



**HAL**  
open science

# The use of FDG-PET in assessing axillary lymph node status in breast cancer: a systematic review and meta-analysis of the literature

Rebecca Peare, R. T. Staff, S. D. Heys

## ► To cite this version:

Rebecca Peare, R. T. Staff, S. D. Heys. The use of FDG-PET in assessing axillary lymph node status in breast cancer: a systematic review and meta-analysis of the literature. *Breast Cancer Research and Treatment*, 2010, 123 (1), pp.281-290. 10.1007/s10549-010-0771-9 . hal-00535444

**HAL Id: hal-00535444**

**<https://hal.science/hal-00535444>**

Submitted on 11 Nov 2010

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

## The use of FDG-PET in assessing axillary lymph node status in breast cancer: a systematic review and meta-analysis of the literature

Rebecca Peare · R. T. Staff · S. D. Heys

Received: 15 April 2009 / Accepted: 22 January 2010 / Published online: 7 February 2010  
© Springer Science+Business Media, LLC. 2010

**Abstract** Axillary lymph node status is the most powerful prognostic indicator in patients with breast cancer. FDG-PET has been suggested as a non-invasive method of staging the axilla. The aim of this study was to review and aggregate all studies that measured the performance of FDG-PET in patients with breast cancer, using surgically obtained axillary histology as a reference, in a meta-analysis. A systematic review of the literature was performed and data extracted from all eligible studies. These were then analysed using meta-analysis software and summary receiver operating characteristic (SROC) curves were plotted for the aggregate data. The data was then tested to determine which parameters impacted on the sensitivity and specificity of the studies. Sensitivities ranging from 20 to 100% and specificities ranging from 65 to 100% have been reported. An aggregated ROC analysis found an area under the curve of 0.95 (95% CI 0.91–0.97) and a  $Q^*$  value of 0.89 (95% CI 0.85–0.92) in a total of 25 studies involving 2,460 patients. The AUC and  $Q^*$  values indicated little difference between the compared study characteristics. The performance of the technique currently remains below, which is required to replace assessment of axillary node status by surgical biopsy and histological assessment. However, sensitivity and specificity are high and FDG-PET may have a role to play under particular circumstances. Moreover, the additional benefit of an

assessment of distal metastatic spread provided by FDG-PET requires further investigation.

**Keywords** Breast cancer · PET · Axilla · Staging

### Introduction

Although a number of prognostic indicators have been identified in patients with breast cancer, the most powerful of all in terms of identifying patients with a poorer prognosis remains histological assessment of axillary lymph node involvement with tumour [1, 2]. At the present time, in all patients with invasive breast cancer, histological assessment of the patient's axillary node(s) is recommended by surgically removing a lymph node(s) by one of three standard procedures; sentinel lymph node biopsy (SNB), axillary sample and axillary clearance. These procedures require a period of hospital admission, a surgical operation with its attendant financial costs and post-operative morbidity (physical and psychological) [3]. Recent interest has focused on SNB as a method to stage the axilla. This technique has resulted in a significant reduction in morbidity compared with axillary lymph node dissection (ALND) and clearance [4]. It is unsuccessful in 3–6% of patients and has a false negative rate of 2.1–2.7% when compared to axillary node dissection and clearance [5, 6]. The nature of SNB means that it has no false positives and consequently no patient is erroneously upstaged in terms of their tumour.

Surgical staging provides essential information in the management of breast cancer, but the disadvantages make it imperfect. A better staging tool would be one with a superior or equivalent diagnostic performance without the disadvantages of surgery. One technique that has attracted

---

R. Peare (✉) · R. T. Staff  
Department of Nuclear Medicine, Aberdeen Royal, Infirmary,  
Aberdeen AB25 2ZD, UK  
e-mail: r.peare@abdn.ac.uk; r.peare@nhs.net

S. D. Heys  
University of Aberdeen, Aberdeen, UK

considerable interest as a possible non-invasive method of staging the axilla has been FDG-PET. Since the initial studies, reported some 20 years ago, further studies have reported a range of sensitivity ranging from 20 to 100% and specificity ranging from 65 to 100% [7–31]. At its best FDG-PET appears good enough to replace surgery in some situations, however, this predominately is not the case. To our knowledge, no study has carried out a systematic review and meta-analysis of FDG-PET in the staging of the axilla. The aim of our study is to review the literature, aggregate the results and to compare the acquisition protocols in order to identify best practice.

Although our focus is on PET it would be imprudent to look at its performance in isolation from other imaging modalities used in staging. It is common clinical practice in patients with breast cancer to carry out ultrasound of the axilla together with fine needle aspiration cytology of any suspicious axillary nodes that are identified (or palpable) prior to patients undergoing SNB [10]. Similarly MRI [11, 32], CT [33, 34], mammography [7, 11] and other non-PET radionuclides (MIBI) [25] have all been suggested as useful in the staging of the axilla. It would not be the aim to perform a complete review of all modalities at this stage, since an extensive number of studies have not been carried out in all modalities. We aim, in this analysis, to review studies where there has been a direct comparison of other staging techniques with PET.

Aggregating findings from imaging studies with and without quantitation is not a simple process due to the lack of homogeneity. Summary receiver operating characteristics (SROC) analysis is a statistical technique that can be applied to meta-analysis of imaging tests. The technique overcomes the limitations associated with simple pooling of sensitivities and specificities of published studies [35]. In this study, we apply this approach to summarise published data describing performance using the area under the ROC curve (AUC) and the point where the sensitivity equals the specificity on the ROC curve ( $Q^*$ ), which are common figures of merit for evaluating performance.

## Method

Data sources consisted of Ovid MEDLINE<sup>®</sup> 1950 to June Week 2 2009, for which we used the exploded MeSH heading ‘breast neoplasm’ combined with the exploded MeSHs ‘positron emission tomography’ or ‘Tomography, emission-computed’ or ‘Fluorodeoxyglucose F18’ (1,067 articles identified). These titles were searched for the following keywords: axilla, staging, nodes, nodal (375 articles identified). Abstracts were then reviewed by one author (R.P.) and all potentially relevant studies retrieved to assess eligibility for inclusion (45 studies). All potentially eligible

studies were reviewed by R.P. and studies not meeting the eligibility criteria were excluded. Studies with questionable eligibility were reviewed by another author (R.S.) and a consensus reached. The eligibility criteria were as follows. We excluded: studies not in English, small studies (<20 patients), studies using gamma camera PET only, studies where the results were included in a follow-up study (i.e. we excluded pilot studies to avoid using data twice), studies that did not perform ALND or SNB and used follow-up or FNAC, etc. for gold standard. Where patients that did not have ALND or SNB could be extracted from the data, the study was included but only the data for the ALND or SNB patients used. Similarly, studies which contained patients with benign tumours were included, but the benign patient data was excluded from the analysis. We also excluded one potentially eligible study as the FP/TN results could not be extracted [36]. Twenty-five studies were found to satisfy these criteria.

The data extraction process was completed independently by two authors (R.P. and R.S.) discrepancies were settled by consensus. All discrepancies were settled this way and it was not necessary to use a third reviewer. TP, TN, FN and FP values were extracted from the data given. If TP, TN, FN and FP figures were not explicitly given, these were calculated from the available data, e.g. sensitivity, specificity, PPV and NPV. If a range of values was given (e.g. if a range of different SUV cut-offs were used), then the value reported as ‘optimum’ by the authors was used. If data from different readers was given then the combined result as reported by the authors was used. If the resolution was not reported in the article then previous articles from the same group quoting resolution were sought [37]. If this could not be found then resolution measurements from other groups using the same system, or articles reporting the performance of the system were used [38–40]. A qualitative study was defined as the one that used purely visual analysis to determine if a result was positive or negative. A quantitative study was defined as the one that used an SUV cut-off as the primary determination of positivity. A semi-quantitative study was defined as the one that used quantification of any kind, but interpreted it in conjunction with visual analysis.

All data available regarding the type of PET system, including whether it was PET/CT or stand-alone PET, its resolution, use of attenuation correction, time to imaging, administered activity, scanning position, and whether quantitative assessment was used, was collated either directly from the article or inferred based on the type of system used. Sensitivities, specificities and diagnostic odds ratios (DOR) for each study were calculated. The *DOR* expresses how much greater the odds of having the disease are for the people with a positive test result than for the people with a negative test result. It is a single measure of

diagnostic test performance that combines both likelihood ratios [41].

The data was entered into meta-analysis software [41] and analysis performed using a weighted least squares method, weighting by study size and using a restricted maximum likelihood technique. The results of this analysis were used to plot SROC curves. The AUC were found using the trapezoidal method as described previously [42]. The  $Q^*$  value represents the point on the SROC curve, where the sensitivity equals the specificity. AUC and  $Q^*$  are figures of merit used to describe ROC curves.

## Results

A summary of the studies that satisfied the inclusion criteria detailed above are shown in Table 1, giving details of the key characteristics of the study. By definition all studies had staging the axilla as part of their aims. This was either stated as the primary aim of the study or as a parallel aim along with the assessment of the primary and comparison with other modalities. Inclusion and exclusion criteria for the studies were varied. In general, previous treatment/surgery was an exclusion criterion. Some studies define the primary tumour by size and/or type and where stated a significant proportion of the studies excluded patients with confirmed disease spread. The age ranges of the samples were similar; the studies were predominately prospective and involved consecutive recruitment. All studies performed ALND in a proportion of the patients and most PET data was analysed independently of other information by more than one reader. The predominate reason given by the authors for the false negative results was the presence of micro-metastases.

A summary of the technical variables and performance is shown in Table 2. The later studies, as expected, used PET/CT. The median time to imaging was 1 h, the median activity used was 370 MBq and the majority of studies imaged in the supine position. Analysis of the data was evenly split between quantitative and qualitative approaches and Filtered Back Projection was the predominate method of reconstruction. Table 2 shows the DOR for each study demonstrating the range of performance levels achieved with this technique.

The combined patient population was 2,460 with 703 true positives, 1,288 true negatives, 339 false negatives and 130 false positives. The AUC of the SROC curve (AUC) was 0.95 (95% CI 0.91–0.97) and the  $Q^*$  value was 0.89 (95% CI 0.85–0.92). The AUC is a figure of merit for ROC curves, 1 represents perfect performance and 0.5 represents random performance. A SROC curve weighted by the number of patients in the study is shown in Fig. 1.

The data set was then analysed by splitting the sample, where possible, by key study characteristics and comparing them using the SROC software. The sample was split into subsets with at least four studies and containing at least 200 patients. The study characteristics compared were: (i) resolution  $<5$  mm compared to resolution  $\geq 5$  mm (FWHM); (ii) studies using attenuation correction compared to studies without attenuation correction; (iii) studies using a prone imaging position compared to studies using a supine position; (iv) studies using only visual interpretation of the image compared to studies using a quantitative approach; (v) studies using PET/CT systems compared to studies that used PET only and (vi) studies that included SNB compared to studies that used ALND only. The results are given in Table 3. The AUC values indicated little difference between the compared characteristics. The  $Q^*$  value showed significant overlap between the 95% CI in each case. There is a trend towards the studies with the poorer resolution having a better AUC and  $Q^*$ . The same is true for studies that use CT and studies that included a SNB biopsy. This is predominantly the same group of studies, i.e. studies that had CT also had a better resolution and used SNB.

Those studies that compared different imaging techniques and clinical assessment with PET (/CT) using pathology as the gold standard are shown in Table 4. Data for MRI, mammography, MIBI and gamma camera PET is sparse. The general indication is that PET is superior to mammography, MIBI and gamma camera PET and is similar to MRI. Clinical comparisons and ultrasound comparisons allow a limited statistical comparison. The summary ROC curve AUC and  $Q^*$  for the clinical assessment was 0.78 (CI 0.38–0.96) and 0.83 (CI 0.73–0.90), respectively, for eight studies using 755 patients. Using the same sub group of studies and the PET performances an AUC of 0.94 (CI 0.86, 0.98) and a  $Q^*$  of 0.88 (CI 0.80, 0.94) were calculated. Similarly ultrasound produced an AUC of 0.90 (CI 0.79, 0.96) and a  $Q^*$  of 0.83 (CI 0.73, 0.90) using four studies and 348 patients. Using the same sub group of studies and the PET performances an AUC of 0.93 (CI 0.72, 0.98) and a  $Q^*$  of 0.87 (CI 0.67, 0.95) were calculated. In each case, the performance of the techniques was lower than the complete PET sample performance and the sub samples for each comparison, although, there is overlap between the CI for each estimate.

## Discussion

We have calculated the aggregated performance of FDG-PET (/CT) by examining 25 studies which involved a total of 2,460 patients. A  $Q^*$  (sensitivity = specificity) of 0.89 (95% CI 0.85–0.92) and an AUC of 0.95 (95% CI

**Table 1** Key characteristics of the included studies

Study	Year	Study aim <sup>a</sup>	Study population: Inclusion Exclusion	Age mean, median, (range), $\pm$ SD	Prospective design?	Sample consecutive recruitment?	Reference gold standard <sup>c</sup>	Read independently?	No. of readers	Reasons for false negatives <sup>d</sup>
Chae [7]	2009	Cf. sonography and mammography	BC, -ve ALN Clin <i>Micro-mets, isolated tumour cell cases</i>	48.6 (27–75)	N	Y	ALND and SNB	Y	1	
Taira [8]	2009	Screen for SNB	BC <i>Chemo/HT/RT, DCIS</i>	54.6 (21–82)	N	Y	ALND and/ or SNB	NS	2	Micro-mets
Fuster [9]	2008	Staging	New BC >3 cm <i>Chemo/RT, prior surgery, IBC</i>	57 $\pm$ 13	Y	Y	ANLD	Y	2	Micro-mets
Ueda [10]	2008	Cf. sonography	BC <i>Primary systemic therapy, excisional biopsy, distal mets</i>	57 (32–81)	Y	NS	ALND and/ or SNB	NS	2	
Yang [11]	2008	Staging cf. MRI, mammography and sonography	IBC	51 51 (25–78)	N	N	ANLD	NS	1	
Veronesi [12]	2007		BC, -ve ALN Clin <i>Chemo DCIS</i>	49 (24–79)	Y	Y	ALND and/ or SNB	Y	3	Micro-mets, small mets
Chung [13]	2006		Invasive BC (28–88)	54 (24–79)	N	N	ALND and/ or SNB	N	1	
Gil-Rendo [14]	2006		BC, -ve ALN Clin and US <i>Previous surgery or biopsy</i>	50.6 (24–87)	Y	Y	ALND and/ or SNB	Y	2	Micro-mets, fewer nodes, primary characteristics
Kumar [15]	2006		New BC -ve ALN Clin (32–79)	52 (32–79) $\pm$ 11	Y	Y	ALND and SNB	NS	NS	Micro-mets
Fehr [16]	2004		BC <3 cm -ve ALN Clin $\pm$ 10.8	56 $\pm$ 10.8	Y	NS	ALND and SNB	Y	2	Micro-mets
Lovrics [17]	2004		New BC <i>T1T1, T1V, multiple multicentric BC, IBC</i>	56.4 $\pm$ 11.1	Y	Y	ALND and SNB	Y	1	
Wahl [18]	2004		New invasive BC <i>RT, chemo</i>	52.3 52 (27–82)	Y	Y	ALND and/ or SNB	NS	3	Larger patients, small mets, fewer nodes
Zornoza [19]	2004		BC, -ve ALN Clin (25–74)	52.2 (25–74)	Y	Y	ALND and/ or SNB	Y	2	Micro-mets, primary characteristics

Table 1 continued

Study	Year	Study aim <sup>a</sup>	Study population: criteria <sup>b</sup> Inclusion Exclusion	Age mean, median, (range), $\pm$ SD	Prospective design?	Sample consecutive recruitment?	Reference gold standard <sup>c</sup>	Read independently?	No. of readers	Reasons for false negatives <sup>d</sup>
Guller [20]	2002		New BC, -ve ALN Clin	64.8 (47–88)	Y	Y	ALND and/ or SNB	Y	NS	Micro-mets
Van der Hoeven [21]	2002		New BC	58 $\pm$ 13	Y	NS	ALND and/ or SNB	Y	3	Fewer nodes, primary characteristics
Greco [22]	2001		T1/T2 BC, N0 or N1	54 (28–84)	Y	Y	ALND	Y	3	Micro-mets, diabetic
Schirmeister [23]	2001	Detection and staging of BC cf. other methods	Suspicious X-ray, US or clinical exam	56.8 (28–86)	Y	Y	ANLD	Y	2	Treatment
Ohta [24]	2000	cf. sonography	NS	50 (36–79)	NS	NS	ALND	NS	1	Primary tumour characteristics
Yutani [25]	2000	Detection and staging of BC cf. MIBI	Suspicion of BC	51 (25–86)	Y	Y	ALND	Y	2	Small mets
Yutani [26]	1999	Detection and staging of BC cf. gamma camera PET	Suspicion of BC, PET available	51 (32–78)	Y	Y	ALND	Y	2	
Smith [27]	1998		BC	67 (26–89)	Y	Y	ALND	Y	2	Position of patient
Adler [28]	1997		<i>Surgery or chemo, tumour &lt;5 mm</i>	(36–79)	Y	Y	ANLD	Y	2	Larger patients
Avril [29]	1996		New BC <i>Locally advanced</i>	49.9 (18–74) $\pm$ 10.3	Y	Y	ANLD	Y	2	Micro-mets, fewer nodes
Utech [30]	1996		New BC	59 (32–94)	Y	NS	ANLD	Y	3	No false negatives

Table 1 continued

Study	Year	Study aim <sup>a</sup>	Study population: criteria <sup>b</sup> Inclusion Exclusion	Age mean, median, (range), $\pm$ SD	Prospective design?	Sample consecutive recruitment?	Reference gold standard <sup>c</sup>	Read independently?	No. of readers	Reasons for false negatives <sup>d</sup>
Crowe [31]	1994	Detection and staging of BC cf. clinical and mammography	BC $\geq$ 1 cm	55 (35–79)	Y	Y	ANLD	N	2	Fewer nodes

Y yes, N no, NS not stated, *met*s metastases, *ALN* axillary lymph nodes, *US* ultrasound, *HT* hormone therapy, *RT* radio therapy, *BC* breast cancer, *DCIS* ductal carcinoma in situ, *IBC* inflammatory breast cancer, *Clin* clinical, *SND* sentinel lymph node dissection, *ALND* axillary lymph node dissection

1. By definition all studies had as part of their aims, to stage the axilla. 2. The majority of studies excluded pregnant women, under 18 and known diabetics as is standard in PET research trials. 3. Where the reference standard is ALND and SNB, ALND was performed regardless of SNB result, where it is ALND and/or SNB, ALND was not performed in patients with both SNB and PET negative results. 4. Suggested reasons for false negative results: micro-mets: studies indicated that PET is not sensitive enough to detect micro-metastases. Fewer nodes: studies suggested that if only one node was affected the sensitivity of PET to detect metastases was reduced. Smaller metastases: studies suggested that smaller metastases are more difficult to detect. Treatment: one study suggested an association between the use of neo-adjuvant chemotherapy and a reduced sensitivity. Position of the patient: one study suggested that patients positioned with arms by their sides made it difficult to differentiate level I axillary lymph nodes from breast tumours in the upper outer quadrant. Primary tumour characteristics: several studies found an association between the characteristics of the primary tumour such as tumour type and grade, SUV and avidity, and the sensitivity. Larger patients: some studies found that larger patients were associated with a decrease in sensitivity

0.91–0.97) were calculated. In radiological terms this would normally be considered a ‘good’ performance. In this clinical setting, however, the performance is not sufficiently good enough to replace surgical biopsy of the axillary lymph node(s) and histological evaluation to determine the presence or absence of metastatic disease in axillary lymph nodes.

It is clear from our results that FDG-PET (/CT) is not the panacea for staging breast cancer and in particular the axilla. It is difficult to assess study quality although certain features such as multiple readers, images independently read, consecutive recruitment, a prospective design and ALND as the gold standard would be qualities of merit. 18 (72%) of the studies in this analysis had four or more of these features. The AUC for these 18 studies was 0.95 (95% CI 0.91–0.97) and the  $Q^*$  was 0.90 (95% CI 0.85–0.93). Three studies (12%) had only one of these features, although in one of these cases the appropriate information was not given in the manuscript. Given these attributes the sample of studies was of acceptable quality. The distribution of study data points in our SROC curve (Fig. 1) was relatively even along our fitted ROC line indicating that a range of confidence thresholds were sampled and that the ROC line was a reasonable estimate of the techniques performance for a range of confidences.

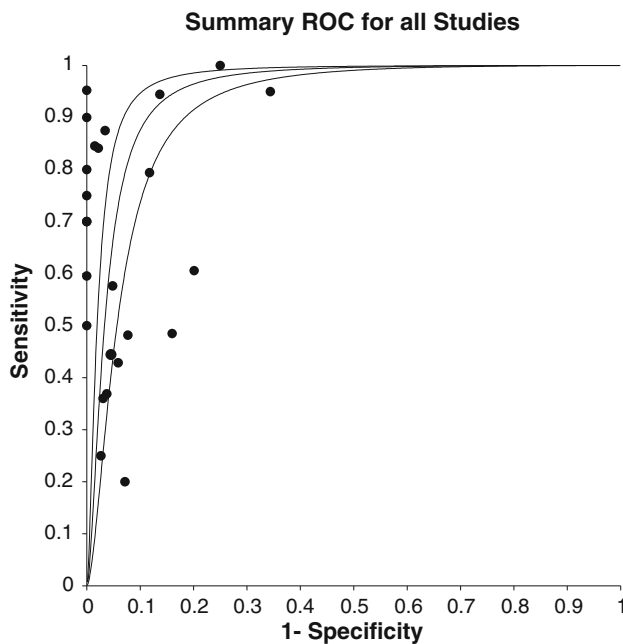
The reasons for this less than perfect performance are varied. Most studies report false negative results. The main reason given for these findings is the inability of PET, PET/CT to detect small metastatic deposits (micro-metastases). It is interesting to note that this is more often the case in the later, more modern, studies and may be a result of better pathological detection. This would represent a shift in the gold standard rather than a shift in the performance of the technique. Authors have also postulated that primary tumour characteristics, such as avidity, are associated with the metastatic avidity [14, 19, 21, 24]. Therefore, a primary with a low avidity will result in difficult to detect nodal metastases. Similarly, authors have also identified larger patients as a cause for false negative results [18]. This is a common obstacle faced in nuclear medicine, where obese patients produce images of inferior quality. Future studies may well consider adjusting protocols when imaging such patients, such as extending acquisition time. One study, Schirrmeister et al. [23], imaged a portion of patients after chemotherapy and concluded that imaging patients post therapy increased the number of false negatives. Similarly, studies also report false positives and have identified previous biopsy [15, 17] other tumours (e.g. non-Hodgkin lymphoma [20]) and infective and inflammatory conditions (a rabies vaccination [29], nodes affected by tuberculosis [14] and an acute infection of the contra lateral hand [43]) as the origins of these false positives. Therefore, patients who had prior chemotherapy, previous biopsy, other

**Table 2** Technical variables and performance of the included studies

Study	Year	Resolution (mm)	Attenuation correction	Time to imaging (min)	Administered activity (MBq)	Scanning position	Assessment method	Reconstruction method	N	TP	TN	FP	FN	Sensitivity	Specificity	DOR
Chae	2009	3.8	CT	60	185	Both	Quant	NS	108	16	63	12	17	0.48	0.84	4.94
Taira	2009	4.5	CT	60	210	Supine	Semi-quant	NS	92	13	60	5	14	0.48	0.92	11.14
Fuster	2008	4.5	CT	60	740	Supine	Quant	IT	52	14	32	0	6	0.70	1.00	145.00
Ueda	2008	6.5	CT	60	259	NS	Semi-quant	FBP	183	34	118	6	25	0.58	0.95	26.75
Yang	2008	6.1	CT	60–90	462	Supine	Semi-quant	NS	24	20	3	0	1	0.95	1.00	95.67
Veronesi	2007	3.8	T	45	371	Supine	Quant	NS	236	38	128	5	65	0.37	0.96	14.97
Chung	2006	5.6	T	100	481	Both	Quant	NS	60	25	18	0	17	0.60	1.00	53.91
Gil-Rendo	2006	4.5	T	40–60	370	Supine	Semi-quant	NS	275	120	131	2	22	0.85	0.98	357.27
Kumar	2006	5.6	T	60	364	NS	Qual	IT	80	16	42	2	20	0.44	0.95	16.80
Fehr	2004	3.8	T	40–50	650	Prone	Qual	NS	24	2	13	1	8	0.20	0.93	3.25
Lovrics	2004	6	N	45	185	NS	Semi-quant	NS	90	9	63	2	16	0.36	0.97	17.72
Wahl	2004	<5	T	50	629	NS	Semi-quant	FBP	308	66	159	40	43	0.61	0.80	6.10
Zornoza	2004	4.5	T	45	370	Supine	Semi-quant	NS	200	90	91	2	17	0.84	0.98	240.88
Guller	2002	5.4	N	90	350	Prone	Qual	NS	31	6	16	1	8	0.43	0.94	12.00
Van der Hoeven	2002	7	N	60	370	Supine	Semi-quant	FBP	70	8	37	1	24	0.25	0.97	12.33
Greco	2001	5.3	T	45–60	400	Supine	Qual	FBP	167	68	82	13	4	0.94	0.86	107.23
Schirmeister	2001	4.2	N	45–60	370	Prone	Qual	NS	85	27	45	6	7	0.79	0.88	28.93
Ohta	2000	5.4	T	45–60	300	Supine	Qual	NS	33	14	13	0	6	0.70	1.00	60.23
Yutani	2000	4	T	60	370	NS	Semi-quant	FBP	38	8	22	0	8	0.50	1.00	45.00
Yutani	1999	4	T	60	370	Supine	Semi-quant	FBP	26	8	16	0	2	0.80	1.00	112.20
Smith	1998	6	N	40	185	Prone	Qual	NS	45	14	28	1	2	0.88	0.97	196.00
Adler	1997	6	T	60	740	Supine	Semi-quant	FBP	52	19	21	11	1	0.95	0.66	36.27
Avril	1996	5	T	60–80	330	Prone	Semi-quant	FBP	37	15	17	0	5	0.75	1.00	98.64
Utech	1996	10.2	T	60	370	NS	Qual	FBP	124	44	60	20	0	1.00	0.75	262.66
Crowe	1994	5	T	40	518	NS	Semi-quant	FBP	20	9	10	0	1	0.90	1.00	133.00

*N* No AC, *T* transmission source, *CT* computed tomography, *NS* not stated, *FBP* Filtered Back Projection, *IT* iterative methods, *Quant* quantitative method of assessment, *Qual* qualitative method of assessment, *DOR* diagnostic odd ratio, *TP* true positive, *TN* true negative, *FP* false positive, *FN* false negative





**Fig. 1** The SROC for all studies included in the meta-analysis

tumours or infective and inflammatory conditions would not be suitable for this approach.

Taken collectively, the reasons given for the false negatives identify a limit to the sensitivity of PET (/CT). The limited number of comparison studies between PET (/CT) and other modalities, shown in Table 4, indicated that it has a similar performance to ultrasound and MRI and probably better than mammography, gamma camera PET and radionuclide imaging with MIBI. Given the expense of PET (/CT) and radiation dose these findings would indicate the ultrasound would be the modality of choice. However,

no single modality provides complete characterisation of the axilla in breast cancer and complementary information is gained for both structural and metabolic imaging. The current management challenge is to devise a management protocol that optimises all available modalities on an individual patient basis.

One possible management protocol would be to perform an FDG-PET (/CT) study prior to surgery if no other diagnostic test, such as ultrasound, has identified the axilla as positive for metastatic involvement. PET would then identify a group of patients not requiring SNB. These patients could proceed immediately to a more appropriate management of their axilla, i.e. axillary clearance without the need for the intervening step of SNB. This potential tool should also be seen in the context of the additional information that modern PET/CT provides with respect to the presence or absence of additional primary sites, detectable distant metastatic disease and internal mammary lymph gland involvement by tumour. Eight of the studies in this analysis included patients in whom distant disease was detected by PET, either in the internal mammary chain, distant metastases or synchronous tumours. In most cases, this would not have been detected by conventional staging procedures. One study, Fuster et al. [9], found that PET led to a change in the initial staging in 42% of patients. These advantages have been recently described in a study by Heusner et al. [36] who has suggested that whole body PET/CT can provide a ‘one stop’ staging examination for patients with breast cancer. If used in this way then the emphasis when staging the axilla would be on a low false positive rate, which would limit the number of unnecessary axillary clearances.

None of the imaging protocol characteristic comparisons indicated a superiority of one protocol aspect over another.

**Table 3** Comparison of selected study characteristic

	Number of studies	Number of patients	AUC	95% CI	$Q^*$	95% CI
Resolution <5 mm	11	1,444	0.91	0.62–0.98	0.84	0.59–0.95
Resolution $\geq$ 5 mm	14	1,016	0.95	0.93–0.96	0.90	0.87–0.92
AC	20	2,139	0.95	0.91–0.97	0.90	0.85–0.93
No AC	5	321	0.95	0.84–0.98	0.91	0.79–0.96
Prone or both	7	390	0.94	0.73–0.99	0.90	0.69–0.97
Supine only	11	1,291	0.95	0.90–0.98	0.90	0.84–0.93
Qualitative	7	544	0.93	0.89–0.95	0.88	0.85–0.92
Quantitative/semi	18	1,916	0.94	0.88–0.98	0.88	0.81–0.93
Qualitative/semi	21	2,004	0.95	0.91–0.97	0.89	0.85–0.93
Quantitative (SUV cut-off used)	4	456	0.53	0.07–0.95	0.53	0.06–0.95
PET/CT	5	459	0.92	0.68–0.99	0.86	0.64–0.95
PET only	20	2,001	0.95	0.92–0.97	0.90	0.86–0.93
SNB	13	1,757	0.91	0.65–0.98	0.85	0.62–0.95
ALND only	12	703	0.96	0.94–0.97	0.90	0.88–0.92

AUC area under the ROC curve, CI confidence interval

**Table 4** Comparison between PET and other techniques

Study	Year	PET		Clinical examination		Ultrasound		MRI		Mammography		MIBI		Gamma Camera PET	
		Sens	Spec	Sens	Spec	Sens	Spec	Sens	Spec	Sens	Spec	Sens	Spec	Sens	Spec
Chae	2009	0.48	0.84			0.52	0.83			0.33	0.96				
Ueda	2008	0.58	0.95			0.54	0.99								
Yang	2008	0.95	1.00			1	1	0.91	1	0.48	1				
Chung	2006	0.60	1.00	0.6	1										
Greco	2001	0.94	0.86	0.42	0.92										
Schirrmmeister	2001	0.79	0.88	0.41	0.96										
Ohta	2000	0.70	1.00	0.55	1	0.65	1								
Yutani	2000	0.50	1.00									0.38	1		
Yutani	1999	0.80	1.00											0.4	1
Smith	1998	0.88	0.97	0.57	0.90										
Avril	1996	0.75	1.00	0.58	0.85										
Crowe	1994	0.90	1.00	0.40	0.90										

*Sens* sensitivity, *Spec* specificity

This was surprising as it could be expected at modern PET/CT systems would be superior over PET only systems due to the additional anatomical information gained. This was not the case. The trend was in the opposite direction with the older systems having a superior AUC and  $Q^*$ . This may have been brought about by a shift in the gold standard described above. The results suggest that minor imaging protocol variations do not have a significant impact on performance.

However, the literature reviewed suggests that FDG-PET has potential as a complementary test. To truly assess the potential of this technique studies must move beyond a relatively simple assessment of accuracy in isolation of other relevant clinical information. Future studies should investigate the complementary role of FDG-PET (/CT) in breast cancer, combined with other modalities, in terms of its cost/benefit and impact on the ‘patient journey’.

## References

- Carter CL, Allen C, Henson DE (1989) Relation of tumor size, lymph node status, and survival in 24,740 breast cancer cases. *Cancer* 63:181–187
- Shek LLM, Godolphin W (1988) Model for breast cancer survival: relative prognostic roles of axillary nodal status, TNM stage, estrogen receptor concentration, and tumor necrosis. *Cancer Res* 48:5565–5569
- Ververs JM, Roumen RM, Vingerhoets AJ et al (2001) Risk, severity and predictors of physical and psychological morbidity after axillary lymph node dissection for breast cancer. *Eur J Cancer* 37:991–999
- Purushotham AD, Upponi S, Klevesath MB et al (2005) Morbidity after sentinel lymph node biopsy in primary breast cancer: results from a randomized controlled trial. *J Clin Oncol* 23:4312–4321
- Jakub JW, Pendas S, Reintgen DS (2003) Current status of sentinel lymph node mapping and biopsy: facts and controversies. *Oncologist* 8:59–68
- Martin RC II, Chagpar A, Scoggins CR et al (2005) Clinicopathologic factors associated with false-negative sentinel lymph node biopsy in breast cancer. *Ann Surg* 241:1005–1012
- Chae BJ, Bae JS, Kang BJ et al (2009) Positron emission tomography-computed tomography in the detection of axillary lymph node metastasis in patients with early stage breast cancer. *Jpn J Clin Oncol* 39:284–289
- Taira N, Ohsumi S, Takabatake D et al (2009) Determination of indication for sentinel lymph node biopsy in clinical node-negative breast cancer using preoperative 18F-fluorodeoxyglucose positron emission tomography/computed tomography fusion imaging. *Jpn J Clin Oncol* 39:16–21
- Fuster D, Duch J, Paredes P et al (2008) Preoperative staging of large primary breast cancer with [18F]Fluorodeoxyglucose positron emission tomography/computed tomography compared with conventional imaging procedures. *J Clin Oncol* 26(29):4746–4751
- Ueda S, Tsuda H, Asakawa H et al (2008) Utility of 18F-fluorodeoxyglucose emission tomography/computed tomography fusion imaging (18F-FDG PET/CT) in combination with ultrasonography for axillary staging in primary breast cancer. *BMC Cancer* 8:165
- Yang WT, Le-Petross HT, Macapinlac H et al (2008) Inflammatory breast cancer: PET/CT, MRI, mammography, and sonography findings. *Breast Cancer Res Treat* 109:417–426
- Veronesi U, De Cicco C, Galimberti VE et al (2007) A comparative study on the value of FDG-PET and sentinel node biopsy to identify occult axillary metastases. *Ann Oncol* 18:473–478
- Chung A, Liou D, Karlan S et al (2006) Preoperative FDG-PET for axillary metastases in patients with breast cancer. *Arch Surg* 141:783–788
- Gil-Rendo A, Zornoza G, Garcia-Velloso MJ et al (2006) Fluorodeoxyglucose positron emission tomography with sentinel lymph node biopsy for evaluation of axillary involvement in breast cancer. *Br J Surg* 93:707–712
- Kumar R, Zhuang H, Schnall M et al (2006) FDG PET positive lymph nodes are highly predictive of metastasis in breast cancer. *Nucl Med Commun* 27:231–236

16. Fehr MK, Hornung R, Varga Z et al (2004) Axillary staging using positron emission tomography in breast cancer patients qualifying for sentinel lymph node biopsy. *Breast J* 10:89–93
17. Lovrics PJ, Chen V, Coates G et al (2004) A prospective evaluation of positron emission tomography scanning, sentinel lymph node biopsy, and standard axillary dissection for axillary staging in patients with early stage breast cancer. *Ann Surg Oncol* 11:846–853
18. Wahl RL, Siegel BA, Coleman RE et al (2004) Prospective multicenter study of axillary nodal staging by positron emission tomography in breast cancer: a report of the staging breast cancer with PET Study Group. *J Clin Oncol* 22:277–285
19. Zornoza G, Garcia-Velloso MJ, Sola J et al (2004) 18F-FDG PET complemented with sentinel lymph node biopsy in the detection of axillary involvement in breast cancer. *Eur J Surg Oncol* 30:15–19
20. Guller U, Nitzsche EU, Schirp U et al (2002) Selective axillary surgery in breast cancer patients based on positron emission tomography with 18F-fluoro-2-deoxy-D-glucose: not yet!. *Breast Cancer Res Treat* 71:171–173
21. van der Hoeven JJ, Hoekstra OS, Comans EF et al (2002) Determinants of diagnostic performance of [F-18]fluorodeoxyglucose positron emission tomography for axillary staging in breast cancer. *Ann Surg* 236:619–624
22. Greco M, Crippa F, Agresti R et al (2001) Axillary lymph node staging in breast cancer by 2-fluoro-2-deoxy-D-glucose-positron emission tomography: clinical evaluation and alternative management. *J Natl Cancer Inst* 93:630–635
23. Schirmeister H, Kuhn T, Guhlmann A et al (2001) Fluorine-18 2-deoxy-2-fluoro-D-glucose PET in the preoperative staging of breast cancer: comparison with the standard staging procedures. *Eur J Nucl Med* 28:351–358
24. Ohta M, Tokuda Y, Saitoh Y et al (2000) Comparative efficacy of positron emission tomography and ultrasonography in preoperative evaluation of axillary lymph node metastases in breast cancer. *Breast Cancer* 7:99–103
25. Yutani K, Shiba E, Kusuoka H et al (2000) Comparison of FDG-PET with MIBI-SPECT in the detection of breast cancer and axillary lymph node metastasis. *J Comput Assist Tomogr* 24:274–280
26. Yutani K, Tatsumi M, Shiba E et al (1999) Comparison of dual-head coincidence gamma camera FDG imaging with FDG PET in detection of breast cancer and axillary lymph node metastasis. *J Nucl Med* 40:1003–1008
27. Smith IC, Ogston KN, Whitford P et al (1998) Staging of the axilla in breast cancer: accurate in vivo assessment using positron emission tomography with 2-(fluorine-18)-fluoro-2-deoxy-D-glucose. *Ann Surg* 228:220–227
28. Adler LP, Faulhaber PF, Schnur KC et al (1997) Axillary lymph node metastases: screening with [F-18]2-deoxy-2-fluoro-D-glucose (FDG) PET. *Radiology* 203:323–327
29. Avril N, Dose J, Janicke F et al (1996) Assessment of axillary lymph node involvement in breast cancer patients with positron emission tomography using radiolabeled 2-(fluorine-18)-fluoro-2-deoxy-D-glucose. *J Natl Cancer Inst* 88:1204–1209
30. Utech CI, Young CS, Winter PF (1996) Prospective evaluation of fluorine-18 fluorodeoxyglucose positron emission tomography in breast cancer for staging of the axilla related to surgery and immunocytochemistry. *Eur J Nucl Med* 23:1588–1593
31. Crowe JP Jr, Adler LP, Shenk RR et al (1994) Positron emission tomography and breast masses: comparison with clinical, mammographic, and pathological findings. *Ann Surg Oncol* 1:132–140
32. Murray AD, Staff RT, Redpath TW et al (2002) Dynamic contrast enhanced MRI of the axilla in women with breast cancer: comparison with pathology of excised nodes. *Br J Radiol* 75:220–228
33. Uematsu T, Sano M, Homma K (2001) In vitro high-resolution helical CT of small axillary lymph nodes in patients with breast cancer: correlation of CT and histology. *Am J Roentgenol* 176:1069–1074
34. March DE, Wechsler RJ, Kurtz AB et al (1991) CT-pathologic correlation of axillary lymph nodes in breast carcinoma. *J Comput Assist Tomogr* 15:440–444
35. Chappell FM, Raab GM, Wardlaw JM (2009) When are summary ROC curves appropriate for diagnostic meta-analyses? *Stat Med* 28:2653–2668
36. Heusner TA, Kuemmel S, Umutlu L et al (2008) Breast cancer staging in a single session: whole-body PET/CT mammography. *J Nucl Med* 49:1215–1222
37. Crippa F, Agresti R, Donne VD et al (1997) The contribution of positron emission tomography (PET) with 18F-fluorodeoxyglucose (FDG) in the preoperative detection of axillary metastases of breast cancer: the experience of the National Cancer Institute of Milan. *Tumori* 83:542–543
38. DeGrado TR, Turkington TG, Williams JJ et al (1994) Performance characteristics of a whole-body PET scanner. *J Nucl Med* 35:1398–1406
39. Mawlawi O, Podoloff DA, Kohlmyer S et al (2004) Performance characteristics of a newly developed PET/CT scanner using NEMA standards in 2D and 3D modes. *J Nucl Med* 45:1734–1742
40. Surti S, Karp JS (2004) Imaging characteristics of a 3-dimensional GSO whole-body PET camera. *J Nucl Med* 45:1040–1049
41. Zamora J, Abairra V, Muriel A et al (2006) Meta-DiSc: a software for meta-analysis of test accuracy data. *BMC Med Res Methodol* 6:31
42. Walter S (2002) Properties of the summary receiver operating characteristic (SROC) curve for diagnostic test data. *Stat Med* 21:1237–1256
43. Scheidhauer K, Scharl A, Pietrzyk U et al (1996) Qualitative [18F]FDG positron emission tomography in primary breast cancer: clinical relevance and practicability. *Eur J Nucl Med* 23:618–623