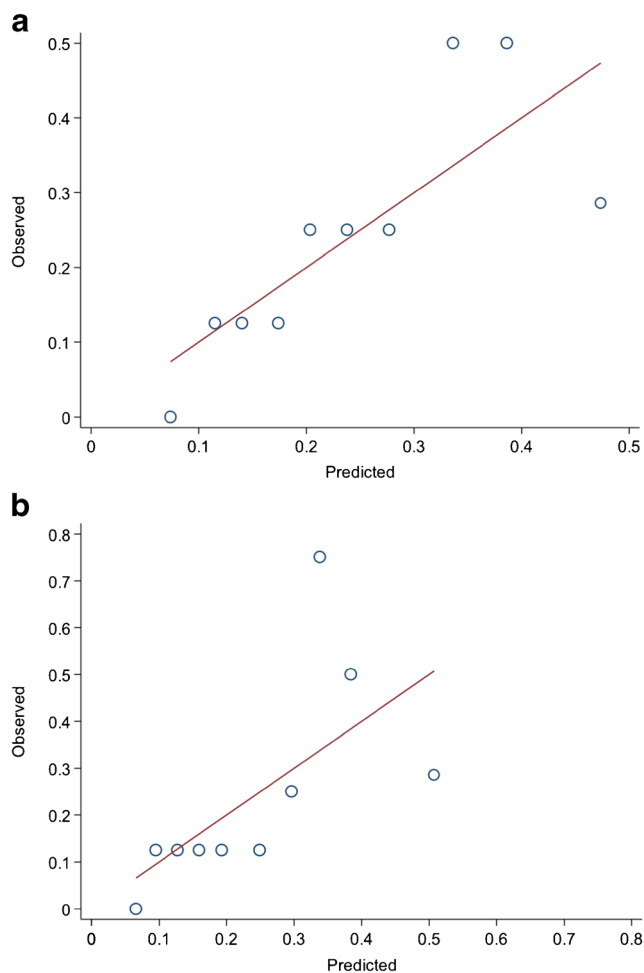


Fig. 5 Calibration of the logistic regression models including the variables selected over 50 imputed datasets using threshold set at **a** 0.5 and **b** 0.4

PET-derived textural features were able to predict responders to NAC and survival in locally advanced breast cancer. A possible explanation for the discrepancy between these findings and our own lies in the study population and in the methods used for the analysis. We evaluated a similar number of cases (79 versus 68 [22] and 83 [23], respectively), but unlike Molina-Garcia et al. [22] we excluded patients with distant metastases (chemotherapy is not a neoadjuvant treatment in the presence of distant metastases [3]), and unlike Yoon et al. [23] we considered only patients presenting pCR as responders since, as mentioned above, this is the only class with a major clinical prognostic relevance. Additionally, in our series, second- and higher-order imaging features were not available for 8/79 patients.

We could not rule out model misspecification issues, with major regard to variable selection. Therefore, in the second approach (based on imputed datasets with complete information, i.e., sensitivity analysis), an alternative specification strategy was used that favored variables with a large effect on the model's discrimination [36]. The lasso approach

allowed an additional issue of our dataset to be addressed (i.e., the combination of a limited number of observations and a large number of potential predictors). Two models (models 3 and 4) were generated using distinct thresholds for predictor selection based on the frequency of variable selection in the imputed datasets. In addition to patient characteristics (i.e., age in both models), tumor markers (i.e., molecular subtype in both models, and estrogen receptor and Ki67 in the 0.4 threshold model only), and type of NAC (for the model 4 only), models 3 and 4 had second- and higher-order PET imaging features as predictors (correlation and coarseness_{NGLDM} for both, GLNU_{GLZLM} for the second only). These models had actual AUC values of 0.70 and 0.73 respectively, which were close to those of models 1 (0.71) and 2 (0.72). This is not surprising, given that all models shared the variables with the largest predictive weight (i.e., age and tumor molecular subtype), regardless of the strategy that was used for variable selection. Despite poor discrimination, this might indicate the appropriateness of model specification.

Overall, all models displayed limited discrimination performances, probably due to the small sample size, which was further reduced due to the missing data and the complete-case approach used in the main analysis. However, we performed a sensitivity analysis using multiple imputation with chained equation, which led to very similar results. The combination of a large number of predictors with the small sample size could have led to the sparse cell problem if more typical, although highly criticized, methods for model selection had been used, such as backward or stepwise selection. Indeed, in our dataset, we were not able to use these methods for model selection due to the convergence issues. However, univariable models for predictor identification performed at least as well as the more complex models used for dimension reduction, such as lasso regression. Furthermore, due to the small sample size we could not divide our population into training and testing sets, which is recommended for model development and validation. Moreover, alternative validation methods (e.g., leave-pair-out cross-validation) might improve the AUC estimates [37].

In our series, the only characteristic that had some predictive weight was the tumor molecular subtype. This is probably due to the small sample size and the limited number of events (i.e., pCR), which represent limitations both for model building and for cross-validation. It should be noted that the models predict that patients with subtype 2 (HER2+) are more likely to have a pCR to NAC than those with subtype 1 (luminal) (OR = 6.59, CI = 1.22–35.47 in the first model), which is consistent with previous knowledge [38].

To summarize, our study has some limitations. First of all, all clinical data were not available for all patients due to the retrospective nature of the study design. Also, as mentioned before, a greater sample size would improve the results and allow a potentially clinically relevant sub-analysis according

to the molecular subtypes. Moreover, second- and higher-order features were not extracted for all patients for technical reasons. However, patient selection and the methods used for data analysis were extremely rigorous. Only patients with locally advanced breast cancer eligible for NAC were included in the analysis. Accordingly, patients with a previous history of breast cancer or metastases (either suspected or confirmed) were excluded. pCR was used as the primary endpoint since it is the only parameter with clinical prognostic significance. Imputation analysis was performed for all missing data.

In conclusion, our findings suggest that radiomics PET-derived features may be associated with pCR to NAC. Therefore, advanced image analysis should be further evaluated in a large population to better clarify and quantitate the potential predictive role of textural features in locally advanced breast cancer patients.

Clinical relevance

Breast cancer patients with pCR to NAC have a better prognosis than partial and non-responders. Predictors of pCR would be of key relevance in daily clinical practice in order to better guide treatment decisions. In this regard, the extraction of biologically relevant information from staging images is increasingly acknowledged as a non-invasive prognostic approach, but its effectiveness is not established yet. Although image-derived biomarkers require a clinical validation before entering routinely into clinical practice, our data confirmed that they are a promising prognostic tool in locally advanced breast cancer.

Acknowledgements We would like to acknowledge Dr. Katia Marzo and Mrs. Elena Bissolotti for database preparation. The contribution of Drs. Elisa Agostinetti, Jelena Jandric, and Giulia Vatteroni to data collection is greatly acknowledged. We thank also Dr. Bethania Fernandes for histological analyses. We also acknowledge Drs. Alberto Testori and Valentina Errico for patients enrollment. We thank Mr. Lorenzo Leonardi for technical support and image acquisition.

Authors' contributions The manuscript has been seen and approved by all authors, whose individual contributions were as follows:

Conception and design: LA, RDS, MS

Patient management and referral: RDS, ASagona, CT, RT, ASantoro

Acquisition of data: LA, MK, RDS

Image analysis: LA, MS, MK, LC

Statistical analysis: LA, RZ, RDS

Interpretation of data: LA, RDS, ASagona, MK, RZ

Drafting the article: LA, RDS, MS

Final approval of the revised manuscript: LA, RDS, LC, ASagona, CT, RT, ASantoro, MK, AC, RZ, MS.

Raw data are available on specific request to the corresponding author.

Compliance with ethical standards

Conflict of interest A. Chiti received speaker honoraria from General Electric, Blue Earth Diagnostics, and Sirtex Medical, acted as scientific

advisor for Blue Earth Diagnostics and Advanced Accelerator Applications, and benefited from an unconditional grant from Sanofi to Humanitas University. A. Santoro received speaker honoraria from Mundipharma, Takeda, Glaxo, Celgene, Roche, Teva, Arqule, Amgen, Eisai, BMS, Bayer, MSD, Astra Zeneca, Gilead, Sandoz, Servier, and Novartis, acted as scientific advisor for Gilead, Pfizer, Eisai, Servier, MSD, Bayer, Eli Lilly, Amgen, Marck Serono, Takeda, Ariad, Italfarmaco, and Celgene. L. Cozzi acts as Scientific Advisor to Varian Medical Systems and is Clinical Research Scientist at Humanitas Cancer Center.

All honoraria and grants are outside the scope of the submitted work. All other authors have no conflict of interest.

Ethics approval The study was approved by the Local Ethics Committee (authorization number 1591). Specific informed consent was not required according to Local Ethics Committee rules for retrospective study design.

References

- Perou CM, Sørile T, Eisen MB, Van De Rijn M, Jeffrey SS, Renshaw CA, et al. Molecular portraits of human breast tumours. *Nature*. 2000;406:747–52.
- Tyagi NK, Dhesy-Thind S. Clinical practice guidelines in breast cancer. *Curr Oncol*. 2018;25:S151–60.
- Cardoso F, Senkus E, Costa A, Papadopoulos E, Aapro M, André F, et al. 4th ESO–ESMO international consensus guidelines for advanced breast cancer (ABC 4). *Ann Oncol*. 2018;29:1634–57.
- Pinder SE, Provenzano E, Earl H, Ellis IO. Laboratory handling and histology reporting of breast specimens from patients who have received neoadjuvant chemotherapy. *Histopathology*. 2007;50:409–17.
- von Minckwitz G, Untch M, Blohmer J-U, 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.
- 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.
- Sella T, Gal Yam EN, Levanon K, Rotenberg TS, Gadot M, Kuchuk I, et al. Evaluation of tolerability and efficacy of incorporating carboplatin in neoadjuvant anthracycline and taxane based therapy in a BRCA1 enriched triple-negative breast cancer cohort. *Breast*. 2018;40:141–6.
- Hurtz H-J, Tesch H, Göhler T, Hutzschenreuter U, Harde J, Kruggel L, et al. Persistent impairments 3 years after (neo)adjuvant chemotherapy for breast cancer: results from the MaTox project. *Breast Cancer Res Treat*. 2017;165:721–31.
- Chakraborty D, Basu S, Ulaner GA, Alavi A, Kumar R. Diagnostic role of fluorodeoxyglucose PET in breast cancer: a history to current application. *PET Clin*. 2018;13:355–61.
- Kaida H, Toh U, Hayakawa M, Hattori S, Fujii T, Kurata S, et al. The relationship between 18F-FDG metabolic volumetric parameters and clinicopathological factors of breast cancer. *Nucl Med Commun*. 2013;34:562–70.
- García Vicente AM, Soriano Castrejón Á, León Martín A, Chacón López-Muñoz I, Muñoz Madero V, Muñoz Sánchez MDM, et al. Molecular subtypes of breast cancer: metabolic correlation with 18F-FDG PET/CT. *Eur J Nucl Med Mol Imaging*. 2013;40:1304–11.
- Koo HR, Park JS, Kang KW, Cho N, Chang JM, Bae MS, et al. 18F-FDG uptake in breast cancer correlates with immunohistochemically defined subtypes. *Eur Radiol*. 2014;24:610–8.
- Kajáry K, Tóké T, Dank M, Kulka J, Szakáll S, Lengyel Z. Correlation of the value of 18F-FDG uptake, described by SUVmax, SUVavg, metabolic tumour volume and total lesion glycolysis, to clinicopathological prognostic factors and biological subtypes in breast cancer. *Nucl Med Commun*. 2015;36:28–37.
- Kitajima K, Fukushima K, Miyoshi Y, Nishimukai A, Hirota S, Igarashi Y, et al. Association between 18F-FDG uptake and molecular subtype of breast cancer. *Eur J Nucl Med Mol Imaging*. 2015;42:1371–7.
- Lee SS, Bae SK, Park YS, Park JS, Kim TH, Yoon HK, et al. Correlation of molecular subtypes of invasive ductal carcinoma of breast with glucose metabolism in FDG PET/CT: based on the recommendations of the St. Gallen Consensus Meeting 2013. *Nucl Med Mol Imaging* (2010). 2017;51:79–85.
- García Vicente AM, Soriano Castrejón A, Amo-Salas M, Lopez Fidalgo JF, Muñoz Sanchez MM, Alvarez Cabellos R, et al. Glycolytic activity in breast cancer using 18F-FDG PET/CT as prognostic predictor: a molecular phenotype approach. *Rev Esp Med Nucl Imagen Mol*. 2016;35:152–8.
- Valdora F, Houssami N, Rossi F, Calabrese M, Tagliafico AS. Rapid review: radiomics and breast cancer. *Breast Cancer Res Treat*. 2018;169:217–29.
- Son SH, Kim D-H, Hong CM, Kim C-Y, Jeong SY, Lee S-W, et al. Prognostic implication of intratumoral metabolic heterogeneity in invasive ductal carcinoma of the breast. *BMC Cancer*. 2014;14:585.
- Soussan M, Orhac F, Boubaya M, Zelek L, Ziolo M, Eder V, et al. Relationship between tumor heterogeneity measured on FDG-PET/CT and pathological prognostic factors in invasive breast cancer. *PLoS One*. 2014;9:1–7.
- Yoon HJ, Kim Y, Kim BS. Intratumoral metabolic heterogeneity predicts invasive components in breast ductal carcinoma in situ. *Eur Radiol*. 2015;25:3648–58.
- Antunovic L, Gallivanone F, Sollini M, Sagona A, Invento A, Manfrinato G, et al. [18F]FDG PET/CT features for the molecular characterization of primary breast tumors. *Eur J Nucl Med Mol Imaging*. 2017;44:1945–54.
- Molina-García D, García-Vicente AM, Pérez-Beteta J, Amo-Salas M, Martínez-González A, Tello-Galán MJ, et al. Intratumoral heterogeneity in 18F-FDG PET/CT by textural analysis in breast cancer as a predictive and prognostic surrogate. *Ann Nucl Med*. 2018;32:379–88.
- Yoon H-J, Kim Y, Chung J, Kim BS. Predicting neo-adjuvant chemotherapy response and progression-free survival of locally advanced breast cancer using textural features of intratumoral heterogeneity on F-18 FDG PET/CT and diffusion-weighted MR imaging. *Breast J*. 2018:1–8. [Epub ahead of print]
- Azad GK, Cousin F, Siddique M, Taylor B, Goh V, Cook GJR. Does measurement of first-order and heterogeneity parameters improve response assessment of bone metastases in breast cancer compared to SUVmax in [18F]fluoride and [18F]FDG PET? *Mol Imaging Biol*. 2018. [Epub ahead of print]. <https://doi.org/10.1007/s11307-018-1262-3>.
- Gong C, Ma G, Hu X, Zhang Y, Wang Z, Zhang J, et al. Pretreatment 18F-FDG uptake heterogeneity predicts treatment outcome of first-line chemotherapy in patients with metastatic triple-negative breast cancer. *Oncologist*. 2018;23(10):1144–52. <https://doi.org/10.1634/theoncologist.2018-0001>.
- Zwanenburg A, Leger S, Vallières M, Löck S, for the Image Biomarker Standardisation Initiative (IBSI). Image biomarker standardisation initiative — feature definitions 2016. <https://arxiv.org/abs/1612.07003>.

27. Boellaard R, Delgado-Bolton R, Oyen WJG, Giammarile F, Tatsch K, Eschner W, et al. FDG PET/CT: EANM procedure guidelines for tumour imaging: version 2.0. *Eur J Nucl Med Mol Imaging*. 2014;42:328–54.
28. Lemarignier C, Martineau A, Teixeira L, Vercellino L, Espié M, Merlet P, et al. Correlation between tumour characteristics, SUV measurements, metabolic tumour volume, TLG and textural features assessed with 18F-FDG PET in a large cohort of oestrogen receptor-positive breast cancer patients. *Eur J Nucl Med Mol Imaging*. 2017;44:1145–54.
29. Garcia-Vicente AM, Pérez-Beteta J, Amo-Salas M, Molina D, Jimenez-Londoño GA, Soriano-Castrejón AM, et al. Predictive and prognostic potential of volume-based metabolic variables obtained by a baseline 18F-FDG PET/CT in breast cancer with neoadjuvant chemotherapy indication. *Rev Esp Med Nucl Imagen Mol*. 2018;37:73–9.
30. Hanley JA, McNeil BJ. The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology*. 1982;143:29–36.
31. Arboretti R, Salmaso L. Model performance analysis and model validation in logistic regression. *Statistica*. 2003;63:375–96.
32. Donders ART, van der Heijden GJMG, Stijnen T, Moons KGM. Review: a gentle introduction to imputation of missing values. *J Clin Epidemiol*. 2006;59:1087–91.
33. Van Buuren S, Oudshoorn K. Flexible multivariate imputation by MICE TNO prevention and health. Leiden: TNO Prevention and Health; 1999.
34. Beukinga RJ, Hulshoff JB, Mul VEM, Noordzij W, Kats-Ugurlu G, Slart RHJA, et al. Prediction of response to neoadjuvant chemotherapy and radiation therapy with baseline and restaging ¹⁸F-FDG PET imaging biomarkers in patients with esophageal cancer. *Radiology*. 2018;287:983–92.
35. Groheux D, Biard L, Lehmann-Che J, Teixeira L, Bouhidel FA, Poirot B, et al. Tumor metabolism assessed by FDG-PET/CT and tumor proliferation assessed by genomic grade index to predict response to neoadjuvant chemotherapy in triple negative breast cancer. *Eur J Nucl Med Mol Imaging*. 2018;45:1279–88.
36. Wood AM, White IR, Royston P. How should variable selection be performed with multiply imputed data? *Stat Med*. 2008;27:3227–46.
37. Airola A, Pahikkala T. A comparison of AUC estimators in small-sample studies. *JMLR - Work Mach Learn Syst Biol*. 2009; 8:3–13.
38. Haque W, Verma V, Hatch S, Suzanne Klimberg V, Brian Butler E, Teh BS. Response rates and pathologic complete response by breast cancer molecular subtype following neoadjuvant chemotherapy. *Breast Cancer Res Treat*. 2018;170:559–67.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.