Preoperative FDG-PET for Axillary Metastases in Patients With Breast Cancer

Alice Chung, MD; Doug Liou, BA; Scott Karlan, MD; Alan Waxman, MD; Kayo Fujimoto, PhD; Masanobu Hagiike, MD; Edward H. Phillips, MD

Hypothesis: Fludeoxyglucose F 18 (FDG) positron emission tomography (PET) can be used to predict axillary node metastases.

Design: Case series.

Setting: Comprehensive breast care center.

Patients: Fifty-one women with 54 biopsy-proven invasive breast cancers.

Intervention: Whole-body FDG-PET performed before axillary surgery and interpreted blindly.

Main Outcome Measures: Axillary FDG activity, quantified by standardized uptake value (SUV); axillary metastases, quantified histologically; and tumor characteristics.

Results: There was PET activity in 32 axillae (59%). The SUVs ranged from 0.7 to 11.0. Twenty tumors had an SUV of 2.3 or greater, and 34 had an SUV of less than 2.3. There were no significant differences between these 2 groups except in axillary metastasis size (SUV ≤ 2.2 vs SUV ≥ 2.3): mean age, 53 vs 58 years (P = .90); mean modified Bloom-Richardson score, 7.7 vs 7.6 (P = .20); lymphovascular invasion present, 25% vs 36% (P = .40); mean Ki-67 level, 25% vs 32% (P = .20); mean tumor size, 2.9 vs 3.2 cm (P = .05); and axillary metastasis size, 0.9 vs 1.7 (P = .001). By adopting an SUV threshold of 2.3, FDG-PET had a sensitivity of 60%, a specificity of 100%, and a positive predictive value of 100%.

Conclusions: Patients with an SUV greater than 2.3 had axillary metastases. This finding obviates the need for sentinel lymph node biopsy or needle biopsy to diagnose axillary involvement. Surgeons can proceed to axillary node dissection to assess the number of nodes involved, eliminate axillary disease, or perhaps provide a survival benefit if preoperative FDG-PET has an SUV greater than 2.3.

©2006 American Medical Association. All rights reserved.
FDG-PET. We investigated the use of FDG-PET as part of a strategy for evaluating the axilla in patients with breast cancers of variable sizes and variable characteristics before SNB or AND.

METHODS

A retrospective review of medical records between November 1, 2001, and August 31, 2005, found 462 women with invasive breast cancer who underwent FDG-PET for staging at the Saul and Joyce Brandman Breast Center. The diagnosis of invasive breast cancer was confirmed by fine-needle aspiration, core-needle biopsy (stereotactic, ultrasound guided, or by palpation), or excisional biopsy. Fifty-one of these women (54 invasive cancers) underwent scanning before axillary surgery or chemotherapy. Eighteen of these patients had clinically palpable axillary lymph nodes, and 1 of these patients had fine-needle aspiration–proven metastases before FDG-PET. Patients with sarcoma, phyllodes, and lymphoma were excluded.

The indication for FDG-PET was not specifically noted on the medical record. Because not all patients with breast cancer during this period underwent preoperative FDG-PET, one can assume that there was a higher suspicion of metastatic disease in patients who did undergo this procedure.

Patients were given a median dose of 13 mCi of FDG in the arm contralateral to the primary tumor. After a mean 100-minute delay, images were obtained using an Allegro or Gemini camera (Phillips/ADAC Corporation, Milpitas, Calif). The FDG-PET scan was reviewed and interpreted by a board-certified nuclear medicine physician. The SUVs of the primary tumor and the ipsilateral axilla were recorded but were not used for interpretation. All visible axillary activity was reported to the surgeon.

The patients were discussed at a multidisciplinary conference. However, all surgical decisions, including the choice of AND without SNB, were made by the treating surgeon. Sentinel node biopsy was most commonly performed using a combination of isosulfan blue dye and technetium Tc 99m–labeled sulfur colloid. Lymph nodes were examined by a breast pathologist, and sentinel lymph nodes underwent serial sectioning and immunohistochemical analysis. Pathologic test results were correlated with the FDG-PET scan interpretation.

To determine the optimal SUV threshold, receiver operating characteristic (ROC) analysis was performed on SUV thresholds from 0.5 to 4.0 in 0.2 to 0.3 increments. The SUVs did not subtract for background activity, which is 0.2 to 0.4. The FDG-PET results (positive vs negative), based on each SUV threshold, were compared with pathologic findings to identify the true-positive, true-negative, false-positive, and false-negative rates. The sensitivity, specificity, positive predictive value (PPV), negative predictive value (NPV), and accuracy were calculated for each SUV threshold. Once the optimal SUV threshold was identified, patients were divided into 3 groups: SUV at the optimal threshold or greater, SUV less than the optimal threshold, and SUV equivalent to tissue background (0-0.4).

These groups were compared with respect to patient age, lymph node status, and primary tumor characteristics, including tumor size, tumor grade based on the modified Bloom-Richardson score, estrogen receptor/progesterone receptor status, the presence of lymphovascular invasion, and Ki-67 level. We then examined the effect of these primary tumor characteristics on the SUV by conducting multivariate linear regression analysis. We also performed a logistic regression analysis to calculate the probability of axillary node positivity based on SUV. This study was approved by the Cedars-Sinai Medical Center institutional review board.

RESULTS

Fifty-four breast cancers were initially diagnosed in 51 patients by means of fine-needle aspiration or core biopsy (n=25; 46%) or excisional biopsy (n=29; 54%). Forty-six tumors (85%) were invasive ductal carcinoma, 5 (9%) were invasive lobular carcinoma, and 3 (6%) were a combination of invasive ductal carcinoma and invasive lobular carcinoma. According to the TNM staging system, there were 22 T1 tumors, 24 T2 tumors, and 8 T3 tumors. Forty tumors (74%) were estrogen receptor positive, and 14 (26%) were estrogen receptor negative. The mean age of patients was 54 years (range, 28-88 years), the mean tumor size was 3.0 cm (range, 0.1-11.0 cm), and the mean modified Bloom-Richardson score was 7.6 (range, 3.0-9.0).

Forty-two patients (82%) underwent FDG-PET in the supine and prone positions and 9 (18%) underwent the procedure in the supine position only. Thirty-two patients (59%) had axillary activity. In 3 patients with bilateral breast cancer, each side was managed separately and considered a “patient” for statistical purposes. There were 60 “hot” axillary sites with SUVs ranging from 0.7 to 11.0. In 22 patients (41%), no axillary activity was detected. The mean interval between FDG-PET and axillary surgery was 29 days. Twenty-five patients (46%) underwent SNB. A mean of 1.5 sentinel nodes were removed. Six (11%) had positive SNB results, and all of these patients underwent subsequent AND. Nineteen patients (35%) had negative SNB results, and none had further axillary surgery. Twenty-nine of 51 patients (54%) underwent AND without SNB. A mean of 16 nodes were removed. Twenty-four of these patients (44%) had at least 1 positive axillary lymph node, and 5 (9%) had a negative axilla. Axillary metastases averaged 1.4 cm (range, 0.1-3.0 cm). Eighteen patients (35%) had clinically suspicious axillary lymph nodes; all were eventually found to have axillary metastases.

ROC ANALYSIS

The ROC analysis, with SUV cutoff points ranging from 0.5 to 4.0, is given in Table 1. With SUV thresholds of 0.5 to 1.0, the maximum sensitivity, specificity, PPV, and NPV were 79%, 74%, 86%, and 60%, respectively. With SUV cutoff points between 1.0 and 2.0, specificity and PPV rates were higher (74% to 89% and 87% to 93%, respectively), but sensitivity was lower (64% to 76%). The NPV did not vary with these SUV thresholds. When the SUV threshold was 2.3, specificity and PPV were 100% and sensitivity and NPV were 60% and 53%, respectively. The SUV cutoff points from 2.5 to 4.0 yielded equal rates of specificity and PPV (100%) but lower NPV. On the basis of this analysis, 2.3 was the optimal SUV threshold. Using this threshold, PET identified 60 axillary sites. There were 25 true positives (42%), 18 true negatives (30%), 0 false positives, and 17 false negatives (28%). Accuracy was 72%. The ROC curve, with an area of 0.85, is shown in Figure 1.

There were 20 patients (37%) with SUVs of 2.3 or greater and 34 (63%) with SUVs less than 2.3, 22 (41%)

Table 1

<table>
<thead>
<tr>
<th>SUV Threshold</th>
<th>Sensitivity (%)</th>
<th>Specificity (%)</th>
<th>PPV (%)</th>
<th>NPV (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.5-1.0</td>
<td>74</td>
<td>86</td>
<td>79</td>
<td>60</td>
</tr>
<tr>
<td>1.0-2.0</td>
<td>87</td>
<td>93</td>
<td>74</td>
<td>64</td>
</tr>
<tr>
<td>2.3</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>60</td>
</tr>
<tr>
<td>2.5-4.0</td>
<td>87</td>
<td>93</td>
<td>89</td>
<td>53</td>
</tr>
</tbody>
</table>

Figure 1

[Graph showing ROC analysis results]

©2006 American Medical Association. All rights reserved.
of which had SUVs equivalent to background. The mean ages, tumor sizes, tumor grades, and Ki-67 levels for these patients are given in Table 2.

**LINEAR AND LOGISTIC REGRESSION MODEL**

We used a multivariate linear regression analysis to examine the effects of age, tumor size, modified Bloom-Richardson score, the presence of lymphovascular invasion, Ki-67 level, and size of axillary metastasis on SUVs. Controlling for other effects in the model, we found that SUVs increased with the size of axillary metastases ($b = 2.8; P < .001$) and that there may be a relationship with the size of the primary tumor ($P = .05$). None of the other variables seemed to be related: age ($P = .90$), tumor grade ($P = .20$), the presence of lymphovascular invasion ($P = .40$), or Ki-67 level ($P = .20$) (Table 3). This model accounted for 60% of the SUV variance ($R^2 = 0.60$).

As a separate analysis, we used a logistic regression model and found that SUVs could predict node positivity with statistical significance ($P = .01$). Axillary nodes with SUVs of 2.3 or greater were 15 times more likely to contain metastasis than nodes with SUVs less than 2.3.
In Figure 2, we treated the SUV as a continuous variable to graphically demonstrate that based on the logit model, \( \text{logit}(p) = b_0 + b_1(\text{SUV}) \), a predicted probability of axillary node metastases for each observation (SUV) could be generated. We used Stata statistical software (StataCorp, College Station, Tex) to perform these analyses.

The role of PET has evolved as most human cancers have demonstrated an affinity for FDG. Minn and Sioni were the first to report the clinical utility of PET in patients with primary breast cancer. In patients with tumors larger than 5 cm, PET had 100% sensitivity. Early studies on smaller tumors were much less successful; however, advances in technology have enabled FDG-PET to more reliably detect such tumors. Most recently, PET has been applied to the evaluation of the axilla. Studies have reported sensitivities ranging from 20% to 100% (Table 4). In our series, using an SUV threshold of 2.3, PET had a sensitivity of 60% and a specificity of 100%, suggesting that the clinical value of PET for noninvasively staging the axilla lies in its specificity.

Greco and colleagues studied 167 patients with tumors smaller than 5 cm. Although PET identified axillary disease with sensitivity, specificity, and accuracy of 94%, 86%, and 90%, respectively, false-negative results were observed in patients with "limited numbers of metastatic cells in the axillary lymph nodes." Barranger et al reported a much lower sensitivity (20%), with PET missing all cases involving axillary micrometastases. That series included smaller tumors (mean, 17.9 mm), 30% of which did not accumulate FDG, and a high incidence of micrometastases (50%). In our series, there were 17 false negatives. The mean size of the false-negative tumors was 0.7 cm. As in other series, PET could not reliably detect small lymph node metastases.

Some studies relate the diagnostic accuracy of PET to primary tumor characteristics. Utech et al looked at a variety of variables, including tumor size, type, grade, estrogen receptor/progesterone receptor status, and number of dissected nodes. They found weak correlations between axillary uptake and tumor size and the number of cells in S-phase. Greco et al reported a higher sensitivity for T2 tumors but a higher specificity for T1 tumors. In our study, larger axillary metastases were more likely to be PET positive. The size of the primary tumor may weakly affect SUVs, but other tumor characteristics did not have any correlation with PET activity. Sensitivity has been reported to vary with tumor type and is in general higher with invasive ductal carcinoma than with invasive lobular carcinoma. Although a few of our patients had invasive lobular cancer, all of the false-negative results occurred in patients with invasive ductal carcinoma.

In most studies, visual analysis (not SUV) is used to interpret PET scans, and true-positive rates range from 55% to 100%. In a meta-analysis of 18 studies using PET to diagnose breast cancer, only 4 used any quantitative (SUV) assessment. The true-positive rates of these 4 studies ranged from 86% to 100%. Most studies did not clearly state their definition for positive PET findings. Our study did not use visual information. If the SUV exceeded 2.3, the finding was positive. This threshold resulted in a specificity of 100%, whereas sensitivity was only 60%. In cases with an SUV less than 2.3, visual PET interpretation might have improved sensitivity.

Four patients with axillary activity but an SUV less than 2.3 were found to have negative axillary lymph nodes. The reason for the PET activity is unknown, but false positives have been reported after breast biopsies. Reactive lymphadenopathy can lead to FDG uptake, mimicking malignant disease. None of the patients in this study had undergone previous axillary surgery, although each patient had a preceding breast biopsy (2 had excisional and 2 had core biopsies). These would have been false positives if we had used visual interpretation rather than an SUV threshold. As the role of PET in breast cancer staging evolves, surgeons are likely to see patients with abnormal scans (interpreted visually) and should be mindful to avoid concluding that the axilla is definitely involved.

This study was open to any patient with breast cancer who had not had previous axillary surgery or chemotherapy. Compared with an average cohort of patients with breast cancer, our patients had slightly larger primary tumors (3 cm), a high rate of positive axillary lymph nodes (67%), a high rate of clinically palpable axillary lymph nodes (35%), and a low rate of SNB (46%), demonstrating a se-
Positron emission tomography is only one of several means of preoperatively staging the axilla. Breast magnetic resonance imaging is increasingly being performed and was reported to have sensitivities ranging from 73% to 100%, with specificities of 56% to 100%.25,26 The results have not found magnetic resonance imaging to be reliable enough to replace axillary staging. The recent development of ultrasmall superparamagnetic iron oxide enhancement20 in magnetic resonance imaging may increase sensitivity and specificity, but this approach has not been widely tested.

Ultrasound-guided core biopsy or fine-needle aspiration of axillary lymph nodes has been found reliable when the results are positive. Sensitivity greater than 90% and specificity of 100% have been reported,27-29 although this approach is limited to patients with lymph nodes that are abnormal (enlarged, loss of fatty hilum, etc) on ultrasound. Patients with multiple palpable or imaged enlarged nodes can be problematic if the first needle biopsy result is negative.

Physical examination is the least invasive method for assessing the axilla. In this series, 18 patients (35%) had clinically palpable lymph nodes, and all of them had at least 1 positive axillary lymph node. However, the false-negative rate was 25%, and other studies have reported rates as high as 33%.30 The false-positive rate of physical examination was reported to range from 30% to 41%.31 Many patients will present with palpable nodes that prove to be negative, particularly after surgery or core biopsies of the primary tumor. As with magnetic resonance imaging and ultrasound, large axillary metastases are easier to identify with physical examination. Smaller metastases can be difficult to distinguish from reactive lymph nodes.

The purpose of this article is not to advocate using PET to identify axillary metastasis but to educate physicians that if PET is performed before surgery for systemic staging, an SUV should be calculated for axillary activity. If greater than 2.3 (at our institution), the lymph nodes are involved. On a cautionary note, PET should be performed and interpreted in the proper clinical context, and SUVs should be used as an adjunct to clinical judgment. In addition, SUVs will vary among PET centers by 10% to 15%, even with the same acquisition protocol, owing to technical and calibration factors. Each PET center needs to develop its own reference values. However, if a PET SUV is validated and can predict node positivity with 100% specificity, chemotherapy can be initiated or a surgeon can proceed directly to AND for locoregional control.

Accepted for Publication: April 19, 2006.

Correspondence: Edward H. Phillips, MD, 8635 W Third St, Suite 795-W, Los Angeles, CA 90048 (edward.phillips@chs.org).

Previous Presentation: This paper was presented at the 77th Annual Meeting of the Pacific Coast Surgical Association; February 18, 2006; San Francisco, Calif; and is published after peer review and revision. The discussions that follow this article are based on the originally submitted manuscript and not the revised manuscript.

REFERENCES

Discussion

Armando E. Giuliano, MD, Los Angeles, Calif: Whereas most surgeons, including me, are daunted by the authors' statistical analysis, their ROC, and even their SUVs, their main point is quite simple and true. In plain English, what they are saying is that if the PET scan shows a high degree of uptake in the axilla, the axillary nodes are involved, and you can probably avoid SNB or any other staging short of AND.

This information can be relied on for operative planning. The authors use a quantitative analysis, SUV, to eliminate some subjectivity. I think we can all understand and accept this concept and should not be surprised by it, no more so than we would be surprised by exploring a patient in whom we felt matted axillary nodes and found that there was cancer. But as the PET scan becomes more widely used, we will all see patients with PET, hot spots, in the axilla.

How can the operating surgeon use this information reliably? I ask several questions, not to criticize this fine paper, but to help me and others interpret it and understand the important points and use this information on our own patients. First, how were these patients chosen? Fifty-four patients in 4 years is not many from one of the largest hospitals in the West. Notably, these patients had very large tumors and large axillary metastases. How did you define a positive node? Was it a macrometastasis greater than 2 mm, a micrometastasis less than 2 mm, or isolated tumor cells? If you did not consider micrometastases as positive, would the inclusion of this information alter your SUV findings and your conclusions?

How reproducible is SUV? Can we use 2.3 at our institution, or do we need a different number? Finally, with only 20 patients having SUVs above 2.3, how sure are you of this cutoff? Do you believe this cutoff predicts 100% axillary positivity? With only 20 patients, could the number be higher or even lower?

I totally agree with your concept, and I think it is an important point to bring to this Association. Strong PET activity in the axilla is a reliable sign of axillary metastases. I suspect this information will apply to only a small portion of patients with operable breast cancer. In this study, 40% of your patients with metastases had negative PET scans. The real problem remains detecting metastases in patients with small primary tumors who may benefit from axillary surgery and most likely have a negative PET scan. Do you anticipate changes in imaging that will enable us to detect these small metastases in patients with smaller primary tumors?

Howard Silberman, MD, Los Angeles: You did not indicate in your presentation the clinical N-stage of the patients in your series. Were patients who had palpable adenopathy excluded? It is our policy at USC [University of Southern California] to preoperatively biopsy with FNA [fine-needle aspiration] or core needle nodes that are either palpable or suspicious on axillary ultrasound, which we perform routinely in patients with N0 invasive cancer. Finding positive nodes would eliminate the necessity for SNB and PET scan. It seems then that the PET scan data you present would be particularly valuable in patients with clinical N0 disease with negative axillary ultrasound results. I ask the authors, therefore, whether they excluded patients with palpable adenopathy and whether ultrasound was performed.

John T. Veteto, MD, Portland, Ore: I want to note that the authors are basically studying a moving target; when PET scan first came out it could not detect lesions less than a centimeter in size. Now, PET scan can “see” lesions about 5 mm in size, and the newer-generation scans will have higher sensitivity and be able to detect even smaller lesions. Eventually the technology may reach a point where the sensitivity of PET scanning and the sensitivity of SNB begin to actually converge (around 2 mm, where PET scan would actually detect micrometastatic disease). So, I would encourage the authors to repeat this study in the future.

Lawrence D. Wagman, MD, Duarte, Calif: I have 2 questions. First, did the primary tumor show up in all of the patients, particularly if the axillary node also showed up? The second question is more complex. How many of the patients had additional findings on their PET scan that either delayed their therapy or required extensive workup that was either positive or negative and potentially changed their staging and prognosis?

Rodney F. Pommier, MD, Portland: I was a little surprised to see that you applied the multivariate regression analysis to the effects of classic tumor prognostic factors on SUV rather than to their effect on axillary nodal status. Even though each of those tumor factors independently doesn’t predict the presence of axillary nodal metastases very accurately, together they can be more powerful. Did you perform a multivariate regression analysis on the effects of these tumor prognostic factors on axillary node status? It could be that in your highly selected group of patients, a weighted scoring system using each of those tumor factors, all of which would have been determined by the biopsy alone, might have predicted the presence of positive axillary nodes just as accurately as did the PET scan.

Dr Phillips: Dr Giuliano asked how the patients in this study were chosen. We really don’t know how they were chosen because this was a retrospective review of patients who had biopsy-proven breast cancer, no chemotherapy, no axillary surgery, and a PET scan. The ordering physicians didn’t indicate in the medical record why they ordered a PET scan. However, we know from practical clinical experience that when patients have either a large tumor and/or the axillary nodes are palpable, or the patient has a symptom or other test that suggests metastasis, a PET scan may be ordered. So we suspect that these patients were at much higher risk of axillary involvement than other patients who didn’t have preoperative PET scans.

Dr Giuliano also asked, “How reliable is FDG-PET scan?” and “Can surgeons use it?” The answer is that it is not a perfect test, but by quantifying the activity, it can become a more accurate test. It is more useful to surgeons in its specificity than its sensitivity. There were patients who had PET activity in the axilla, and the surgeon assumed positive nodes and proceeded right to AND. In fact, 5 of 20 patients I won’t say had an un-
had a negative AND based on visual interpretation of the PET scan. So I think this new push toward quantification will make this a more accurate test.

How were positive nodes defined? This is an excellent question. We did consider micrometastasis as positive axillae. We also had IHC [immunohistochemistry] data. We didn’t include positive IHC as positive axillae. There was only 1 patient with a micrometastasis in the PET SUV under 2.3 category, and it was interpreted visually as a positive scan result.

How reproducible is SUV? I think that this is the take-home message for all of us. There is variability among institutions in methods, the timing of the injection, and the timing of the scan. The hardware varies, but the software now is pretty much consistent. Those variables allow for a 10% to 15% variation in SUV values between centers. Therefore, each center has to develop its own SUV ROC curve. So, if the SUV is 11, that node is going to be positive, no matter which center did the study, but if the SUV is 2.3 and the ROC curves have not been validated, that could be a 2.1 at our institution. Each center should calculate its own ROC curve. That’s not difficult.

Even though we had just 20 patients with an SUV over 2.3, we do believe that in the proper clinical setting one can proceed directly to AND. However, there weren’t enough patients between baseline activity and 2.3 to give an accurate probability of nodal involvement. In other words, if someone has an axillary SUV of 1.8, we should be able to tell them that they have an 80% chance of having an involved node, but we just didn’t have the right sample size. So we hope as more of these studies are done, we can actually pool our data and calculate risk that should help patients and physicians make better decisions.

Both Dr Giuliano and Dr Vetto asked if the FDG-PET would ever be sensitive enough to identify a 2-mm lesion. I hope it will. We continue to see advances in diagnostic technology not dreamed about when I was in medical school. I remember the sestamibi scans that have now given way to FDG-PET. So I think by combining PET and CT [computed tomography] with visual and quantitative data, we will be able to identify smaller and smaller tumors. Will we ever be able to identify small groups of cells (the IHC-positive nodes)? Nobody knows what is going to be happening in the future.

Dr Silberman asked what was the clinical end stage. Did we exclude patients with clinically positive axillae? We did exclude patients who had needle or core biopsies of their axillae before PET or surgery of their axillae before PET scanning. We do perform, as you do, Howard, needle biopsies of clinically positive axillary nodes. But those patients were excluded from our study because that can affect PET activity.

Dr Wagman asked, “Did the primary breast tumor show in all patients?” The answer is yes. However, I am aware of patients who have had a positive, palpable axillary node without a known primary tumor where it has not shown the primary tumor in the breast. Dr Wagman asked, “How many additional findings delayed treatment or required other biopsies?” This is an important question. I am sure all of you noticed that the average length of time to axillary surgery was 29 days, and you probably all wondered why it was so long. Well, that’s the reason. [The] FDG-PET can show bone metastasis, lung primaries and metastasis, arthritis, and trauma. Many of our patients have had to have other biopsies and other investigations before proceeding with surgery. I am unable to tell you how many false positives there were in these other investigations. We don’t know the false positives in the nonbreast areas.

Regarding Dr Pommier’s queries about statistical analysis, I wish our statistician were here to answer. After multiple discussions and multiple different statistical analyses, she felt this was the fairest way to look at the data and the safest. Could a weighted scoring system of tumor characteristics have identified patients with positive axillae? I think it’s possible, but to reach a specificity of 100%, fewer patients would have been identified. Quantification of SUV is still the best way to maximize sensitivity with specificity.