FDG POSITRON EMISSION TOMOGRAPHY FOR EVALUATING BREAST CANCER

May 2001

David Samson
Associate Director
Technology Evaluation Center
Blue Cross and Blue Shield Association
Chicago, Illinois

Carole Redding Flamm, M.D., M.P.H.
Senior Assessor
Technology Evaluation Center
Blue Cross and Blue Shield Association
Chicago, Illinois

and

Naomi Aronson, Ph.D.
Executive Director
Technology Evaluation Center
Blue Cross and Blue Shield Association
Chicago, Illinois

FDG POSITRON EMISSION TOMOGRAPHY FOR EVALUATING BREAST CANCER

EXECUTIVE SUMMARY

Positron emission tomography (PET) imaging uses radiotracers that can reveal both anatomical and metabolic information. The glucose analog, 2-[fluorine-18]-fluoro-2-deoxy- D-glucose (FDG), is potentially useful in cancer imaging because tumor cells show increased utilization of glucose. This technology assessment reviews evidence on the use of FDG PET imaging in breast cancer and focuses on 4 specific clinical settings: (1) initial diagnosis of breast cancer; (2) staging axillary lymph nodes; (3) detection of locoregional recurrence or distant metastasis/recurrence; and (4) evaluating response to treatment. For each clinical indication, the evidence describing diagnostic performance of FDG PET will be evaluated. When the diagnostic performance data permit further analysis, the effect of FDG PET on health outcomes will be modeled through analysis of PET's effect on patient management and the effects of those management changes on health outcomes.

In order to be reviewed in this assessment, articles had to meet all of the following criteria:

- the study was published or accepted for publication as a full article in a peer-reviewed journal (studies published only in abstracts were excluded due to insufficient details for meaningful analysis);
- when an institution published multiple articles, it was represented for the purpose of quantitative data synthesis by the report with the largest patient series;
- the study sample included at least 10 patients and did not mix results in breast cancer patients with those of patients who have other tumor types;
- the study performed tomographic, not planar, imaging with FDG as the radiotracer; and
- the article described the correlation of FDG PET findings with data from an appropriate reference standard, for both diseased and nondiseased patients (permitting calculation of both sensitivity and specificity).

A total of 32 articles met these criteria. More specific study selection criteria were applied to each of the 4 main clinical indications reviewed.

Each article was assessed for study quality based on the guidelines of the Cochrane Methods Working Group on Systematic Review of Screening and Diagnostic Tests. Key study quality characteristics include: use of a valid reference standard; blinded test interpretation; avoidance of verification bias (e.g., by selecting patients consecutively); clear description of the spectrum of disease and other characteristics of the patient population; and prospective design.

This assessment is organized into 4 parts, as follows:

- Part I: Initial Diagnosis of Breast Cancer;
- Part II: Initial Staging of Axillary Lymph Nodes;
- Part III: Detection of Locoregional Recurrence or Distant Metastasis/Recurrence; and
- Part IV: Evaluating Response to Treatment.

For each indication, 2 questions were addressed: 1) does the available evidence permit conclusions about the diagnostic performance of PET; and 2) if there is adequate evidence on diagnostic performance, does the use of PET improve health outcomes?

PART I – INITIAL DIAGNOSIS OF BREAST CANCER

Two potential roles for PET were considered: (a) using PET in patients who have been referred for breast biopsy in order to avoid biopsy when PET results are negative, and (b) using PET in patients who have been referred for short-interval imaging follow-up due to low suspicion findings on mammography.

Indication Ia

Among patients who have been referred for biopsy, a true-negative PET finding would result in a patient safely avoiding a painful invasive biopsy and its consequences; while a false-negative PET finding could result in delayed or missed diagnosis and treatment. Patients with positive PET scans would presumably undergo biopsy confirmation; thus there would be no change in the net health outcome from using PET compared with not using PET prior to biopsy.

Evidence on Diagnostic Performance.

In studies of PET for differential diagnosis of breast lesions, patients were selected for having suspicious mammograms or palpable masses. These study samples have a notably higher prevalence of malignancy than that reported for the general population and a relatively large average tumor size at initial diagnosis. These studies represent the upper part of the biopsy population spectrum. No published studies are available on the diagnostic performance of PET in the lower part of the biopsy population, comprising a range of prevalence between 20% and 50%. This group consists of patients with indeterminate mammograms and smaller, nonpalpable lesions. Without of evidence on diagnostic performance of PET in the lower portion of the biopsy population, no conclusions can be reached and it would be imprudent to generalize from the studied population.

Thirteen studies (total n=606) met study selection criteria for inclusion in the data synthesis. The prior probability of malignancy in the study samples ranged between 53% and 95%, compared to 20% to 30% in the general population. Mean tumor size across studies was relatively large, ranging from about 2 cm to 4 cm.

Sensitivity estimates in all 13 studies ranged from 79% to 100% and specificity estimates were between 50% and 100%. Meta-analysis was first performed using a random effects model. The pooled sensitivity estimate was 88% (95% CI: 83%, 92%) and the pooled specificity estimate was 79% (95% CI: 71%, 85%). Then a summary receiver operating characteristic (ROC) curve was constructed which accounts for the dependent relationship between sensitivity and specificity. A point on the summary ROC curve was selected which reflected average performance, with an estimated sensitivity of 89% and a specificity of 80%.

Sensitivity analysis based on higher quality studies, defined as prospective, free of verification bias and used blinded interpretation of PET, was initially planned. However, only 1 study met these qualifications (n=40), thus precluding the planned analysis.

Analysis of Effect on Health Outcomes.

In order to be used to avoid biopsy, PET should provide a highly sensitive evaluation for malignancy. The rate of false negative PET results weighs heavily in considering whether the risk of delayed or missed diagnosis of breast cancer is worth the benefit of avoiding biopsy of a benign lesion. The risks and benefits of PET were analyzed using two perspectives: (1) the entire population of patients undergoing PET and (2) the individual patient who has a negative PET result.

For both analyses, sensitivity of PET was assumed to be 89% and specificity was 80%. The prevalence (i.e., pre-test probability) of malignancy was assumed to range from 50% to 75%. Evidence is lacking about PET's diagnostic performance for smaller tumors and in patient populations with disease prevalence lower than 50%. As the prevalence of malignancy rises from 50% to 75%, the false-negative risks also rise, making the probabilities of harm from delayed diagnosis and treatment higher

The population perspective assumes that the results of PET are not yet known. All PET results (both positive and negative) are considered and the proportions of the entire population deriving benefits and harms can be estimated. When the prevalence of malignancy is 50%, 40% of all patients would benefit by avoiding the harms of negative biopsy. The risk of a false-negative result, leading to delayed diagnosis and treatment, is 5.5%. When the prevalence of malignancy is 75%, 20% of patients avoid biopsy of a benign lesion; and the risk of delayed treatment is 8.25%.

From the perspective of an individual patient with a negative PET scan, the risk of a false-negative result is higher than for the entire population undergoing PET scanning. When the prevalence of malignancy is 50%, the NPV is 87.9%, thus, the false-negative risk is 12.1%. For an individual with a negative PET scan, a 12% chance of missed or delayed diagnosis of breast cancer is most likely too high to make the 88% chance of avoiding an negative biopsy of a benign lesion worthwhile. When the prevalence of malignancy is 75%, there is a 29.2% risk of missed or delayed diagnosis, which is surely unacceptable in order to avoid a biopsy.

Evidence is lacking to assess the negative predictive value of PET in the population of patients referred for biopsy with indeterminate mammograms and smaller, nonpalpable lesions. Such patients would have a prevalence of malignancy from 20% to 50%.

Summary.

Evidence on the diagnostic performance of PET for differential diagnosis of breast lesions among patients with abnormal mammograms or palpable masses is lacking for patients with indeterminate mammograms and small, nonpalpable lesions (low prevalence for malignancy). Among study populations of patients with higher prevalence of malignancy, risk of a false-

negative diagnosis is likely too high relative to the benefit of avoiding biopsy of a benign lesion. A false-negative PET result may cause a missed or delayed diagnosis of breast cancer and associated delay in treatment.

From the perspective of an individual patient with a prior probability of malignancy of 50% and a negative PET result, the risk of a false-negative result PET is 12.1%. At the 75% prevalence, there is a 29.2% risk of a false-negative finding. Evidence on PET diagnostic performance is unavailable to permit estimation of the risk of a false-negative PET result in the patients with a prevalence of malignancy from 20% to 50%.

Indication Ib

FDG-PET may also be used as a diagnostic aid in patients with low suspicion mammographic findings who have been referred for short interval mammographic follow-up. Positive PET results may help to select patients who should be referred for biopsy while negative PET results might enable the frequency of follow-up to be reduced. Selective biopsy might achieve earlier diagnosis of breast cancer than short-interval mammographic follow-up, which is presently recommended in this patient population.

Evidence on Diagnostic Performance.

No studies meeting selection criteria included a patient population to address this question. Performance of PET in the available studies in patients who have been referred for biopsy due to an abnormal mammogram or palpable mass cannot be generalized to patients with low suspicion findings on mammography referred for short interval follow-up.

Analysis of Effect on Health Outcomes.

This question cannot be addressed in the absence of data on the diagnostic performance of PET in the population of interest.

PART II – STAGING AXILLARY LYMPH NODES

The proposed role for PET for this indication is to select which patients need to undergo axillary lymph node dissection (ALND) among the subset of patients who have no clinically palpable axillary adenopathy. If the PET scan correctly suggested no spread of tumor to the axillary lymph nodes, the patient could avoid the pain and other complications associated with ALND (e.g., chronic lymphedema). However, a false-negative PET result could lead to harm if a patient with undetected axillary involvement chose to forego adjuvant systemic therapy.

Adjuvant systemic therapy has been reported to reduce recurrence and improve survival in patients with breast cancer. Improved outcomes occur in patients with positive axillary nodes and also in patients without axillary involvement. However, the absolute magnitude of the reduction in recurrence rate or mortality is greater for those with axillary nodal disease. Compared to patients without axillary involvement, node-positive patients are at a greater baseline risk of recurrence and disease-related mortality and thus, have greater potential for

benefit, based on the Early Breast Cancer Trialists' Collaborative Group overview of 133 randomized clinical trials. However, decisions on the use of adjuvant therapy in patients with node-negative disease is complicated by uncertainties in balancing potential benefits and toxicity of systemic therapy, as well as by variation in patient preferences.

Chronic lymphedema is common following ALND, and strategies for reducing the morbidity of axillary node staging are being developed. Sentinel node biopsy (SNB), a more limited surgical approach to axillary lymph node staging, has been introduced as an alternative surgical technique. More recently, PET has been proposed as a noninvasive method for determining the presence of axillary lymph node involvement and for selecting patients for ALND.

Evidence on Diagnostic Performance

The available body of literature is too sparse to draw conclusions regarding the diagnostic performance of PET for staging of axillary lymph node metastases in patients without palpable adenopathy. Only 4 studies reported on patients with nonpalpable axillary lymph nodes, with a total pool of 203 patients. A random effects meta-analysis was performed using this data. The estimates for sensitivity and specificity were 80% (95% CI: 46%, 95%) and 89% (95% CI: 83%, 94%), respectively. The width of the confidence interval for sensitivity is almost 50 percentage points.

In contrast, data from more than 3,000 patients is available for sentinel node biopsy performed in patients with nonpalpable nodes. The 95% confidence interval for SNB has a width of only 5 percentage points.

Analysis of Effect on Health Outcomes

In the absence of adequate evidence to estimate diagnostic performance, the outcomes of using PET to decide whether to perform axillary lymph node dissection cannot be determined. However, for illustrative purposes, this assessment estimated the probabilities of outcomes using diagnostic performance data from the available studies.

Taking the perspective of an individual patient with a known negative PET scan, the negative predictive value of PET is 92.1%, given a prevalence for node-positive disease of 30%. Thus, the risk of undertreatment in this situation would be 7.9%. As prevalence for node-positive disease goes to 50%, the false-negative risk rises to 16.7%. This range of risk for undertreatment appears to be unacceptably high.

Undertreatment in this case would be associated with an absolute difference in 10-year survival of 8.2%. Comparison of median survival rates in recent trials indicates about a 2 year average prolongation in life for node positive patients treated with systemic adjuvant therapy.

Summary

The available body of literature is too sparse to draw conclusions regarding the diagnostic performance of PET for staging of axillary lymph node metastases. Estimated diagnostic

performance based on available data, suggests that the false-negative rate of PET in detecting axillary lymph node metastasis is too high to support a favorable risk/benefit ratio from using PET to avoid axillary lymph node dissection.

PART III: DETECTION OF LOCOREGIONAL RECURRENCE OR DISTANT METASTASIS/RECURRENCE

For this indication, PET may serve either as an adjunct to other imaging tests or as a replacement. Because systemic therapy appears to provide a small but significant survival benefit, accurate diagnostic assessment and identification of metastatic disease is essential. Similarly, it is necessary for the staging evaluation to rule out sites of metastatic disease not amenable to local therapy so that optimal treatment decisions can be made.

Evidence on Diagnostic Performance

The evidence is insufficient to permit conclusions about the diagnostic performance of PET in detecting locoregional recurrence, which includes recurrence at the brachial plexus. Two studies reporting on a total of 85 patients met study selection criteria for this review. One study included only 10 patients, while the larger study of 75 patients does not provide enough details about the reference standard to determine the validity of diagnostic performance estimates. The larger study was also one of 2 key studies providing evidence on detection of distant recurrence or metastasis.

The evidence on how well PET detects distant recurrence or metastasis to bone is currently insufficient. The 2 key studies included 75 and 34 patients. As noted previously, the larger study did not provide sufficient detail about whether and how a histologic reference standard was applied, creating uncertainty about the diagnostic performance estimates. The report also gave no information about discordances or correct changes in stage classification when PET is added to CT/MRI. The findings of the smaller study suggest that PET could have replaced radionuclide bone scan, but this single study of 34 patients is not sufficient to draw such a conclusion. Additional data are needed with respect to diagnostic performance, discordance rates, frequency of PET and other tests giving correct results among discordances, and frequency of PET correctly changing stage classification when added to other tests.

Little data have been reported on use of PET to detect recurrence or metastasis in sites other than bone. One study reported PET and CT/MRI findings in 2 patients with liver metastasis and another study reported on a single case of liver metastasis. A single study addressed 6 patients with confirmed lung metastases. A metastasis to pericardium was reported in 1 patient. These data are clearly insufficient to permit conclusions about the diagnostic performance of PET in detecting recurrence or metastasis in bone, lung, liver or other distant sites.

Analysis of Effect on Health Outcomes

As the available data are insufficient to determine the diagnostic performance of PET in detecting recurrence or metastasis, it is not possible to reliably determine the effect diagnostic information might have on management decisions and patient health outcomes.

7

PART IV: EVALUATING RESPONSE TO TREATMENT

The proposed role for PET for this indication is to provide a more accurate or earlier determination of tumor response to treatment to facilitate treatment decisions (e.g., to discontinue ineffective systemic therapy).

Evidence on Diagnostic Performance

Four studies with a total of 103 patients have addressed whether PET imaging early in the course of treatment predicts response to treatment evaluated at its conclusion. All 4 studies were prospective. Treatment was neoadjuvant chemotherapy in 2 studies, chemohormonotherapy in one and hormone therapy in one. The available evidence is of limited quantity, quality, and consistency and is insufficient to permit conclusions about the diagnostic performance of PET in evaluating response to treatment.

Analysis of Effect on Health Outcomes

Additional evidence is needed to determine diagnostic performance of PET, as well as to assess whether health outcomes would be improved by using PET response as a guide to patient management. Two of the 4 studies reported sensitivities that would lead to substantial undertreatment, if a finding of nonresponse on PET were used to guide treatment. Inappropriate discontinuation of systemic therapy would have occurred in 10% (n=30) of patients in one study and 17% (n=22) in the other study.

FDG POSITRON EMISSION TOMOGRAPHY FOR EVALUATING BREAST CANCER

ASSESSMENT OBJECTIVE

Positron emission tomography (PET) imaging uses radiotracers that can reveal both anatomical and metabolic information. The glucose analog, 2-[fluorine-18]-fluoro-2-deoxy- D-glucose (FDG), is potentially useful in cancer imaging because tumor cells show increased utilization of glucose. This technology assessment reviews evidence on the use of FDG PET imaging in breast cancer and focuses on 4 specific clinical settings: (1) initial diagnosis of breast cancer; (2) staging axillary lymph nodes; (3) detection of locoregional recurrence or distant metastasis/recurrence; and (4) evaluating response to treatment. For each clinical indication, the evidence describing diagnostic performance of FDG PET will be evaluated. When the diagnostic performance data permit further analysis, the effect of FDG PET on health outcomes will be modeled through analysis of PET's effect on patient management and the effects of those management changes on health outcomes.

INTRODUCTION

Positron Emission Tomography

Positron emission tomography (PET) is an imaging technology that can reveal both anatomical and metabolic information in various tissue sites. The metabolic information is what distinguishes it from other imaging modalities such as magnetic resonance imaging (MRI) and computed tomography (CT) that provide primarily anatomic information. PET uses radiotracers composed of organic compounds (e.g., glucose, ammonia, water) labeled with positron-emitting isotopes. These radiotracers can be metabolized in certain physiological processes that are correlated with disease states. Several radiotracers have been used in cancer imaging, but by far the most common is the glucose analog FDG. Once injected into a patient, it is metabolized in both normal and cancerous tissue in proportion to the rate of glucolysis. FDG is potentially useful in cancer imaging because tumor cells show increased utilization of glucose (Wahl 1995).

After IV injection, PET tracer isotopes undergo a process of decay, emitting positrons that lose energy in a so-called "annihilation" reaction as they pass through tissue and combine with electrons. During this reaction, the total mass is converted into energy, which is released in the form of two high-energy photons. A detection device registers the photons simultaneously and localizes the annihilation event because the photons travel at an angle of approximately 180 degrees from their origin (Schelbert 1991).

Given that the majority of published studies have used the radiotracer FDG, and that FDG is the only radiotracer used in cancer imaging that has received FDA approval, this assessment will focus on the use of FDG PET imaging of breast cancer.

After the patient fasts for at least 4 hours, a dose of FDG (usually 10 mCi) is injected intravenously. A transmission scan is conducted, either before tracer injection or after FDG imaging, with an external ring of a positron-emitting isotope (e.g., gallium-68), to provide data for correction of photon attenuation. Some equipment is able to obtain attenuation correction

and emission scans simultaneously. Attenuation-corrected PET images provide better data for quantitative analysis and qualitative interpretation; however, attenuation correction techniques have not always been consistently available and applied in clinical investigations of PET. Image acquisition usually begins from 30–60 minutes after injection of FDG and continues for a period of between 10 and 20 minutes. Scanning may be limited to particular body regions or can be performed over the whole body.

The most common method of PET image analysis is visual interpretation, which is qualitative and based on recognition of areas of increased tracer uptake relative to background. Quantitative interpretation begins after the operator draws a region-of-interest (ROI) over the suspicious area, sometimes based on other imaging. Tracer uptake counts within the ROIs are usually corrected for injected dose and body weight. Quantitative indices mentioned in published reports, include: standardized uptake value (SUV); distribution absorption ratio (DAR); differential uptake ratio (DUR); tumor-to-normal tissue (TNT) ratio; and regional metabolic rate of glucose (rMRglu). Tracer uptake varies across different types of tissue, and some quantitative indices compare uptake in the suspicious area with "normal background" uptake in either a contralateral comparison site or a predetermined distal site common to all patients.

Quantitative interpretation would be expected to be more precise and possibly more accurate than qualitative interpretation, because it is more objective and relies less on operator judgment. With quantitative analysis, the criterion for interpretation of a positive test is a numerical value along a continuous function. A threshold can be established for a particular quantitative index to separate benign from malignant lesions much the same way a blood chemistry test separates normal and abnormal results using a numerical cut-off. Thus, different operating points on the receiver operating characteristic (ROC) curve can be achieved simply by setting different numerical thresholds.

FDA Status. The U.S. Food and Drug Administration (FDA) has granted marketing clearance for various PET scanners through the 510(k) process. Clearance has been granted for the general indication of taking an image created by a radiopharmaceutical. The FDA requires PET radiotracers to be approved through a new drug approval (NDA) process. Because PET radiotracers have an extremely short half-life, they must be produced in the clinical setting or express delivered from another facility and used within a short period of time; the FDA also intends to regulate drug manufacturing processes in PET facilities.

In March 2000, the FDA issued a *Federal Register* notice stating it has concluded that FDG can be found to be a safe and effective drug when used to assess abnormal glucose metabolism to assist in evaluation of malignancy in patients with known or suspected abnormalities found by other testing or in patients with an existing diagnosis of cancer. The FDA has also concurrently issued draft guidances to assist manufacturers to submit applications, but applications are not required to be submitted until 2 years after it has established approval procedures and current good manufacturing practice (CGMP) guidelines.

Systematic Review and Meta-Analysis of Studies of Diagnostic Tests

Guidelines for the systematic assessment and meta-analysis of studies of diagnostic tests have been developed and reviewed by Irwig et al. (1994) and the Cochrane Methods Working Group on Systematic Review of Screening and Diagnostic Tests (1996). These guidelines describe key information needed from studies and focus in particular on issues relating to quality of methods. The following criteria should be considered in assessing study quality:

- whether a valid reference standard was used;
- whether the test was interpreted blindly with respect to reference standard and vice versa;
- whether verification bias (test results influencing performance of reference standard) was avoided;
- whether a clear description of spectrum of disease including diseased and non-diseased individuals was provided;
- whether a clear description of other patient characteristics was provided;
- whether a clear description was provided of how the test was performed including interpretation and estimates of reproducibility;
- whether study design was prospective or retrospective; and
- if alternative tests were being compared, whether a valid design was used (tests done independently on each patient).

Meta-analysis of diagnostic test performance characteristics may be approached in a variety of ways. Diagnostic performance is commonly reported as test sensitivity and specificity. However, these two parameters are inter-related and thus adjusting the test interpretation criteria to increase sensitivity will produce a decrease in specificity. One method of summarizing diagnostic performance data is to quantitatively pool the estimates for sensitivity and separately pool the estimates for specificity using fixed-effects or random-effects statistical modeling. Because this method ignores the inter-relationship and inherent trade-offs between sensitivity and specificity estimates across studies, the summary pooled estimates of sensitivity and specificity tend to be somewhat underestimated. When only studies reporting dichotomous test results are available, pooling results using a summary receiver operating characteristic curve (summary ROC) is the preferred method, as the ROC curve takes into consideration the interdependency of sensitivity and specificity. Additional diagnostic test performance characteristics that may be used in quantitative comparisons include likelihood ratios and odds ratios.

METHODS

Search Methods

Searches of the MEDLINE® database, using PubMed, and the CANCERLIT database were performed, using the subject heading "radionuclide imaging" (exploded) and the text words "positron" and "PET." The intersection of these terms and references indexed under the subject heading "neoplasms" served as the initial pool from which articles concerned with breast cancer were searched. The search was limited to articles written in English. The dates covered by the current search include references entered between January 1966 and March 2001. In addition, reference lists of key articles and *Current Contents* were searched for additional citations. The search identified a total of 163 references concerning use of PET for breast cancer.

General Study Selection Criteria

In order to be reviewed in this assessment, articles had to meet all of the following criteria:

- the study was published or accepted for publication as a full article in a peer-reviewed journal (studies published only in abstracts were excluded due to insufficient details for meaningful analysis);
- when an institution published multiple articles, it was represented for the purpose of quantitative data synthesis by the report with the largest patient series;
- the study sample included at least 10 patients and did not mix results in breast cancer patients with those of patients who have other tumor types;
- the study performed tomographic, not planar, imaging with FDG as the radiotracer; and
- the article described the correlation of FDG PET findings with data from an appropriate reference standard, for both diseased and nondiseased patients (permitting calculation of both sensitivity and specificity).

A total of 32 articles met these criteria. More specific study selection criteria were applied to each of the 4 main clinical indications reviewed.

Rating of Study Quality

Each of the studies included in the evidence tables was classified according to a prespecified set of characteristics which are related to study quality. The technique of PET interpretation used in the study was described as to whether quantitative and/or qualitative techniques were used and whether attenuation correction (AC) was employed. Whether the study controlled for verification bias was evaluated. Avoidance of verification bias was considered to be true (coded as "Y" for "yes" in the tables) when consecutively enrolled patients were subjected to PET and the reference standard was obtained independent of the results of the PET study, and all enrolled patients were included in the final analysis. Studies were coded with a question mark when insufficient information was provided to make this determination. Whether studies reported using blinding of study investigators was also recorded. Interpretation of PET studies blinded to the results of the reference standard was rated as "Y" for yes; "N" for no; or "?" when there was insufficient information to determine. Interpretation of the reference standard blinded to the results of the PET findings was rated using the same categories.

Organization of the Assessment

This technology assessment is organized into 4 parts, as follows:

- Part I: Initial Diagnosis of Breast Cancer;
- Part II: Initial Staging of Axillary Lymph Nodes;
- Part III: Detection of Locoregional Recurrence or Distant Metastasis/Recurrence; and
- Part IV: Evaluating Response to Treatment.

PART I: INITIAL DIAGNOSIS OF BREAST CANCER

BACKGROUND – PART I

Screening mammography has improved detection of primary breast cancer. Earlier diagnosis of malignant lesions at earlier stages reduces breast cancer mortality (NIH Consensus Statement 1997). However, the majority of abnormalities identified on screening mammography and referred for biopsy are benign (NIH Consensus Statement 1997; Bassett et al. 1991). While estimates of negative breast biopsy rates vary across different settings and patient populations, the rate of negative biopsy commonly falls in the range of 70 to 90% (Bassett et al. 1991). Concern over the number of biopsies resulting from false-positive mammograms has stimulated efforts to improve the selection of patients for biopsy diagnosis.

Mammographic findings are frequently reported according to the Breast Imaging Reporting and Data System (BI-RADS) established by the American College of Radiology. BI-RADS classifies mammographic results into 5 categories, with Category 5 indicating the highest likelihood of malignancy (Lacquement et al. 1999; Liberman et al. 1998). Category 1 is defined as a normal mammogram; Category 2 is defined as a benign finding; Category 3 is defined as a probably benign finding and short-term follow-up may be suggested; Category 4 is defined as a suspicious abnormality where biopsy should be considered; and Category 5 is defined as highly suggestive of malignancy where appropriate action should be taken (Liberman et al. 1998; American College of Radiology 1995).

Follow-up studies using the BI-RADS system demonstrate that the spectrum of patients undergoing breast biopsy is heterogeneous with regard to pre-biopsy probability of malignancy. In one study of 688 radiographically guided biopsies (Lacquement et al. 1999), 96.2% of biopsies were performed for BI-RADS category 3, 4, or 5 findings. Interestingly, category 3 findings, which accounted for almost half of all biopsies, were associated with a positive predictive value (PPV) of only 3%, and category 4 findings (34% of biopsies) yielded only a 23% PPV. Category 5 findings (15.4% of biopsies) were highly predictive of malignancy with PPV of 92%. Similar findings were reported in other series (Liberman et al. 1998; Orel et al. 1997).

Core needle biopsy techniques provide a minimally invasive alternative to conventional surgical biopsy for many patients (Blue Cross Blue Shield Association Technology Evaluation Center 1995), nevertheless, undergoing a breast biopsy may be psychologically burdensome for the patient, and the scarring and tissue distortion resulting from a biopsy may complicate subsequent mammographic evaluation in that region of the breast. Thus, FDG-PET imaging is proposed as an additional diagnostic test following a suspicious mammogram in order to reduce the number of breast biopsies. FDG-PET may also be used as a diagnostic aid in patients with low suspicion mammographic findings who have been referred for short interval mammographic follow-up. Positive PET results may help to select patients who should be referred for biopsy while negative PET results might enable the frequency of follow-up to be reduced. Selective biopsy might achieve earlier diagnosis of breast cancer than short-interval mammographic follow-up, which is presently recommended in this patient population.

FORMULATION OF THE ASSESSMENT – PART I: INITIAL DIAGNOSIS OF BREAST CANCER

Patient Indications

Indication Ia. Patients have an abnormal mammogram or a palpable breast mass and are recommended to undergo biopsy diagnosis.

Indication Ib. Patients have a low suspicion finding on mammography (and other routine imaging procedures) and are referred for short-interval (i.e., 3–6 month) imaging follow-up.

Technologies to be Compared

A definitive tissue diagnosis is established by cytologic (core needle or fine needle aspiration) or histologic (surgical) sampling, which would serve as reference standards for evaluating the diagnostic performance of PET. PET will not be compared to routine imaging technologies (e.g., mammography, ultrasound) because PET is proposed for use in addition to these technologies and not as a replacement for them.

The comparison of interest for Indication Ia is between using negative PET results to avoid a biopsy compared to performing biopsy in all patients.

The relevant comparison for Indication Ib is between using PET results to elect early biopsy or avoid short-interval imaging follow-up, versus performing short-interval follow-up on all patients.

Health Outcomes

Benefits.

Indication Ia. In the setting of using PET to avoid biopsy, true negative PET results benefit the patient through avoiding the pain and anxiety associated with biopsy. True-positive results on PET do not result in any additional benefit since patients would still undergo biopsy and receive accurate diagnosis.

Indication Ib. When PET results are used to elect early biopsy, a true positive leads to earlier detection and treatment of malignancy. A true negative PET result could permit a patient to forgo frequent short-interval follow-up and revert to a routine screening schedule.

Harms.

Indication Ia. False negative PET results may cause harm from a missed or delayed diagnosis and delayed treatment. A false positive PET result would not expose the patient to any additional harm compared with not using PET since the patient would undergo biopsy in both cases.

Indication Ib.   A false negative PET result may be harmful to the extent that a patient would forgo short-interval imaging follow-up and the potential benefit of earlier detection and treatment. False positive PET results expose the patient to the adverse effects of biopsy.

Causal Chain.

Figure 1 displays causal chains for 2 separate management paths, both concerned with Indication Ia. The first path entails using PET to decide whether to perform a biopsy. The second path entails sending all patients directly to biopsy.

Specific Assessment Questions, Part I, Indication Ia

1. Does the available evidence permit conclusions about the diagnostic performance of PET for differential diagnosis of breast lesions among patients with abnormal mammograms or palpable masses?
2. Can the use of PET improve outcomes by obviating biopsy in patients who have been referred for biopsy due to an abnormal mammogram or palpable mass?

Specific Assessment Questions, Part I, Indication Ib

1. Does the available evidence permit conclusions about the diagnostic performance of PET for differential diagnosis of breast lesions among patients with low suspicion finding on mammography and other routine imaging procedures who have been referred for short-interval follow-up?

2. Can the use of PET improve outcomes by leading to earlier and more accurate diagnosis of breast cancer compared to short-interval mammographic follow-up in patients with low suspicion findings on mammography and other routine imaging procedures?

Specific Study Selection Criteria

- Patients of interest meet the descriptions given under "Formulation of the Assessment – Part I: Patient Indications."
- For indication Ia, all patients in the study must receive PET and biopsy.
- For indication Ib, all patients must receive PET and short-interval imaging follow-up as well as imaging with an adequate reference standard.

Figure 1: Causal Chain, Relative Outcomes for Differential Diagnosis of Suspicious Breast Lesions With and Without PET

**Outcomes of Path 1
Relative to Path 2**

Benefits	Harms

Positive
Biopsy → Confirm tumor present, treatment → Same as Path 2A   Same as Path 2A
(PET True
Positive)

Positive
PET → Perform Biopsy
Result

Negative → Confirm no tumor present, resume screening schedule → Same as Path 2B   Same Path 2B
Biopsy
(PET False
Positive)

**Path 1:
With
PET**   Palpable Mass/ Abnormal Mammogram → Perform PET

PET True → No tumor present, resume screening schedule → Differs from Path 2: avoid pain, anxiety of biopsy
Negative

Negative
PET → Forgo Biopsy, Wait and
Result   Watch

PET False → Undetected tumor present, resume screening schedule → Differs from Path 2: delayed treatment (lose advantage of early treatment)
Negative

Outcomes of Path 2

Benefits	Harms

Positive → Confirm tumor present, initiate treatment → Path 2A: survival associated with early treatment   Path 2A: pain, anxiety of biopsy, treatment-related adverse effects
Biopsy

**Path 2:
Without
PET**   Palpable Mass/ Abnormal Mammogram → Perform Biopsy →

Negative → Confirm no tumor present, resume screening schedule → Path 2B: reassurance   Path 2B: pain, anxiety of negative biopsy
Biopsy

16

REVIEW OF EVIDENCE – INDICATION Ia: INITIAL DIAGNOSIS OF BREAST CANCER
IN PATIENTS WITH ABNORMAL MAMMOGRAMS OR PALPABLE MASSES

Evidence on Diagnostic Performance of PET

Study populations.

The patient populations in studies of PET for differential diagnosis of breast lesions have a
notably higher prevalence of malignancy than that reported for the general population. In studies
meeting the selection criteria for this systematic review, the prevalence of biopsy-confirmed
malignancy was between 53% and 95%. In contrast, 20% to 30% positive biopsies are
commonly reported in the literature (Basset et al. 1991). PET studies consistently described that
patients were selected for having suspicious mammograms or palpable masses. These patients
are in the upper part of the spectrum of the overall biopsy population. This characterization is
supported by evidence from the PET studies that mean tumor size across studies ranged from 2
cm to 4 cm (i.e., relatively large tumors). The lower part of the biopsy population spectrum
corresponds with prevalence for malignancy between 20% and 50%, in which tumors are smaller
and nonpalpable. Such patients would often be referred for biopsy with indeterminate
mammograms.

The published studies that are so far available on the diagnostic performance of PET omit a
critical segment of the biopsy population: those with indeterminate mammograms and small
nonpalpable lesions. The sensitivity of PET in such patients, compared with patients who have
suspicious mammograms or palpable masses, might be as high, but is likely to be lower. Given
the lack of data, however, estimating PET's diagnostic performance in the lower portion of the
biopsy population spectrum would be speculative. In the analysis that follows, assumptions and
extrapolations that are not evidence-based are avoided. In the absence of evidence on diagnostic
performance in this important segment of the population, it is not possible to generalize from the
study populations to those with prevalence of malignancy lower than 50%.

Summary of Available Evidence

Table 1 provides details about the methods and results of the 13 articles that are included in the
data synthesis. Five additional articles met study selection criteria, but were excluded from data
synthesis because the patient population may overlap with a later report from the same institution
(Avril et al. 1997; 1996b; Yutani et al. 1999; Adler et al. 1993; Tse et al. 1992).[1] Two reports
from the M.D. Anderson Cancer Center are included for synthesis because the more recent report
had fewer patients than the earlier one and patient selection appeared different (Bassa et al.
1996; Nieweg et al. 1993) .

Table 2 summarizes information across the 13 studies included in the data synthesis (total
n=606). Three studies used the lesion as the unit of analysis (n=191 patients, 238 lesions), while

---

[1] These 5 studies reported on a total of 145 to 196 patients. Inclusion of the 5 studies does not change the results,
but confidence intervals are slightly narrower when the larger population is included in the synthesis.

Table 1. Differential Diagnosis of Suspicious Breast Lesions

Study	N	Design	Patient Selection	Mean Age (SD)	Mean T Size (SD)	PET Interp	Avoid Verif Bias	PET: RS Blind	RS: PET Blind	Ref Stand	Test/ Site	UA	TP	FN	FP	TN	Prev Dis	Sens	Spec
Schirrmeister et al. (2001) Ulm, Germany	117	pro	palpable breast masses/suspicious MM/US	56.8		qual	?	Y	?	hist	PET	P	83	6	7	21	76%	93%	75%
Avril et al. (2000) Munich, Germany	144 185 L	pro	abn MM or palpable mass, no prev tx < 3 mo	50.6 (10.3)	3.1 (2.1)	qual, AC	?	Y	?	hist	PET-conventl	L	85	47	3	50	71%	64%	94%
											PET-sensitive	L	106	26	13	40	71%	80%	75%
Murthy et al. (2000) McGill University, Montreal	16	pro	suggestive mass	54.8 med		qual, semi-quant	?	?	?	hist	MM	P	9	1	2	2	71%	90%	50%
											PET	P	8	2	0	4	71%	80%	100%
Yutani et al. (2000) Osaka, Japan	40	pro	consec pts w/ suspicious lesions on PE, MM, US	50.9 (13.4)	2.1 (1.0) T1a-3% T1b-18% T1c-29% T2-50%	qual, AC	Y	Y	?	hist	MM	P	28	10				74%	
											US	P	33	5				87%	
											MIBI-SPECT	P	29	9	1	1	95%	76%	50%
											PET	P	30	8	0	2	95%	79%	100%
Rostom et al. (1999) Saudi Arabia	93	retro	consec pts attending breast clinic	40.3		qual, AC in half	Y	Y	?	hist (86%) FNAB (14%)	PET	P	68	7	3	15	81%	91%	83%
Noh et al. (1998) Seoul, South Korea	27 31 L	?	breast mass, pts underwent PET and had hist/cytol data		2.0 (med)	?, AC	?	?	?	hist (26), FNAB (1)	Palpation	L	20	2	5	4	71%	91%	44%
											MM	L	15	6	4	2	78%	71%	33%
											PET	L	22	0	1	8	71%	100%	89%

Abbreviations Key: See Appendix

Table 1. Differential Diagnosis of Suspicious Breast Lesions (cont'd)

Study	N	Design	Patient Selection	Mean Age (SD)	Mean T Size (SD)	PET Interp	Avoid Verif Bias	PET: RS Blind	RS: PET Blind	Ref Stand	Test/Site	UA	TP	FN	FP	TN	Prev Dis	Sens	Spec
Kole et al. (1997) Groningen, Netherlands	13	pro	abn MM & palpable breast mass		3.8 (2.3) T1-10% T2-60% T4-30%	quant, AC in 7	?	?	?	hist	PET	P	10	0	1	2	77%	100%	67%
Palmedo et Al. (1997) Bonn, Germany	20 22 L	pro	abn MM or palpable Mass	58.4	2.8 (1.6) T1b-14% T1c-29% T2-43% T3-14%	qual, quant, AC	?	Y	?	hist	MM	L	10	3	2	4	68%	77%	67%
											MIBI-SMM	L	12	3	1	5	71%	80%	83%
											PET	L	12	3	1	5	71%	80%	83%
Bassa et al. (1996) MD Anderson, Houston	16	retro	consec pts, locally advanced breast ca to receive neoadju-vant CTX (15), no hepatic mets on CT	43.8 (9.5)	T2-12% T3-50% T4-38%	qual, AC	Y	?	?	hist	MM-pretx	P	10	6	0	1	94%	63%	100%
											US-pretx	P	14	2	0	1	94%	88%	100%
											PET-pretx	P	16	0	0	1	94%	100%	100%

Abbreviations Key: See Appendix

Table 1. Differential Diagnosis of Suspicious Breast Lesions (cont'd)

Study	N	Design	Patient Selection	Mean Age (SD)	Mean T Size (SD)	PET Interp	Avoid Verif Bias	PET: RS Blind	RS: PET Blind	Ref Stand	Test/ Site	UA	TP	FN	FP	TN	Prev Dis	Sens	Spec
Scheidhauer Et al. (1996) Cologne, Germany	30	pro	suspicion of breast ca based on palpation, MM, US; scheduled for surgery	57.0	T1-52% T2-9% T3/4-39%	qual, AC	?	Y	?	hist	Palpation MM/US PET	P P P	17 18 21	6 3 2	2 5 1	5 1 6	77% 78% 77%	74% 86% 91%	71% 17% 86%
Crowe et al. (1994) University Hospitals, Cleveland	37	pro	breast lesion ≥ 1 cm on palpation or MM, requiring pathologic dx	55.0 (14.0)	2.9 (1.5)	qual, AC	?	Y	?	hist	PE MM PET	P P P	19 14 23	4 9 0	5 0 0	5 10 10	70% 70% 70%	83% 61% 100%	50% 100% 100%
Hoh et al. (1993) UCLA	34	retro	underwent whole-body PET and had correlative tissue Biopsy			qual	?	?	?	hist	PET	P	24	2	2	6	76%	92%	75%
Nieweg et al. (1993) MD Anderson, Houston	19	pro	pts w/ breast ca, no evidence of malignancy, fibrocystic disease or healthy Volunteers	49.0 (med)	3.6 (2.8) T1c-18% T2-64% T3-18%	?, AC	?	?	?	hist	PET	P	10	1	0	8	58%	91%	100%

Abbreviations Key: See Appendix

Table 2. Differential Diagnosis of Suspicious Breast Lesions, Summary

Unit	Study n	Pts n	Design			Avoid Verif Bias			Blinding PET: RS			Blinding RS:PET			Sensitivity Range		Specificity Range		Random Effects Meta-Analysis					
			Pro	Retro	?	Y	N	?	Y	N	?	Y	N	?					Sensitivity			Specificity		
			Study n			Study n			Study n			Study n								95% CI			95% CI	
L	3	191	2	-	1	-	-	3	2	-	1	-	-	3	80%	100%	75%	89%						
		238 L																						
P	10	415	7	3	-	3	-	7	5	-	5	-	-	10	79%	100%	75%	100%	89%	84%	93%	80%	70%	87%
All	13	606	9	3	1	3	-	10	7	-	6	-	-	13	79%	100%	50%	100%	88%	83%	92%	79%	71%	85%

Abbreviations Key: See Appendix

10 studies used the patient as the unit of analysis (n=415 patients). Study design was prospective in 9 studies, retrospective in 3 studies and unclear in 1.

The assessment of study quality suggested that 3 studies were free of verification bias, and 10 provided insufficient information to make a determination. In 7 studies, it was clear that interpreters of PET images were blinded to reference standard results, and it was unclear in 6 studies. None of the 13 studies mentioned whether investigators who assessed the reference standard were blinded to PET results. Attenuation correction (AC) was used for all patients in 8 studies and in half the patients in 2 studies, representing 63% of all patients. In 12 studies, at least 90% of the patients in the study sample had a histopathologic reference standard, while the proportion was 86% in the remaining study.

The spectrum of disease addressed in the available studies tends toward relatively large tumors. Mean tumor size across studies ranged from about 2 cm to 4 cm. In addition, all studies included a majority of patients with malignancies. The prior probability of malignancy in the study samples ranged between 53% and 95%. As noted above, these observations suggest that estimates of diagnostic performance derived from these studies may not be applicable to populations with smaller tumors.

Sensitivity estimates in all 13 studies ranged from 79% to 100% and specificity estimates were between 50% and 100%.

Meta-analysis

Meta-analysis was performed with Meta-Test software (Lau 1997). A random effects model was used (Figure 2, Table 2), producing a pooled sensitivity estimate of 88% (95% CI: 83%, 92%) and a pooled specificity estimate of 79% (95% CI: 71%, 85%). Sensitivity analysis based on higher quality studies, defined as prospective, free of verification bias and used blinded interpretation of PET, was initially planned. However, only 1 study met these qualifications (Yutani et al. 2000, n=40, 3 benign cases), thus meta-analysis was not possible using this definition of a high-quality study.

Random effects meta-analysis (REM) summary estimates do not take into account the dependency between sensitivity and specificity. In contrast, a summary receiver operating characteristic (ROC) curve is a meta-analytic tool that does account for the dependent relationship (Figure 3, Meta-Test software, Lau 1997). Random effects meta-analysis underestimates diagnostic performance compared with a summary ROC curve.   The summary ROC curve derived from inputs weighted by the inverse of study variance is used in this analysis.

The summary ROC curve technique assumes that different points on the curve represent different thresholds for test positivity. While it would be tempting to assume that an ROC curve point with very high sensitivity could be selected, such a point has relevance only if clinicians can actually achieve such high sensitivity in practice. A test such as PET is generally interpreted qualitatively and relies heavily on the binary determination of lesion detection. It may be

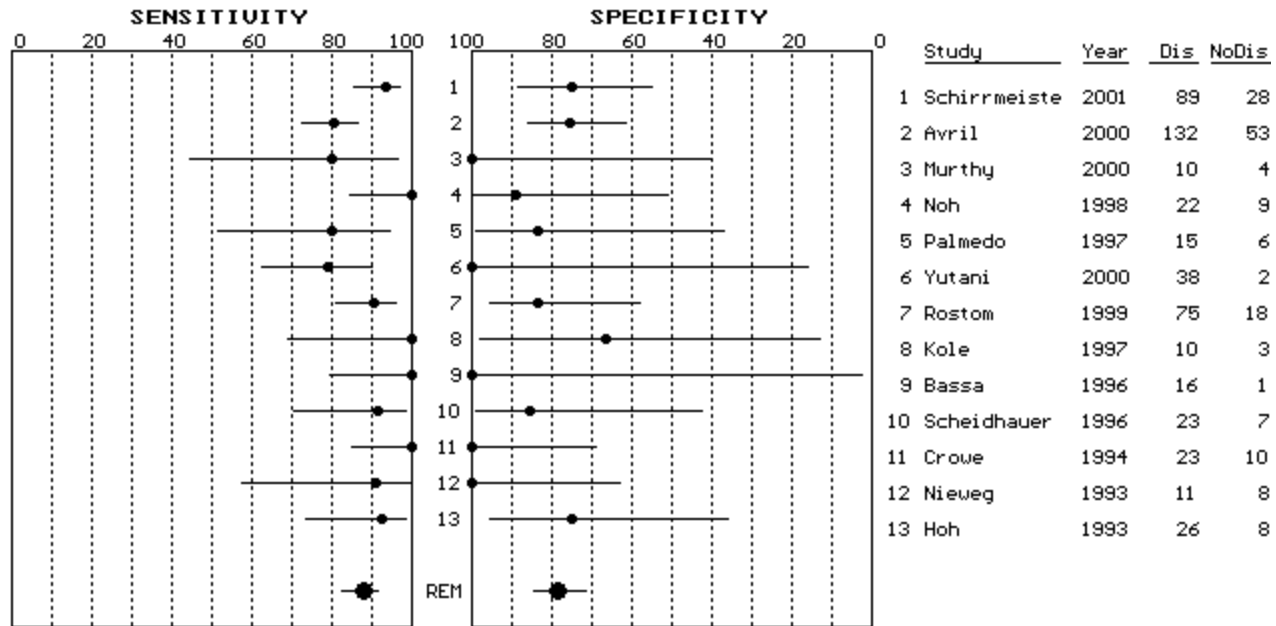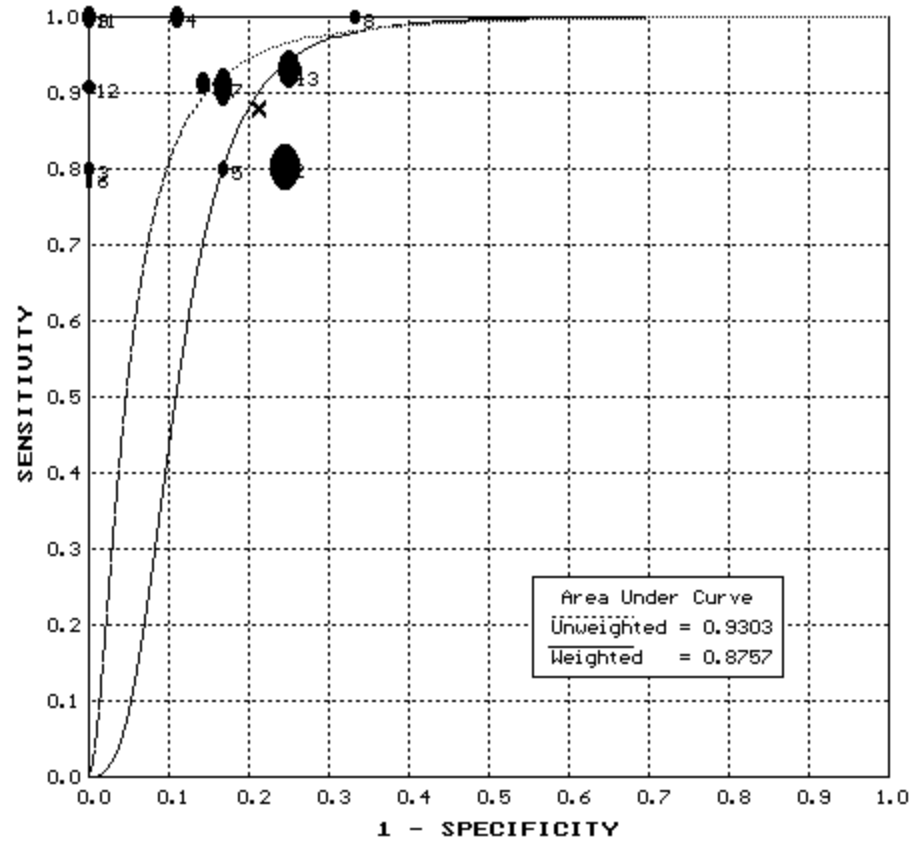**Figure 2: Meta-analysis, Differential Diagnosis of Suspicious Breast Lesions, Random Effects Model**

Figure 3: Summary ROC Curve, Differential Diagnosis of Suspicious Breast Lesions

X: Random Effects Meta-Analysis Estimates

difficult for clinicians to adjust qualitative PET criteria to achieve high sensitivity. The feasibility of adjusting the interpretation criteria for PET to achieve highest sensitivity must first be empirically demonstrated before assuming that extreme points of the ROC curve can be achieved. Therefore, it is critical to select points on the summary ROC curve with caution.

An average point on the summary ROC curve, in the region where multiple studies cluster, would be more representative of the performance achievable in actual practice. The REM findings can be considered averages, but again they underestimate diagnostic performance. This can be seen in Figure 3, which shows that the point representing the REM estimates is located slightly below the summary ROC curve. To find a representative point that does not underestimate diagnostic performance, a logical choice is the point on the summary ROC curve nearest to the REM point. This point on the curve has a sensitivity of 89% and a specificity of 80%.

Analysis of Outcomes of Using PET to Select Patients for Biopsy

In order to be used to avoid biopsy, PET should provide a highly sensitive evaluation for malignancy. The rate of false-negative PET results weighs heavily in considering whether the risk of delayed or missed diagnosis of breast cancer is worth the benefit of avoiding biopsy of a benign lesion. The risks and benefits of PET can be viewed from two perspectives: (1) the entire population of patients undergoing PET and (2) the individual patient who has a negative PET result.

The population perspective can be appreciated by imagining that a patient is given a choice between allowing the PET result to guide whether to perform biopsy (i.e., if PET is positive, do biopsy; if PET is negative, avoid biopsy), versus proceeding directly to biopsy, without using PET. If the patient chooses to allow PET to guide the biopsy decision, Figure 1 shows that the only outcomes that would differ from those of the path involving no PET scanning are those associated with PET true negatives and false negatives. A true-negative PET result would allow a patient to safely avoid the pain and anxiety of biopsy. A false-negative PET result would mean an undetected malignancy and delayed diagnosis and treatment. The patient would have to decide whether a given probability of benefit (avoiding the harms of biopsy) is worth a given probability of harm (delayed diagnosis/treatment). In other words, the patient must decide whether there is an acceptable risk-benefit trade-off. The patient may decide that the risk-benefit trade-off is unacceptable and will need no more information, from the perspective of an individual with a negative PET scan, to decide against using PET to decide whether to biopsy. Instead, the patient may decide to proceed directly to biopsy.

From the perspective of an individual with a negative PET scan, the patient has already made the choice to undergo PET scanning and is considering using its results to guide whether to undergo biopsy. Once the negative result is known, the probabilities of true-negative and false-negative PET results change, compared with the probabilities viewed from the population perspective. The reason for the change can be understood by focusing on the denominator used in calculating these probabilities. From the population perspective, probabilities of true negatives and false negatives are calculated as proportions of the entire group of patients who undergo PET scanning, regardless of whether patients test positive or negative on PET (i.e., a bigger

denominator). From the perspective of a patient with a known negative PET scan, the probabilities of true negatives and false negatives are calculated as proportions of only the patients with negative PET scans (a smaller denominator). From either perspective, the absolute numbers of true negatives and false negatives remain the same, however the dominators differ. Thus, if the probability of a false negative result has a certain value from the population perspective, the probability of a false negative from the perspective of an individual with a negative PET scan will be higher, given the smaller denominator. Although a patient may decide that the risk-benefit trade-off is acceptable from the population perspective, the same patient later found to have a negative PET scan may decide that the increased false negative probability is unacceptably high, thus choosing to undergo biopsy.

For both the population and the individual perspectives, we assume that sensitivity of PET is 89%, and specificity is 80%, derived from the summary ROC curve in Figure 3. It is also assumed that the prevalence (i.e., pre-test probability) of malignancy ranges from 50% to 75%. As noted previously, evidence is lacking about PET's diagnostic performance for smaller tumors and in patient populations with disease prevalence lower than 50%. An upper prevalence limit of 75% was selected because it seems unlikely that patients with higher pretest probability of malignancy would forgo biopsy, even if PET results were negative. As the prevalence of malignancy rises from 50% to 75%, the false negative risks also rise, making the probabilities of harm from delayed diagnosis and treatment higher.

Population Perspective

The population perspective (Table 3) estimates the probabilities of outcomes of using PET to decide whether to perform biopsy, assuming that the results of PET are not yet known. All PET results (both positive and negative) are considered and the proportions of the entire population deriving benefits and harms can be estimated.

The columns headed "Biopsy (No PET)" and "PET" represent the management strategies described in Figure 1. The rows define prevalences of malignancy from 50% to 75%. The intersection of a specified prevalence of malignancy and the management strategies compares the outcomes of each management path given that prevalence of malignancy. Of the 4 outcomes listed, the false-negative and true-negative outcomes are of most interest. The false-negative outcome reflects the proportion of patients who would be exposed to the harms of delayed treatment. The true-negative outcome reflects the proportion of patients who would benefit by avoiding biopsy of a benign lesion.

When the prevalence of malignancy is 50% and PET is used in deciding whether to perform a biopsy, 40% of all patients would benefit by avoiding the harms of a negative biopsy (when PET is truly negative). The risk of a false-negative result, leading to delayed diagnosis and treatment, is 5.5%. When the prevalence is 75%, 20% of patients avoid biopsy of a benign lesion; and the risk of delayed treatment is 8.25%.

Table 3. Summary of Benefits and Harms, Population Perspective, Differential Diagnosis of Suspicious Breast Lesions

Prev Malign	Health Outcome		Biopsy (No Pet)	PET
50%	Benefit	PET TP/biopsy positive: receive appropriate treatment	50.0%	44.50%
	Harm	**PET FN: delay treatment**		**5.50%**
	Benefit	**PET TN: avoid biopsy morbidity**		**40.00%**
	Harm	PET FP/biopsy negative: exposed to biopsy morbidity	50.0%	10.00%
55%	Benefit	PET TP/biopsy positive: receive appropriate treatment	55.0%	48.95%
	Harm	**PET FN: delay treatment**		**6.05%**
	Benefit	**PET TN: avoid biopsy morbidity**		**36.00%**
	Harm	PET FP/biopsy negative: exposed to biopsy morbidity	45.0%	9.00%
60%	Benefit	PET TP/biopsy positive: receive appropriate treatment	60.0%	53.40%
	Harm	**PET FN: delay treatment**		**6.60%**
	Benefit	**PET TN: avoid biopsy morbidity**		**32.00%**
	Harm	PET FP/biopsy negative: exposed to biopsy morbidity	40.0%	8.00%
65%	Benefit	PET TP/biopsy positive: receive appropriate treatment	65.0%	57.85%
	Harm	**PET FN: delay treatment**		**7.15%**
	Benefit	**PET TN: avoid biopsy morbidity**		**28.00%**
	Harm	PET FP/biopsy negative: exposed to biopsy morbidity	35.0%	7.00%
70%	Benefit	PET TP/biopsy positive: receive appropriate treatment	70.0%	62.30%
	Harm	**PET FN: delay treatment**		**7.70%**
	Benefit	**PET TN: avoid biopsy morbidity**		**24.00%**
	Harm	PET FP/biopsy negative: exposed to biopsy morbidity	30.0%	6.00%
75%	Benefit	PET TP/biopsy positive: receive appropriate treatment	75.0%	66.75%
	Harm	**PET FN: delay treatment**		**8.25%**
	Benefit	**PET TN: avoid biopsy morbidity**		**20.00%**
	Harm	PET FP/biopsy negative: exposed to biopsy morbidity	25.0%	5.00%

Individual Perspective

The individual patient perspective (Table 4) estimates the probabilities of outcomes of using PET to decide whether to perform biopsy, assuming that the patient has had a negative PET scan. From this patient's perspective, the crucial measure is PET's negative predictive value (NPV). The NPV is the probability that a negative test result has correctly assessed that the patient has no disease.

Table 4 shows NPVs across the selected range for prevalence of malignancy. At a prevalence of 50%, the NPV is 87.9%, thus, the false-negative risk is 12.1%. For an individual with a negative PET scan, a 12% chance of missed/delayed diagnosis of breast cancer is most likely too high to make the 88% chance of avoiding an negative biopsy of a benign lesion worthwhile. At the 75% prevalence, a 29.2% risk of missed or delayed diagnosis is surely unacceptable in order to avoid a biopsy.

Recall that false-negative rates differ from the population perspective and the perspective of an individual with a negative PET scan. Thus, the rates of false negatives differ between Table 3 and Table 4, given the same diagnostic performance and prevalence of malignancy. The rates from Table 3 assume that PET results are not known (population perspective). The rates from Table 4 assume a patient with a known negative PET result, and are calculated according to Bayes' theorem.

CONCLUSIONS - INDICATION Ia: INITIAL DIAGNOSIS OF BREAST CANCER IN PATIENTS WITH ABNORMAL MAMMOGRAMS OR PALPABLE MASSES

1. **Does the available evidence permit conclusions about the diagnostic performance of PET for differential diagnosis of breast lesions among patients with abnormal mammograms or palpable masses?**

In studies of PET for differential diagnosis of breast lesions, patients were selected for having suspicious mammograms or palpable masses. These study samples have a notably higher prevalence of malignancy than that reported for the general population and a relatively large average tumor size at initial diagnosis. These studies represent the upper part of the biopsy population spectrum. No published studies are available on the diagnostic performance of PET in the lower part of the biopsy population, comprising a range of prevalence between 20% and 50%. This group consists of patients with indeterminate mammograms and smaller, nonpalpable lesions. Without of evidence on diagnostic performance of PET in the lower portion of the biopsy population, no conclusions can be reached and it would be imprudent to generalize from the studied population.

Thirteen studies (total n=606) met study selection criteria for inclusion in the data synthesis. The prior probability of malignancy in the study samples ranged between 53% and 95%, compared to 20% to 30% in the general population. Mean tumor size across studies was relatively large, ranging from about 2 cm to 4 cm.

Table 4. PET Negative Predictive Values by Prevalence of Malignancy and Diagnostic Performance Estimates, Individual Perspective, Differential Diagnosis

$$NPV = \frac{(prev\ benign)(spec)}{((prev\ benign)(spec) + (prev\ malign)(1\text{-}sens))}$$

Sens	Spec
89%	80%

Prevalence of Malignancy	NPV	FN Rate
50%	87.9%	12.1%
55%	85.6%	14.4%
60%	82.9%	17.1%
65%	79.7%	20.3%
70%	75.7%	24.3%
75%	70.8%	29.2%

Sensitivity estimates in all 13 studies ranged from 79% to 100% and specificity estimates were between 50% and 100%. Meta-analysis was first performed using a random effects model. The pooled sensitivity estimate was 88% (95% CI: 83%, 92%) and the pooled specificity estimate was 79% (95% CI: 71%, 85%). Then a summary receiver operating characteristic (ROC) curve was constructed which accounts for the dependent relationship between sensitivity and specificity. A point on the summary ROC curve was selected which reflected average performance, with an estimated sensitivity of 89% and a specificity of 80%.

Sensitivity analysis based on higher quality studies, defined as prospective, free of verification bias and used blinded interpretation of PET, was initially planned. However, only 1 study met these qualifications (n=40), thus precluding the planned analysis.

2. **Can the use of PET improve outcomes by obviating biopsy in patients who have been referred for biopsy due to an abnormal mammogram or palpable mass?**

In order to be used to avoid biopsy, PET should provide a highly sensitive evaluation for malignancy. The rate of false negative PET results weighs heavily in considering whether the risk of delayed or missed diagnosis of breast cancer is worth the benefit of avoiding biopsy of a benign lesion. The risks and benefits of PET were analyzed using two perspectives: (1) the entire population of patients undergoing PET and (2) the individual patient who has a negative PET result.

For both analyses, sensitivity of PET was assumed to be 89% and specificity was 80%. The prevalence (i.e., pre-test probability) of malignancy was assumed to range from 50% to 75%. Evidence is lacking about PET's diagnostic performance for smaller tumors and in patient populations with disease prevalence lower than 50%. As the prevalence of malignancy rises from 50% to 75%, the false-negative risks also rise, making the probabilities of harm from delayed diagnosis and treatment higher

The population perspective assumes that the results of PET are not yet known. All PET results (both positive and negative) are considered and the proportions of the entire population deriving benefits and harms can be estimated. When the prevalence of malignancy is 50%, 40% of all patients would benefit by avoiding the harms of negative biopsy. The risk of a false-negative result, leading to delayed diagnosis and treatment, is 5.5%. When the prevalence of malignancy is 75%, 20% of patients avoid biopsy of a benign lesion; and the risk of delayed treatment is 8.25%.

From the perspective of an individual patient with a negative PET scan, the risk of a false-negative result is higher than for the entire population undergoing PET scanning. When the prevalence of malignancy is 50%, the NPV is 87.9%, thus, the false-negative risk is 12.1%. For an individual with a negative PET scan, a 12% chance of missed or delayed diagnosis of breast cancer is most likely too high to make the 88% chance of avoiding an negative biopsy of a benign lesion worthwhile. When the prevalence of malignancy is 75%, there is a 29.2% risk of missed or delayed diagnosis, which is surely unacceptable in order to avoid a biopsy.

Evidence is lacking to assess the negative predictive value of PET in the population of patients referred for biopsy with indeterminate mammograms and smaller, nonpalpable lesions. Such patients would have a prevalence of malignancy from 20% to 50%.

Overall Conclusion

Evidence on the diagnostic performance of PET for differential diagnosis of breast lesions among patients with abnormal mammograms or palpable masses is lacking for patients with indeterminate mammograms and small, nonpalpable lesions (low prevalence for malignancy). Among study populations of patients with higher prevalence of malignancy, risk of a false-negative diagnosis is likely too high relative to the benefit of avoiding biopsy of a benign lesion. A false-negative PET result may cause a missed or delayed diagnosis of breast cancer and associated delay in treatment.

From the perspective of an individual patient with a prior probability of malignancy of 50% and a negative PET result, the risk of a false-negative result PET is 12.1%. At the 75% prevalence, there is a 29.2% risk of a false-negative finding. Evidence on PET diagnostic performance is unavailable to permit estimation of the risk of a false-negative PET result in the patients with a prevalence of malignancy from 20% to 50%.

REVIEW OF EVIDENCE – INDICATION Ib: INITIAL DIAGNOSIS OF BREAST CANCER IN PATIENTS WITH LOW SUSPICION FINDINGS WHO HAVE BEEN REFERRED FOR SHORT-INTERVAL FOLLOW-UP

Evidence on Diagnostic Performance of PET

No studies meeting selection criteria included a population of patients who were selected because of referral for short-interval mammographic follow-up due to low suspicion mammograms. The diagnostic performance characteristics achieved by FDG-PET in the available study patient populations referred for biopsy with mammographic findings that are suspicious or highly suggestive of malignancy, should not be generalized to the setting of low suspicion mammographic findings referred for short-interval mammographic follow-up. This is primarily because the spectrum of lesions and imaging abnormalities in the low suspicion, short-interval follow-up setting may be considerably more subtle and smaller in size compared to the biopsy referral setting where most patients had relatively large tumors, averaging 2–4 cm in size.

Analysis of Outcomes of Using PET to Select Patients for Biopsy

A direct evaluation of the diagnostic performance of FDG-PET in the appropriately selected population of patients is necessary to determine its efficacy in guiding decisions to perform short-interval follow-up or biopsy. Furthermore, a direct comparison of diagnostic findings and management decisions guided by short-interval mammographic follow-up compared directly with FDG-PET in the same group of patients is needed to determine which diagnostic strategy is preferred.

CONCLUSIONS – INDICATION Ib: INITIAL DIAGNOSIS OF BREAST CANCER IN PATIENTS WITH LOW SUSPICION FINDINGS WHO HAVE BEEN REFERRED FOR SHORT-INTERVAL FOLLOW-UP

1. **Does the available evidence permit conclusions about the diagnostic performance of PET for differential diagnosis of breast lesions among patients with low suspicion finding on mammography and other routine imaging procedures who have been referred for short-interval follow-up?**

No studies meeting selection criteria included a patient population to address this question. Performance of PET in the available studies in patients who have been referred for biopsy due to an abnormal mammogram or palpable mass cannot be generalized to patients with low suspicion findings on mammography referred for short interval follow-up.

2. **Can the use of PET improve outcomes by leading to earlier and more accurate diagnosis of breast cancer compared to short-interval mammographic follow-up in patients with low suspicion findings on mammography and other routine imaging procedures?**

This question cannot be addressed in the absence of data on the diagnostic performance of PET in the population of interest.

PART II: INITIAL STAGING OF AXILLARY LYMPH NODES

BACKGROUND – PART II

In patients with an initial diagnosis of primary breast cancer, staging evaluation of the axillary lymph nodes is used to define prognosis and to determine appropriate therapy. In addition to providing information on nodal status, axillary lymph node dissection (ALND) may be therapeutic, if removing tumor-involved nodes improves local control. This hypothesis is being investigated by ongoing clinical trials. Results of the National Surgical Adjuvant Breast and Bowel Project (NSABP) B-04 trial did not find that axillary dissection improves survival, although the data were insufficient to rule out the possibility of a small, but clinically meaningful, benefit (Rockette et al. 1982).

The presence of tumor in the axillary nodes frequently influences the decision to use adjuvant systemic chemotherapy and/or hormonal therapy, although patients with high-risk primary lesions may be offered adjuvant therapy regardless of axillary node status. Furthermore, the number of positive axillary nodes also influences the selection of more aggressive treatment. The presence of 4 or more positive axillary nodes may indicate the need for radiation therapy. However, the evidence on axillary node status and treatment is evolving; at present, there is uncertainty as to optimal treatment and considerable variation in current practice (personal communication, Henderson C and Abrams J).

Chronic lymphedema is common following ALND, and strategies for reducing the morbidity of axillary node staging are being developed. Sentinel node biopsy (SNB), a more limited surgical approach to axillary lymph node staging, has been introduced as an alternative surgical technique. More recently, PET has been proposed as a noninvasive method for determining the presence of axillary lymph node involvement and for selecting patients for ALND.

Following is background information on the outcomes of adjuvant chemotherapy and hormonal therapy. SNB, as well as recent reports on the diagnostic performance of SNB, are also described. Finally, a brief overview of the morbidity of ALND and of SNB is also presented. This background provides context for considering the potential role of PET in axillary lymph node staging.

Adjuvant Chemotherapy or Hormonal Therapy

Adjuvant systemic therapy has been reported to reduce recurrence and improve survival in patients with breast cancer. Improved outcomes occur in patients with positive axillary nodes and also in patients without axillary involvement. However, the absolute magnitude of the reduction in recurrence rate or mortality is greater for those with axillary nodal disease. Compared to patients without axillary involvement, node-positive patients are at a greater baseline risk of recurrence and disease-related mortality and thus, have greater potential for benefit (Henderson 1994). In 1992, the Early Breast Cancer Trialists' Collaborative Group published an overview of 133 randomized clinical trials results on the effect of adjuvant chemotherapy or hormonal therapy (Early Breast Cancer Trialists' Collaborative Group 1992).

<u>Patients with Positive Axillary Nodes</u>. For women with positive axillary nodes, receiving chemotherapy or hormonal therapy reduces the odds of recurrence by approximately one-third and improves survival. Data on median survival suggests about a 2-year average prolongation in life for treated patients compared to controls. For premenopausal women younger than age 50, the absolute differences in 10-year survival rates for treated and control patients are 12% for disease-free survival and 10% for overall survival. A separate analysis including all ages found a 6.8% absolute difference in 10-year survival associated with chemotherapy.

For postmenopausal women older than age 50 with positive nodes, tamoxifen reduces the odds of recurrence by $\geq$29% and the odds of death by $\geq$20%. Absolute differences in 10-year survival rates for treated and control patients are 9% for disease-free survival and 7% for overall survival. When age is not restricted, absolute difference in 10-year survival is 8.2%.

<u>Patients with Negative Axillary Nodes</u>. For women without involved axillary nodes, chemotherapy resulted in a 29% reduction in annual odds of recurrence and 16% reduction in annual odds of death. Absolute difference in survival at 10 years was 4%. Similarly, tamoxifen treatment resulted in a 27% reduction in annual odds of recurrence and 17% reduction in annual odds of death. Absolute difference in survival at 10 years was 3.5%.

Decisions on the use of adjuvant therapy in patients with node-negative disease is complicated by uncertainties in balancing potential benefits and toxicity of systemic therapy, as well as by variation in patient preferences (Fisher 1999). Controlled studies in node-negative patients have found an overall treatment benefit in survival and disease-free survival with systemic therapy. There is continued interest in determining whether some subgroups of patients benefit while others do not, as reliable predictors for treatment response are not well established.

<u>Patient Preferences</u>. Simes and Margrie (1991, n=104; cited in Henderson 1994) studied patient preferences regarding adjuvant chemotherapy. Patients who had all experienced the toxicity of chemotherapy were asked if they would undergo chemotherapy again under hypothetical circumstances defined by expected likelihood of recurrence and/or death. A majority of patients felt the benefits of therapy outweighed the toxicity, but rates of favorable responses were sensitive to differences in estimated prognosis. For example, 77% of women would choose chemotherapy if it were associated with an increase in survival from 5 to 6 years; while 52% would choose chemotherapy if it were associated with an increase in survival from 15 to 16 years. Even when told that chemotherapy would improve expected survival from 5 years up to 10 years, a small minority of patients (2%) would not choose chemotherapy again (Henderson 1994).

Sentinel Node Biopsy

Sentinel node biopsy is an emerging technology that has been used as an alternative to performing complete ALND in patients requiring axillary staging. SNB may be an attractive alternative to ALND, in that patients undergo a procedure that is less extensive and less morbid. The indications for SNB are evolving, but certain patients may continue to be candidates for full ALND, such as those with advanced disease, palpable axillary lymph nodes and multicentric disease.

Morbidity of ALND. While there is uncertainty about whether axillary lymph node dissection produces intrinsic therapeutic benefit, its potential morbidity is well established. The morbidity of axillary lymph node dissection depends on the extent of removal of lymph nodes, which are categorized into three levels: I (most superficially located), II, and III (most deeply located). Most staging axillary lymph node dissections involve removal of at least 10 nodes from levels I and II, which is called a partial ALND.

Reported procedure-related morbidity from partial ALND varies. Chronic lymphedema is estimated to occur in 8–25% of patients, with half of patients experiencing chronic pain. Other adverse effects of axillary lymph node dissection include wound complications (8%) and limitations in shoulder movement (2%). Major complications such as injury to axillary motor nerves or the axillary vein are infrequent (Tasmuth et al. 1999; Spillane and Sacks 1999; Winer et al. 2001).

Sentinel node biopsy. SNB uses a tracer injected into the breast tissue around the primary tumor, which drains through the lymphatics toward the axilla. The earliest lymph node identified by the tracer is designated the "sentinel" node. The sentinel node is surgically removed and directly analyzed for presence of tumor. It is postulated that the absence of tumor in the sentinel node can reliably predict the absence of tumor in the axilla. Unlike axillary node dissection, SNB does not seem to be associated with chronic lymphedema and has a very low morbidity (Giuliano et al. 2000).

Commonly used tracers include a visual blue dye that is visible to the surgeon or a radioactive colloid tracer that can be detected with a gamma camera for overall imaging or hand-held gamma detector probe for localization. Either one of these tracers or both together have been used in the available studies of SNB, and there is continued evolution as to the optimal technique of performing the procedure.

Diagnostic Performance of SNB. A review of the available literature reporting diagnostic performance of SNB in patients with no evidence of palpable axillary adenopathy was conducted to provide supplemental information for this assessment. The included studies were restricted to those that performed both SNB and ALND on at least 30 patients with nonpalpable nodes and provided sufficient information to identify the numbers of joint events between SNB and reference standard (ALND) results. Institutions publishing multiple articles are represented by the report with the largest patient series.

Table 5 provides information on 21 studies (total n=3,021) identified by the search. Studies used a variety of sentinel node mapping techniques. Of the 21, 5 used a radiocolloid (RC) tracer along with a blue dye in all patients. Another 14 studies used the RC tracer alone, 1 used blue dye alone and 1 study used a mix of techniques. Of the 20 studies that used a RC tracer, 19 tracked it with a hand-held gamma detector probe (GDP) and 12 used lymphoscintigraphy.

Table 5. Axillary Sentinel Lymph Node Biopsy

Study	n	Patient Selection	Mean Age	Mean Size/ T Stage	Mapping Tracer/Dye	Injection Site	Pathology Method	% Succ	TP	FN	FP	TN	Prev LN+	Sens	Spec	NPV
Altinyollar et al. (2000) Ankara, Turkey	60	stage I/II br ca, cN0	51 (med)	T1-32% T2-68%	Blue dye	4 IP sites around bx cavity	frozen sect, H&E, IHC	81.7%	19	3	0	27	44.9%	86.4%	100.0%	90.0%
Bianchi et al. (2000) Genoa, Italy	30	consec pts, cN0	62.5 (med)		99m-Tc HSA microcolloids; LS, 6 injection site grps	SD or peritu-moral	6 frozen sects, H&E	100.0%	10	2	0	18	40.0%	83.3%	100.0%	90.0%
Casalegno et al. (2000) Turin, Italy	102	T1/T2 N0 br ca	57.3 (11.2)	1.9 (0.8)	99m-Tc HSA colloid, LS, GDP	SD	4 stand sects, H&E , ≤ 20 serial sects if stand neg	86.3%	35	2	0	51	42.0%	94.6%	100.0%	96.2%
Doting et al. (2000) Groningen/Amsterdam, Netherlands	136	palpable br ca, cN0	59	T1-52% T2-45% T3-4%	99m-Tc nan-ocolloid, LS, GDP, blue dye	intratu-moral	3 sects, H&E, IHC	92.6%	56	3	0	67	46.8%	94.9%	100.0%	95.7%
Galli et al. (2000) Biella, Italy	46	T1/T2 N0 br ca	63.5	1.4	99m-Tc HSA, LS, GDP	SD		95.7%	12	3	0	29	34.1%	80.0%	100.0%	90.6%
Gucciardo et al. (2000) Rome, Italy	50	monofocal < T3 br ca, cN0			99m-Tc microcolloids, LS, GDP	SD		86.0%	13	5	0	25	41.9%	72.2%	100.0%	83.3%

Abbreviations Key: See Appendix

Table 5. Axillary Sentinel Lymph Node Biopsy (cont'd)

Study	n	Patient Selection	Mean Age	Mean Size/ T Stage	Mapping Tracer/Dye	Injection Site	Pathology Method	% Succ	TP	FN	FP	TN	Prev LN+	Sens	Spec	NPV
Kollias et al. (2000) Adelaide, Australia	169	consec pts, invasive br ca ≤ 5 cm, cN0	60 (med)		99m-Tc sulfur colloid, GDP, LS, blue dye; RC (51), BD (19), RC+BD (99)	4 IP sites		84.0%	48	5	0	89	37.3%	90.6%	100.0%	94.7%
Liu et al. (2000) Taichung, Taiwan	62	operable T1/ T2 cN0 br ca	49.2	T1b-7% T1c-36% T2-52%	1000 nm 99m-Tc sulfur colloid, GDP, LS	SD, 4 sites		93.5%	14	3	0	41	24.1%	82.4%	100.0%	89.8%
Martin et al. (2000) University of Louisville	758	invasive br ca, T1-2 N0	61 (med)	T1-71% T2-26% T3-3%	200 nm 99m-Tc sulfur col-loid, GDP, blue dye	peritu-moral	serial sect, H&E, IHC in some	88.7%	195	12	0	465	30.8%	94.2%	100.0%	97.5%
Olson et al. (2000) Memorial-Sloan Kettering, New York	224	cN0		T1a-8% T1b-17% T1c-47% T2-29%	unfiltered 99m-Tc sulfur colloid, GDP, blue dye	ID/IP	serial sects, H&E, IHC	91.1%	83	9	0	112	45.1%	90.2%	100.0%	92.6%
Pizzocaro et al. (2000) Brescia, Italy	83	consec pts w/ monofocal T1/T2 br ca, cN0		T1a-5% T1b-27% T1c-46% T2-23%	99m-Tc col-loid, 50 nm sulfide < 80/ /200-300 nm HSA (67), LS, GDP	peritu-moral	H&E, immuno-staining	90.4%	23	5	0	47	37.3%	82.1%	100.0%	90.4%

Table 5. Axillary Sentinel Lymph Node Biopsy (cont'd)

Study	n	Patient Selection	Mean Age	Mean Size/ T Stage	Mapping Tracer/Dye	Injection Site	Pathology Method	% Succ	TP	FN	FP	TN	Prev LN+	Sens	Spec	NPV
Burak et al. (1999) Ohio State University	50	br ca, to undergo ALND, cN0		1.6	220 nm 99m-Tc sulfur colloid, LS (24), GDP, blue dye	IP, 4 sites		90.0%	14	0	0	31	31.1%	100.0%	100.0%	100.0%
Czerniecki et al. (1999) University of Pennsylvania	43	br ca, cN0			220 nm 99m-Tc sulfur colloid or albumin, LS, GDP, blue dye	IP	2 sect, H&E, neg had IHC	95.3%	15	0	0	26	36.6%	100.0%	100.0%	100.0%
Moffat et al. (1999) University of Miami/Jackson Memorial	70	unifocal br ca, TMx/SMx + ALND, KPS > 70; cN0	54 (10)	1.8 (1.2) T1-64% T2-33% T3-3%	unfiltered 99m-Tc sulfur colloid, GDP	4 IP sites around tumor, bx cavity	1-2 sect, H&E	88.6%	18	2	0	50	28.6%	90.0%	100.0%	96.2%
Morgan et al. (1999) University of Washington	44	invasive br ca, palpable, operable, cN0	65		blue dye	IP	serial froz paraffin sects, IHC	72.7%	10	2	0	20	37.5%	83.3%	100.0%	90.9%
Veronesi et al. (1999) Milan, Italy	376	operable br ca, cN0	52		99m-Tc HSA colloid, LS, GDP	SD/peri-tumoral	stand hist (60); froz sect, IHC for neg (311)	98.7%	168	12	0	191	48.5%	93.3%	100.0%	94.1%

Table 5. Axillary Sentinel Lymph Node Biopsy (cont'd)

Study	n	Patient Selection	Mean Age	Mean Size/ T Stage	Mapping Tracer/Dye	Injection Site	Pathology Method	% Succ	TP	FN	FP	TN	Prev LN+	Sens	Spec	NPV
Crossin et al. (1998) Indiana University	50	invasive br ca, cN0		T1a/b-20% T1c-50% T2-30%	99m-Tc sulfur colloid, GDP	IP, 4 sites	usual procedure	84.0%	7	1	0	34	19.0%	87.5%	100.0%	97.1%
Krag et al. (1998) NCI Multicenter Study	443	invasive br ca, cN0	56 (12)	1.9 (1.3)	99m-Tc sulfur colloid, GDP	4 IP sites	H&E; IHC not routine	93.2%	101	13	0	291	28.1%	88.6%	100.0%	95.7%
Snider et al. (1998) Montgomery, Alabama; Chicago	80	invasive br ca, cN0	62	1.3	450 nm/un-filtered 99m-Tc sulfur colloid, GDP	IP	2/serial perm sects, H&E	87.5%	13	1	0	56	20.0%	92.9%	100.0%	98.2%
Roumen et al. (1997) Veldhoven, Netherlands	83	potentially curable T1/T2 cN0 br ca	59	2.1	99m-Tc colloidal albumin, LS, GDP	peritu-moral	H&E	68.7%	22	1	0	34	40.4%	95.7%	100.0%	97.1%
Albertini et al. (1996) University of South Florida	62	invasive br ca, cN0	60	2.2	filtered 99m-Tc sulfur colloid, GDP, blue dye	peritu-moral	1-2 sect, paraffin, H&E	91.9%	18	0	0	39	31.6%	100.0%	100.0%	100.0%

Success rates for sentinel node identification ranged from 68.7% to 100%. The average success rate, weighted by study sample size, is 90.1%. The pooled prevalence for node-positive disease in these studies is 35.9% (range for individual studies: 19–48.5%).

Diagnostic performance estimates were calculated for the group of patients for whom the sentinel node could be successfully localized. The reported estimates of diagnostic performance of SNB in this setting show a range of sensitivity between 72% and 100%. Because a positive result on SNB is based on an pathologic finding of malignant tissue in the sentinel node, a positive test is always a true positive; there are no false-positive results and specificity for SNB is always 100%. The study with the lowest sensitivity estimate (Gucciardo et al. 2000), was described as a training series. Using a random-effects meta-analysis model (Meta-Test software, Lau 1997), the summary estimate of SNB sensitivity is 89% with a 95% confidence interval of 86% to 91% (Figure 4).

FORMULATION OF THE ASSESSMENT – PART II: INITIAL STAGING OF AXILLARY LYMPH NODES

Patient Indications

Patients have a confirmed primary breast malignancy, no palpable axillary lymph node metastases and no evidence of distant metastases.

Technologies to be Compared

The reference standard diagnosis for axillary lymph node status is axillary lymph node dissection (ALND). The proposed role for PET is to allow patients who are candidates for breast-conserving surgery and who have negative PET results to avoid ALND. PET will not be compared to clinical exam because PET is proposed for use in addition to clinical exam and not as a replacement for it.

Health Outcomes

Benefits. If the PET scan correctly suggests no spread of tumor to the axillary lymph nodes, the patient could avoid the pain and other potential complications associated with ALND (e.g., lymphedema). True-positive results on PET would not result in additional benefit since patients would still undergo ALND and receive accurate staging diagnosis.

Harms. False-negative PET results would be associated with a harm because the patient would not have the benefit of accurate staging information and might, therefore, not receive adjuvant systemic therapy for node-positive disease. Such undertreatment would reduce the probability of 10-year survival by about 8%, or by about 2 years on average. In addition, a patient with a false-negative PET result who forgoes ALND could not derive any therapeutic benefit that may exist from surgically removing involved lymph nodes, potentially increasing the risk of regional failure.

Figure 4: Meta-analysis, Staging of Axillary Lymph Nodes with Sentinel Lymph Node Biopsy, Random Effects Model
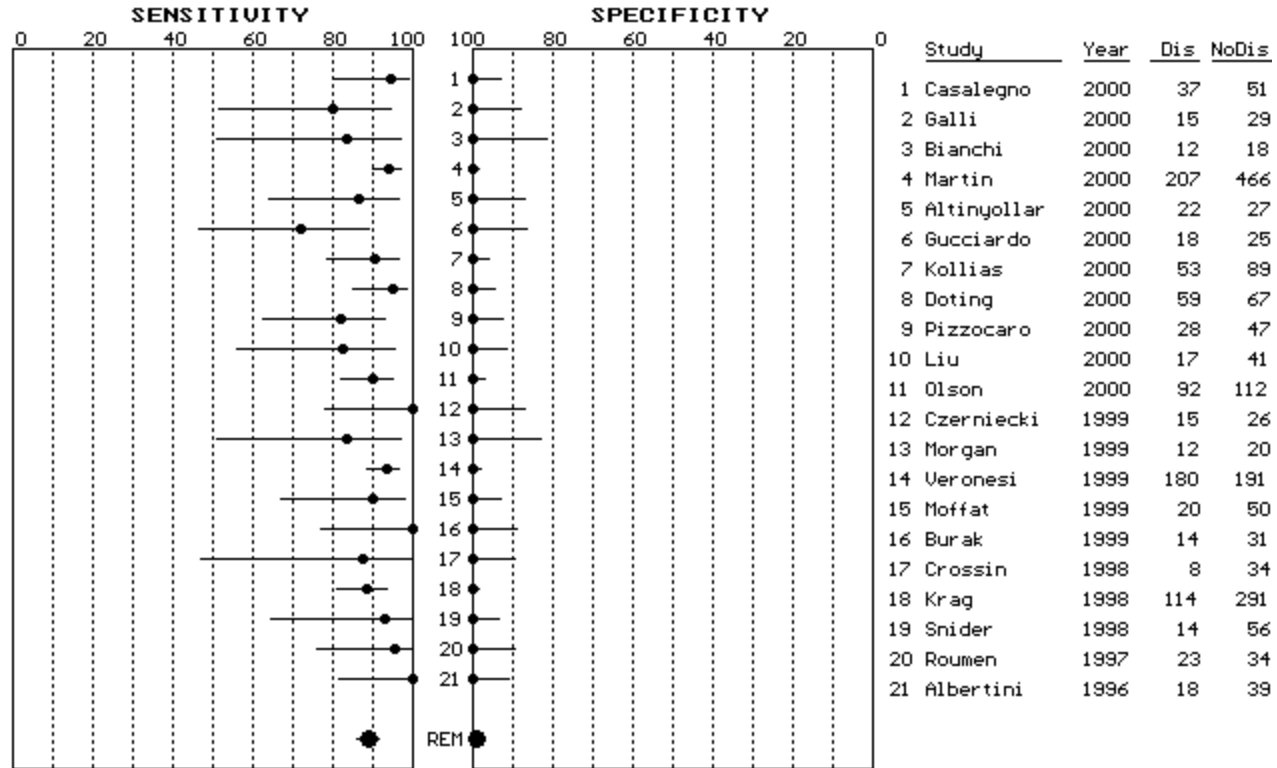
A false-positive PET result would not expose the patient to any additional harm compared with not using PET, assuming that the patient would undergo ALND and receive an accurate staging diagnosis.

Causal Chain

Figure 5 displays causal chains for 3 management paths. The first path uses PET to decide whether to perform ALND. The second path uses SNB to decide which patients receive ALND. When sentinel node mapping is unsuccessful (i.e., the sentinel node is not identified), it is assumed that patients would be referred for ALND. The third path refers all patients directly to ALND.

The following review of evidence focuses on the first and third management paths. The causal chain focuses on relative outcomes, comparing the use of PET to select patients for ALND, compared to referring all patients to ALND. The second path, for SNB, is included for illustrative purposes only because it is an emerging technology that, like PET, is proposed for selecting which patients need ALND. Furthermore it should be noted that, for any path, the causal chain assumes that knowing that positive lymph nodes are present could influence selecting of adjuvant therapy (e.g., chemotherapy, hormonal therapy, radiotherapy). The causal chain does not apply to patients for whom nodal status would not influence whether to select adjuvant therapy, including patients with negative nodes who choose to undergo adjuvant treatment.

Specific Assessment Questions

1. Does the available evidence permit conclusions about the diagnostic performance of PET for staging of axillary lymph node metastases?
2. Does the use of PET to decide whether to perform axillary lymph node dissection improve outcomes?

Specific Study Selection Criteria

- Patients of interest meet the descriptions given under "Formulation of the Assessment – Part II: Patient Indications."
- All patients must receive PET and ALND.

REVIEW OF EVIDENCE – PART II: INITIAL STAGING OF AXILLARY LYMPH NODES

Evidence on Diagnostic Performance of PET

Study Populations

This part of the assessment focuses on studies with specific data for patients with nonpalpable nodes. Differences in the spectrum of lymphatic disease between groups with palpable versus nonpalpable nodes may correspond to differences in diagnostic performance for PET. Patients with nonpalpable nodes may have smaller foci of lymphatic metastasis than patients with

palpable nodes, which may result in lower sensitivity for PET in the group with nonpalpable nodes. Moreover, in practice, patients with palpable nodes would be quite likely to undergo ALND even if imaging suggests no metastasis. Thus, PET results may have no impact on management in patients with palpable nodes.

Figure 5: Causal Chain, Relative Outcomes for Staging Axillary Lymph Nodes With and Without PET (See Table 9 for Probabilities)

Outcomes of Paths 1 and 2 Relative to Path 3

					Benefits	Harms
Path 1: With PET	Perform PET	Positive PET Result → Perform ALND	ALND Positive (PET True Positive) →	Confirm LN+s present, initiate adjuvant CTX →	Same as Path 3A	Same as Path 3A
			ALND Negative (PET False Positive) →	Prove no LN+s present, monitor for recurrence →	Same as Path 3B	Same as Path 3B
		Negative PET Result → Forgo ALND	PET True Negative →	No LN+s present, monitor for recurrence →	Differs from Path 3: avoid adverse effects of ALND	
			PET False Negative →	Undetected LN+s, monitor for recurrence →		Differs from Path 3: undertreat LN+ disease.
Path 2: With SNB	Perform SNB	Successful SN Mapping → Positive SNB Result (SNB True Positive) → Perform ALND →	ALND Positive →	Determine number of LN+s present, initiate adjuvant CTX →	Same as Path 3A	Differs from Path 3: adverse effects of SNB and ALND, undertreat LN+ disease.
		Negative SNB Result → Forgo ALND	SNB True Negative →	No LN+s present, monitor for recurrence →	Differs from Path 3: avoid adverse effects of ALND	Differs from Path 3: adverse effects of SNB
			SNB False Negative →	Undetected LN+s, monitor for recurrence →		Differs from Path 3: adverse effects of SNB, undertreat LN+ disease.
		Unsuccessful SN Mapping → → Perform ALND →	ALND Positive →	Initiate adjuvant CTX →	Same as Path 3A	Same as Path 3A
			ALND Negative →	Monitor for recurrence →	Same as Path 3B	Same as Path 3B

Outcomes of Path 3

					Benefits	Harms
Path 3: Without PET/SNB	Perform ALND →	LN+s found →	Initiate adjuvant CTX →		Path 3A: survival associated with adjuvant CTX	Path 3A: adverse effects of ALND, adjuvant CTX
		No LN+s found →	Monitor for recurrence →		Path 3B: prognostic information	Path 3B: adverse effects of ALND

45

Four key studies that provide specific data for patients with nonpalpable nodes will be summarized below (Table 7). However, it is worth noting that these studies also provide data for patients with palpable nodes and their findings suggest, as previously hypothesized, that PET is less sensitive for detecting lymphatic metastasis among patients with nonpalpable nodes, compared to patients with palpable nodes. The random effects meta-analysis (REM) sensitivity for the group of patients with palpable nodes is 93% (Appendix Table A1), compared with a sensitivity of 80% for patients with nonpalpable nodes (Table 8). This is consistent with PET having greater sensitivity in detecting larger tumor foci that would be present in palpable lymph nodes.

Summary of Available Evidence

The four studies which provided diagnostic performance data specifically in patients who have no palpable axillary lymph nodes are summarized in Tables 7 and 8, and Figures 6 and 7. If clinical nodal status is ignored, a total of 15 studies are available that meet all other study selection criteria (Appendix Tables A2 and A3, Appendix Figures A1 and A2). Five additional articles met study selection criteria, but were excluded from the data synthesis because the patient population may overlap with a later report from the same institution (Yutani et al. 1999a; Crippa et al. 1998; Crowe et al. 1994; Adler et al. 1993; Tse et al. 1992). Table 6 shows that, of 15 studies with nonduplicative data (total n=809), 10 studies provide information about the clinical nodal status of all patients, while 5 studies provide no information about it at all.

The 4 studies described in Table 8 included a total of 269 patients (270 axillary regions, each patient can potentially contribute 2 regions). Of these, 203 regions had nonpalpable nodes and 67 had palpable nodes. All 4 studies were prospective designs. One study (Greco et al. 2001, n=167) clearly avoided verification bias by selecting a consecutive series of patients. In 3 studies, PET images were interpreted blindly with respect to the reference standard. No study specified whether the reference standard was evaluated blindly with respect to PET. Attenuation correction was used in 2 studies, representing 74% of all patients.

The available body of literature is too sparse to draw conclusions regarding the diagnostic performance of PET for staging of axillary lymph node metastases. In addition, no studies reported on whether PET was able to predict the extent of nodal involvement (i.e., ≥ 4 nodes), which could be useful in selecting patients for radiation therapy.

Although the data on performance of PET in this clinical setting is sparse, meta-analysis and modeling were attempted. The purpose of this analysis to illustrate the potential effects of PET on health outcomes, assuming the diagnostic performance estimates that have been reported in the literature.

Among patients with nonpalpable nodes, sensitivities from individual studies ranged from 40% to 93% and specificities were between 87% and 100%. The REM estimates for sensitivity and specificity (Table 8, Figures 6 and 7) were 80% (95% CI: 46%, 95%) and 89% (95% CI: 83%,

Table 6. Available Information about Clinical Nodal Status in PET Studies, Staging Axillary Lymph Nodes

Author	Year	% Nonpalpable		n		
Yutani et al.	2000	?		38		
Rostom et al.	1999	?		74		
Adler et al.	1997	?		50		
Scheidhauer et al.	1996	?		18		
Hoh et al.	1993	?		14		
5 studies			**Subgroup n**	**194**	**% of Total n**	**24.0%**
Greco et al.	2001	77.2%		167		
Schirrmeister et al.	2001	14.8%		113		
Yang et al.	2001	94.4%		18		
Ohta et al.	2000	69.7%		32		
Noh et al.	1998	70.4%		24		
Smith et al.	1998	70.0%		50		
Palmedo et al.	1997	65.0%		20		
Avril et al.	1996a	45.1%		51		
Bassa et al.	1996	12.5%		16		
Utech et al.	1996	63.7%		124		
10 studies	**Pooled Proportion**	**57.5%**	**Subgroup n**	**615**	**% of Total n**	**76.0%**
			Total n	**809**		

Table 7. Staging Axillary Lymph Nodes, Data Stratified by Clinical Nodal Status

| Study	N	Design	Patient Selection	Mean Age (SD)	Mean T Size (SD)	PET Interp	Avoid Verif Bias	PET: RS Blind	RS: PET Blind	Ref Stand	Test/ Site	UA	TP	FN	FP	TN	Prev Met	Sens	Spec
Greco et al. (2001) Milan, Italy	167	pro	T1/T2 breast ca, scheduled to receive ALND	54.0	2.1 T1-59% T2-41%	qual, AC	Y	Y	?	hist	PET	P	68	4	13	82	43%	94%	86%
											PET-cN0	P	39	3	11	76	33%	93%	87%
											PET-cN+	P	29	1	2	6	79%	97%	75%
Ohta et al. (2000) 33 R Isehara, Japan	32	pro	underwent PET, US, ALND in 30, ALN sampling in 1, cN0 (70%), cN+ (30%)	50.0 med		qual, AC	?	?	?	hist	palpation	R	11	9	0	13	61%	55%	100%
											US	R	13	7	0	13	61%	65%	100%
											PET	R	14	6	0	13	61%	70%	100%
											PET+US	R	15	5	0	13	61%	75%	100%
											PET-cN0	R	4	6	0	13	43%	40%	100%
											PET-cN+	R	10	0	0	0	100%	100%	
Smith et al. (1998) Aberdeen, Scotland	50	pro	dx of breast ca; cN0 (70%), cN+ (30%)	67.0	T1-20% T2-42% T3-18% T4-20%	qual	?	Y	?	cytol (5), hist (45)	palpation	P	12	9	3	26	42%	57%	90%
											PET	P	19	2	1	28	42%	90%	97%
											PET-cN0	P	9	1	1	25	28%	90%	96%
											PET-cN+	P	10	1	0	3	79%	91%	100%
Crowe et al. (1994) University Hospitals, Cleveland	20	pro	breast lesion ≥ 1 cm on palpation or MM, path proven breast ca; cN0 (75%), cN+ (25%)	55.0 (14.0)		qual	?	Y	?	hist	PE	P	4	6	1	9	50%	40%	90%
											PET	P	9	1	0	10	50%	90%	100%
											PET-cN0	P	5	1	0	9	40%	83%	100%
											PET-cN+	P	4	0	0	1	80%	100%	100%

Abbreviations Key: See Appendix

**Table 8. Staging Axillary Lymph Nodes, Random Effects Meta-Analysis**

Study	Total n	Nonpalpable Subgroup n	TP	FN	FP	TN	Sens	Spec
Greco et al. (2001)	167	129	39	3	11	76	93%	87%
Ohta et al. (2000)	32 33 R	23	4	6	0	13	40%	100%
Smith et al. (1998)	50	36	9	1	1	25	90%	96%
Crowe et al. (1994)	20	15	5	1	0	9	83%	100%
Totals	**269 270 R**	**203**						

	Sens	95% CI		Spec	95% CI	
Random Effects Meta-Analysis Results:	80%	46%	95%	89%	83%	94%

Figure 6: Meta-analysis, Staging of Axillary Lymph Nodes, Clinically Negative Nodes, Random Effects Model
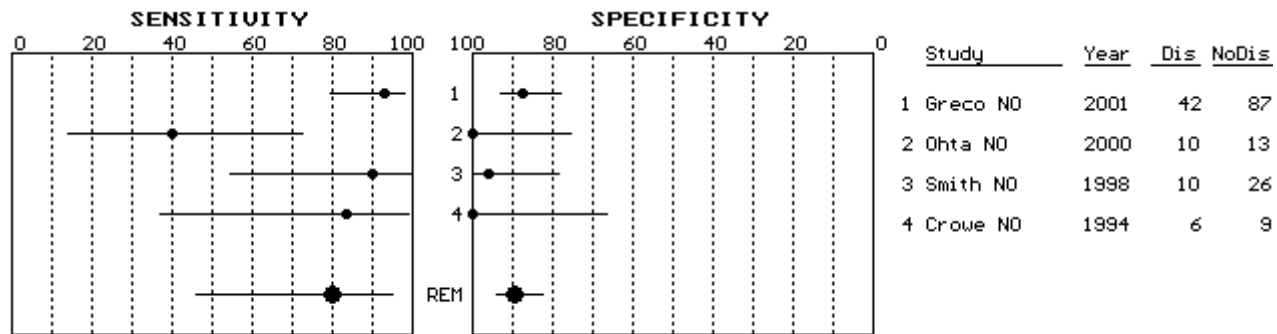
Figure 7: Summary ROC Curve, Staging of Axillary Lymph Nodes, Clinically Negative Nodes
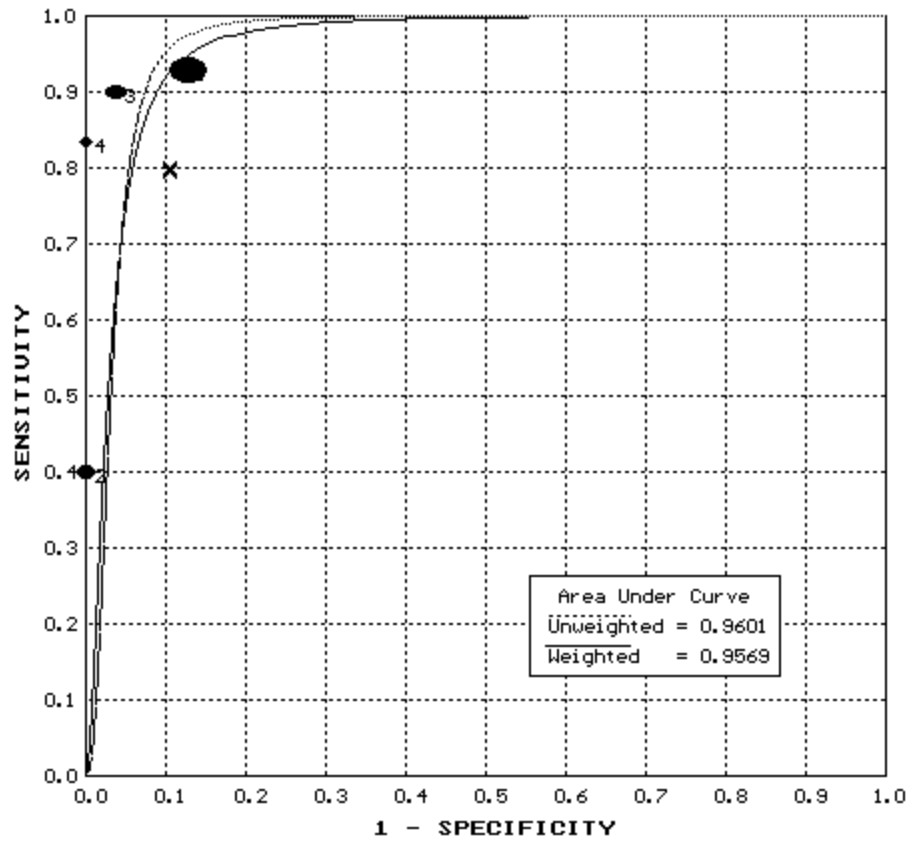
X: Random Effects Meta-Analysis Estimates

94%), respectively. Figure 7 displays the summary ROC curve derived from data for nonpalpable nodes. The Figure shows that the REM estimates underestimate diagnostic performance somewhat relative to the summary ROC curve. Thus, the point on the summary ROC curve nearest to the REM point, with a sensitivity of 81% and a specificity of 95% was chosen to estimate the average diagnostic performance for these 4 studies.

Analysis of Outcomes of Using PET to Select Patients for Axillary Lymph Node Dissection

In order to be used to avoid biopsy, PET should provide a highly sensitive evaluation for axillary node involvement. The rate of false-negative PET results weighs heavily in considering whether the risk of undertreatment by forgoing adjuvant therapy is worth the benefit of avoiding axillary node dissection. The risks and benefits of PET can be viewed from two perspectives: (1) the entire population of patients undergoing PET and (2) the individual patient who has a negative PET result.

The population perspective can be understood as follows: a patient is given a choice between allowing the PET result to guide whether to perform ALND (i.e., if PET is positive, do ALND; if PET is negative, avoid ALND), versus proceeding directly to ALND, without using PET. If the patient chooses to allow PET to guide the biopsy decision, Figure 4 shows that the only outcomes that would differ from those of the path involving no PET scanning are those associated with PET true-negative results and false-negative results. A true-negative PET result would allow a patient to safely avoid the adverse effects of ALND. A false-negative PET result would mean undetected node-positive disease and forgoing adjuvant chemotherapy (undertreatment). The patient must decide whether there is an acceptable risk-benefit tradeoff between avoiding ALND's adverse effects and being undertreated. The same type of tradeoff must be addressed by a patient with a negative PET scan.

Table 9 takes the population perspective of all patients, assuming that the results of PET are unknown. Table 10 takes the perspective of an individual patient with a known negative PET scan. The rates of false-negative results (leading to undertreatment) differ in the 2 tables, given the same prevalence of node-positive disease. From the population perspective, probabilities of true negatives and false negatives are calculated as proportions of the entire group of patients who undergo PET scanning, regardless of whether patients test positive or negative on PET (i.e., a bigger denominator). From the perspective of a patient with a known negative PET scan, the probabilities of true negatives and false negatives are calculated as proportions of only the patients with negative PET scans (a smaller denominator). From either perspective, the absolute numbers of true negatives and false negatives remain the same, so as the denominator shrinks in the group with negative PET scans, the false negative risk rises.

For both the population and the individual perspectives, we assume a sensitivity of 81% and a specificity of 95% derived from the summary ROC curve in Figure 7. It is also assumed that the prevalence (i.e., pre-test probability) of malignancy ranges from 30% to 50%. The prevalence for node-positive disease among patients with nonpalpable nodes ranged from 28% to 43% in the 4 studies included in the review of evidence. The pooled prevalence is 33%. It has been reported that the prevalence of nodal disease in early breast cancer is 40–50% (Spillane and Sacks 2000), perhaps lower (30%) for those patients with nonpalpable nodes.

Population Perspective

The population perspective (Table 9) estimates the probabilities of outcomes of using PET to decide whether to perform ALND. All PET results (both positive and negative) are considered and the proportions of the entire population deriving benefits and harms can be estimated.

The columns headed "ALND (No PET/SNB)" and "PET" and "SNB" represent the management strategies described in Figure 5. The rows define prevalences of malignancy from 30% to 50%. The intersection of a specified prevalence of malignancy and the management strategies compares the outcomes of each management path given that prevalence of malignancy. Of the 4 outcomes listed, the false-negative and true-negative outcomes are of most interest. Patients with a false-negative result would not undergo ALND; they would have undetected lymph node metastasis and would presumably be harmed by forgoing adjuvant systemic therapy (undertreatment). The true negative outcome is the proportion of patients who would benefit by avoiding ALND morbidity in the absence of nodal involvement.

At a prevalence of 30%, use of PET to decide whether to perform ALND, compared with doing ALND in all patients, would allow 66.5% of patients to benefit by avoiding the morbidity of ALND. On the population level, the false-negative risk, or risk of undertreatment, would be 5.7%. As prevalence rises from 30% to 50%, the false-negative risk for PET goes from 5.7% to 9.5%.

Individual Perspective

Table 10 takes the perspective of an individual patient with a known negative PET scan. At a prevalence for node-positive disease of 30%, the NPV for PET is 92.1%. Thus, the risk of undertreatment in this situation would be 7.9%. As prevalence for node-positive disease goes to 50%, the false-negative risk rises to 16.7%. This range of risk for undertreatment appears to be unacceptably high. Undertreatment in this case would be associated with an absolute difference in 10 year survival of 8.2%. Comparison of median survival rates in recent trials indicates about a 2-year average prolongation in life for node-positive patients treated with systemic adjuvant therapy.

Comparison with Sentinel Node Biopsy

The SNB strategy is included in Tables 9 and 10 for illustrative purposes only. It should be noted that the outcome in the third row is not entirely comparable for PET and SNB because PET is noninvasive, while SNB is an operative procedure (albeit likely less morbid than ALND). So while PET patients would avoid ALND morbidity and any axillary surgical procedure, SNB patients would undergo a limited surgical procedure to search for sentinel nodes. Both PET patients and SNB patients would face the same type of consequences of undertreatment from false negative results: failure to initiate adjuvant systemic therapy in the presence of node-positive disease.

Table 9. Benefits and Harms with and without PET, Population Perspective, by Prevalence of Nodal Disease

Prev LN+		Health Outcome	ALND (No PET/SNB)	PET	SNB
30%	Benefit	PET TP/ALND positive: undergo ALND, LN+, receive adjuvant treatment	30.0%	24.30%	27.03%
	Harm	**PET FN: do not undergo ALND, LN+, undertreat**		**5.70%**	**2.97%**
	Benefit	**PET TN: do not undergo ALND, LN-, avoid ALND morbidity**		**66.50%**	**63.07%**
	Harm	PET FP/ALND negative: undergo ALND, LN-, exposed to ALND morbidity	70.0%	3.50%	6.93%
35%	Benefit	PET TP/ALND positive: undergo ALND, LN+, receive adjuvant treatment	35.0%	28.35%	31.53%
	Harm	**PET FN: do not undergo ALND, LN+, undertreat**		**6.65%**	**3.47%**
	Benefit	**PET TN: do not undergo ALND, LN-, avoid ALND morbidity**		**61.75%**	**58.56%**
	Harm	PET FP/ALND negative: undergo ALND, LN-, exposed to ALND morbidity	65.0%	3.25%	6.44%
40%	Benefit	PET TP/ALND positive: undergo ALND, LN+, receive adjuvant treatment	40.0%	32.40%	36.04%
	Harm	**PET FN: do not undergo ALND, LN+, undertreat**		**7.60%**	**3.96%**
	Benefit	**PET TN: do not undergo ALND, LN-, avoid ALND morbidity**		**57.00%**	**54.06%**
	Harm	PET FP/ALND negative: undergo ALND, LN-, exposed to ALND morbidity	60.0%	3.00%	5.94%
45%	Benefit	PET TP/ALND positive: undergo ALND, LN+, receive adjuvant treatment	45.0%	36.45%	40.54%
	Harm	**PET FN: do not undergo ALND, LN+, undertreat**		**8.55%**	**4.46%**
	Benefit	**PET TN: do not undergo ALND, LN-, avoid ALND morbidity**		**52.25%**	**49.55%**
	Harm	PET FP/ALND negative: undergo ALND, LN-, exposed to ALND morbidity	55.0%	2.75%	5.45%
50%	Benefit	PET TP/ALND positive: undergo ALND, LN+, receive adjuvant treatment	50.0%	40.50%	45.04%
	Harm	**PET FN: do not undergo ALND, LN+, undertreat**		**9.50%**	**4.96%**
	Benefit	**PET TN: do not undergo ALND, LN-, avoid ALND morbidity**		**47.50%**	**45.05%**
	Harm	PET FP/ALND negative: undergo ALND, LN-, exposed to ALND morbidity	50.0%	2.50%	4.95%

Abbreviations Key: See Appendix

Table 10: Negative Predictive Values by Prevalence of Malignancy, Individual Perspective, Staging Axillary Nodes

$$NPV = \frac{(prev\ benign)(spec)}{((prev\ benign)(spec) + (prev\ malign)(1\text{-}sens))}$$

PET

	Sens	Spec
	81%	95%

Prevalence LN+	NPV	FN Rate
30%	92.1%	7.9%
35%	90.3%	9.7%
40%	88.2%	11.8%
45%	85.9%	14.1%
50%	83.3%	16.7%

SNB

	Sens	Spec
	89%	100%

Prevalence LN+	NPV	FN Rate
30%	95.5%	4.5%
35%	94.4%	5.6%
40%	93.2%	6.8%
45%	91.7%	8.3%
50%	90.1%	9.9%

For SNB, the following diagnostic performance values were used: a success rate for mapping of 90.1%, a sensitivity of 89% and a specificity of 100%. From the population perspective, at all levels of prevalence for node-positive disease, the proportion of patients avoiding ALND would be similar for PET and SNB. However, the level of false-negative risk in each case would be greater for PET compared with SNB. At a node-positive prevalence of 30%, the risk of undertreatment would be 2.97% for SNB, versus 5.7% for PET. At 50% prevalence of malignancy, undertreatment resulting from false negatives on PET is estimated at 9.5% versus 4.96% for SNB. At all relevant levels of prevalence, PET has a false-negative risk that is almost 2 times higher than that of SNB. The false-negative risk is also higher for PET when viewed from the perspective of an individual with a negative PET scan.

Tables 9 and 10 must be interpreted cautiously given the small quantity of available evidence on the diagnostic performance of PET in patients with nonpalpable axillary lymph nodes. The pool of patients from 4 PET studies totals 203. In contrast, data from more than 3,000 patients is available for sentinel node biopsy performed in patients with nonpalpable nodes. The 95% confidence interval for the REM sensitivity estimate has a width of approximately 50 percentage points for PET, compared with only 5 percentage points for sentinel node biopsy.

CONCLUSIONS – PART II: INITIAL STAGING OF AXILLARY LYMPH NODES

1. **Does the available evidence permit conclusions about the diagnostic performance of PET for staging of axillary lymph node metastases?**

The available body of literature is too sparse to draw conclusions regarding the diagnostic performance of PET for staging of axillary lymph node metastases in patients without palpable adenopathy. Only 4 studies reported on patients with nonpalpable axillary lymph nodes, with a total pool of 203 patients. A random effects meta-analysis was performed using this data. The estimates for sensitivity and specificity were 80% (95% CI: 46%, 95%) and 89% (95% CI: 83%, 94%), respectively. The width of the confidence interval for sensitivity is almost 50 percentage points.

In contrast, data from more than 3,000 patients is available for sentinel node biopsy performed in patients with nonpalpable nodes. The 95% confidence interval for SNB has a width of only 5 percentage points.

2. **Does the use of PET to decide whether to perform axillary lymph node dissection improve outcomes?**

In the absence of adequate evidence to estimate diagnostic performance, the outcomes of using PET to decide whether to perform axillary lymph node dissection cannot be determined. However, for illustrative purposes, this assessment estimated the probabilities of outcomes using diagnostic performance data from the available studies.

Taking the perspective of an individual patient with a known negative PET scan, the negative predictive value of PET is 92.1%, given a prevalence for node-positive disease of 30%. Thus, the risk of undertreatment in this situation would be 7.9%. As prevalence for node-positive

disease goes to 50%, the false-negative risk rises to 16.7%. This range of risk for undertreatment appears to be unacceptably high.

Undertreatment in this case would be associated with an absolute difference in 10-year survival of 8.2%. Comparison of median survival rates in recent trials indicates about a 2 year average prolongation in life for node positive patients treated with systemic adjuvant therapy.

Overall Conclusion

The available body of literature is too sparse to draw conclusions regarding the diagnostic performance of PET for staging of axillary lymph node metastases. Estimated diagnostic performance based on available data, suggests that the false-negative rate of PET in detecting axillary lymph node metastasis is too high to support a favorable risk/benefit ratio from using PET to avoid axillary lymph node dissection.

PART III: DETECTION OF LOCOREGIONAL RECURRENCE OR DISTANT METASTASIS/RECURRENCE

BACKGROUND – PART III

A complete staging evaluation is recommended if clinical suspicion for metastatic disease is high at initial diagnosis or when recurrent breast cancer is suspected (Winer et al. 2001). Conventional staging evaluation includes physical exam and patient history, computed tomography (CT) scan of the chest and abdomen and radionuclide bone scan. Other imaging tests such as magnetic resonance imaging (MRI) may be used for body imaging as well. Sites of metastatic breast cancer include bone, soft tissue, viscera and central nervous system; however, CNS involvement is rare at initial diagnosis. There is considerable variation in the patterns of metastasis and the aggressiveness of metastatic progression of disease (Winer et al. 2001).

In the setting of an isolated metastasis, local therapy may be considered. For example, an isolated pulmonary nodule may be surgically resected, an isolated bone metastasis may be irradiated, or an isolated chest wall recurrence may be treated with definitive local therapy. Systemic therapy is not universally elected in the setting of metastatic disease that can be approached locally, and combining systemic and local therapy raises issues of toxicity and tolerability. Thus, providing a complete and accurate assessment of the extent of metastatic disease is important in guiding treatment.

Because systemic therapy appears to provide a small but significant survival benefit, accurate diagnostic assessment and identification of metastatic disease is essential. Similarly, it is necessary for the staging evaluation to rule out sites of metastatic disease not amenable to local therapy so that optimal treatment decisions can be made.

FORMULATION OF THE ASSESSMENT – PART III: DETECTION OF LOCOREGIONAL RECURRENCE OR DISTANT METASTASIS/RECURRENCE

Patient Indications

Patients have a diagnosis of breast cancer and are undergoing staging evaluation for locoregional recurrence or distant metastasis/recurrence. Patients being assessed for locoregional recurrence may include those who present with arm pain or other symptoms referable to the brachial plexus. Distant disease spread can be evaluated at the initial discovery of disease or after previous treatment.

Technologies to be Compared

Routine tests used to assess for metastasis and monitor for recurrence after treatment include: physical examination, chest X-ray; CT or MRI of the body; and radionuclide bone scan. Performance of PET and alternatives must be compared by anatomic site, as the preferred imaging test varies by anatomic site.

PET is proposed to provide earlier and more accurate information compared to standard staging tests. There are two potential roles for PET in relation to standard tests:

Comparison IIIa. PET is used as an adjunct to standard staging tests. The focus of this comparison is whether PET can correctly upstage or downstage disease when used in addition to standard tests.

Comparison IIIb. PET is used as a replacement for standard staging tests. The focus of this comparison is whether PET, when used in place of standard staging tests, produces similar or more accurate diagnostic information.

Health Outcomes

Benefits. If the PET scan correctly determines the presence and extent of metastatic disease, then the patient could receive initial treatment that is appropriate for that stage. If PET correctly detects recurrence, follow-up treatment could be initiated. If PET detects recurrence earlier than other tests, then follow-up treatment can be initiated at an earlier time. If PET correctly suggests the absence of distant recurrence or metastasis when other tests have been falsely positive, patients could avoid the morbidity of unneeded treatment.

Harms. If the PET scan falsely suggests that distant metastases or recurrent lesions are present, patients may undergo unneeded biopsy and potentially harmful and unnecessary treatment if biopsy confirmation is not conducted. When PET gives a false indication that recurrence or metastases are absent, patients would forgo the potential benefits of timely initiation of treatment.

Specific Assessment Questions

1) Does the available evidence permit conclusions about the diagnostic performance of PET for detecting locoregional recurrence or distant metastasis/recurrence, in either of 2 roles?
   a) When used as an adjunct to standard staging tests
   b) When used as a replacement for standard staging tests

2) Does the available evidence permit conclusions that use of PET alters patient management in a way that improves health outcomes?

Specific Study Selection Criteria

Comparison IIIa.

For use of PET as an adjunct to standard imaging tests, the added diagnostic value of PET is studied in a population of patients selected according to findings on the battery of standard staging tests given to all patients. The diagnostic performance of the standard battery only is compared to the standard battery plus PET.

Comparison IIIb.

For use of PET as an alternative to standard staging tests, all subjects would need to have PET and the standard staging test that PET is proposed to replace.

REVIEW OF EVIDENCE – PART III: DETECTION OF LOCOREGIONAL RECURRENCE OR DISTANT METASTASIS/RECURRENCE

Evidence on Diagnostic Performance

Comparative Diagnostic Studies in Advanced Disease

When whole-body imaging is used to detect distant metastasis and determine disease stage, data analysis and calculation of diagnostic performance characteristics may be performed separately on a variety of tissue locations (e.g., liver metastasis, chest wall metastasis, bone metastasis) or may focus on presence or absence of metastasis in the patient as a whole. Appropriate reference standards may vary depending on the unit of analysis being used and also whether the finding to be confirmed is positive or negative. For example, assume a patient undergoes whole body PET which identifies an area of uptake in the liver and another focus of uptake in one vertebral body. The most expedient clinical work-up might involve biopsy of whichever one of these lesions is the more clinically accessible in order to confirm the presence of distant metastasis for staging purposes and to provide the minimum information needed to guide treatment decisions. Most times it would not be clinically appropriate to biopsy each and every tissue area being analyzed in calculating the diagnostic performance of PET.

In the example case, it would be inappropriate to expect that biopsy confirmation of each of the negative vertebral bodies would be performed, and it would be inappropriate to expect, for example, that multiple random lung biopsies be performed to confirm that the lungs are truly free from metastasis. In these situations, there is no true gold standard and sole reliance on histologic proof is unrealistic. Instead a flexible approach to appropriate reference standards can be taken where clinical follow-up of an adequate duration and periodic follow-up imaging with other conventional imaging can frequently confirm with reasonable certainty whether metastatic disease was correctly ascertained by the PET results.

With respect to PET's role in relation to other tests of recurrence or metastasis, if PET is intended as an adjunct to other tests, the following data are needed:

1. diagnostic performance of PET in the relevant population;
2. the proportion of test results in which PET and other tests are discordant;
3. the frequency with which PET or the other tests are correct when discordant; and
4. the frequency with which PET correctly upstages or downstages extent of disease when it is added to other tests.

When PET is intended to replace alternative tests, a rigorous assessment of the diagnostic performance of PET and the alternative test in a representative population is essential. Concordance of PET results with results of the alternative test is of additional interest to

determine whether each test yields valuable and complementary information or whether one test can adequately replace the other.

Detection of Locoregional Recurrence

Overview

Table 11 presents information from 2 studies meeting selection criteria for detection of locoregional recurrence (Hathaway et al. 1999; Bender et al. 1997) . These 2 studies include a total of 85 patients, 75 of which were reported by Bender. A third study (Ahmad et al. 1999) selected 19 patients with signs or symptoms (e.g., arm pain) residing in the brachial plexus. However, this study is excluded from Table 11 because the authors did not clearly define a reference standard by which to evaluate the diagnostic performance of PET. A fourth study included a small subset of patients (n=3) presenting with axillary pain or lymphedema (Lonneaux et al. 2000). Three additional studies are also excluded from Table 11 because they did not perform an alternative test on all patients (Rostom et al. 1999, n=19; Moon et al. 1998, n=57; Scheidhauer et al. 1996, n=23).

Hathaway et al. (1999) retrospectively selected 10 patients who had signs or symptoms suspicious of recurrence in the axilla or surrounding regions. Recurrent tumor in 9 patients was demonstrated in the axilla (including 5 patients with brachial plexus involvement), supraclavicular region or chest wall. PET had a sensitivity of 100%, compared to 56% for MRI. Both tests had a specificity of 100% (1 case of benign adenopathy was included and correctly classified by both tests).

A prospective study by Bender et al. (1997, n=75) described diagnostic performance in patients referred for confirmation of a suspicion of tumor recurrence or systemic disease in equivocal cases. Results were reported separately for 5 disease sites, 2 of which are relevant to locoregional recurrence (local site and lymph nodes), and 3 distant sites (bone, lung and liver). For local recurrence, PET had lower sensitivity than CT/MRI (80% vs. 93%) and similar specificity (96% vs. 98%). In lymph nodes, PET had higher sensitivity than CT/MRI (97% vs. 74%) and similar specificity (91% vs. 95%).

Although Bender et al. state that results were confirmed by histology in 71 patients and clinical follow-up in 4 patients, the paper does not clarify whether histological reference standards were applied in all 5 sites of analysis. Details are lacking from this article about how patients were histologically sampled, and it seems unlikely that large numbers of patients without suggestion of metastasis on CT, MRI, or PET would undergo a histologic sampling procedure in each region of analysis.

Since there is uncertainty about application of a reference standard in the Bender study, the diagnostic performance estimates may not be valid. In addition, this study provides no information about discordance or frequency of correct upstaging/downstaging for PET compared to CT and MRI. Finally, Bender at al. provide no direct information about how extent of disease classification is changed when PET is added to CT/MRI.

Table 11. Detecting Locoregional Recurrence

Study	N	Design	Patient Selection	Mean Age (SD)	PET Interp	Avoid Verif Bias	PET: RS Blind	RS: PET Blind	Ref Stand	Test/ Site	UA	TP	FN	FP	TN	Prev Met	Sens	Spec
Hathaway et al. (1999)	10	retro	Consec pts w/ Breast ca, sx/signs Suspicious of Recurrence in Axilla/surrounding Regions	58.4 (10.3)	qual/ quant	?	?	?	hist (4), f/u (6)	MR-local	P	5	4	0	1	90%	56%	100%
										PET-local	P	9	0	0	1	90%	100%	100%
Bender et al. (1997)	75	pro	Suspected recur-Rent or metastatic Breast ca; $\geq$ 6 mo f/u	46.0	qual	?	N	?	hist (71), f/u (4)	CT/MR-local	P	14	1	1	47	24%	93%	98%
										PET-local	P	16	4	2	53	27%	80%	96%
										CT/MR-LNs	P	17	6	2	38	37%	74%	95%
										PET-LNs	P	28	1	4	42	39%	97%	91%

Abbreviations Key: See Appendix

Another study (Lonneux et al. 2000) included 39 patients, most of whom had elevated serum tumor markers; this study is reported in Table 12, as patients were mainly not selected for symptoms of locoregional recurrence. Of the 5 patients who had symptoms of recurrence, 3 had axillary pain or lymphedema. This study found 8 confirmed sites of logoregional recurrence, including 3 at the axillary lymph nodes, 4 at the internal mammary lymph nodes and 1 axillary subcutaneous metastasis (which was falsely negative on PET).

Summary

Overall, the available evidence is insufficient to permit conclusions about the diagnostic performance of PET in detecting locoregional recurrence, which includes recurrence at the brachial plexus. Two studies reporting on a total of 85 patients met study selection criteria for this review. One study included only 10 patients, while the larger study does not provide enough details about the reference standard to determine the validity of diagnostic performance estimates.

Distant Metastasis or Recurrence

Overview

Table 12 summarizes 5 studies (total n=196) reporting on performance of PET in detecting distant metastasis or recurrence. Study designs were prospective in 3 and unclear in 2 in studies. There was insufficient information to determine whether verification bias occurred in any study. PET was interpreted blindly to the reference standard in 2 studies. One study stated that images were interpreted by investigators who were not blinded to available data. One study used 1 blinded observer and 1 unblinded observer; there was 100% agreement between them. A fifth study did not make clear whether PET was interpreted blindly.

Of the 5 studies, all reported information about bone metastases, 3 reported on liver metastases, 2 reported on lung metastases, 1 study provided data on distant lymph nodes, 1 described a case of diffuse peritoneal metastasis and 1 noted a case of metastasis to the pericardium. No studies provided direct data on the frequency of correctly upstaging or downstaging extent of disease when PET is added to other tests

Bone Metastases. Use of PET to detect bone metastases the most commonly reported anatomic sites among studies of the performance of PET for detection of distant metastases or recurrence. Among studies on bone metastasis, Lonneux et al. (2000) provided data on 11 patients, from a study of 39 patients presenting with elevated serum tumor markers (87%) or symptoms (13%). The reference standard was variable in this study, including biopsy in some cases and imaging follow-up in others. There were no false-negative results on PET among 10 confirmed recurrences in bone or bone marrow. One bone site was false positive on PET. Overall, the sensitivity and specificity for PET were 94% and 50%, respectively. Sensitivity and specificity data were not provided by lesion site to compare PET and conventional imaging, which included chest X-ray, radionuclide bone scan, ultrasound or CT of the liver, and symptom-oriented CT/MRI. Conventional imaging was positive in 6 of 33 patients with confirmed recurrence (18% sensitivity).

The study by Bender et al. (1997) included 75 patients. PET sensitivity and specificity for detecting bone metastases was reported as 100% and 98%, while CT/MRI was reported to have sensitivity and specificity of 46% and 98%. As noted previously, the high rates of absence of metastasis included in the analyses in this study raise questions about whether and how histology was performed in patients without metastases. For example, 60 of 75 patients assessed with PET for bone involvement had no metastases; rates of no metastasis were even higher at other sites. As discussed previously, it seems unlikely that so many patients without suggestion of metastasis underwent a histologic sampling procedure for confirmation and the reference standard in such negative cases was not otherwise specified.. Since there is uncertainty about application of a reference standard in the Bender study, the diagnostic performance estimates may not be valid .

Schirrmeister et al. (1999, n=34) selected patients referred for radionuclide bone scan (RBS), for suspicion of bone metastasis in 28 patients or after treatment of previously diagnosed bone metastases in 6 patients. The reference standard was vertebral MRI in 28, spiral CT of the spine in 4 and follow-up of 1 year or more in 2. Compared with RBS, PET had better sensitivity (100% vs. 83%) and better specificity (94% vs. 69%). ROC curve analysis showed that PET had significantly greater area under the curve compared with RBS. The article gives no direct information about discordances. However, the authors noted that among the 6 previously diagnosed patients, PET found additional metastases in 5, compared to 2 by RBS. Among patients with suspected bone metastases, PET correctly identified the extent of bone disease in 100% of 17 patients with such disease, compared to 35% by RBS. PET resulted in a change in management in 4 patients (12%).

Cook et al. (1998, n=23) selected patients who had undergone RBS showing evidence of bone metastases which had been confirmed by at least 1 other test. The reference standard was CT/MRI in 21 patients and biopsy in 2. Overall, PET detected a higher mean number of lesions per patient (14.1) than RBS (7.8). However, this finding was due to PET detecting greater numbers of osteolytic lesions, while RBS was better at detecting osteoblastic lesions. This article does not provide information on specificity, discordances, or frequency of changes in disease stage when PET is added to RBS.

Mortimer et al. (1996) selected 25 patients with locally advanced breast cancer who had lesions assessable for response after chemotherapy or hormone therapy. Of these, 10 developed recurrence with follow-up ranging from 12–58 months. Conventional imaging was negative when PET suggested recurrence in all 10. Distant recurrences included 2 patients with spinal metastases. Thus, PET identified distant recurrence earlier than other imaging tests.

The evidence on how well PET detects recurrence or metastasis to bone is currently insufficient to permit conclusions. The 2 key studies were conducted by Bender et al. (n=75)and Schirrmeister et al. (n=34). The Bender et al. study did not provide sufficient detail about whether and how a histologic reference standard was applied, creating uncertainty about the reported diagnostic performance estimates. The study further gave no information about discordances or correct change in stage classification when PET is added to CT/MRI.

Table 12. Detecting Distant Recurrence/Metastasis

Study	N	Design	Patient Selection	Mean Age (SD)	PET Interp	Avoid Verif Bias	PET: RS Blind	RS: PET Blind	Ref Stand	Test/ Site	UA	TP	FN	FP	TN	Prev Met	Sens	Spec
Lonneux et al. (2000)	39	?	pts treated by surgery +/- adjuvant CTX/RT; elevated serum markers (34), symptoms of recurrence (5)	57.0 (10)	qual, AC	?	?	?	imaging, bx, f/u ≥ 1 yr	PET	P	31	2	3	3	85%	94%	50%
										8/39 (21%) proven sites locoregional, 31/39 (79%) distant; conventional Imaging (CXR, RBS, liver US/CT, oriented CT/MRI) was positive in 6/33 (sensitivity 18%)								
Schirrmeister et al. (1999)	34	pro	RBS done after tx of previously dx of bone mets (6) or for suspected bone mets (28)	52.3	Qual	?	Y	?	MR, spiral CT, f/u ≥ 1 yr, X-ray	RBS-bone	P	15	3	5	11	53%	83%	69%
										PET-bone	P	18	0	1	15	53%	100%	94%
										ROC showed greater area under curve for PET (p<0.05)								
Cook et al. (1998)	23	?	Hx of breast ca, evidence of bone mets on RBS & ≥ 1 other test	52.0	Quant SUV	?	Y	?	CT/MR (21), bx (2)	PET detected higher mean number of lesions per pt (14.1) than RBS (7.8), pts w/ SUVs above median did not have significantly different survival than pts w/ SUVs below median								
Bender et al. (1997)	75	pro	Suspected recurrent or metastatic breast ca; undecided/ equivocal cases; ≥ 6 mo f/u	46.0	Qual	?	N	?	hist (71), f/u (4)	CT/MR-bone	P	6	7	1	49	21%	46%	98%
										PET-bone	P	15	0	1	59	20%	100%	98%
										CT/MR-lung	P	5	1	2	55	10%	83%	96%
										PET-lung	P	5	1	2	67	8%	83%	97%
										CT/MR-liver	P	1	1	1	20	9%	50%	95%
										PET-liver	P	2	0	2	71	3%	100%	97%
Mortimer et al. (1996)	25	pro	Locally advanced breast ca; lesions assessable for response, underwent CTX or hormone therapy	56.0 (med)	Qual, 1 observer Blinded, 1 observer Not blinded, 100% agreement	?	Y/N	?	f/u	10 of 25 pts developed recurrence w/ f/u from 12 to 58 mo; conventional imaging was negative in all 10 but PET identified abnormalities in 4 at distant sites (pericardium, liver, spine) later detected as recurrence by other tests								

Abbreviations Key: See Appendix

The findings of the Schirrmeister et al. study suggest that PET could have replaced RBS; but this single study of 34 patients is not sufficient to draw such a conclusion. Additional data are needed with respect to diagnostic performance, discordance rates, frequency of PET and other tests giving correct results among discordances, and frequency of PET correctly changing stage classification when added to other tests. Thus far, investigators have been unclear about whether PET's role is to serve as an adjunct to other tests or as a replacement.

Other Anatomic Sites.

Little data have been reported on use of PET to detect recurrence or metastasis in sites other than bone. Due to limitations in both the quantity and quality of the available evidence, these data are insufficient to permit conclusions about the diagnostic performance of PET in detecting recurrence or metastasis in the lung, liver or other distant sites

Liver. The studies by Lonneux et al. Bender et al. and Mortimer et al. give data on liver involvement. There were 6 sites of liver metastasis detected in the Lonneux study, and 1 PET scan was falsely positive. The Bender et al. study reported that PET found 2 of 2 liver metastases, while CT/MRI found 1 of 2. There were 2 false positive PET tests of 73 patients with no liver involvement, compared with 1 false positive for CT/MRI. However, the total number of liver-negative patients included in the analysis for CT/MRI (n=21) was inexplicably much lower than for PET (n=73). As noted before, it is unclear whether and how patients had histologic confirmation of absent liver disease in the Bender et al. report. The study by Mortimer et al. reported that among 4 distant sites in which PET suggested disease earlier than other tests, 1 had liver metastasis.

Lung. Lonneux et al. detected 5 sites of metastasis in the lungs or pleura. One PET scan was falsely positive for lung metastasis. The study by Bender et al. reported that both PET and CT/MRI correctly identified 5 patients with lung metastasis, and both missed 1 patient. Both tests also had 2 false positive results, although again, there is uncertainty about the reference standard used to rule out the presence of tumor in this study.

Other. In addition to liver and lung, a metastasis to pericardium was reported in 1 patient in the Mortimer et al. study. Lonneux et al. noted that PET detected 9 sites of distant lymph node metastasis, along with 1 false negative for lymphatic metastasis. These authors also reported a false negative PET scan for diffuse peritoneal metastasis.

CONCLUSIONS – PART III: DETECTION OF LOCOREGIONAL RECURRENCE OR DISTANT METASTASIS/RECURRENCE

1) **Does the available evidence permit conclusions about the diagnostic performance of PET for detecting locoregional recurrence or distant metastasis/recurrence, in either of 2 roles?**
   a) **When used as an adjunct to standard staging tests.**
   b) **When used as a replacement for standard staging tests.**

The evidence is insufficient to permit conclusions about the diagnostic performance of PET in detecting locoregional recurrence, which includes recurrence at the brachial plexus. Two studies reporting on a total of 85 patients met study selection criteria for this review. One study included only 10 patients, while the larger study of 75 patients does not provide enough details about the reference standard to determine the validity of diagnostic performance estimates. The larger study was also one of 2 key studies providing evidence on detection of distant recurrence or metastasis.

The evidence on how well PET detects distant recurrence or metastasis to bone is currently insufficient. The 2 key studies included 75 and 34 patients. As noted previously, the larger study did not provide sufficient detail about whether and how a histologic reference standard was applied, creating uncertainty about the diagnostic performance estimates. The report also gave no information about discordances or correct changes in stage classification when PET is added to CT/MRI. The findings of the smaller study suggest that PET could have replaced radionuclide bone scan, but this single study of 34 patients is not sufficient to draw such a conclusion. Additional data are needed with respect to diagnostic performance, discordance rates, frequency of PET and other tests giving correct results among discordances, and frequency of PET correctly changing stage classification when added to other tests.

Little data have been reported on use of PET to detect recurrence or metastasis in sites other than bone. One study reported PET and CT/MRI findings in 2 patients with liver metastasis and another study reported on a single case of liver metastasis. A single study addressed 6 patients with confirmed lung metastases. A metastasis to pericardium was reported in 1 patient. These data are clearly insufficient to permit conclusions about the diagnostic performance of PET in detecting recurrence or metastasis in bone, lung, liver or other distant sites.

2.   **Does the available evidence permit conclusions that use of PET alters patient management in a way that improves health outcomes?**

The available data are insufficient to determine the diagnostic performance of PET in detecting recurrence or metastasis, and thus, it is not possible to reliably determine the effect diagnostic information might have on management decisions and patient health outcomes.

PART IV: EVALUATING RESPONSE TO TREATMENT

BACKGROUND – PART IV

Decision making in breast cancer treatment is often guided by prognostic information and the presence of predictive factors associated with response to a particular treatment regimen. Predictive factors associated with the tumor such as estrogen receptor status are useful in guiding hormonal therapy. There is research interest as to whether tumor proliferation characteristics such as S-phase fraction may be able to predict response to chemotherapy. The proposed role for PET is to provide a more accurate or earlier determination of response to therapy than is possible with conventional modalities, in order to facilitate treatment decisions (e.g., to discontinue ineffective systemic therapy).

Studies have also explored the role that FDG PET tumor activity may have in defining prognosis or predicting tumor response to treatment. Oshida et al. (1998) prospectively studied 70 subjects treated for breast cancer by measuring a baseline ratio of tumor activity on FDG PET and correlating the quantitative level of the ratio with clinical outcome. These authors found that overall survival rates and disease-free survival rates were significantly better for subjects with low ratios of tumor activity compared to those with higher ratios. Multivariate regression analysis suggested that tumor activity ratio, histologic tumor grade, and the number of positive lymph nodes were all significant independent prognostic factors.

Other investigators have used FDG PET assessment before and after courses of systemic therapy to determine if PET can provide an earlier and more accurate indicator of tumor response to therapy. Conventional clinical assessment of tumor response relies on decrease in tumor size and is generally evaluated on physical exam or radiological anatomical imaging. FDG PET evaluates changes in the level of metabolic activity in the tumor. Whether the functional assessment of metabolic activity afforded by FDG PET can reliably identify subgroups of patients who will not respond to chemotherapy, thereby potentially sparing such patients the repeated toxicity of multiple courses of chemotherapy, remains a research question.

FORMULATION OF THE ASSESSMENT – PART IV: EVALUATING RESPONSE TO TREATMENT

Patient Indications

Patients are those undergoing multicourse treatment for breast cancer.

Technologies to be Compared

Routine tests used to evaluate response to treatment for breast cancer vary depending on the type of surgery or other treatment given, but generally include: physical examination; mammography; X-ray; CT; MRI; and bone scan. PET will be compared to the conventional criteria for response, based on routine tests, used in each study.

For evaluating tumor size, PET can be compared to routine imaging. However, using PET to evaluate a tumor's metabolic activity is a unique feature of PET that cannot be compared directly to routine imaging.

Health Outcomes

Benefits. If a PET scan accurately reflects and predicts the response to treatment, results could aid in deciding to initiate new treatment, continue effective treatment, discontinue ineffective treatment or to identify disease-free patients who simply need monitoring. Timing of treatment decisions may be affected if PET can diagnose the response earlier than other methods. Earlier discontinuation of ineffective treatment might improve quality of life; earlier initiation of a new treatment might improve treatment outcomes.

Harms. If a PET scan falsely reflects a patient's response to treatment, the patient could face the following consequences: continued harmful side effects of ineffective treatment when PET falsely suggested a response to treatment or forgoing the benefits of additional treatment when PET falsely suggested a complete response to previous treatment.

Specific Assessment Questions

1. Does the available evidence permit conclusions about the diagnostic performance of PET for evaluating or predicting response to treatment?

2. Does PET improve outcomes by providing a more accurate or earlier determination of tumor response to treatment compared to use of conventional response criteria?

REVIEW OF EVIDENCE – PART IV: EVALUATING RESPONSE TO TREATMENT

Four studies with a total of 103 patients have addressed whether PET imaging performed early in the course of treatment predicts response to treatment evaluated at its conclusion (Table 13; Mortimer et al. n=40, 2001; Schelling et al. 2000; n=22; Smith et al. 2000, n=30; Wahl et al. 1993, n=11). All 4 studies were prospective. No studies provided sufficient information to determine whether verification bias was avoided. In 1 study, it was clear that interpreters of PET images were blinded to reference standard results, and it was unclear in 3 studies. None of the studies mentioned whether investigators who assessed the reference standard were blinded to PET results. Treatment was neoadjuvant chemotherapy in 2 studies (Schelling et al. 2000; Smith et al. 2000), chemohormonotherapy in 1 (Wahl et al. 1993) and hormone therapy in 1 (Mortimer et al. 2001).

Evidence on Diagnostic Performance of PET

Mortimer et al. (2001) gave tamoxifen to 40 estrogen receptor-positive patients with locally advanced, recurrent or metastatic breast cancer. After 7–10 days of treatment, patients were assessed for a temporary flare response on a quantitative PET measure ($\geq 10\%$ rise in FDG). The presence or absence of flare was highly predictive of response assessed by standard criteria

69

on follow-up: sensitivity was 95% and specificity was 89%. Percent change in FDG was also a significant predictor of overall response in multivariate regression analysis.

Schelling et al. (2000) selected 22 patients with newly diagnosed locally advanced breast cancer who were to undergo neoadjuvant chemotherapy with either epirubicin and cyclophosphamide or epirubicin and paclitaxel. The response criteria assessed whether there was histologically gross versus minimal disease present at the time of surgery. The results of PET after the first of 3 or 4 courses of chemotherapy was assessed in 16 of 22 patients. Sensitivity and specificity for histologic response were 100% and 85%. All 22 patients were assessed on PET after the second course of chemotherapy. At this point, sensitivity was 83% and specificity was 94%. The proportion of patients missing from the data after the first course of chemotherapy was 27%, so it is unclear whether the first sensitivity estimate of 100% is reliable.

Smith et al. (2000) chose patients with newly diagnosed noninflammatory breast cancer with a large primary (>3 cm) or locally advanced disease. These 30 patients were enrolled in a trial comparing 2 neoadjuvant chemotherapy regimens: cyclophosphamide, vincristine, doxorubicin and prednisolone; or primary docetaxel. The key response criteria were histology at the time of surgery: pathologic complete response (pCR) or pathologic partial response (pPR). At the end of the first pulse of chemotherapy, PET result was assessed quantitatively. A drop of 10% or more in the PET measure had the following diagnostic performance for predicting either pCR or pPR: a sensitivity 82% and a specificity of 67%. A drop of 20% or more in the PET measure had a sensitivity of 90% and a specificity of 74% for predicting a pCR. The diagnostic performance of this PET measure was improved by using a higher threshold change in PET measure and by using pathologic complete response as the outcome of interest.

Wahl et al. (1993) included 11 patients who were given a nonstandard regimen for locally advanced disease, which included cyclophosphamide, doxorubicin, methotrexate, fluorouracil, tamoxifen, and conjugated estrogens (i.e., Premarin®). After 1 cycle of treatment, all patients responding by clinical criteria had significant decreases in a PET quantitative measure. No nonresponders had a significant decrease in the PET measure.

Conclusion. Due to limitations in its quantity, quality, and consistency the available evidence is insufficient to permit conclusions about the diagnostic performance of PET in evaluating response to treatment. The 4 available studies selected small patient samples, and together reported on a total of only 103 patients

Outcomes of Using PET

Additional evidence is needed to better define diagnostic performance, as well as to assess whether health outcomes would be improved by using PET response as a guide to patient management. Two of the 4 studies report sizable rates of false negatives that would lead to undertreatment, if a finding of nonresponse on PET inappropriately led to discontinuation of systemic therapy.

Table 13. Evaluating Response to Treatment

Study	N	Design	Patient Selection	Mean Age (SD)	PET Interp	Avoid Verif Bias	PET: RS Blind	RS: PET Blind	Ref Stand	Test/ Site	UA	TP	FN	FP	TN	Sens	Spec
Mortimer et al. (2001)	40	pro	postmenopausal, locally advanced, recurrent, or, metastatic breast ca ER+, given tamoxifen	58.0 (med)	qual, semi-quant	?	?	?	standard response criteria	PET change at 7-10 d, flare ($\geq$ 10% rise in	P	20	1	2	17	95%	89%
Schelling et al. (2000)	22 24 L	pro	newly diagnosed locally advanced breast ca, sched-uled for neoadjuvant CTX (CE or CT), no prev tx < 3 mo	50.0 (med)	quant SUV ROC	?	?	?	histopath response gross vs. minimal residual disease	PET-response after 1st CTX course	P	3	0	2	11	100%	85%
										PET-response after 2nd CTX course	P	5	1	1	15	83%	94%
Smith et al. (2000)	30 31 L	pro	newly diagnosed, noninflammatory, large (> 3 cm) or locally advanced breast ca, enrolled in trial comparing 2 neoadjuvant CTX regimens (CVAP vs. Doc)	49.0	quant	?	Y	?	clinical response (IUAC cri-teria), histopath response	PET-pPR/pCR after 1st CTX $\geq$ 10% fall in DUR	P	14	3	4	8	82%	67%
										PET-pCR after 1st CTX $\geq$ 20% fall in DUR	P	9	1	5	14	90%	74%
Wahl et al. (1993)	11	pro	newly diagnosed locally advanced breast ca, > 3 cm, given C, A, M, 5-FU, T, Premarin	47.4 (13.7)	quant	?	?	?	clinical response	all responding pts had significant decreases in FDG SUV after 1 cycle of tx; no nonresponding pts had a significant decrease in SUV							

Abbreviations Key: See Appendix

In the study (n=22) by Schelling et al. (2000), PET sensitivity was 83% and specificity was 94% following the second course of chemotherapy. Based on this performance data, using a PET result suggesting no response to therapy to discontinue chemotherapy would inappropriately deprive 17% of all patients from receiving chemotherapy to which they were truly responsive. An additional 6% of patients would have PET results suggesting tumor response when in fact their tumor was not responsive to the chemotherapy.

Smith et al. (2000, n=30) reported that drop of 20% or more in the PET measure had a sensitivity of 90% and a specificity of 74% for predicting a pathologic complete response. Based on this performance data, using PET results to discontinue treatment would potentially deprive 10% of patients from receiving beneficial chemotherapy and would provide potentially misleading and overly optimistic information in those with false positive PET results (26%).

CONCLUSIONS – PART IV: EVALUATING RESPONSE TO TREATMENT

1. **Does the available evidence permit conclusions about the diagnostic performance of PET for evaluating or predicting response to treatment?**

The proposed role for PET is to provide a more accurate or earlier determination of tumor response to treatment to facilitate treatment decisions (e.g., to discontinue ineffective systemic therapy). Four studies with a total of 103 patients have addressed whether PET imaging early in the course of treatment predicts response to treatment evaluated at its conclusion. All 4 studies were prospective. Treatment was neoadjuvant chemotherapy in 2 studies, chemohormonotherapy in one and hormone therapy in one. The available evidence is of limited quantity, quality, and consistency and is insufficient to permit conclusions about the diagnostic performance of PET in evaluating response to treatment.

2. **Does PET improve outcomes by providing a more accurate or earlier determination of tumor response to treatment compared to use of conventional response criteria?**

Additional evidence is needed to determine diagnostic performance of PET, as well as to assess whether health outcomes would be improved by using PET response as a guide to patient management. Two of the 4 studies reported sensitivities that would lead to substantial undertreatment, if a finding of nonresponse on PET were used to guide treatment. Inappropriate discontinuation of systemic therapy would have occurred in 10% (n=30) of patients in one study and 17% (n=22) in the other study.

ASSESSMENT CONCLUSIONS

These assessment conclusions are organized into 4 parts, as follows:

- Part I: Initial Diagnosis of Breast Cancer;
- Part II: Initial Staging of Axillary Lymph Nodes;
- Part III: Detection of Locoregional Recurrence or Distant Metastasis/Recurrence; and
- Part IV: Evaluating Response to Treatment.

For each indication, 2 questions were addressed: 1) does the available evidence permit conclusions about the diagnostic performance of PET; and 2) if there is adequate evidence on diagnostic performance, does the use of PET improve health outcomes?

PART I – INITIAL DIAGNOSIS OF BREAST CANCER

Two potential roles for PET were considered: (a) using PET in patients who have been referred for breast biopsy in order to avoid biopsy when PET results are negative, and (b) using PET in patients who have been referred for short-interval imaging follow-up due to low suspicion findings on mammography.

Indication Ia

Among patients who have been referred for biopsy, a true-negative PET finding would result in a patient safely avoiding a painful invasive biopsy and its consequences; while a false-negative PET finding could result in delayed or missed diagnosis and treatment. Patients with positive PET scans would presumably undergo biopsy confirmation; thus there would be no change in the net health outcome from using PET compared with not using PET prior to biopsy.

Evidence on Diagnostic Performance.

In studies of PET for differential diagnosis of breast lesions, patients were selected for having suspicious mammograms or palpable masses. These study samples have a notably higher prevalence of malignancy than that reported for the general population and a relatively large average tumor size at initial diagnosis. These studies represent the upper part of the biopsy population spectrum. No published studies are available on the diagnostic performance of PET in the lower part of the biopsy population, comprising a range of prevalence between 20% and 50%. This group consists of patients with indeterminate mammograms and smaller, nonpalpable lesions. Without of evidence on diagnostic performance of PET in the lower portion of the biopsy population, no conclusions can be reached and it would be imprudent to generalize from the studied population.

Thirteen studies (total n=606) met study selection criteria for inclusion in the data synthesis. The prior probability of malignancy in the study samples ranged between 53% and 95%, compared to 20% to 30% in the general population. Mean tumor size across studies was relatively large, ranging from about 2 cm to 4 cm.

Sensitivity estimates in all 13 studies ranged from 79% to 100% and specificity estimates were between 50% and 100%. Meta-analysis was first performed using a random effects model. The pooled sensitivity estimate was 88% (95% CI: 83%, 92%) and the pooled specificity estimate was 79% (95% CI: 71%, 85%). Then a summary receiver operating characteristic (ROC) curve was constructed which accounts for the dependent relationship between sensitivity and specificity. A point on the summary ROC curve was selected which reflected average performance, with an estimated sensitivity of 89% and a specificity of 80%.

Sensitivity analysis based on higher quality studies, defined as prospective, free of verification bias and used blinded interpretation of PET, was initially planned. However, only 1 study met these qualifications (n=40), thus precluding the planned analysis.

Analysis of Effect on Health Outcomes.

In order to be used to avoid biopsy, PET should provide a highly sensitive evaluation for malignancy. The rate of false negative PET results weighs heavily in considering whether the risk of delayed or missed diagnosis of breast cancer is worth the benefit of avoiding biopsy of a benign lesion. The risks and benefits of PET were analyzed using two perspectives: (1) the entire population of patients undergoing PET and (2) the individual patient who has a negative PET result.

For both analyses, sensitivity of PET was assumed to be 89% and specificity was 80%. The prevalence (i.e., pre-test probability) of malignancy was assumed to range from 50% to 75%. Evidence is lacking about PET's diagnostic performance for smaller tumors and in patient populations with disease prevalence lower than 50%. As the prevalence of malignancy rises from 50% to 75%, the false-negative risks also rise, making the probabilities of harm from delayed diagnosis and treatment higher

The population perspective assumes that the results of PET are not yet known. All PET results (both positive and negative) are considered and the proportions of the entire population deriving benefits and harms can be estimated. When the prevalence of malignancy is 50%, 40% of all patients would benefit by avoiding the harms of negative biopsy. The risk of a false-negative result, leading to delayed diagnosis and treatment, is 5.5%. When the prevalence of malignancy is 75%, 20% of patients avoid biopsy of a benign lesion; and the risk of delayed treatment is 8.25%.

From the perspective of an individual patient with a negative PET scan, the risk of a false-negative result is higher than for the entire population undergoing PET scanning. When the prevalence of malignancy is 50%, the NPV is 87.9%, thus, the false-negative risk is 12.1%. For an individual with a negative PET scan, a 12% chance of missed or delayed diagnosis of breast cancer is most likely too high to make the 88% chance of avoiding an negative biopsy of a benign lesion worthwhile. When the prevalence of malignancy is 75%, there is a 29.2% risk of missed or delayed diagnosis, which is surely unacceptable in order to avoid a biopsy.

Evidence is lacking to assess the negative predictive value of PET in the population of patients referred for biopsy with indeterminate mammograms and smaller, nonpalpable lesions. Such patients would have a prevalence of malignancy from 20% to 50%.

Summary.

Evidence on the diagnostic performance of PET for differential diagnosis of breast lesions among patients with abnormal mammograms or palpable masses is lacking for patients with indeterminate mammograms and small, nonpalpable lesions (low prevalence for malignancy). Among study populations of patients with higher prevalence of malignancy, risk of a false-negative diagnosis is likely too high relative to the benefit of avoiding biopsy of a benign lesion. A false-negative PET result may cause a missed or delayed diagnosis of breast cancer and associated delay in treatment.

From the perspective of an individual patient with a prior probability of malignancy of 50% and a negative PET result, the risk of a false-negative result PET is 12.1%. At the 75% prevalence, there is a 29.2% risk of a false-negative finding. Evidence on PET diagnostic performance is unavailable to permit estimation of the risk of a false-negative PET result in the patients with a prevalence of malignancy from 20% to 50%.

Indication Ib

FDG-PET may also be used as a diagnostic aid in patients with low suspicion mammographic findings who have been referred for short interval mammographic follow-up. Positive PET results may help to select patients who should be referred for biopsy while negative PET results might enable the frequency of follow-up to be reduced. Selective biopsy might achieve earlier diagnosis of breast cancer than short-interval mammographic follow-up, which is presently recommended in this patient population.

Evidence on Diagnostic Performance.

No studies meeting selection criteria included a patient population to address this question. Performance of PET in the available studies in patients who have been referred for biopsy due to an abnormal mammogram or palpable mass cannot be generalized to patients with low suspicion findings on mammography referred for short interval follow-up.

Analysis of Effect on Health Outcomes.

This question cannot be addressed in the absence of data on the diagnostic performance of PET in the population of interest.

PART II – STAGING AXILLARY LYMPH NODES

The proposed role for PET for this indication is to select which patients need to undergo axillary lymph node dissection (ALND) among the subset of patients who have no clinically palpable axillary adenopathy. If the PET scan correctly suggested no spread of tumor to the axillary

lymph nodes, the patient could avoid the pain and other complications associated with ALND (e.g., chronic lymphedema). However, a false-negative PET result could lead to harm if a patient with undetected axillary involvement chose to forego adjuvant systemic therapy.

Adjuvant systemic therapy has been reported to reduce recurrence and improve survival in patients with breast cancer. Improved outcomes occur in patients with positive axillary nodes and also in patients without axillary involvement. However, the absolute magnitude of the reduction in recurrence rate or mortality is greater for those with axillary nodal disease. Compared to patients without axillary involvement, node-positive patients are at a greater baseline risk of recurrence and disease-related mortality and thus, have greater potential for benefit, based on the Early Breast Cancer Trialists' Collaborative Group overview of 133 randomized clinical trials. However, decisions on the use of adjuvant therapy in patients with node-negative disease is complicated by uncertainties in balancing potential benefits and toxicity of systemic therapy, as well as by variation in patient preferences.

Chronic lymphedema is common following ALND, and strategies for reducing the morbidity of axillary node staging are being developed. Sentinel node biopsy (SNB), a more limited surgical approach to axillary lymph node staging, has been introduced as an alternative surgical technique. More recently, PET has been proposed as a noninvasive method for determining the presence of axillary lymph node involvement and for selecting patients for ALND.

Evidence on Diagnostic Performance

The available body of literature is too sparse to draw conclusions regarding the diagnostic performance of PET for staging of axillary lymph node metastases in patients without palpable adenopathy. Only 4 studies reported on patients with nonpalpable axillary lymph nodes, with a total pool of 203 patients. A random effects meta-analysis was performed using this data. The estimates for sensitivity and specificity were 80% (95% CI: 46%, 95%) and 89% (95% CI: 83%, 94%), respectively. The width of the confidence interval for sensitivity is almost 50 percentage points.

In contrast, data from more than 3,000 patients is available for sentinel node biopsy performed in patients with nonpalpable nodes. The 95% confidence interval for SNB has a width of only 5 percentage points.

Analysis of Effect on Health Outcomes

In the absence of adequate evidence to estimate diagnostic performance, the outcomes of using PET to decide whether to perform axillary lymph node dissection cannot be determined. However, for illustrative purposes, this assessment estimated the probabilities of outcomes using diagnostic performance data from the available studies.

Taking the perspective of an individual patient with a known negative PET scan, the negative predictive value of PET is 92.1%, given a prevalence for node-positive disease of 30%. Thus, the risk of undertreatment in this situation would be 7.9%. As prevalence for node-positive

disease goes to 50%, the false-negative risk rises to 16.7%. This range of risk for undertreatment appears to be unacceptably high.

Undertreatment in this case would be associated with an absolute difference in 10-year survival of 8.2%. Comparison of median survival rates in recent trials indicates about a 2 year average prolongation in life for node positive patients treated with systemic adjuvant therapy.

Summary

The available body of literature is too sparse to draw conclusions regarding the diagnostic performance of PET for staging of axillary lymph node metastases. Estimated diagnostic performance based on available data, suggests that the false-negative rate of PET in detecting axillary lymph node metastasis is too high to support a favorable risk/benefit ratio from using PET to avoid axillary lymph node dissection.

PART III: DETECTION OF LOCOREGIONAL RECURRENCE OR DISTANT METASTASIS/RECURRENCE

For this indication, PET may serve either as an adjunct to other imaging tests or as a replacement. Because systemic therapy appears to provide a small but significant survival benefit, accurate diagnostic assessment and identification of metastatic disease is essential. Similarly, it is necessary for the staging evaluation to rule out sites of metastatic disease not amenable to local therapy so that optimal treatment decisions can be made.

Evidence on Diagnostic Performance

The evidence is insufficient to permit conclusions about the diagnostic performance of PET in detecting locoregional recurrence, which includes recurrence at the brachial plexus. Two studies reporting on a total of 85 patients met study selection criteria for this review. One study included only 10 patients, while the larger study of 75 patients does not provide enough details about the reference standard to determine the validity of diagnostic performance estimates. The larger study was also one of 2 key studies providing evidence on detection of distant recurrence or metastasis.

The evidence on how well PET detects distant recurrence or metastasis to bone is currently insufficient. The 2 key studies included 75 and 34 patients. As noted previously, the larger study did not provide sufficient detail about whether and how a histologic reference standard was applied, creating uncertainty about the diagnostic performance estimates. The report also gave no information about discordances or correct changes in stage classification when PET is added to CT/MRI. The findings of the smaller study suggest that PET could have replaced radionuclide bone scan, but this single study of 34 patients is not sufficient to draw such a conclusion. Additional data are needed with respect to diagnostic performance, discordance rates, frequency of PET and other tests giving correct results among discordances, and frequency of PET correctly changing stage classification when added to other tests.

Little data have been reported on use of PET to detect recurrence or metastasis in sites other than bone. One study reported PET and CT/MRI findings in 2 patients with liver metastasis and another study reported on a single case of liver metastasis. A single study addressed 6 patients with confirmed lung metastases. A metastasis to pericardium was reported in 1 patient. These data are clearly insufficient to permit conclusions about the diagnostic performance of PET in detecting recurrence or metastasis in bone, lung, liver or other distant sites.

Analysis of Effect on Health Outcomes

As the available data are insufficient to determine the diagnostic performance of PET in detecting recurrence or metastasis, it is not possible to reliably determine the effect diagnostic information might have on management decisions and patient health outcomes.

PART IV: EVALUATING RESPONSE TO TREATMENT

The proposed role for PET for this indication is to provide a more accurate or earlier determination of tumor response to treatment to facilitate treatment decisions (e.g., to discontinue ineffective systemic therapy).

Evidence on Diagnostic Performance

Four studies with a total of 103 patients have addressed whether PET imaging early in the course of treatment predicts response to treatment evaluated at its conclusion. All 4 studies were prospective. Treatment was neoadjuvant chemotherapy in 2 studies, chemohormonotherapy in one and hormone therapy in one. The available evidence is of limited quantity, quality, and consistency and is insufficient to permit conclusions about the diagnostic performance of PET in evaluating response to treatment.

Analysis of Effect on Health Outcomes

Additional evidence is needed to determine diagnostic performance of PET, as well as to assess whether health outcomes would be improved by using PET response as a guide to patient management. Two of the 4 studies reported sensitivities that would lead to substantial undertreatment, if a finding of nonresponse on PET were used to guide treatment. Inappropriate discontinuation of systemic therapy would have occurred in 10% (n=30) of patients in one study and 17% (n=22) in the other study.

REFERENCES

Adler LP, Crowe JP, al-Kaisi NK et al. (1993). Evaluation of breast masses and axillary lymph nodes with [F-18] 2- deoxy-2-fluoro-D-glucose PET. *Radiology,* 187(3):743-50.

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(2):323-7.

Ahmad A, Barrington S, Maisey M et al. (1999). Use of positron emission tomography in evaluation of brachial plexopathy in breast cancer patients. *Br J Cancer,* 79(3-4):478-82.

Albertini JJ, Lyman GH, Cox C et al. (1996). Lymphatic mapping and sentinel node biopsy in the patient with breast cancer. *JAMA,* 276(22):1818-22.

Altinyollar H, Kapucuoglu N, Pak I et al. (2000). Lymphatic mapping and sentinel lymphadenectomy in early stage breast carcinoma. *J Exp Clin Cancer Res,* 19(2):141-4.

American College of Radiology (1995). Breast imaging reporting and data system (BI-RADS), 2nd ed. Reston, VA: American College of Radiology.

Avril N, Rose CA, Schelling M et al. (2000). Breast imaging with positron emission tomography and fluorine-18 fluorodeoxyglucose: use and limitations. *J Clin Oncol,* 18(20):3495-502.

Avril N, Bense S, Ziegler SI et al. (1997). Breast imaging with fluorine-18-FDG PET: quantitative image analysis. *J Nucl Med,* 38(8):1186-91.

Avril N, Dose J, Janicke F et al. (1996a). 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(17):1204-9.

Avril N, Dose J, Janicke F et al. (1996b). Metabolic characterization of breast tumors with positron emission tomography using F-18 fluorodeoxyglucose. *J Clin Oncol,* 14(6):1848-57.

Bassa P, Kim EE, Inoue T et al. (1996). Evaluation of preoperative chemotherapy using PET with fluorine-18- fluorodeoxyglucose in breast cancer. *J Nucl Med,* 37(6):931-8.

Bassett LW, Liu TH, Giuliano AE. (1991). The prevalence of carcinoma in palpable vs. impalpable, mammographically detected lesions. *AJR,* 157:21-24.

Bender H, Kirst J, Palmedo H et al. (1997). Value of 18 fluoro-deoxyglucose positron emission tomography in the staging of recurrent breast carcinoma. *Anticancer Res,* 17(3B):1687-92.

Bianchi P, Villa G, Buffoni F et al. (2000). Different sites and modes of tracer injection for mapping the sentinel lymph node in patients with breast cancer. *Tumori,* 86(4):307-8.

Blue Cross Blue Shield Association Technology Evaluation Center. (1995). Management of nonpalpable breast lesions: biopsy techniques. *TEC Assessment Program*, 10(32):1-12.

Burak WE, Walker MJ, Yee LD et al. (1999). Routine preoperative lymphoscintigraphy is not necessary prior to sentinel node biopsy for breast cancer. *Am J Surg,* 177(6):445-9.

Casalegno PS, Sandrucci S, Bello M et al. (2000). Sentinel lymph node and breast cancer staging: final results of the Turin Multicenter Study. *Tumori,* 86(4):300-3.

Cochrane Methods Working Group on Systematic Review of Screening and Diagnostic Tests: Recommended Methods, updated 6 June 1996. Available online at http://www.cochrane.org/cochrane/sadtdoc1.htm. Last accessed: 10 April 2001.

Cook GJ, Houston S, Rubens R et al. (1998). Detection of bone metastases in breast cancer by 18FDG PET: differing metabolic activity in osteoblastic and osteolytic lesions. *J Clin Oncol,* 16(10):3375-9.

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(2):542-3.

Crippa F, Agresti R, Seregni E et al. (1998a). Prospective evaluation of fluorine-18-FDG PET in presurgical staging of the axilla in breast cancer. *J Nucl Med,* 39(1):4-8.

Crippa F, Seregni E, Agresti R et al. (1998b). Association between [18F]fluorodeoxyglucose uptake and postoperative histopathology, hormone receptor status, thymidine labelling index and p53 in primary breast cancer: a preliminary observation. *Eur J Nucl Med,* 25(10):1429-34.

Crossin JA, Johnson AC, Stewart PB et al. (1998). Gamma-probe-guided resection of the sentinel lymph node in breast cancer. *Am Surg,* 64(7):666-8; discussion 669.

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(2):132-40.

Czerniecki BJ, Scheff AM, Callans LS et al. (1999). Immunohistochemistry with pancytokeratins improves the sensitivity of sentinel lymph node biopsy in patients with breast carcinoma. *Cancer,* 85(5):1098-103.

Dehdashti F, Flanagan FL, Mortimer JE et al. (1999). Positron emission tomographic assessment of "metabolic flare" to predict response of metastatic breast cancer to antiestrogen therapy. *Eur J Nucl Med,* 26(1):51-6.

Doting MHE, Jansen L, Nieweg OE et al. (2000). Lymphatic mapping with intralesional tracer administration in breast carcinoma patients. *Cancer*, 88(11):2546-52.

Early Breast Cancer Trialists' Collaborative Group. (1992). Systematic treatment of early breast cancer by hormonal, cytotoxic, or immune therapy: 133 randomized trials involving 31,000 recurrences and 24,000 deaths among 75,000 women. *Lancet*, 339:1-15, 71-85.

Fisher B. (1999). Highlights from recent National Surgical Adjuvant Breast and Bowel Project studies in the treatment and prevention of breast cancer. *CA Cancer J Clin,* 49:159-77.

Galli A, Massaza L, Chiappo L et al. (2000). Study of sentinel lymph node in breast cancer: our experience in 46 cases. *Tumori,* 86(4):317-9.

Giuliano AE, Haigh PI, Brennan MB et al. (2000). Prospective observational study of sentinel lymphadenectomy without further axillary dissection in patients with sentinel node-negative breast cancer. *J Clin Oncol,* 18(13):2553-9.

Greco M, Crippa F, Agresti R, Seregni E et al. (2001). Axillary lymph node staging in breast cancer by 2-fluoro-2deoxy-D-glucose-positron emission tomography: Clinical evaluation and alternative management. *J Natl Cancer Inst*, 93(8):630-5.

Gucciardo G, Schiavo R, Grillo LR et al. (2000). The success rate of identification of the sentinel lymph node in breast cancer: our training series. *Tumori,* 86(4):312-3.

Gulec SA, Moffat FL, Carroll RG et al. (1998). Sentinel lymph node localization in early breast cancer. *J Nucl Med,* 39(8):1388-93.

Harris JR, Morrow M, Norton L. (1997). Malignant tumors of the breast. In DeVita VT, Hellman S, Rosenberg SA (eds.), Cancer: Principles and Practice of Oncology, Fifth Edition, Lippincott-Raven, Philadelphia, 1557-1616.

Hathaway PB, Mankoff DA, Maravilla KR et al. (1999). Value of combined FDG PET and MR imaging in the evaluation of suspected recurrent local-regional breast cancer: preliminary experience. *Radiology,* 210(3):807-14.

Henderson IC. (1994). Adjuvant systemic therapy for early breast cancer. *Cancer,* 74:401-9.

Hoh CK, Hawkins RA, Glaspy JA et al. (1993). Cancer detection with whole-body PET using 2-[18F]fluoro-2-deoxy-D- glucose. *J Comput Assist Tomogr,* 17(4):582-9.

Irwig L, Tosteson ANA, Gatsonis C et al. (1994). Guideline for meta-analysis evaluating diagnostic tests. *Ann Intern Med,* 120:667-676.

Jansson T, Westlin JE, Ahlstrom H et al. (1995). Positron emission tomography studies in patients with locally advanced and/or metastatic breast cancer: a method for early therapy evaluation? *J Clin Oncol,* 13(6):1470-7.

Kapteijn BAE, Nieweg OE, Petersen JL et al. (1998). Identification and biopsy of the sentinel

lymph node in breast cancer. *Eur J Surg Oncol*, 24:427-30.

Kole AC, Nieweg OE, Pruim J et al. (1997). Standardized uptake value and quantification of metabolism for breast cancer imaging with FDG and L-[1-11C]tyrosine PET. *J Nucl Med,* 38(5):692-6.

Kollias J, Gill PG, Coventry BJ et al. (2000). Clinical and histological factors associated with sentinel node identification in breast cancer. *Aust NZ J Surg,* 70(7):485-9.

Krag D, Weaver D, Ashikaga T et al. (1998). The sentinel node in breast cancer. A multicenter study. *N Engl J Med*, 339(14):941-6.

Lacquement MA, Mitchell D, Hollingsworth AB. (1999). Positive predictive value of the Breast Imaging Reporting and Data System. *J Am Coll Surg,* 189(1):34-40.

Lau J. (1997). Meta-Test version 0.6. New England Medical Center; Boston, Mass.

Liberman L, Abramson AF, Squires FB et al. (1998). The Breast Imaging Reporting and Data System: positive predictive value of mammographic features and final assessment categories. *AJR,* 171:35-40.

Liu TJ, Yeh DC, Wu CC et al. (2000). Selective sentinel lymph node dissection in breast cancer: experiences from Taiwan. *Surg Clin North Am,* 80(6):1779-86.

Lonneux M, Borbath I, Berliere M et al. (2000). The place of whole-body PET FDG for the diagnosis of distant recurrence of breast cancer. *Clin Positron Imaging*, 3(2):45-9.

Martin RC, Edwards MJ, Wong SL et al. (2000). Practical guidelines for optimal gamma probe detection of sentinel lymph nodes in breast cancer: results of a multi-institutional study. For the University of Louisville Breast Cancer Study Group. *Surgery,* 128(2):139-44.

Moffat FL, Gulec SA, Sittler SY et al. (1999). Unfiltered sulfur colloid and sentinel node biopsy for breast cancer: technical and kinetic considerations. *Ann Surg Oncol,* 6(8):746-55.

Moon DH, Maddahi J, Silverman DH et al. (1998). Accuracy of whole-body fluorine-18-FDG PET for the detection of recurrent or metastatic breast carcinoma. *J Nucl Med,* 39(3):431-5.

Morgan A, Howisey RL, Aldape HC et al. (1999). Initial experience in a community hospital with sentinel lymph node mapping and biopsy for evaluation of axillary lymph node status in palpable invasive breast cancer. *J Surg Oncol,* 72(1):24-30; discussion 30-1.

Mortimer JE, Dehdashti F, Siegel BA et al. (1996). Positron emission tomography with 2-[18F]Fluoro-2-deoxy-D-glucose and 16alpha-[18F]fluoro-17beta-estradiol in breast cancer: correlation with estrogen receptor status and response to systemic therapy. *Clin Cancer Res,* 2(6):933-9.

Mortimer JE, Dehdashti F, Siegel BA et al. (2001). Metabolic flare: An indicator of hormone responsiveness in advanced breast cancer. *(In press).*

Murthy K, Aznar M, Bergman AM et al. (2000). Positron emission mammographic instrument: initial results. *Radiology,* 215(1):280-5.

Nieweg OE, Kim EE, Wong WH et al. (1993). Positron emission tomography with fluorine-18-deoxyglucose in the detection and staging of breast cancer. *Cancer,* 71(12):3920-5.

NIH Consensus Statement. (1997). Breast cancer screening for women ages 40-49. *NIH Consens Statement,* 15(1):1-35.

Noh DY, Yun IJ, Kim JS et al. (1998). Diagnostic value of positron emission tomography for detecting breast cancer. *World J Surg,* 22(3):223-7; discussion 227-8.

O'Hea BJ, Hill AD, El-Shirbiny AM et al. (1998). Sentinel lymph node biopsy in breast cancer: initial experience at Memorial Sloan-Kettering Cancer Center. *J Am Coll Surg,* 186(4):423-7.

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(1):99-103.

Olson JA, Fey J, Winawer J et al. (2000). Sentinel lymphadenectomy accurately predicts nodal status in T2 breast cancer. *J Am Coll Surg,* 191(6):593-9.

Orel SG, Sullivan DC, Dambro TJ. (1997). BI-RADS categorization as a predictor of malignancy (abstr). *Radiology,* 205(P):447.

Oshida M, Uno K, Suzuki M et al. (1998). Predicting the prognoses of breast carcinoma patients with positron emission tomography using 2-deoxy-2-fluoro[18F]-D-glucose. *Cancer,* 82(11):2227-34.

Palmedo H, Bender H, Grunwald F et al. (1997). Comparison of fluorine-18 fluorodeoxyglucose positron emission tomography and technetium-99m methoxyisobutylisonitrile scintimammography in the detection of breast tumours. *Eur J Nucl Med,* 24(9):1138-45.

Pijpers R, Meijer S, Hoekstra OS et al. (1997). Impact of lymphoscintigraphy on sentinel node identification with technetium-99m-colloidal albumin in breast cancer. *J Nucl Med,* 38(3):366-8.

Pizzocaro C, Rossini PL, Terzi A et al. (2000). Sentinel node biopsy in breast cancer: the experience of Brescia Civic Hospital. *Tumori,* 86(4):309-11.

Rockette HE, Redmond CK, Fisher B et al. (1982). Impact of randomized clinical trials on therapy of primary breast cancer: the NSABP overview. *Control Clin Trials,* 3:209-225

Rostom AY, Powe J, Kandil A et al. (1999). Positron emission tomography in breast cancer: a clinicopathological correlation of results. *Br J Radiol,* 72(863):1064-8.

Roumen RMH, Valkenburg JGM, Geuskens LM. (1997). Lymphoscintigraphy and feasibility of sentinel node biopsy in 83 patients with primary breast cancer. *Eur J Surg Oncol,* 23:495-502.

Rubio IT, Korourian S, Cowan C et al. (1998). Sentinel lymph node biopsy for staging breast cancer. *Am J Surg,* 176(6):532-7.

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(6):618-23.

Schelbert HR. (1991). Principles of positron emission tomography. In: *Cardiac Imaging: A Companion to Braunwald's Heart Disease.* Marcus ML, Schelbert HR, Skorton DJ et al. eds. Philadelphia: W.B. Saunders, pp. 1140-68.

Schelling M, Avril N, Nahrig J et al. (2000). Positron emission tomography using [(18)F]Fluorodeoxyglucose for monitoring primary chemotherapy in breast cancer. *J Clin Oncol,* 18(8):1689-95.

Schirrmeister H, Guhlmann A, Kotzerke J et al. (1999). Early detection and accurate description of extent of metastatic bone disease in breast cancer with fluoride ion and positron emission tomography. *J Clin Oncol,* 17(8):2381-9.

Schirrmeister 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 imaging procedures. *Eur J Nucl Med,* 28(3):351-8.

Simes RJ, Margrie SJ. (1991). Patient preferences for adjuvant chemotherapy in breast cancer: report #8901. NHMRC Clinical Trials Center: Sydney, Australia.

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(2):220-7.

Smith IC, Welch AE, Hutcheon AW et al. (2000). Positron emission tomography using [(18)F]-fluorodeoxy-D-glucose to predict the pathologic response of breast cancer to primary chemotherapy. *J Clin Oncol,* 18(8):1676-88.

Snider H, Dowlatshahi K, Fan M et al. (1998). Sentinel node biopsy in the staging of breast cancer. *Am J Surg,* 176(4):305-10.

Spillane AJ, Sacks NP. (1999). Which patients with early breast cancer need axillary clearance? *Swiss Surg,* 5(5):205-13.

Tasmuth T, Blomqvist C, Kalso E. (1999). Chronic post-treatment symptoms in patients with breast cancer operated in different surgical units. *Eur J Surg Oncol,* 25:38-43.

Tse NY, Hoh CK, Hawkins RA et al. (1992). The application of positron emission tomographic imaging with fluorodeoxyglucose to the evaluation of breast disease. *Ann Surg,* 216(1):27-34.

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(12):1588-93.

Veronesi U, Paganelli G, Viale G et al. (1999). Sentinel lymph node biopsy and axillary dissection in breast cancer: Results in a large series. *J Natl Cancer Inst*, 91(4):368-73.

Wahl RL. (1995). Positron emission tomography: applications in oncology. In: *Nuclear Medicine in Clinical Diagnosis and Treatment.* Murray IPC and Ell PJ, eds. New York: Churchill Livingstone, pp. 801-20.

Wahl RL, Cody RL, Hutchins GD et al. (1991). Primary and metastatic breast carcinoma: initial clinical evaluation with PET with the radiolabeled glucose analogue 2-[F-18]-fluoro-2-deoxy- D-glucose. *Radiology,* 179(3):765-70.

Wahl RL, Zasadny K, Helvie M et al. (1993). Metabolic monitoring of breast cancer chemohormonotherapy using positron emission tomography: initial evaluation. *J Clin Oncol,* 11(11):2101-11.

Winer WP, Morrow M, Osborne CK et al. (2001). Malignant tumors of the breast. Chapter 37.2. In: DeVita VT Jr, Hellman S, Rosenberg SA, eds. *Cancer: Principles and Practice of Oncology, 6th ed.* Lippincott Williams and Wilkins: Philadelphia; 1651-1717.

Yang JH, Nam SJ, Lee TS et al. (2001). Comparison of intraoperative frozen section analysis of sentinel node with preoperative positron emission tomography in the diagnosis of axillary lymph node status in breast cancer patients. *Jpn J Clin Oncol*, 31(1):1-6.

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(2):274-80.

Yutani K, Tatsumi M, Shiba E et al. (1999a). 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(6):1003-8.

Yutani K, Tatsumi M, Uehara T et al. (1999b). Effect of patients' being prone during FDG PET for the diagnosis of breast cancer. *AJR Am J Roentgenol,* 173(5):1337-9.

APPENDIX

Abbreviations Key

# LN+	number of positive lymph nodes
# Succ	number of successful sentinel node mappings
(CE or CT)	cyclophosphamide plus epirubicin or paclitaxel
99m Tc	radioactive technetium
abn	abnormal
AC	attenuation correction used
ALND	axillary lymph node dissection
BD	blue dye
bx	biopsy
C, A, M, 5-FU, T, Premarin	cyclophosphamide, doxorubicin (Adriamycin®), methotrexate, fluorouracil, tamoxifen, Premarin®
ca	cancer
clin	clinical
cN+	clinically node positive
cNO	clinically node negative
Coincid FDG	coincidence detection FDG imaging
consec	consecutive
Corr.d for p vol & gluc	corrected for partial volume and glucose
CT	computed tomography
CTX	chemotherapy
CVAP	cyclophosphamide, vincristine, doxorubicin and prednisolone
CXR	chest X-ray
cytol	cytology
DA	diagnostic accuracy
DM	diabetes mellitus
Doc	docetaxel
DUR	differential uptake ratio
dx	diagnosis
ER+	estrogen receptor positive
Eval	evaluable
excl	exclusion
f/u	follow-up
FN	false negative
FNAB	fine needle aspiration biopsy
FP	false positive
GDP	gamma detecting probe
grps	groups
H & E	hematoxylin and eosin
hist	histology
HSA	human serum albumin
hx	history
ID	intradermal
IHC	immunohistochemistry
Interp	interpretation
IP	intraparenchymal
IUAC	International Union Against Cancer
KPS	Karnovsky Performance Scale
L	lesion
LN	lymph node
LS	lymphoscintigraphy
med	median

Abbreviations Key (cont'd)

mets	metastases
MIBI-SPECT	99-technetium sestamibi single photon emission computed tomography
MM	mammography
mo	month
MR	magnetic resonance imaging
n	number of subjects
neg	negative
nm	nanometer
NPV	negative predictive value
P	patient
path	pathology
pCR	pathologic complete response
PE, phys exam	physical examination
perm sect	permanent sections
PET	positron emission tomography
PET:RS Blind	PET interpreted blindly with respect to reference standard
pos	positive
pPR	pathologic partial response
PPV	positive predictive value
pregn	pregnancy
preop	preoperative
pretx	pretreatment
prev benign	prevalence of benign condition
Prev Dis	prevalence of disease
Prev LN+	prevalence of lymph node positive disease
Prev Malign	prevalence of malignancy
Prev Met	prevalence of metastasis
prev tx	previous treatment
pro	prospective
pts	patients
qual	qualitative
quant	quantitative
R	region
RBS	radionuclide bone scan
RC	radiocolloid
Ref Stand	reference standard
retro	retrospective
ROC	receiver operating characteristic
RS: PET Blind	reference standard interpreted blindly with respect to PET
RT	radiation therapy
SC	subcutaneous
SD	standard deviation
sect(s)	section(s)
SENS	sensitivity
SPEC	specificity
stand	standard
SUV	standardized uptake value
sx	symptoms
TMx/SMx	total or segmental mastectomy
TN	true negative
TP	true positive
tx	therapy
US	ultrasound
Verif Bias	verification bias
w/	with
w/o	without

Table A1: Staging Axillary Lymph Nodes, Studies Reporting Data for Patients with Palpable Lymph Nodes

Study	Total n	Palpable Subgroup n	TP	FN	FP	TN	Sens	Spec
Greco et al. (2001)	167	38	29	1	2	6	97%	75%
Ohta et al. (2000)	32 33 R	10	10	0	0	0	100%	
Smith et al. (1998)	50	14	10	1	0	3	91%	100%
Crowe et al. (1994)	20	5	4	0	0	1	100%	100%
Totals	**269** **270 R**	**67**						

	Sens	95% CI		Spec	95% CI	
Random Effects Meta-Analysis Results:	**93%**	**81%**	**98%**	**78%**	**49%**	**93%**

Table A2: Staging Axillary Lymph Nodes, All Studies Irrespective of Clinical Nodal Status

Study	N	Design	Patient Selection	Mean Age (SD)	Mean T Size (SD)	PET Interp	Avoid Verif Bias	PET: RS Blind	RS: PET Blind	Ref Stand	Test/ Site	UA	TP	FN	FP	TN	Prev Met	Sens	Spec
Greco et al. (2001) Milan, Italy	167	pro	T1/T2 breast ca, Scheduled to receive ALND	54.0	2.1 T1-59% T2-41%	qual, AC	Y	Y	?	hist	PET	P	68	4	13	82	43%	94%	86%
											PET-cN0	P	39	3	11	76	33%	93%	87%
											PET-cN+	P	29	1	2	6	79%	97%	75%
Schirrmeister et al. (2001) Ulm, Germany	113	pro	palpable breast masses/suspicious lesions on MM/US	56.8		qual	?	Y	?	hist	PE	P	14	20	3	78	30%	41%	96%
											PET	P	27	7	6	73	30%	79%	92%
Yang et al. (2001) Seoul, South Korea	18	?	breast ca, underwent ALND; cN0 (94%), cN+ (6%)	44.7	3.5 (2.0) T1-31% T2-38% T3-31%	quant	?	?	?	hist	SNB	P	5	1	0	12	33%	83%	100%
											PET	P	3	3	0	12	33%	50%	100%
Ohta et al. (2000) Isehara, Japan	32 33 R	pro	underwent PET, US, ALND in 30, ALN sampling in 1, cN0 (70%), cN+ (30%)	50.0 med		qual, AC	?	?	?	hist	palpation	R	11	9	0	13	61%	55%	100%
											US	R	13	7	0	13	61%	65%	100%
											PET	R	14	6	0	13	61%	70%	100%
											PET+US	R	15	5	0	13	61%	75%	100%
											PET-cN0	R	4	6	0	13	43%	40%	100%
											PET-cN+	R	10	0	0	0	100%	100%	
Yutani et al. (2000) Osaka, Japan	38	pro	consect pts w/ suspicious lesions on PE, MM, US; path proven breast ca	50.9 (13.4)	2.1 (1.0) T1a-3% T1b-18% T1c-29% T2-50%	qual, AC	Y	Y	?	hist	MIBI-SPECT	P	6	10	0	22	42%	38%	100%
											PET	P	8	8	0	22	42%	50%	100%

Table A2: Staging Axillary Lymph Nodes, All Studies Irrespective of Clinical Nodal Status (cont'd)

Study	N	Design	Patient Selection	Mean Age (SD)	Mean T Size (SD)	PET Interp	Avoid Verif Bias	PET: RS Blind	RS: PET Blind	Ref Stand	Test/ Site	UA	TP	FN	FP	TN	Prev Met	Sens	Spec
Rostom et al. (1999) Saudi Arabia	74	retro	consec pts attending breast clinic; path proven breast ca	40.3		qual, AC in half	Y	Y	?	hist	PET	P	42	7	0	25	66%	86%	100%
Noh et al. (1998) Seoul, South Korea	24 27 R	?	breast ca		2.0 (med)	?, AC	?	?	?	hist	palpation MM PET	R R R	8 4 14	7 10 1	0 0 0	12 12 12	56% 54% 56%	53% 29% 93%	100% 100% 100%
Smith et al. (1998) Aberdeen, Scotland	50	pro	dx of breast ca; cN0 (70%), cN+ (30%)	67.0	T1-20% T2-42% T3-18% T4-20%	qual	?	Y	?	cytol (5), hist (45)	palpation PET PET-cN0 PET-cN+	P P P P	12 19 9 10	9 2 1 1	3 1 1 0	26 28 25 3	42% 42% 28% 79%	57% 90% 90% 91%	90% 97% 96% 100%
Adler et al. (1997) University Hospitals, Cleveland	50 52 R	pro	interpretable ALN PET scans available, operable breast ca, ≥ level 2 ALND ≤3 mo of PET, ≥ 10 LNs dissected; tumor >5 mm		T1b-18% T1c-43% T2-33% T3-6%	qual	?	Y	?	hist	PET	R	19	1	11	21	38%	95%	66%
Palmedo et al. (1997) Bonn, Germany	20	pro	abn MM or palpable mass; cN0 (65%), cN+ (35%)	58.4	2.8 (1.6) T1b-14% T1c-29% T2-43% T3-14%	qual, quant, AC	?	Y	?	hist	MIBI-SMM PET	P P	4 5	2 1	0 0	14 14	30% 30%	67% 83%	100% 100%

Table A2: Staging Axillary Lymph Nodes, All Studies Irrespective of Clinical Nodal Status (cont'd)

Study	N	Design	Patient Selection	Mean Age (SD)	Mean T Size (SD)	PET Interp	Avoid Verif Bias	PET: RS Blind	RS: PET Blind	Ref Stand	Test/ Site	UA	TP	FN	FP	TN	Prev Met	Sens	Spec
Avril et al. (1996b) Munich, Germany	51	pro	newly discovered breast ca, surgery scheduled, cN0 (55%), cN+ (45%)	49.9 (10.3)		qual, AC	?	Y	?	hist	phys exam	P	14	10	4	23	47%	58%	85%
											PET	P	19	5	1	26	47%	79%	96%
											PET-pT1	P	2	4	0	12	33%	33%	100%
											PET->pT1	P	17	1	0	5	78%	94%	100%
Bassa et al. (1996) MD Anderson, Houston	16	retro	consec pts, locally advanced breast ca to receive neoadju-vant CTX (15), no hepatic mets on CT; cN0 (13%), cN+ (87%)	43.8 (9.5)	T2-12% T3-50% T4-38%	qual, AC	Y	?	?	hist	MM	P	7	3	0	2	83%	70%	100%
											US	P	12	2	0	2	88%	86%	100%
											PET	P	10	3	0	3	81%	77%	100%
Scheidhauer et al. (1996) Cologne, Germany	18	pro	suspicion of breast ca based on palpa-tion, MM, US; under-went surgery	57.0		qual, AC	?	Y	?	hist	PET	P	9	0	1	8	50%	100%	89%
Utech et al. (1996) Peoria, Illinois	124	pro	newly diagnosed breast ca; cN0 (64%), cN+ (36%)	59.0	T1-67% T2-29% T3-4%	qual, AC	?	Y	?	hist	PET	P	44	0	20	60	35%	100%	75%
Hoh et al. (1993) UCLA	14	retro	underwent whole-body PET and had correlative tissue biopsy			qual	?	?	?	hist	PET	P	6	3	0	5	64%	67%	100%

Table A3. Staging Axillary Lymph Nodes, Irrespective of Clinical Nodal Status, Summary

Unit	Study n	Pts n	Design			Avoid Verif Bias			Blinding PET: RS			RS:PET			Sensitivity Range		Specificity Range		Random Effects Meta-Analysis					
			Pro	Retro	?	Y	N	?	Y	N	?	Y	N	?					Sensitivity 95% CI			Specificity 95% CI		
			Study n			Study n			Study n			Study n			Range		Range		95% CI			95% CI		
R	3	106 114 R	2	-	1	-	-	3	1	-	2	-	-	3	70%	95%	66%	100%						
P	11	703	8	3	1	4	-	8	9	-	3	-	-	12	50%	100%	75%	100%						
All	15	809	10	3	2	4	-	11	10	-	5	-	-	15	50%	100%	66%	100%	82%	73%	88%	90%	83%	94%

Figure A1: Meta-analysis, Staging of Axillary Lymph Nodes, Irrespective of Clinical Nodal Status, Random Effects Model
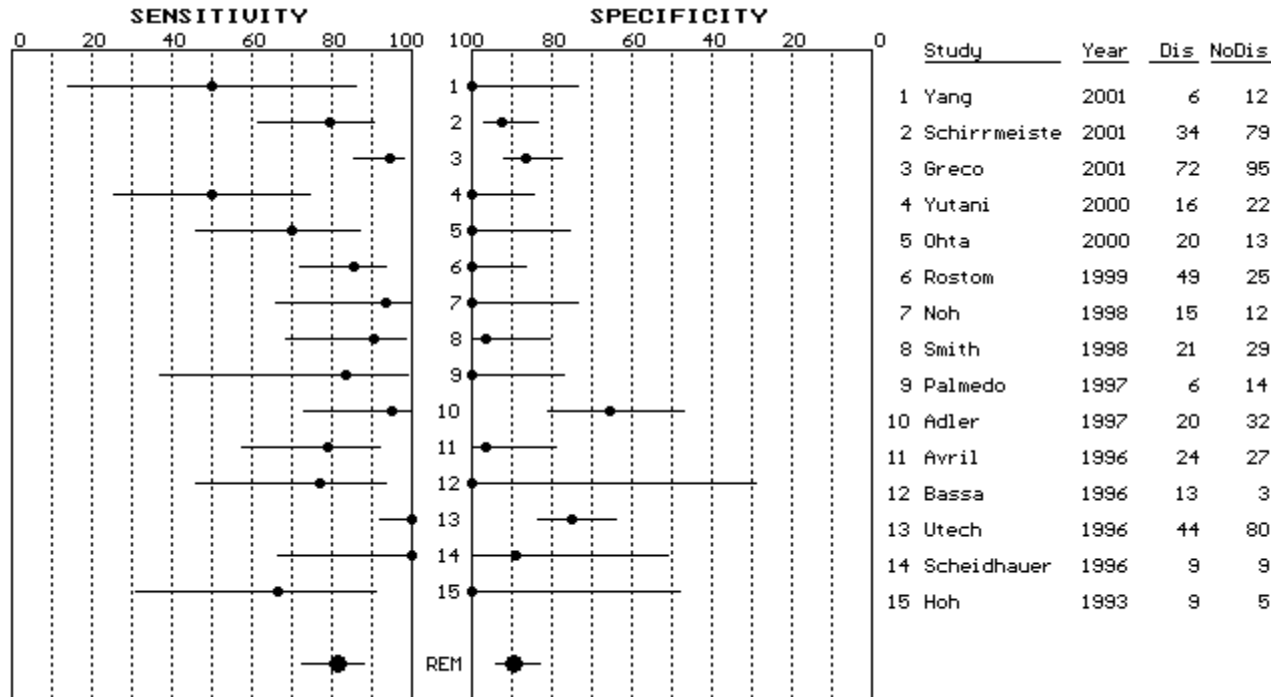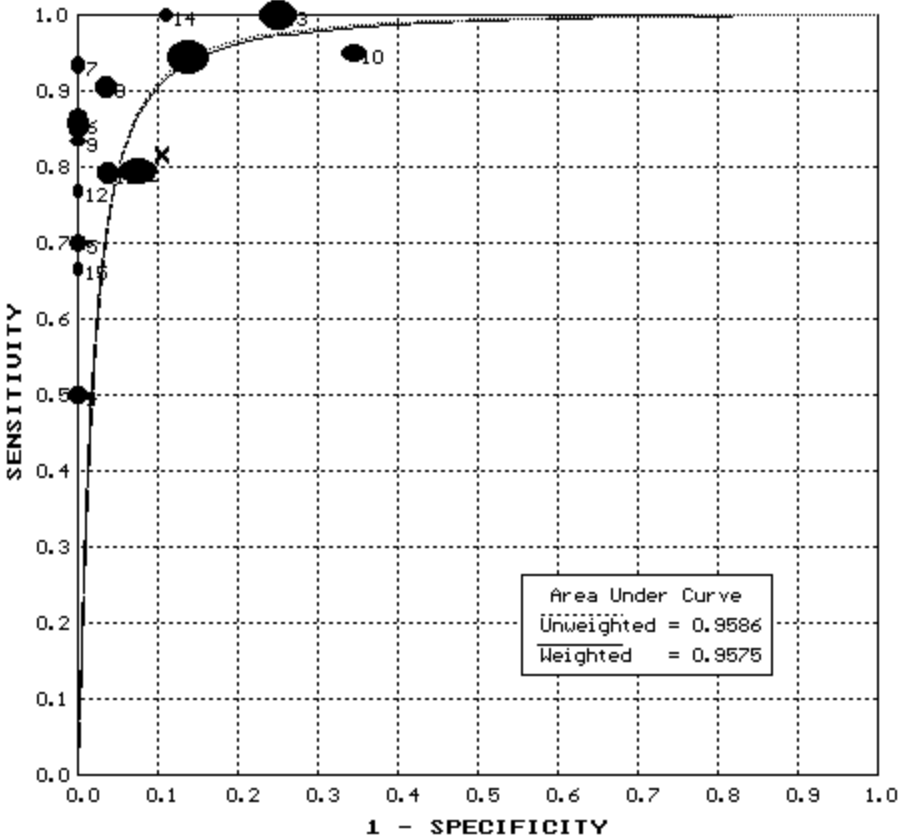
Figure A2: Summary ROC Curve, Staging Axillary Lymph Nodes, Irrespective of Clinical Nodal Status

X: Random Effects Meta-Analysis Estimates