Critical Reviews in Oncology/Hematology 112 (2017) 198–207

Contents lists available at ScienceDirect

Critical Reviews in Oncology/Hematology

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

# Imaging performance in guiding response to neoadjuvant therapy according to breast cancer subtypes: A systematic literature review

Melanie A. Lindenberg[a,b], Anna Miquel-Cases[a,b], Valesca P. Retèl[a,b], Gabe S. Sonke[c], Jelle Wesseling[d], Marcel P.M. Stokkel[e], Wim H. van Harten[a,b,*]

[a] Netherlands Cancer Institute, Division of Psychosocial Research and Epidemiology, Plesmanlaan 121, 1066 CX Amsterdam, The Netherlands
[b] Department of Health Technology and Services Research, MB-HTSR, University of Twente, PO Box 217, 7500 AE Enschede, The Netherlands
[c] Netherlands Cancer Institute, Medical Oncology Division, Plesmanlaan 121, 1066 CX Amsterdam, The Netherlands
[d] Netherlands Cancer Institute, Diagnostic Oncology Division and Division of Molecular Pathology, Plesmanlaan 121, 1066 CX Amsterdam, The Netherlands
[e] Netherlands Cancer Institute, Division of Nuclear Medicine, Plesmanlaan 121, 1066 CX Amsterdam, The Netherlands

**Contents**



ARTICLE INFO

*Article history:*
Received 18 May 2016
Received in revised form 18 January 2017
Accepted 14 February 2017

ABSTRACT

Monitoring therapeutic response to neoadjuvant chemotherapy(NAC) is likely to improve NAC effectiveness in breast cancer(BC). Imaging performance seems to vary per tumour subtype(by ER and HER2 status), therefore we performed a systematic review on subtype specific imaging performance in monitoring NAC in BC.

*Abbreviations:* BC, breast cancer; NAC, neoadjuvant chemotherapy; HER2, human epidermal growth factor 2; ER, oestrogen receptor; PR, progesterone receptor; TNBC, triple negative breast cancer; PET/CT, positron emission tomography computed tomography; pCR, pathologic complete response; PE, physical examination; US, ultrasound; MRI, magnetic resonance imaging; DCE-MRI, dynamic contrast-enhanced magnetic resonance imaging; DW-MRI, diffusion-weighted magnetic resonance imaging; NPV, negative predictive value; PPV, positive predictive value; ROC, receiver operating curve; AUC, area under the curve; QUADAS, quality assessment of diagnostic accuracy studies; TN, true negative; FN, False negative; TP, true positive; FP, false positive.

* Corresponding author at: Netherlands Cancer Institute, Division of Psychosocial Research and Epidemiology, Plesmanlaan 121, 1066 CX Amsterdam, The Netherlands.
*E-mail addresses:* m.lindenberg@nki.nl (M.A. Lindenberg), a.miquel@nki.nl (A. Miquel-Cases), v.retel@nki.nl (V.P. Retèl), g.sonke@nki.nl (G.S. Sonke), j.wesseling@nki.nl (J. Wesseling), m.stokkel@nki.nl (M.P.M. Stokkel), w.v.harten@nki.nl (W.H. van Harten).

http://dx.doi.org/10.1016/j.critrevonc.2017.02.014

*Keywords:*
Neoadjuvant therapy
Breast cancer
ER
HER2
Imaging
Therapeutic response

Studies examining imaging performance in predicting pathologic complete response(pCR) during NAC in BC subtypes were selected. Per study, negative- and positive predictive value, sensitivity(se) and specificity(sp), AUC and accuracy were derived.
Fifteen/106 articles were included. Inter-study variability was revealed in: monitoring interval, response and pCR definitions. In ER-positive/HER2-negative BC, $^{18}$1F FDG-PET/CT showed se/sp of 38%–89%/74%–100%, MRI showed se/sp of 35%–37%/87%–89%. In triple negative BC, $^{18}$1F FDG-PET/CT showed se/sp of 0%–79%/95%–100%. $^{18}$1F FDG-PET/CT showed in ER-positive/HER2-positive BC se/sp of 59%/80% and in ER-negative/HER2-positive 27%/88%.
Evidence on imaging performance in monitoring NAC according BC subtypes is lacking. Consensus should be reached in: definitions of pCR, response and monitoring interval before starting well-designed studies.


## 1. Introduction

In 2012, 1.7 million new cases of breast cancer were diagnosed worldwide. Despite research and improvements in breast cancer treatment, breast cancer is still: one of the most prevalent cancers overall, the most prevalent cancer among women, and one of the main causes of death (WHO, 2012). Research on new treatment approaches is thus of evident interest.

Neoadjuvant chemotherapy(NAC) showed to be at least equally effective as adjuvant chemotherapy (Mauri et al., 2005) while having additional advantages (Fisher et al., 1997; van der Hage et al., 2001), such as the ability to monitor therapeutic response during treatment (Gralow et al., 2008). Early therapeutic response assessed by imaging seems to be a predictor of pathologic complete response(pCR) (Marinovich et al., 2012), usually defined as absence of any residual invasive tumour cells in the original tumour bed and axilla (Kaufmann et al., 2006). PCR itself predicts long-term survival, especially in HER2-positive and triple negative(TN) tumours (Chollet et al., 2002; Von Minckwitz et al., 2012), monitoring early therapeutic response may be used to guide systemic treatment, which is called a response-guided NAC approach (von Minckwitz et al., 2013). Under this scenario, patients could be monitored after a specific number of NAC cycles, and according to their response at imaging, their further systematic treatment could be tailored, i.e. responders continue with the same initial treatment, and non-responders can be switched to a presumably non-cross-resistant regimen(Fig. 1) (von Minckwitz et al., 2013).

Currently, there is no definite guideline to assess response to NAC during treatment. Previous authors proposed physical examination plus mammography and ultrasound, but their performance seems to be limited (Yeh et al., 2005; Hamisa et al., 2015; Londero et al., 2004). Therefore, performance examination of more advanced techniques, i.e. magnetic resonance imaging(MRI) and PET—Computed Tomography(PET/CT) is of interest. So far, meta-analyses have shown sensitivities and specificities of 68% and 91% for dynamic contrast-enhanced(DCE)-MRI (Wu et al., 2012), 93% and 82% for diffusion-weighted(DW)-MRI (Wu et al., 2012) and 84% and 71% for $^{18}$F FDG-PET/CT (Cheng et al., 2012) respectively. On the basis of these findings, MRI is currently the technique mainly used in clinical practice. While these techniques seem to already have better performance, recent studies have shown that breast cancer subtype affects imaging performance (Loo et al., 2011; Hayashi et al., 2013; Ko et al., 2013). Hence, personalizing the use of imaging techniques based on subtypes may further improve their performance in evaluating therapeutic response (Loo et al., 2011; Humbert et al., 2012).

As there is no subtype-specific guidance on imaging techniques to monitor therapeutic response during NAC to guide in further treatment regimen, this paper aims to create an overview of current knowledge on the performance of imaging techniques in breast cancer subtypes based on expression of ER and HER2.

## 2. Methods

We performed a systematic literature search to find studies reporting on the performance of imaging in assessing pCR during NAC for breast cancer subtypes.

### 2.1. Search strategy

For PubMed the terms: “breast cancer”(MeSH: Breast neoplasm); “imaging”(i.e. MRI, PET/CT); “outcome”(pathologic complete response, clinical response); “Neoadjuvant chemotherapy” and “breast cancer subtype”(oestrogen receptor(ER), progesterone receptor(PR), luminal, triple negative(TN) and human endocrine receptor 2(HER2)) were combined for the systematic search(Supplement 1). Snowballing was used to find additional relevant publications.

### 2.2. Selection criteria

The search was limited to studies written in English and published between January 2000 and March 2015. Case studies were excluded. Studies were included if performance data of the imaging technique(s) was reported: before and during NAC, specified to at least one receptor status(ER/HER2) and controlled with pCR as primary outcome. As secondary outcomes the neoadjuvant response index(NRI) (Rodenhuis et al., 2010) and residual cancer burden (Symmans et al., 2007) were accepted as response definition. Finally, studies using FDG-PET without CT were excluded, as this technology is no longer recommended in daily practice.

### 2.3. Data extraction

The first selection was performed based on abstract information and following the inclusion and exclusion criteria by two independent reviewers(AMC and ML). The selected studies were fully read by the same reviewers and were again assessed based on the inclusion and exclusion criteria. Disagreements were first discussed between the two reviewers, and if no agreement was reached, a third reviewer was approached(VR). For each article, the following items were extracted: author, sample size, study design, treatment regimen, breast cancer subtype, clinical stage, age, monitoring technique, cut-off value or response definition at imaging, interval time i.e. number of NAC cycles between baseline and response monitoring, technical settings, pCR definition: pCR or partial response, performance results, i.e. sensitivity, specificity, accuracy, negative and positive predictive values(NPV, PPV) and

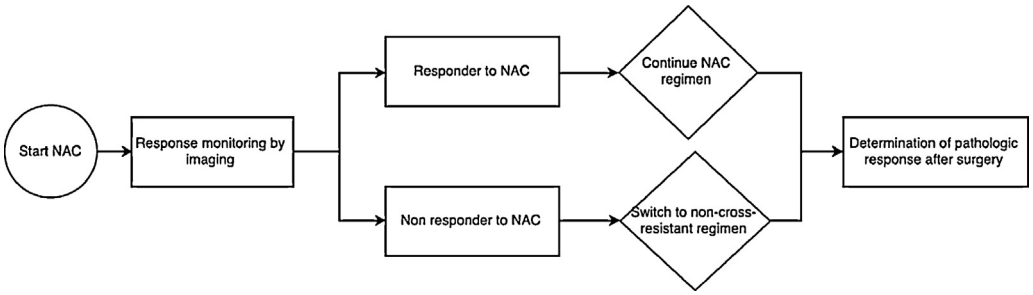


**Fig. 1.** Response-guided neoadjuvant (NAC) approach. Patients start with first-line NAC treatment and after a specific number of cycles, they are monitored by imaging. Patients considered responders of NAC at imaging (according to a pre-defined threshold) continue the same initial treatment, whereas non-responders are switched to a presumably non-cross resistant treatment. Upon NAC finalization, pathologic response is determined at surgery, which is used to determine whether there the imaging results were correct.

Area Under the Curve(AUC) in a Receiver Operating Curve(ROC) and if available false and true positives and negatives. Finally, we requested to authors of included studies in which information was presented towards only ER or HER2, whether performance information on further stratified groups existed.

### 2.4. Quality assessment

Three research design criteria were defined to assess quality of the included articles. Articles were considered of sufficient quality if they complied with 2 of the 3 following criteria: 1) no treatment switch during NAC; 2) score higher than 8 on the Quality Assessment of Diagnostic Accuracy Studies(QUADAS) (Bossuyt et al., 2007); and 3) sample size higher or equal to 20. If more than one subtype was presented in the article, criteria 2 and 3 were assessed per subtype.

### 2.5. Performance of imaging

For articles directly reporting on the number of true/false negative/positive(TN,FN,TP,FP) patients, and studies in which these were possible to derive, $2 \times 2$ contingency tables were constructed. These were used to calculate sensitivity(ability of imaging to identify non-responders with residual tumour tissue after NAC i.e.,TP/TP + FN), specificity (ability of imaging to identify responders achieving a pCR after NAC i.e.,TN/TN + FP), NPV (TN/TN + FN), PPV(TP/TP + FP) and accuracy(TP + TN/all patients). Pooling of these sensitivity and specificity values would be the preferable method to compare different imaging modalities. However, due to substantial heterogeneity across the included studies it was inappropriate to use this method.

### 2.6. Preferred imaging technique per subtype

To score and compare the performance of the imaging techniques a scale was developed. The scale runs from A(perfect performance) to D(insufficient performance), and was applied to the various performance concepts i.e., ROC-AUC value, accuracy and sensitivity/specificity. Based on the coding shown in Table 3, the performance results per breast cancer subtype were placed in order in Table 5, and, if sufficient results were available for different imaging techniques per subtype, the preferred imaging technique in each subtype was chosen.

## 3. Results

Of the initially 229 identified articles, 30 were selected for full reading after removing duplicates. Sixteen articles were further excluded because: 1)response monitoring was performed before or after NAC, 2)did not report performance data or did not specify their results to subtypes and 3)FDG-PET was used without CT. After snowballing one extra article was included, which made a total of 15 articles(Fig. 2).

### 3.1. Study characteristics

Study populations' size ranged from seven to 246 patients (median: 31), and the overall mean age was 50 years. Studies enrolled patients prospectively (9 studies) and retrospectively (6 studies). One of the five contacted authors replied with additional data towards receptor status (Rigter et al., 2013). Nine articles presented results for the subgroup of ER-positive/HER2-negative patients (Loo et al., 2011; Rigter et al., 2013; Koolen et al., 2014; Koolen et al., 2013; Groheux, Hatt et al., 2013; Zucchini et al., 2013; Charehbili et al., 2014; Martoni et al., 2010), nine for TN patients (Loo et al., 2011; Humbert et al., 2012; Koolen et al., 2014; Koolen et al., 2013; Zucchini et al., 2013; Martoni et al., 2010; Hatt et al., 2013; Groheux et al., 2012; Groheux et al., 2014), nine for the whole HER2-positive group (Loo et al., 2011; Humbert et al., 2012; Koolen et al., 2014; Koolen et al., 2013; Zucchini et al., 2013; Martoni et al., 2010; Hatt et al., 2013; Groheux, Giacchetti et al., 2013; Humbert, Cochet et al., 2014) and one for HER2-positive stratified by ER receptor status (Gebhart et al., 2013). The NAC regimen differed per subtype, with ER-positive/HER2-negative patients mainly receiving doxorubicin and cyclophosphamide(AC) and switched to docetaxel and capecitabine(DC) in case of an unfavourable intermediate response (Loo et al., 2011; Rigter et al., 2013; Koolen et al., 2014; Koolen et al., 2013), TNBC patients mainly receiving epirubicin and cyclophosphamide followed by docetaxel(EC-D) (Groheux, Hatt et al., 2013; Groheux et al., 2012; Groheux et al., 2014; Groheux, Giacchetti et al., 2013) or one of the following regimens: intensified EC-D(SIM) (Groheux et al., 2012; Groheux et al., 2014), fluorouracil plus EC(FEC) (Humbert et al., 2012; Humbert, Berriolo-Riedinger et al., 2014), FEC-D (Humbert et al., 2012; Humbert, Berriolo-Riedinger et al., 2014). ER-negative/HER2-positive patients received mainly EC(-D) followed by a combination of trastuzumab and paclitaxel or docetaxel (Hatt et al., 2013; Humbert, Cochet et al., 2014; Gebhart et al., 2013).

Three definitions of pCR were identified (Humbert et al., 2012; Rigter et al., 2013; Koolen et al., 2014; Koolen et al., 2013; Charehbili et al., 2014; Groheux et al., 2012; Groheux et al., 2014; Groheux, Giacchetti et al., 2013; Gebhart et al., 2013; Humbert, Berriolo-Riedinger et al., 2014), and are shown in Table 1. These three definitions will be referred to as pCR in rest of the article. The fourth identified category, described as partial response, was: considerable or partial reduction in tumour cells in breast after completion of neoadjuvant chemotherapy (Loo et al., 2011; Groheux, Hatt et al., 2013; Zucchini et al., 2013; Martoni et al., 2010).

Of the included articles, three were on MRI and 12 on $^{18}$F FDG-PET/CT. A summary of the main settings used in the assessment

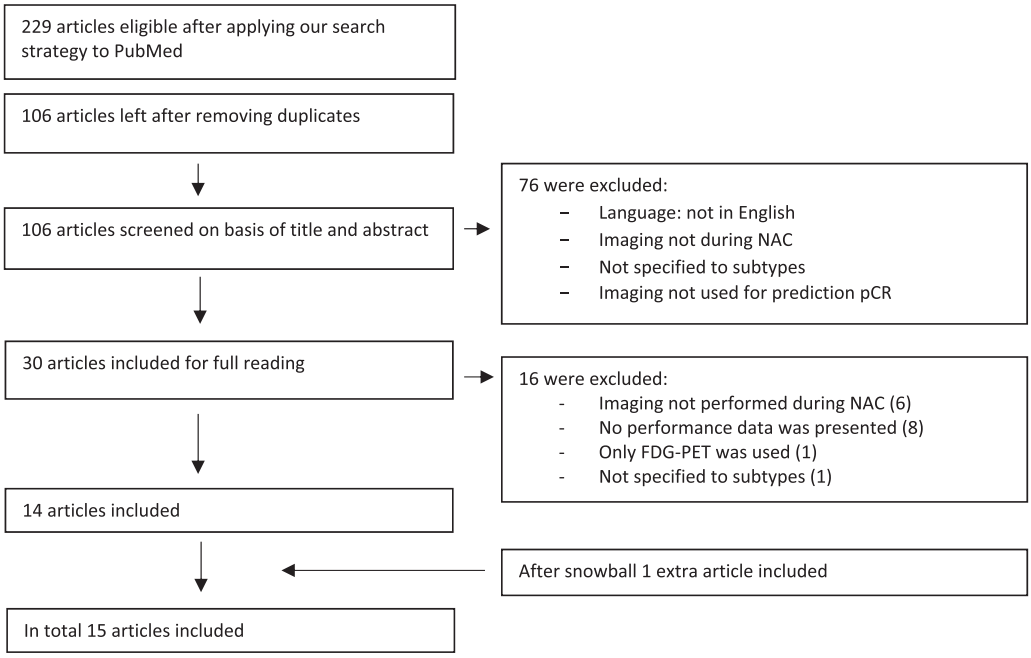


**Fig. 2.** Flow diagram of selection process. Of the 106 identified articles through PubMed, 15 articles were finally included.

**Table 1**
Categorization of different pathologic complete response definitions (pCR).

| Category | Classifications and scales used in literature |
|---|---|
| **Category 1**<br>*Complete absence of invasive tumourcells andductal carcinoma in situ (DCIS) in breast **and** axillary lymph nodes after completion of neoadjuvant chemotherapy* | *- Chevalier classification grade 1* (Chevallier et al., 1993)<br>*- ypT0 ypN0* |
| **Category 2**<br>*Complete absence of invasive tumour cells in the breast **and** axillary lymph nodes after completion of neoadjuvant chemotherapy* | *- Chevalier classification grade 2* (Chevallier et al., 1993)<br>*- ypT0/is ypN0*<br>*- Miller and Payne grade 5 and NRG A or D* (Ogston et al., 2003) |
| ***Category 3***<br>*Complete absence of invasive tumour cells in the breast after completion of neoadjuvant chemotherapy* | *- Miller and Payne grade 5* (Ogston et al., 2003)<br>*- YpT0/is*<br>*- ypT0/is ypN0/+* |
| ***Category 4***<br>*Considerable or partial reduction in tumour cells in breast after completion of neoadjuvant chemotherapy* | *- Sataloff classification T-A* (Sataloff et al., 1995)<br>*- Sataloff classification T-B* (Sataloff et al., 1995)<br>*- Miller and Payne grade 4* (Ogston et al., 2003) |

**Table 2**
Scale to score diagnostic performance. Each performance concept has its sensitivity and specificity data described as $^{(\alpha)}$, ROC-AUC values were presented as $^{(\beta)}$ and accuracy results as $^{(\gamma)}$. The performance scales used per concept are presented in the last three columns of the table, and these are in turn categorized from perfect (A) to insufficient (D) performance by the first column of the table.

| Performance | | Sensitivity/specificity $^{(\alpha)}$ | ROC-AUC value $^{(\beta)}$ | Accuracy $^{(\gamma)}$ |
|---|---|---|---|---|
| A | Good | Both >80% | 0.80–1.00 | 80%–100% |
| B | Sufficient | Both >60% and <80% or one result >60% and <80% and one result >80% | 0.60–<0.80 | 60%–<80% |
| C | Limited | Both >40% and <60% or one result >40% and <60% and one result >60% | 0.40–<0.60 | 40%–<60% |
| D | Insufficient | Both >0% and <40% or one result >0% and <40% and one result >40% | 0.00–<0.40 | 0%–<40% |

General abbreviations: ROC-AUC: Area Under the – Receiver operator curve.

with $^{18}$F FDG-PET/CT and MRI are presented in Table 3. The study characteristics described are presented in Supplement 2. Regarding quality assessment, three subgroups showed a small sample size (Zucchini et al., 2013; Martoni et al., 2010; Hatt et al., 2013), four subgroups had a study design that allowed a switch in treatment during NAC (Loo et al., 2011; Rigter et al., 2013; Koolen et al., 2014; Koolen et al., 2013), and no study showed a score below 8 on the QUADAS list (Supplement 3). Since each subgroup of each article satisfied 2 of the 3 criteria described in the method section, no study or subgroup was excluded from further analysis (Table 4 ).

### 3.2. Performance of imaging techniques per subtype

Results on performance of the various imaging techniques per breast cancer subtype are summarized in the section below and in Table 5. Table 5 shows also the number of NAC cycles between baseline monitoring and response monitoring, the cut-off value that was used, and pCR definition according Table 1. In addition to Table 5, the sensitivity and specificity values identified per subgroup are shown in an AUC plot (Fig. 3).

**Table 3**
Main technical settings of imaging techniques used in response assessment summarized per imaging technique. More details are described in the study characteristics table (Supplement 2).

| Imaging technique | Technology | Contrast (dosage) | Settings | Position |
|---|---|---|---|---|
| MRI (Loo et al., 2011; Rigter et al., 2013; Charehbili et al., 2014) | Philips magnetom vision (Loo et al., 2011; Rigter et al., 2013) 1.5T and 3.0T magnet (Loo et al., 2011; Rigter et al., 2013; Charehbili et al., 2014) | Gadolinium (14 ml of 0.1 mmol/kg)(Loo et al., 2011; Rigter et al., 2013) | – | Use of breast coils (Loo et al., 2011; Rigter et al., 2013; Charehbili et al., 2014) |
| $^{18}$F FDG-PET/CT(Humbert et al., 2012; Koolen et al., 2014; Koolen et al., 2013; Groheux, Hatt et al., 2013; Zucchini et al., 2013; Martoni et al., 2010; Groheux et al., 2012; Groheux et al., 2014; Groheux, Giacchetti et al., 2013; Humbert, Cochet et al., 2014; Gebhart et al., 2013) | **Philips** (Humbert et al., 2012; Koolen et al., 2014; Koolen et al., 2013; Groheux, Hatt et al., 2013; Groheux et al., 2012; Groheux et al., 2014; Groheux, Giacchetti et al., 2013; Gebhart et al., 2013; Humbert, Cochet et al., 2015) **GE medical** (Zucchini et al., 2013; Martoni et al., 2010; Gebhart et al., 2013) **Siemens** (Gebhart et al., 2013) | $^{18}$F-FDG (3.5 MBq/kg – 7.4 MBq/kg) (all articles) Fasted 6 h before injection (all articles) | Scan performed 60–70 min after contrast injection CT: 120 kV and 100mAs (all articles) | Hanging breast method (Koolen et al., 2014; Koolen et al., 2013) |

General abbreviations: MBq MegaBecquerel; mAs: milliampere/second; kV: Kilovolt; T:Tesla.

### 3.2.1. ER-positive/HER2-negative

Six studies assessed performance of $^{18}$F FDG-PET/CT and 3 of MRI. The use of $^{18}$F FDG-PET/CT showed AUC-ROC values of: 0.61(CI 0.37–0.86) after 1 NAC cycle (Koolen et al., 2014), 0.87(CI 0.69–1.00) after 3 NAC cycles (Koolen et al., 2014), 0.77(CI 0.68–0.87) after 3 NAC cycles (Koolen et al., 2013) and 0.88 after 2 NAC cycles (Hatt et al., 2013). An Italian research group described performance of $^{18}$F FDG-PET/CT in 2 articles, showing a sensitivity of 38% and specificity of 100% after 2 NAC cycles using the difference in Standardized Uptake Value (ΔSUVmax) (Zucchini et al., 2013; Martoni et al., 2010). Another study showed sensitivity of 62% and specificity of 78% also after 2 NAC cycles (Groheux, Hatt et al., 2013). Using the difference in Total Lesion Glycolysis(ΔTLG), $^{18}$F FDG-PET/CT showed a sensitivity of 89%, sensitivity of 74%, and an AUC value of 0.81 after 2 NAC cycles (Groheux, Hatt et al., 2013) and 0.96 in case partial responders were included as responder (Hatt et al., 2013). MRI showed sensitivity of 35%–37%, specificity of 87%–89%, accuracy of 39%–45%, NPV of 10%–22% and PPV of 93%–98% after 3 NAC cycles in two different studies (Rigter et al., 2013; Charehbili et al., 2014). Although the trial of Charehbili et al. was in HER2-negative patients, its results were included in this subtype because the majority of patients showed an ER-positive expression (187/222)

**Table 4**
Quality assessment based on three criteria. 1. The treatment was not switched during NAC, 2. Study does not score below 8 on the quality assessment tool for diagnostic accuracy studies (QUADAS), 3. The sample size is above 20 patients.

| Reference | Subtype | Sample size | Criteria 1 Treatment is not switched during NAC | Criteria 2 No risk of bias is present | Criteria 3 Sample size is ≥20 patients | Include? |
|---|---|---|---|---|---|---|
| Charehbili et al. (2014) | ER-positive/HER2-negative | 194 | + | + | + | Yes |
| Gebhart et al. (2013) | ER-negative/HER2-positive | 43 | + | + | + | Yes |
| | ER-positive/HER2-positive | 34 | + | + | + | Yes |
| Groheux et al. (2012) | TN | 20 | + | + | + | Yes |
| Groheux, Hatt et al. (2013) | ER-positive/HER2-negative | 64 | + | + | + | Yes |
| Groheux, Giacchetti et al. (2013) | HER2-positive | 30 | + | + | + | Yes |
| Groheux et al. (2014) | TN | 50 | + | + | + | Yes |
| Hatt et al. (2013) | ER-positive/HER2-negative | 26 | + | + | + | Yes |
| | TN | 13 | + | + | – | Yes |
| | HER2-positive | 12 | + | + | – | Yes |
| Humbert et al. (2012) | ER-positive/HER2-negative | 53 | + | + | + | Yes |
| | TN | 25 | + | + | + | Yes |
| | HER2-positive | 37 | + | + | + | Yes |
| Humbert, Cochet et al. (2014) | HER2-positive | 57 | + | + | + | Yes |
| Koolen et al. (2014) | ER-positive/HER2-negative | 50 | – | + | + | Yes |
| | TN | 31 | + | + | + | Yes |
| | HER2-positive | 26 | + | + | + | Yes |
| Koolen et al. (2013) | ER-positive/HER2-negative | 45 | – | + | + | Yes |
| | TN | 25 | + | + | + | Yes |
| | HER2-positive | 25 | + | + | + | Yes |
| Loo et al. (2011) | ER-positive/HER2-negative | 103 | – | + | + | Yes |
| | TN | 47 | + | + | + | Yes |
| | HER2-positive | 38 | + | + | + | Yes |
| Martoni et al. (2010) | ER-positive/HER2-negative | 16 | + | + | – | Yes |
| | TN | 9 | + | + | – | Yes |
| | HER2-positive | 7 | + | + | – | Yes |
| Rigter et al. (2013) | ER-positive/HER2-negative | 246 | – | + | + | Yes |
| Zucchini et al. (2013) | ER-positive/HER2-negative | 31 | + | + | + | Yes |
| | TN | 15 | + | + | – | Yes |
| | HER2-positive | 14 | + | + | – | Yes |

**Table 5**
Performance of imaging techniques per subtype. **Response definition**: [I] response category 1; [II] response category 2; [III] response category 3; [IIII] response category 4; **Cut-off values**: [1]: cut-off value 25% size reduction; [2]: cut-off value: 30% size reduction; [3]: cut-off value −15% ΔSUVmax; [4]: cut-off value −25% ΔSUVmax; [5]: cut-off value: −38% ΔSUVmax; [6]: cut-off value: −42% ΔSUVmax; [7]: cut-off value: −50% ΔSUVmax; [8]: cut-off value −60% ΔSUVmax; [9]: cut-off value −62% ΔSUVmax; [10]: cut-off value −75% ΔSUVmax; [11]: cut-off value: −71% ΔTLG; **Outcome parameters**: [*]: ΔSUVmax; [Δ]: Different outcome parameters; **Performance score:** [α]: Sensitivity and specificity results; [β]: AUC values; [γ]: Accuracy values; **Other:** [#] = in the original article it was described as administrations instead of cycles; **General abbreviations:** AUC = Area Under the Curve; NPV: Negative Predictive Value; PPV: Positive Predictive Value; SUV: Standard Uptake Value; TLG: Total Lesion Glycolysis; MATV: Metabolic Active Tumour Value.

| Article | Monitoring technique | Monitoring interval | Cut-off value | pCR definition | Sens | Spec | Acc | NPV | PPV | AUC | Performance score |
|---|---|---|---|---|---|---|---|---|---|---|---|
| **ER-positive/HER2-negative** | | | | | | | | | | | |
| Hatt et al. (2013) | $^{18}$1F FDG-PET/CT | After 2 cycles | Several imaging measurement parameters | IV | | | | | | ΔSUVmax: 0.88<br>ΔTLG: 0.96<br>ΔMATV: 0.98 | $A^{(\beta)}$<br>$A^{(\beta)}$<br>$A^{(\beta)}$ |
| Koolen et al. (2014) | $^{18}$1F FDG-PET/CT | After 3 cycles | ΔSUVmax | II | | | | | | 0.87 (0.69–1.00) | $A^{(\beta)}$ |
| Groheux, Hatt et al. (2013) | $^{18}$1F FDG-PET/CT | After 2 cycles | -71% ΔTLG | IV | 89% | 74% | | 31% | 98% | 0.81 | $B^{(\alpha)}A^{(\beta)}$ |
| Groheux, Hatt et al. (2013) | $^{18}$1F FDG-PET/CT | After 2 cycles | -38% ΔSUVmax | IV | 62% | 78% | | 12% | 97% | 0.73 | $B^{(\alpha)}B^{(\beta)}$ |
| Koolen et al. (2013) | $^{18}$1F FDG-PET/CT | After 3 cycles | ΔSUVmax | III | | | | | | 0.77 (0.68–0.87) | $B^{(\beta)}$ |
| Koolen et al. (2014) | $^{18}$1F FDG-PET/CT | After 1 cycle | ΔSUVmax | II | | | | | | 0.61 (0.37–0.86) | $B^{(\beta)}$ |
| Zucchini et al. (2013) | $^{18}$1F FDG-PET/CT | After 2 cycles | -50% ΔSUVmax | IV | 38% | 100% | | 24% | 100% | | $D^{(\alpha)}$ |
| Martoni et al. (2010) | $^{18}$1F FDG-PET/CT | After 2 cycles | -50% ΔSUVmax | IV | 38% | 100% | 50% | 27% | 100% | | $D^{(\alpha)}C^{(\gamma)}$ |
| Charehbili et al. (2014) | DCE MRI | After 3 cycles | 30% size reduction | III | 37% | 87% | 45% | 22% | 93% | 0.55 (0.45–0.65) | $D^{(\alpha)}C^{(\beta)}C^{(\gamma)}$ |
| Rigter et al. (2013) | DCE MRI | After 3 cycles | 25% size reduction | III | 35% | 89% | 39% | 10% | 98% | | $D^{(\alpha)}D^{(\gamma)}$ |
| Loo et al. (2011) | DCE MRI | After 3 cycles | 30% size reduction | IV | Association between BRI and tumour decrease was not significant (p = 0.07) | | | | | | – |
| **Triple negative** | | | | | | | | | | | |
| Groheux et al. (2014) | $^{18}$1F FDG-PET/CT | After 2 cycles | −50% ΔSUVmax | II | 71% | 95% | 80% | 67% | 96% | – | $A^{(\alpha)}A^{(\gamma)}$ |
| Koolen et al. (2013) | $^{18}$1F FDG-PET/CT | After 3 cycles | ΔSUVmax | III | | | | | | 0.85 (0.69–1.00) | $A^{(\beta)}$ |
| Koolen et al. (2014) | $^{18}$1F FDG-PET/CT | After 3 cycles | ΔSUVmax | II | | | | | | 0.87 (0.73–1.00) | $A^{(\beta)}$ |
| Groheux et al. (2012) | $^{18}$1F FDG-PET/CT | After 2 cycles | −50% ΔSUVmax | II | 79% | 100% | 85% | 67% | 100% | 0.881 | $B^{(\alpha)}A^{(\beta)}A^{(\gamma)}$ |
| Groheux et al. (2012) | $^{18}$1F FDG-PET/CT | After 2 cycles | −42% ΔSUVmax | II | 64% | 100% | 75% | 55% | 100% | 0.881 | $B^{(\alpha)}A^{(\beta)}B^{(\gamma)}$ |
| Groheux et al. (2014) | $^{18}$1F FDG-PET/CT | After 2 cycles | −42% ΔSUVmax | II | 58% | 100% | 74% | 59% | 100% | – | $C^{(\alpha)}B^{(\gamma)}$ |
| Koolen et al. (2014) | $^{18}$1F FDG-PET/CT | After 1 cycle | ΔSUVmax | II | | | | | | 0.76 (0.55–0.96) | $B^{(\beta)}$ |
| Zucchini et al. (2013) | $^{18}$1F FDG-PET/CT | After 2 cycles | −50% ΔSUVmax | IV | 0% | 100% | | 27% | 0% | – | $D^{(\alpha)}$ |
| Martoni et al. (2010) | $^{18}$1F FDG-PET/CT | After 2 cycles | −50% ΔSUVmax | IV | 0% | 100% | 33% | 33% | | – | $D^{(\alpha)}D^{(\gamma)}$ |
| Humbert et al. (2012) | $^{18}$1F FDG-PET/CT | After 1 cycle | −75% ΔSUVmax | II | No significant correlation between early metabolic response and pCR | | | | | | – |
| Hatt et al. (2013) | $^{18}$1F FDG-PET/CT | After 2 cycles | Several outcome parameters | IV | Use of different parameters did not improve predictive value of ΔSUVmax | | | | | | – |
| Loo et al. (2011) | DCE MRI | After 3 cycles | 30% size reduction | IV | Association between BRI and largest tumour diameter was significant (p = < 0.001) | | | | | | – |
| **HER2-positive** | | | | | | | | | | | |
| Groheux, Giacchetti et al. (2013) | $^{18}$1F FDG-PET/CT | After 2 cycles | -62% ΔSUVmax | III | 86% | 63% | 73% | 84% | 67% | 0.86 | $B^{(\alpha)}A^{(\beta)}B^{(\gamma)}$ |
| Groheux, Giacchetti et al. (2013) | $^{18}$1F FDG-PET/CT | After 2 cycles | -62% ΔSUVmax | II | 86% | 75% | 80% | 86% | 75% | 0.86 | $B^{(\alpha)}A^{(\beta)}A^{(\gamma)}$ |
| Humbert et al. (2012) | $^{18}$1F FDG-PET/CT | After 1 cycle | -75% ΔSUVmax | II | 64% | 83% | 76% | 79% | 69% | 0.73 | $B^{(\alpha)}B^{(\beta)}B^{(\gamma)}$ |
| Humbert, Cochet et al. (2014) | $^{18}$1F FDG-PET/CT | After 1 cycle | -60% ΔSUVmax | II | 83% | 52% | | 84% | 50% | 0.70 (0.55–0.85) | $C^{(\alpha)}B^{(\beta)}$ |
| Koolen et al. (2014) | $^{18}$1F FDG-PET/CT | After $\frac{3}{8}$ of 1st cycle[#*] | ΔSUVmax | II | | | | | | 0.61 (0.33–0.89) | $B^{(\beta)}$ |
| Zucchini et al. (2013) | $^{18}$1F FDG-PET/CT | After 2 cycles | -50% ΔSUVmax | IV | 20% | 100% | | 33% | 100% | | $D^{(\alpha)}$ |
| Koolen et al. (2014) | $^{18}$1F FDG-PET/CT | After 1 cycle[#] | ΔSUVmax | II | | | | | | 0.59 (0.34–0.85) | $C^{(\beta)}$ |
| Koolen et al. (2013) | $^{18}$1F FDG-PET/CT | After 1 cycle[#] | ΔSUVmax | III | | | | | | 0.41 (0.16–0.67) | $C^{(\beta)}$ |
| Martoni et al. (2010) | $^{18}$1F FDG-PET/CT | After 2 cycles | -50% ΔSUVmax | IV | 17% | 100% | 29% | 17% | 100% | | $D^{(\alpha)}D^{(\gamma)}$ |
| Hatt et al. (2013) | $^{18}$1F FDG-PET/CT | After 2 cycles | Several imaging measurement parameters | IV | Use of different parameters did not improve predictive value of ΔSUVmax | | | | | | – |
| Loo et al. (2011) | DCE MRI [(2)] | After 1 cycle[#] | 30% size reduction | IV | Association between BRI and largest tumour diameter was significant (p = 0.05) | | | | | | – |
| **HER2-positive and ER-positive** | | | | | | | | | | | |
| Gebhart et al. (2013) | $^{18}$1F FDG-PET/CT | After 6 weeks | -25% ΔSUVmax | III | 59% | 80% | 62% | 24% | 95% | – | $C^{(\alpha)}B^{(\gamma)}$ |
| Gebhart et al. (2013) | $^{18}$1F FDG-PET/CT | After 2 weeks | -15% ΔSUVmax | III | 38% | 71% | 44%, | 20% | 86% | – | $D^{(\alpha)}C^{(\gamma)}$ |
| **HER2-positive and ER-negative** | | | | | | | | | | | |
| Gebhart et al. (2013) | $^{18}$1F FDG-PET/CT | After 2 weeks | -15% ΔSUVmax | III | 27% | 88% | 64% | 65% | 60% | – | $D^{(\alpha)}B^{(\gamma)}$ |
| Gebhart et al. (2013) | $^{18}$1F FDG-PET/CT | After 6 weeks | -25% ΔSUVmax | III | 18% | 76% | 54% | 59% | 33% | – | $D^{(\alpha)}C^{(\gamma)}$ |

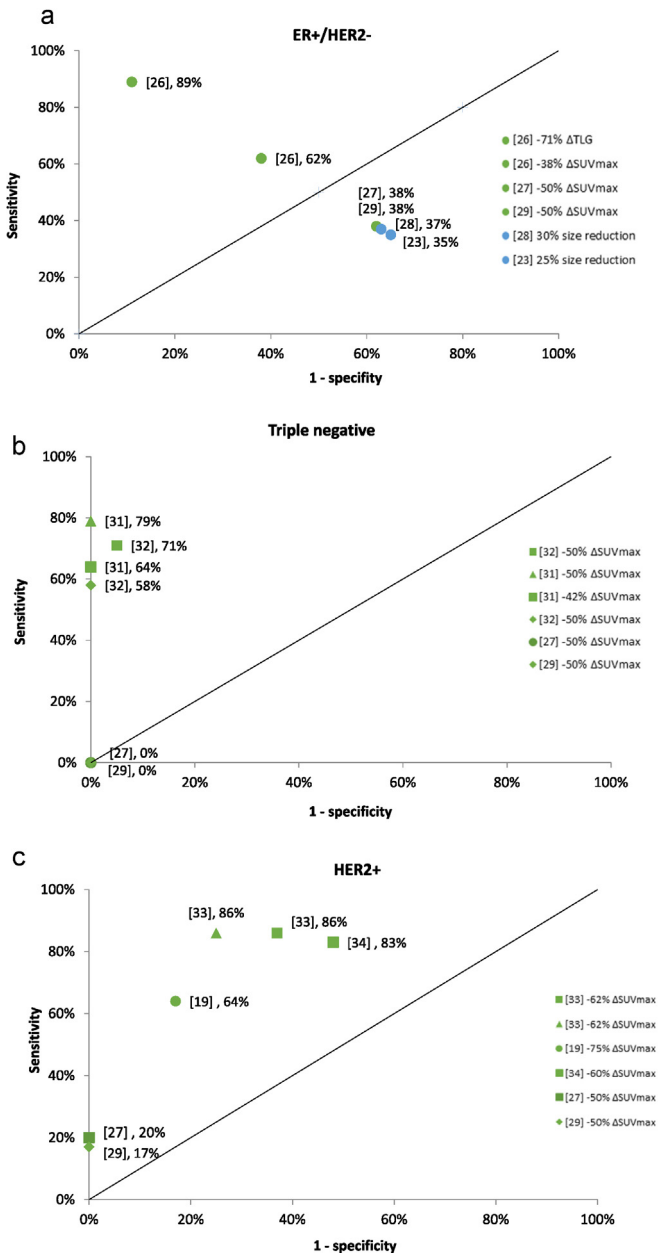


**Fig. 3.** The identified sensitivity and specificity values as shown in Table 5 are presented in an Area Under the Curve plot per subtype. The corresponding imaging settings (threshold, interval time, and pCR definition) are described in Table 5. Fig. 3a. shows sensitivity and specificity values in the subgroup: ER-positive and HER2-negative. The green dots show the values for $^{18}$FDG-PET/CT and the blue ones for MRI. Fig. 3b. shows sensitivity and specificity values in the triple negative subgroup. In this subgroup only sens/spec data was available for $^{18}$FDG-PET/CT (green). Fig. 3c shows sensitivity and specificity values in the HER2positive subgroup. In this subgroup only sens/spec data was found for $^{18}$FDG-PET/CT (green). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

(Charehbili et al., 2014). The final included MRI study did not report specific performance results, but showed no significant association between tumour size decrease and Breast Response Index (BRI; part of the NRI outcome measure (Rodenhuis et al., 2010))(p = 0.07) after 3 NAC cycles (Loo et al., 2011).

### 3.2.2. Triple negative

Eight studies assessed performance of $^{18}$F FDG-PET/CT and one of MRI. The use of $^{18}$F FDG-PET/CT showed AUC-ROC values of 0.76(CI 0.55–0.96) after 1 NAC cycle (Koolen et al., 2014), 0.87(CI 0.73–1.00) after 3 NAC cycles (Koolen et al., 2014) and 0.85(CI 0.68–1.00) also after 3 NAC cycles (Koolen et al., 2013). Two additional studies assessed performance of $^{18}$F FDG-PET/CT and showed high sensitivity of 71%–79%, specificity of 95%–100% and accuracy of **80%–85%** after 2 NAC cycles (Groheux et al., 2012; Groheux et al., 2014). Lowering the cut-off value from 50% to ≥-42% ΔSUVmax improved specificity to 100%, but decreased sensitivity to 58% and 64% (Groheux et al., 2012; Groheux et al., 2014). Two additional studies showed sensitivity of 0% and specificity of 100% after 2 NAC cycles since true or false non-responders were not discovered (Zucchini et al., 2013; Martoni et al., 2010). Of the two final studies, one study showed no significant association (p = 0.50) between ΔSUV and pCR after 1 NAC cycle (Humbert et al., 2012), and another showed no improvement (p > 0.05) in predictive value by using ΔTLG as imaging measurement parameter (Hatt et al., 2013). The only study assessing performance of MRI presented no specific performance results but showed a significant association between tumour size decrease and BRI (p < 0.001)(Loo et al., 2011).

### 3.2.3. HER2-positive

Eight studies assessed the performance of $^{18}$F FDG-PET/CT and 1 of MRI. The use of $^{18}$F FDG-PET/CT showed AUC-ROC values of 0.61(CI 0.33-0.89) after 3 administrations of the 1st NAC cycle(8 in total) (Koolen et al., 2014), 0.59(CI 0.34-0.85) after the 1st completed NAC cycle (Koolen et al., 2014) and 0.41(CI 0.16–0.67) after 1 NAC cycle (Koolen et al., 2013). Two studies also assessed the performance of $^{18}$F FDG-PET/CT and showed sensitivity of 17%- 20%, specificity of 100% (Zucchini et al., 2013; Martoni et al., 2010) and accuracy of **29%** after 2 NAC cycles (Martoni et al., 2010). Three other studies showed also performance results in terms of sensitivity and specificity of $^{18}$F FDG-PET/CT. The first study showed sensitivity of 86% and specificity of 75% after 2 NAC cycles and 86% and 63% also after 2 NAC cycles but using another pCR category (Table 5) (Groheux, Giacchetti et al., 2013). In this study the absolute level of residual SUVmax at PET2 showed even a higher accuracy (AUC = 0.91). The second study showed sensitivity and specificity of 83% and 53% after 1 NAC cycle (Humbert, Cochet et al., 2014) and the third, sensitivity, specificity and accuracy of 64%, 83% and **76%** also after 1 NAC cycle (Humbert et al., 2012). Using ΔTLG showed no improvement in predictive value compared to ΔSUVmax (Hatt et al., 2013). The MRI study presented a significant association with the BRI (p = 0.05)(after 1 NAC cycle) (Loo et al., 2011).

### 3.2.4. ER-positive/HER2-positive

One study assessed the performance of $^{18}$F FDG-PET/CT which showed sensitivity of 38%, specificity of 71%, accuracy of **44%,** NPV of 20% and PPV of 86% after 2 weeks, and improved results with a higher cut-off value after 6 weeks: sensitivity of 59%, specificity of 80%, accuracy of **62%**, NPV of 24% and PPV of 95% (Gebhart et al., 2013).

### 3.2.5. ER-negative/HER2-positive

One study showed sensitivity of 27%, specificity of 88%, accuracy of **64%**, NPV of 65% and PPV of 60% for the use of $^{18}$F FDG-PET/CT after 2 weeks and sensitivity of 18%, specificity of 76%, accuracy of **54%**, NPV of 59% and PPV of 33% after 6 weeks (Gebhart et al., 2013).

## 3.3. Within-study comparisons

Seven of the 15 included studies analysed imaging performance in more than one BC subtype, the full results are shown in Table 5, in this paragraph we highlight some of these findings.

Koolen et al. analysed the performance of $^{18}$F FDG-PET/CT in all three groups in which the best results were found in the ER-positive/HER2-negative and triple negative groups (Koolen et al., 2014; Koolen et al., 2013). Martoni et al. and Zucchini et al.

described also $^{18}$F FDG-PET/CT performance in all groups which showed low sensitivity/specificity values in all groups (Zucchini et al., 2013; Martoni et al., 2010) and Humbert et al. described performance of $^{18}$F FDG-PET/CT in TN and HER2-positive only, in which for the HER2-positive group a sufficient performance was found (Humbert et al., 2012). Hatt et al. described $^{18}$F FDG-PET/CT performance with several imaging measurement parameters in the three subgroups. Only in ER-positive/HER2-negative it showed improved performance when imaging measurement parameters ΔTLG and ΔMetabolic Active Tumour Volume were used (Hatt et al., 2013). Finally, MRI performance was described by Loo et al. in all groups, showing only a significant association in the triple negative group between BRI and tumour decrease ($P < 0.001$) (Loo et al., 2011).

### 3.4. Preferred imaging technique per subtype

We aimed to find the preferred imaging technique per subtype, however due to limited performance results for different imaging techniques within subtypes it was not possible to conclude on the preferred imaging technique.

## 4. Discussion

In view of the potential of response-guided NAC to improve breast cancer survival, we aimed to create an overview of current knowledge on imaging performance to monitor NAC according to breast cancer subtype.

Our results suggest that due to the differences in imaging performance across subtypes, personalizing the monitoring step of response-guided NAC based on these is of relevance. However after reviewing the 15 included articles, we revealed that there is a lack of evidence with enough statistical power to conclude on the preferred imaging technique per subtype. Although, we did identify studies reporting on the performance of MRI and $^{18}$F FDG-PET/CT specified to breast cancer subtypes, all studies were observational, showed a lot of inter study variability, and described only performance of one imaging modality. Thereby, our results should be seen as preliminary and thus be interpreted with caution. This information can nonetheless serve to pinpoint areas of further research.

In the ER-positive/HER2-negative subtype, the best performing technique was $^{18}$F FDG-PET/CT after 2 NAC cycles (Groheux, Hatt et al., 2013), while the use of DCE MRI was limited and insufficient (Loo et al., 2011; Rigter et al., 2013; Charehbili et al., 2014). Worth mentioning is that the performance of $^{18}$F FDG-PET/CT was even higher with the imaging measurement parameter ΔTLG and MATV, than with the standard ΔSUVmax (Groheux, Hatt et al., 2013). However, the performance results of $^{18}$F FDG-PET/CT are based on thresholds that were derived after analysis which might have led to overestimation of these performance results.

In TNBC, $^{18}$F FDG-PET/CT showed also a good performance (Koolen et al., 2014; Koolen et al., 2013; Groheux et al., 2012; Groheux et al., 2014), in which the best results were found after 2 NAC cycles using cut-off value ≥-50% ΔSUVmax (performance: $A^{(\alpha)}A^{(\gamma)}$) (Groheux et al., 2014). The use of MRI seems also promising in this subtype, as size decrease showed a correlation with BRI (Loo et al., 2011).

In the overall HER2-positive group, $^{18}$F FDG-PET/CT showed promising results (Humbert et al., 2012; Koolen et al., 2014; Groheux, Giacchetti et al., 2013; Humbert, Cochet et al., 2014), especially after 2 NAC cycles, using cut-off value ≥-62% ΔSUVmax (performance: $B^{(\alpha)}A^{(\beta)}B^{(\gamma)}$) (Groheux, Giacchetti et al., 2013). However the study of Gebhart et al., which splits this HER2-positive group by ER status, showed limited and insufficient performance results. This might be explained by the use of a lower cut-off value and a different monitoring interval compared to the other $^{18}$F FDG-PET/CT studies (Gebhart et al., 2013). MRI showed also in the overall HER2 positive group an association between tumour size decrease and BRI (Loo et al., 2011). Therefore, further investigation towards the performance of MRI during NAC in TNBC and HER2-positive breast cancer seems relevant.

Previous publications that described and reviewed literature on the assessment of response by imaging during NAC in specific subtypes were in line with our findings. For instance, Lobbes and colleagues showed that MRI was more accurate in HER2-positive tumours than in HER2-negative tumours (Lobbes et al., 2012). Humbert et al. and Groheux et al. presented a good performance of $^{18}$F FDG-PET/CT in HER2-positive breast cancer patients using the difference in SUV uptake (Groheux et al., 2015; Humbert, Cochet et al., 2015). Also in TNBC $^{18}$F FDG-PET/CT showed promising performance results by both ΔSUVmax and ΔTLG with AUC values of 0.86 and 0.88 respectively (Groheux et al., 2015) and an overall accuracy of 75% (Humbert, Riedinger et al., 2015). For $^{18}$F FDG-PET/CT, the potential of ΔTLG as an imaging measurement parameter was confirmed by other research groups, who showed its correlation with survival (Groheux et al., 2015; Ulaner et al., 2013). In addition the use of absolute values of SUVmax and SUVpeak instead of their difference was suggested for its better performance in predicting pCR compared to ΔSUVmax (Humbert, Cochet et al., 2014; Groheux et al., 2015; Groheux et al., 2016). Furthermore, besides $^{18}$F FDG-PET/CT and MRI, FES-PET/CT and DWI-MRI might be of interest for response prediction as well. Since, FES-PET/CT seems a useful tool for response prediction in ER-positive tumours (Currin et al., 2011), and DWI-MRI seems to provide complementary information to DCE-MRI (O'Flynn and deSouza, 2011). Both are being investigated in a trial (NCT02398773;NCT01564368). Finally, two reviews addressed also the importance of reaching consensus about early NAC evaluation: the first was already published in 2012 and focused on early prediction of pathologic response on NAC by MRI, which did not specify to breast cancer subtypes. They pointed at the importance of reaching consensus on pCR definitions and thresholds for response definition, which shows the lack of progress in standardizing methodology over the years (Marinovich et al., 2012). The second review, published in 2016, described a need for standardization when using $^{18}$F FDG PET/CT to evaluate response to NAC in BC patients on: monitoring interval, BC subtype and type of treatment (Groheux et al., 2016).

With regard to the response-guided NAC approach, we identified two studies. One RCT for ER-negative/HER2positive in which patients were being scanned by $^{18}$F FDG-PET after 1 cycle of NAC and bevacizumab was randomly added to the first-line treatment in non-responders(≤ −70% ΔSUVmax) in a 2:1 ratio (Coudert et al., 2014). This study was however excluded from our analysis for using PET. The second study(uncontrolled) used MRI in ER-positive/HER2-negative patients to guide treatment regimen. In patients considered as non-responders a switch from AC to DC showed improved tumour size reduction (Rigter et al., 2013). Since the NPV value in this study was 10%, only 10% of the non-responders were correctly identified. If the study had used $^{18}$F FDG-PET/CT instead, according to our results, NPV could have been increased to 31% which consequently would have increased therapeutic response (Groheux, Hatt et al., 2013). This is of course, assuming that 1)the switch to non-cross resistant therapy DC would be beneficial in non-responders to AC, 2)pCR would correlate to survival in this subtype, and 3)the optimal way to predict therapeutic response had been chosen. This hypothetical scenario illustrates that improved effectiveness of the response-guided NAC approach can be achieved with improved imaging performance, more effective treatments or the combination of both.

This review included a few studies, mainly underpowered, and of heterogeneous study designs and outcome measures. Variability mainly occurred due to differences in interval time between

imaging at baseline and monitoring, cut-off values to define treatment response, and pCR definitions, which are consequence of the lack of consensus on imaging settings and protocols. As we were aware of this, and of its possible influence on our results, we carefully described study differences in our results section. Regarding the cut-off values it should be mentioned that there is a difference in its use: some are set upfront which enable analysing sensitivity and specificity of this imaging modality under this cut-off value, when others were derived after analysis in order to identify the most optimal sensitivity and specificity which certainly could overestimate accuracy. Another issue is the identification of mainly studies examining $^{18}$F FDG-PET/CT and some on MRI. Although the performance of MRI in pathologic response prediction is often investigated, we were unable to include these articles, since they combined performance results of response assessment during and after NAC in their analysis. The lack of results on MRI in the majority of the subtypes made it impossible to compare its performance to $^{18}$F FDG-PET/CT and consequently to conclude on the preferred imaging technique per subtype. The final issues that should be discussed are (1) the inclusion of studies only describing performance results according to one receptor status, as it is known that performance could be affected by the other unknown receptor status. (2) The quality assessment, since we included studies either describing a very small sample size (<20) or studies in which a switch in neoadjuvant treatment regimen has been made, could have biased the performance results and finally (3) the way pathologic complete response was used in the identified articles, since the value of reaching a pCR is different according to subtypes: firstly, some identified articles used a pCR definition for all subtypes (mainly chosen afterwards), some articles describe different pCR definitions per subtype. For instance in ER-positive/HER2-negative subtype is complete pathologic response rare, therefore complete and partial pathologic response are mainly being pooled. Secondly, the three identified pCR categories are not comparable because of varying inclusion of: response in both axillary lymph nodes and breast tissue, and absence of DCIS and invasive tumour cells (Table 1).

Besides, in the ER-positive/HER2-negative group we did not differentiate into luminal A and B tumours, despite knowing that in luminal A tumours pCR does not correlate with survival (Von Minckwitz et al., 2012). Therefore, our conclusions for this subtype may be unlikely. Nonetheless, they serve to illustrate the urgency to reach consensus for a reliable alternative for pCR in this subgroup.

A major limitation of an analysis as presented here, is the small fraction and the insufficient statistical power of the included studies. It shows however, what is needed to decide on the most effective imaging technique per subtype; consensus on several aspects that affect study comparability. Specifically, on 1)the definition for pathologic response, 2)the thresholds to define complete-, near-, partial-, or no- response during NAC in both $^{18}$1F FDG-PET/CT and MRI, 3)the required interval time between baseline and response monitoring, per subtype and imaging technique, and 4)imaging settings. Only then, meaningful well-designed studies which account for various breast cancer subtypes and imaging techniques can be conducted. Whereupon, RCTs such as the AVATACXER trial (Coudert et al., 2014), which mimics the response-guided NAC approach, could be set to also inform on suitable treatment switches. Further, we suggest to conduct further research to: 1)less investigated techniques such as FES-FDG/PET and DWI-MRI, 2)potential predictive biomarkers that could further personalize the response-guided NAC approach i.e. Ki67 and P53 and 3)the association between NAC treatments and imaging performance. Finally, a cost-effectiveness analysis could be interesting to explore the health-economic consequences of the various scenarios for this response-guided NAC approach.

This literature review is unique in the way that it focuses on imaging performance of NAC monitoring specified to breast cancer subtypes. We conclude that current evidence is too low to draw on subtype-specific imaging recommendations, and that these can only occur when consensus on imaging settings and work regulations are reached. Thus, further research on these are necessary to eventually build protocols and use them to conceive comparable study outcomes.

## Competing interests

The authors declare that they have no conflict of interest.

## Funding

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

## Author's contributions

AMC, VR, WvH and ML designed and conceptualized the study together, AMC, VR and ML gathered, selected and interpreted the data and drafted the manuscript. WvH, GS, JW and MS helped drafting the manuscript and critically reviewed it. All authors read and approved the final manuscript.

## Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.critrevonc.2017.02.014.

## References

Bossuyt, P.M., Reitsma, J.B., Bruns, D.E., Gatsonis, C.A., Glasziou, P.P., Irwig, L.M., et al., 2007. Towards complete and accurate reporting of studies of diagnostic accuracy: the STARD initiative. Vet. Clin. Pathol.

Charehbili, A., Wasser, M.N., Smit, V.T.H.B.M., Putter, H., van Leeuwen-Stok, A.E., Meershoek-Klein Kranenbarg, W.M., et al., 2014. Accuracy of MRI for treatment response assessment after taxane- and anthracycline-based neoadjuvant chemotherapy in HER2-negative breast cancer. Eur. J. Surg Oncol. Engl. 40 (October (10)).

Cheng, X., Li, Y., Liu, B., Xu, Z., Bao, L., Wang, J., 2012. 18F-FDG PET/CT and PET for evaluation of pathological response to neoadjuvant chemotherapy in breast cancer: a meta-analysis. Acta Radiol. 53 (6).

Chevallier, B., Roche, H., Olivier, J.P., Chollet, P., Hurteloup, P., 1993. Inflammatory breast cancer. Pilot study of intensive induction chemotherapy (FEC-HD) results in a high histologic response rate. Am. J. Clin. Oncol. 16 (3).

Chollet, P., Amat, S., Cure, H., de Latour, M., Le Bouedec, G., Mouret-Reynier, M.-A., et al., 2002. Prognostic significance of a complete pathological response after induction chemotherapy in operable breast cancer. Br. J. Cancer 86 (7).

Coudert, B., Pierga, J.-Y., Mouret-Reynier, M.-A., Kerrou, K., Ferrero, J.-M., Petit, T., et al., 2014. Use of [(18)F]-FDG PET to predict response to neoadjuvant trastuzumab and docetaxel in patients with HER2-positive breast cancer, and addition of bevacizumab to neoadjuvant trastuzumab and docetaxel in [(18)F]-FDG PET-predicted non-responders (AVATAXHER). Lancet Oncol. Engl. 15 (December (13)).

Currin, E., Linden, H.M., Mankoff, D.A., 2011. Predicting breast cancer endocrine responsiveness using molecular imaging. Curr. Breast Cancer Rep. 3 (4).

Fisher, B., Brown, A., Mamounas, E., Wieand, S., Robidoux, A., Margolese, R.G., et al., 1997. Effect of preoperative chemotherapy on local-regional disease in women with operable breast cancer: findings from National Surgical Adjuvant Breast and Bowel Project B-18. J. Clin. Oncol.

Gebhart, G., Gamez, C., Holmes, E., Robles, J., Garcia, C., Cortes, M., et al., 2013. 18F-FDG PET/CT for early prediction of response to neoadjuvant lapatinib, trastuzumab, and their combination in HER2-positive breast cancer: results from Neo-ALTTO. J. Nucl. Med. 54.

Gralow, J.R., Burstein, H.J., Wood, W., Hortobagyi, G.N., Gianni, L., Von Minckwitz, G., et al., 2008. Preoperative therapy in invasive breast cancer: pathologic assessment and systemic therapy issues in operable disease. J. Clin. Oncol. 26.

Groheux, D., Hindie, E., Giacchetti, S., Delord, M., Hamy, A.-S., de Roquancourt, A., et al., 2012. Triple-negative breast cancer: early assessment with 18F-FDG PET/CT during neoadjuvant chemotherapy identifies patients who are unlikely to achieve a pathologic complete response and are at a high risk of early relapse. J. Nucl. Med. U. S. 53 (February (2)).

Groheux, D., Hatt, M., Hindié, E., Giacchetti, S., De Cremoux, P., Lehmann-Che, J., et al., 2013. Estrogen receptor-positive/human epidermal growth factor receptor 2-negative breast tumors: early prediction of chemosensitivity with

18F-fluorodeoxyglucose positron emission tomography/computed tomography during neoadjuvant chemotherapy. Cancer 119 (11).
Groheux, D., Giacchetti, S., Hatt, M., Marty, M., Vercellino, L., de Roquancourt, A., et al., 2013. HER2-overexpressing breast cancer: FDG uptake after two cycles of chemotherapy predicts the outcome of neoadjuvant treatment. Br. J. Cancer 109 (5).
Groheux, D., Hindie, E., Giacchetti, S., Hamy, A.-S., Berger, F., Merlet, P., et al., 2014. Early assessment with 18F-fluorodeoxyglucose positron emission tomography/computed tomography can help predict the outcome of neoadjuvant chemotherapy in triple negative breast cancer. Eur. J. Cancer Engl. 50 (11).
Groheux, D., Majdoub, M., Sanna, A., de Cremoux, P., Hindié, E., Giacchetti, S., et al., 2015. Early metabolic response to neoadjuvant treatment: FDG PET/CT criteria according to Breast cancer subtype. Radiol. Radiol. Soc. N. Am. 27 (April).
Groheux, D., Mankoff, D., Espié, M., Hindié, E., 2016. 18F-FDG PET/CT in the early prediction of pathological response in aggressive subtypes of breast cancer: review of the literature and recommendations for use in clinical trials. Eur. J. Nucl. Med. Mol. Imaging 43 (5).
Hamisa, M., Dabess, N., Yosef, R., Zakeria, F., Hammed, Q., 2015. Role of breast ultrasound, mammography, magnetic resonance imaging and diffusion weighted imaging in predicting pathologic response of breast cancer after neoadjuvant chemotherapy. Egypt J. Radiol. Nucl. Med. 46 (1).
Hatt, M., Groheux, D., Martineau, A., Espié, M., Hindié, E., Giacchetti, S., et al., 2013. Comparison between 18F-FDG PET image-derived indices for early prediction of response to neoadjuvant chemotherapy in breast cancer. J. Nucl. Med. 54 (3).
Hayashi, Y., Takei, H., Nozu, S., Tochigi, Y., Ichikawa, A., Kobayashi, N., et al., 2013. Analysis of complete response by MRI following neoadjuvant chemotherapy predicts pathological tumor responses differently for molecular subtypes of breast cancer. Oncol. Lett. 5 (January (1)).
Humbert, O., Berriolo-Riedinger, A., Riedinger, J.M., Coudert, B., Arnould, L., Cochet, A., et al., 2012. Changes in 18F-FDG tumor metabolism after a first course of neoadjuvant chemotherapy in breast cancer: influence of tumor subtypes. Ann. Oncol. Engl. 23 (October (10)).
Humbert, O., Cochet, A., Riedinger, J.-M., Berriolo-Riedinger, A., Arnould, L., Coudert, B., et al., 2014. HER2-positive breast cancer: (1)(8)F-FDG PET for early prediction of response to trastuzumab plus taxane-based neoadjuvant chemotherapy. Eur. J. Nucl. Med. Mol. Imaging Germany 41 (August (8)).
Humbert, O., Berriolo-Riedinger, A., Cochet, A., Gauthier, M., Charon-Barra, C., Guiu, S., et al., 2014. Prognostic relevance at 5 years of the early monitoring of neoadjuvant chemotherapy using 18F-FDG PET in luminal HER2-negative breast cancer. Eur. J. Nucl. Med. Mol. Imaging.
Humbert, O., Cochet, A., Coudert, B., Berriolo-Riedinger, A., Kanoun, S., Brunotte, F., et al., 2015. Role of positron emission tomography for the monitoring of response to therapy in breast cancer. Oncologist 20 (February (2)).
Humbert, O., Riedinger, J.-M., Charon-Barra, C., Berriolo-Riedinger, A., Desmoulins, I., Lorgis, V., et al., 2015. Identification of biomarkers including 18FDG-PET/CT for early prediction of response to neoadjuvant chemotherapy in triple negative breast cancer. Clin. Cancer Res. 30 (June).
Kaufmann, M., Hortobagyi, G.N., Goldhirsch, A., Scholl, S., Makris, A., Valagussa, P., et al., 2006. Recommendations from an international expert panel on the use of neoadjuvant (primary) systemic treatment of operable breast cancer: an update. J. Clin. Oncol. 24 (12).
Ko, E.S., Han, B.-K., Kim, R.B., Ko, E.Y., Shin, J.H., Hahn, S.Y., et al., 2013. Analysis of factors that influence the accuracy of magnetic resonance imaging for predicting response after neoadjuvant chemotherapy in locally advanced breast cancer. Ann. Surg. Oncol. U. S. 20 (August (8)).
Koolen, B.B., Pengel, K.E., Wesseling, J., Vogel W, V., Vrancken Peeters, M.-J.T.F.D., Vincent, A.D., et al., 2013. FDG PET/CT during neoadjuvant chemotherapy may predict response in ER-positive/HER2-negative and triple negative, but not in HER2-positive breast cancer. Breast 22 (5).
Koolen, B.B., Pengel, K.E., Wesseling, J., Vogel, W.V., Vrancken Peeters, M.-J.T.F.D., Vincent, A.D., et al., 2014. Sequential 8)F-FDG PET/CT for early prediction of complete pathological response in breast and axilla during neoadjuvant chemotherapy. Eur. J. Nucl. Med. Mol. Imaging Germany 41 (January(1)).
Lobbes, M., Prevos, R., Smidt, M., 2012. Response monitoring of breast cancer patients receiving neoadjuvant chemotherapy using breast MRI —a review of current knowledge. J. Cancer Ther. Res. 1 (34).
Londero, V., Bazzocchi, M., Del, Frate C., Puglisi, F., Di, Loreto C., Francescutti, G., et al., 2004. Locally advanced breast cancer: comparison of mammography, sonography and MR imaging in evaluation of residual disease in women receiving neoadjuvant chemotherapy. Eur. Radiol. 14 (8).
Loo, C.E., Straver, M.E., Rodenhuis, S., Muller, S.H., Wesseling, J., Vrancken Peeters, M.J.T.F.D., et al., 2011. Magnetic resonance imaging response monitoring of breast cancer during neoadjuvant chemotherapy: relevance of breast cancer subtype. J. Clin. Oncol. 29.
Marinovich, M.L., Sardanelli, F., Ciatto, S., Mamounas, E., Brennan, M., Macaskill, P., et al., 2012. Early prediction of pathologic response to neoadjuvant therapy in breast cancer: systematic review of the accuracy of MRI. Breast 21 (5).
Martoni, A.A., Zamagni, C., Quercia, S., Rosati, M., Cacciari, N., Bernardi, A., et al., 2010. Early (18)F-2-fluoro-2-deoxy-d-glucose positron emission tomography may identify a subset of patients with estrogen receptor-positive breast cancer who will not respond optimally to preoperative chemotherapy. Cancer U. S. 116 (February (4)).
Mauri, D., Pavlidis, N., Ioannidis, J.P.A., 2005. Neoadjuvant versus adjuvant systemic treatment in breast cancer: a meta-analysis. J. Natl. Cancer Inst. 97.
O'Flynn, E.A., deSouza, N.M., 2011. Functional magnetic resonance: biomarkers of response in breast cancer. Breast Cancer Res. 13 (1), 204, http://dx.doi.org/10.1186/bcr2815.
Ogston, K.N., Miller, I.D., Payne, S., Hutcheon, A.W., Sarkar, T.K., Smith, I., et al., 2003. A new histological grading system to assess response of breast cancers to primary chemotherapy: prognostic significance and survival. Breast 12 (5).
Rigter, L.S., Loo, C.E., Linn, S.C., Sonke, G.S., van Werkhoven, E., Lips, E.H., et al., 2013. Neoadjuvant chemotherapy adaptation and serial MRI response monitoring in ER-positive HER2-negative breast cancer. Br. J. Cancer Engl. 109 (December (12)).
Rodenhuis, S., Mandjes, I.A.M., Wesseling, J., van de Vijver, M.J., Peeters, M.-J.T.D.F.V., Sonke, G.S., et al., 2010. A simple system for grading the response of breast cancer to neoadjuvant chemotherapy. Ann. Oncol. 21 (3).
Sataloff, D.M., Mason, B.A., Prestipino, A.J., Seinige, U.L., Lieber, C.P., Baloch, Z., 1995. Pathologic response to induction chemotherapy in locally advanced carcinoma of the breast: a determinant of outcome. J. Am. Coll. Surg. 180 (3).
Symmans, W.F., Peintinger, F., Hatzis, C., Rajan, R., Kuerer, H., Valero, V., et al., 2007. Measurement of residual breast cancer burden to predict survival after neoadjuvant chemotherapy. J. Clin. Oncol. 25 (28).
Ulaner, G.A., Eaton, A., Morris, P.G., Lilienstein, J., Jhaveri, K., Patil, S., et al., 2013. Prognostic value of quantitative fluorodeoxyglucose measurements in newly diagnosed metastatic breast cancer. Cancer Med. 2 (5).
van der Hage, J.A., van de Velde, C.J., Julien, J.P., Tubiana-Hulin, M., Vandervelden, C., Duchateau, L., 2001. Preoperative chemotherapy in primary operable breast cancer: results from the European Organization for Research and Treatment of Cancer trial 10902. J. Clin. Oncol.
von Minckwitz, G., Blohmer, J.U., Costa, S.D., Denkert, C., Eidtmann, H., Eiermann, W., et al., 2013. Response-guided neoadjuvant chemotherapy for breast cancer. J. Clin. Oncol. 31.
Von Minckwitz, G., Untch, M., Blohmer, J.U., Costa, S.D., Eidtmann, H., Fasching, P.A., et al., 2012. Definition and impact of pathologic complete response on prognosis after neoadjuvant chemotherapy in various intrinsic breast cancer subtypes. J. Clin. Oncol. 30 (15).
WHO, 2012. Breast Cancer Estimated Incidence, Mortality and Prevalence Worldwide in 2012 [Internet] (cited 2015 Nov 11. Available from) http://globocan.iarc.fr/Pages/fact_sheets_cancer.aspx.
Wu, L.-M., Hu, J.-N., Gu, H.-Y., Hua, J., Chen, J., Xu J-, R., 2012. 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. 135 (1).
Yeh, E., Slanetz, P., Kopans, D.B., Rafferty, E., Georgian-Smith, D., Moy, L., et al., 2005. Prospective comparison of mammography, sonography, and MRI in patients undergoing neoadjuvant chemotherapy for palpable breast cancer. Am. J. Roentgenol. 184 (3).
Zucchini, G., Quercia, S., Zamagni, C., Santini, D., Taffurelli, M., Fanti, S., et al., 2013. Potential utility of early metabolic response byfluoro-2-deoxy-D-glucose-positron emission tomography/computed tomography in a selected group of breast cancer patients receiving preoperative chemotherapy. Eur. J. Cancer 49 (7), 18F–182.