

Lung cancer staging: the value of PET depends on the clinical setting

Frank C. Detterbeck¹, Santiago Figueroa Almanzar²

¹Department of Surgery, Division of Thoracic Surgery, Yale University School of Medicine, New Haven, CT 06520-8062, USA; ²General University Hospital of Valencia, Avda, Tres Cruces 2, Valencia 46014, Spain

Correspondence to: Frank C. Detterbeck, MD. Department of Surgery, Division of Thoracic Surgery, Yale University School of Medicine, PO Box 208062, New Haven, CT 06520-8062, USA. Email: frank.detterbeck@yale.edu; Santiago Figueroa Almanzar, MD. Clinical University Hospital of Valencia, Avda, Blasco Ibañez 17, Valencia 46010, Spain. Email: santiago.figueroa@me.com.

Background: Although positron emission tomography (PET) imaging is widely recommended in the evaluation of patients with lung cancer, randomized controlled trials (RCTs) assessing this have demonstrated inconsistent results. We asked whether differences in the clinical context and endpoints could explain these discrepancies.

Methods: We used realist synthesis methods to analyze how contextual differences among RCTs affected the results. We focused on RCTs to minimize confounding yet permit evaluation of differences by comparing across studies.

Results: This analysis suggests that the impact of PET depends on the clinical setting. PET is of greatest benefit in identifying M1 disease in patients with a high chance of such involvement and when little traditional imaging [e.g., abdominal/pelvis computed tomography (CT) and bone scan] is used. Identification of N2,3 involvement by PET prior to resection is seen primarily when there is at least a moderate probability of such and the rate of invasive staging is high. The rate of N2 disease not identified preoperatively appears to increase if PET is used to avoid invasive mediastinal staging in clinical settings in which the risk of N2,3 involvement is moderately high. There is both a potential benefit in avoiding stage-inappropriate resection as well as a risk of missed (stage-appropriate) resection if PET findings are not evaluated carefully.

Conclusions: A blanket recommendation for PET may be too simplistic without considering nuances of the clinical setting.

Keywords: Non-small cell lung cancer; positron emission tomography (PET); staging

Submitted Jun 28, 2014. Accepted for publication Sep 23, 2014.

doi: 10.3978/j.issn.2072-1439.2014.11.16

View this article at: <http://dx.doi.org/10.3978/j.issn.2072-1439.2014.11.16>

Introduction

Positron emission tomography (PET)—more recently PET with integrated computed tomography (PET/CT)—has become a cornerstone in evaluating patients with lung cancer. This paper focuses on the impact of PET in the pretreatment evaluation of patients with suspected lung cancer. One can look at specific components of this process, such as diagnosis or identification of nodal or distant metastases, but the clinical value is determined by the overall impact of PET on the accuracy of pre-treatment

patient evaluation (i.e., definition of disease extent).

Randomized controlled trials (RCTs) assessing the value of PET in patients with known or suspected lung cancer have yielded somewhat inconsistent results. The RCTs suggest that upfront PET is similarly efficient compared to traditional staging (1) and that adding PET to traditional staging does (2) or does not (3) identify more metastases, and does (1,4,5) or does not (3) reduce the number of so-called “futile” thoracotomies. PET does (1) or does not (2,4,5) reduce the rate of invasive mediastinal staging. Furthermore, PET does not seem to affect the overall

survival of patients (2,5).

The inconsistent RCT results leave the exact value of PET unclear. There are significant differences in the design of these trials, the patients included, and the outcomes assessed. This paper explores the results and differences between these trials in order to achieve a better understanding of the role of PET and factors that influence this in patients with lung cancer.

Methods

We chose a realist synthesis method as most appropriate to develop an understanding of factors influencing the impact of PET (what works, for whom, in which setting, why and how?) (6-8). The realist method considers that implementation of an intervention may yield slightly different results depending on the context. This approach combines theoretical understanding and available data to reach a deeper understanding of how an intervention produces the observed outcomes, with a focus on explaining the relationship between the context in which the intervention is applied, the mechanisms by which it works and the outcomes which are produced. This review was conducted following the RAMESES publication standards for realist syntheses (*Figure S1*) (7).

Preliminary scoping of the literature used the recent extensive systematic review conducted for the ACCP Lung Cancer Guidelines (9). We limited the focus to RCT because the impact of PET *vs.* traditional staging is not confounded, yet allows analysis of contextual factors by comparing across studies.

We performed a MEDLINE search [1990-2013] for English language papers assessing the utility of PET or PET/CT in the pre-treatment evaluation of patients with known or suspected lung cancer in a RCT (details available on request). We excluded studies of PET for chemotherapy response, RT treatment planning or restaging after induction therapy. We did not attempt to identify unpublished studies. There was no funding support and no involvement of other people or organizations.

We found five RCTs, involving 1,362 patients (*Table S1*) (1-5,10). Data was abstracted regarding study design, end-points, PET technology, scan interpretation, patient characteristics, and details of pre-enrollment and post-enrollment further studies. Slight discrepancies were noted between two papers reporting on one study; data from the later publication was chosen after communication with the authors (2,10).

The selected studies used various denominators and endpoints, including detection of mediastinal or distant metastases, early recurrence and appropriateness of resection. To enhance comparability, we abstracted raw data and calculated outcomes consistently across the studies on an intent-to-treat basis, for all enrolled eligible patients.

We calculated results according to parameters that could be affected by PET (i.e., whether PET identifies more patients with benign disease, N2,3 or M1 involvement preoperatively *vs.* intra-operatively or within 1 year). We excluded endpoints which PET is unlikely to impact (e.g., “unresectability” or T4) although we show these results when reported. We excluded outcomes such as unrelated death within 1 year, because it seems inappropriate to expect PET to predict unrelated events.

We defined stage-inappropriate resection as surgery for patients with benign lesions or with N2,3 or M1 involvement (although we acknowledge that exceptions exist and sometimes surgery may be considered appropriate). The Viney *et al.*'s study (3) results are presented as if N2 involvement was a contraindication for surgery (contrary to the study authors' policy) in order to be consistent with the general view and with the other studies. Avoidance of inappropriate surgery is important, but so is inappropriately missed resection due to falsely interpreted preoperative staging. Because this was not explicitly reported, we estimated the risk of missed stage-appropriate resection from the incidence of positive PET results subsequently shown to be false-positives.

We assigned a qualitative assessment in each study to patient characteristics and the extent of pre- or post-enrollment but preoperative testing for N2,3 or M1 disease in order to facilitate evaluation of how these factors influenced the study outcomes (a quantitative assessment was not possible). Because of heterogeneity in patient and study design characteristics, a formal meta-analysis or calculation of summary statistics across all studies is not appropriate.

Results

Study characteristics

End-points

The primary outcome in the PLUS and Fischer studies was “futile thoracotomies”, defined as preoperatively unrecognized benign disease, N2,3 or M1 involvement, or recurrence or any death within one year (related or not) (4,5).

Viney *et al.* also lists the thoracotomy rate as the primary endpoint, but effectively it is the identification of distant metastases (those with suspected N2 involvement underwent thoracotomy nevertheless) (3). The Maziak study assessed the percentage correctly and incorrectly upstaged, as well as incorrect understaging (but not correct downstaging) by PET/CT *vs.* traditional staging (2). Herder assessed whether PET as the first test reduced the number of tests/procedures to finalize staging and define operability, with secondary endpoints of work-up duration, morbidity and costs (1).

Patient characteristics

The RCTs all enrolled patients deemed potential candidates for surgical resection (*Table 1*), but according to varying criteria. Some studies (1,4) included many patients with significant weight loss—generally considered a marker of distant metastasis. Some studies (4,5) included many patients with clinical evidence of mediastinal node involvement. Merging such features yields a qualitative assessment of the risk of advanced disease, which varies markedly (*Table 1*).

In most studies, the vast majority of patients had lung cancer—either as demonstrated by subsequent work-up or mandated by biopsy prior to study entry. Only the Herder study included many patients (47%) who did not have lung cancer (1).

Patient entry into the RCTs varied markedly: from being referred by a general practitioner (GP) on the basis of a chest radiograph (CXR) alone (1), to a prerequisite of biopsy proof, specialist evaluation and extensive traditional imaging (3). A qualitative assessment of the extent of pre-enrollment evaluation can be assigned (see the last column in *Table 1*).

Thoroughness of preoperative staging

The extensiveness of staging procedures to rule out mediastinal and distant metastases differ markedly between the studies (*Table 2*). The value of PET may be different if all or few patients undergo further testing for N2,3 or M1 disease. The extent of further testing was generally similar between the PET and traditional work-up arms with 2 exceptions. In the Herder study mediastinoscopies were done less often in the PET arm (13% *vs.* 34%) (1). The Maziak study mandated liver/adrenal CT and bone scans only in the traditional arm (2).

Potential factors influencing the impact of PET include scanner technology, interpretation quality, and the extent of confirmation of abnormal findings. These factors also

varied between studies (*Table 2*). The thoroughness of the PET interpretation is patterned according to a proposed scale (11). In most studies only a few institutions performed PET scans for all participating sites. This concentrated experience may affect the quality of the interpretation. Only in the Maziak study was PET performed in a more disseminated fashion. The thoroughness of the PET and of preoperative testing for N2,3/M1 disease is qualitatively summarized in the last three columns of *Table 2*.

Outcomes

Preoperative identification of benign disease

The RCTs demonstrate little difference for PET *vs.* traditional evaluation in identifying benign disease preoperatively *vs.* intraoperatively (*Table 3*), but most studies included few patients with benign disease. PET identified more benign disease preoperatively only in the Herder study, which involved patients referred by the GP for potential lung cancer resection based on only a CXR (1). However, this observation is weakened by the fact that there were also more patients overall in the PET arm with benign disease (19% *vs.* 13%) despite randomization (1).

These results suggest that if the diagnosis of lung cancer is fairly certain, either due to a biopsy result or evaluation by a specialist, there is little benefit to PET to identify benign lesions. However, PET can be helpful to evaluate the primary lesion when there is little specialist involvement and limited diagnostic evaluation.

Preoperative identification of mediastinal node involvement

In a RCT the total number of patients with N2,3 involvement should be similar, but PET might identify N2,3 disease more often preoperatively rather than intra- or post-operatively. These results are summarized in *Table S2* and *Figure 1*. While PET seems beneficial in the Fisher and PLUS studies (4,5), the opposite was true in the Herder study (1). The Viney study provided no data for the traditional staging arm (3). The Maziak study observed fewer patients with intraoperative identification of N2 involvement in the PET arm, but with no corresponding increase of preoperatively identified N2 disease by PET (2,10). Therefore, the imbalance in the overall number of patients with N2 disease between the arms (despite randomization) seems to account for the lower rate of intraoperative discovery of N2, rather than PET imaging.

Comparing studies suggests that PET is beneficial in

Table 1 Patient characteristics at study entry

Study	N	Referred by	% biopsy proven	Pre-enrollment studies	PS ≥ 2 (%)	>5% weight loss (%)	Clinical stage at entry ^a (%)				Risk of advanced stage ^b	Extent of pre-enrollment evaluation
							I	II	III	IV		
Fisher	189	Pulm ^c	–	CT x2, Bronch	1	–	28	6	62	4	+++	Moderate
Plus	188	GP	52	Local routine	–	15	65	5	29	1	+++	Minimal
Herder	465	GP	–	CXR	4	31	Oper (38) ^d	(20) ^d	(20) ^d	(16) ^d	++++	Minimal
Maziak	337	Specialist	100	CT	4	–	78	12	10	0	++	Moderate
Viney	183	T Surg	98	CT x2, (Bn) ^e , Br	7	10	92	8	0	0	+	Good

Bn, bone scan; Br, brain scan [either computed tomography (CT) or magnetic resonance imaging (MRI)]; Bronch, fiberoptic bronchoscopy; CT, chest CT; CT x2, chest and upper abdomen CT; CXR, chest radiograph; GP, general practitioner; Oper, patient deemed candidates for resection (i.e., “operable”); Pulm, Pulmonologist; PS, performance status; T Surg, thoracic surgeon.^a, as reported in study (the low rate of stage IV in some studies appears in conflict with the proportion of patients with significant weight loss); ^b, risk based on presence of clinical markers of advanced disease (>5% weight loss, performance status ≥ 2); ^c, specialists who practice both pulmonary medicine and medical oncology for lung cancer; ^d, final clinical stage after completion of all post-enrollment studies (not the stage at enrollment); ^e, done selectively if symptoms or signs of bone metastases.

Table 2 Post-enrollment conduct of study

Study	Extent of work-up (%) ^a			PET technique	Thoroughness of PET read	Confirmation of suspicious PET findings		Thoroughness of preoperative staging			Quality/thoroughness of PET assessment
	Invasive MLN st.	Brain CT/MRI	Bone scan	Liver, adrenal							
Fisher	94	0 ^b	–	(100) ^c	PET/CT	High	All	Loose	High	Low	Good
Plus	69	4	27	47	PET + CT	Good	Most	Most	Mod	Low	Good
Herder	24 ^d	High ^e	High ^e	High ^e	PET ^f	Fair	–	–	Low	High	Fair
Maziak	50	100	100 ^g	100 ^g	PET/CT	High	All	All	Mod	High	High
Viney	5	(100) ^c	(Few) ^c	(100) ^c	PET ^h	Fair	All	All	Low	High	Good

CT, computed tomography; F/U, follow-up; MLN st., mediastinal lymph node staging (i.e., mediastinoscopy, endobronchial ultrasound and aspiration); MRI, magnetic resonance imaging; PET, positron emission tomography; PET/CT, integrated PET/CT scanner; PET + CT, PET scan with visual correlation with a separate CT.^a, in both arms equally unless indicated; ^b, except 1 patient in PET arm; ^c, done pre-enrollment; ^d, mediastinoscopy done significantly less often in PET arm (13% vs. 34%); ^e, not specified but patients underwent and average of 7.9 additional tests (equal in both arms); ^f, PET read in isolation without correlation with a CT; ^g, in conventional work-up arm; ^h, CT scan available but was not used during PET interpretation; PET reader knew that the mediastinum was deemed normal on all CT scans.

Table 3 Study endpoints-benign disease

Study	Referred by	% biopsy proven	Pre-enrollment studies	% with lung cancer of all enrolled	Extent of Pre-enrollment evaluation	% benign				Impact of PET on identifying benign Dx
						Preop		Intraop		
						Trad	PET	Trad	PET	
Herder	GP	–	CXR	53	Minimal	12	18	1	1	(†) ^a
Plus	GP	52	Local routine	93	Low	2	3	7	2	(†) ^a
Fisher	Pulm ^b	–	CT x2, Bronch	99	Moderate	–	–	3	0	–
Maziak	Specialist	100	CT	100	Moderate	NA	NA	NA	NA	–
Viney	T Surg	98	CT x2, (Bn) ^c , Br	99	Good	NA	2	NA	2	–

Note: all results calculated on intent to treat basis for entire cohort (arm) in each study. Bn, bone scan; Br, brain scan [either computed tomography (CT) or magnetic resonance imaging (MRI)]; Bronch, fiberoptic bronchoscopy; CT, chest CT; CT ×2, chest and upper abdomen CT; CXR, chest radiograph; Dx, diagnosis; GP, general practitioner; intraop, intraoperatively identified; NA, not applicable; PET, positron emission tomography; preop, preoperatively identified; Pulm, pulmonologist; T Surg, thoracic surgeon; Trad, traditional evaluation arm. ^a, The “benefit” of PET in pre- vs. intra-operative identification of benign disease is largely due to an discrepancy in the overall incidence of benign disease despite randomization; ^b, specialists who practice both pulmonary medicine and medical oncology for lung cancer; ^c, done selectively if there were symptoms or signs of bone metastases.

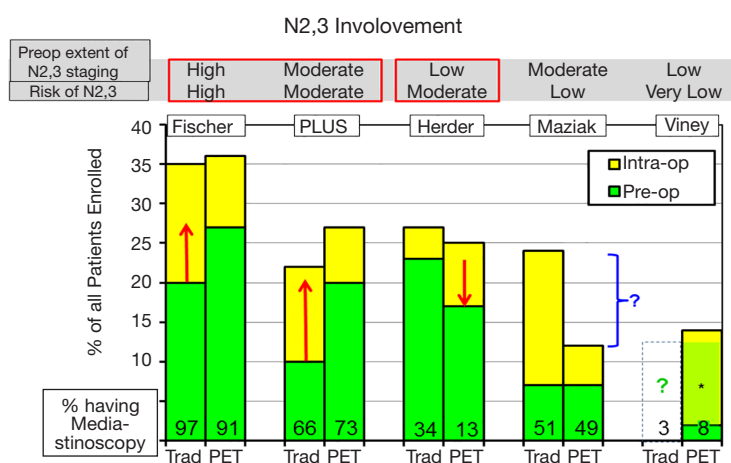


Figure 1 Rate of N2,3 involvement. Rate of N2,3 node involvement diagnosed pre- and intra-operatively. Upwards red arrows indicate a benefit to PET, downward red arrow indicates a detriment. Values represent the percent of all patients enrolled in the respective study arm. Intraop, intraoperatively; preop, preoperatively; Trad, traditional.

preoperatively identifying N2,3 involvement when the risk is high and there is a high rate of invasive staging. PET seems to help identify N2,3 disease even when almost all patients undergo mediastinoscopy-perhaps by directing attention to suspicious nodes. When PET is used to decrease the rate of invasive staging, it appears that the risk of intraoperative discovery of N2 involvement is increased, at least in a patient cohort with a moderate incidence of mediastinal nodal disease. When the risk of N2,3 disease is low, PET has little impact on preoperative identification of nodal involvement.

Preoperative identification of distant metastases

Does PET identify more patients with distant metastases preoperatively and reduce the number found to have M1 disease or recurrence within 1 year? This was seen in the Fisher and PLUS studies but not clearly so in the others (Table S3, Figure 2) (1-5).

An unexplained difference between the randomized arms exists in the Maziak study in the number of patients with M1 disease (identified at any time) (2). The discrepancy is in the opposite direction as for N2,3 involvement; however, these differences are not explained by postulating that

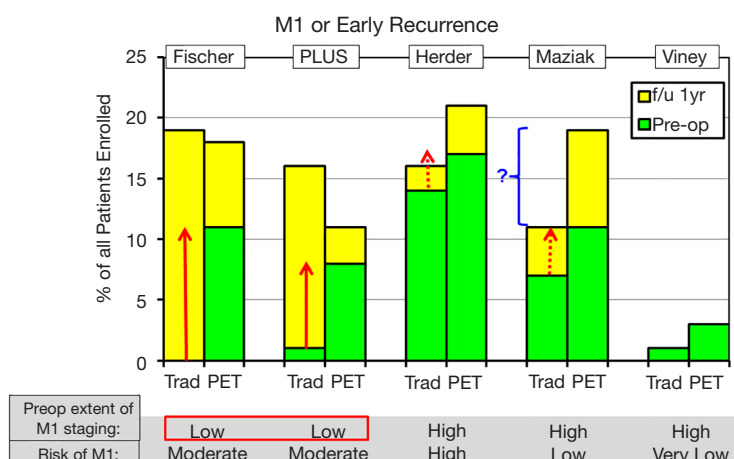


Figure 2 Rate of M1 disease or early recurrence. Rate of M1 disease diagnosed pre-operatively and the rate of early recurrence (within 1 year). Upwards solid red arrows indicate a benefit to PET, dashed red arrows indicate a trend towards a benefit. f/u, follow-up; preop, preoperatively; Trad, traditional; yr, year.

preoperative identification of M1 disease would obviate the need to identify N2,3 involvement. Furthermore, the incidence of M1 disease is higher than expected from the patients' characteristics. The reason for these findings is unclear.

Thus the RCTs suggest that PET helps identify M1 disease preoperatively when the risk is at least moderate and the thoroughness of searching for M1 disease without PET is low. However, PET has little additional impact if the risk of advanced disease is low or when extensive investigation for M1 disease is already being done.

Avoidance of stage-inappropriate resection

Table S4 summarizes the rate of pre- vs. intra-/post-operative identification of N2,3/M1 involvement. Figure 3 shows the rate of stage-inappropriate resection (defined as surgery for something other than stage I, II NSCLC). PET was beneficial by this assessment in the Fischer and PLUS studies and to a lesser extent in the Maziak trial (2,4,5). In the Fischer and PLUS studies PET lowered the overall rates of surgery, but the other studies found no difference. The discrepancy in the Maziak study between a similar overall rate of surgery, yet fewer stage-inappropriate resections with PET, appears to be due to unequal overall rates of N2,3 and M1 involvement between the arms (which should be similar in a randomized study).

The potential of PET to reduce stage-inappropriate surgery is seen in patients with a high risk of advanced cancer and with relatively little investigation for this without PET. PET has little impact when the risk of M1 disease is

low or the rate of traditional investigation is high.

Missed stage-appropriate resection

A balanced assessment requires that the rate of missed stage-appropriate resection is assessed. Unfortunately, none of the RCTs addressed this directly. The potential for missed resection can be estimated from the rate of PET findings suggesting N2,3/M1 disease subsequently found to be false positive. This assessment suggests this risk is not minor, although the rate varies markedly (from 1% to 42%) (Table S4).

Discussion

The literature on PET for lung cancer staging has progressed from dramatic anecdotal PET images of metastases (not necessarily otherwise undetected) to series comparing PET to historical studies and finally RCTs. Clinical guidelines recommended PET in the evaluation of most lung cancer patients (9,12-14).

However, the RCT results are not consistent with respect to the identification of metastasis, the rate of so-called "futile" thoracotomies and the need to perform invasive mediastinal staging. However, the RCT have differed in terms of the patients included, the endpoints and the context of the studies. We hoped to clarify how details of the design and conduct of the RCTs affected the results. The realist method is specifically designed to explore differences in context to develop a deeper understanding of what works, for whom, in which setting, why and how (6-8).

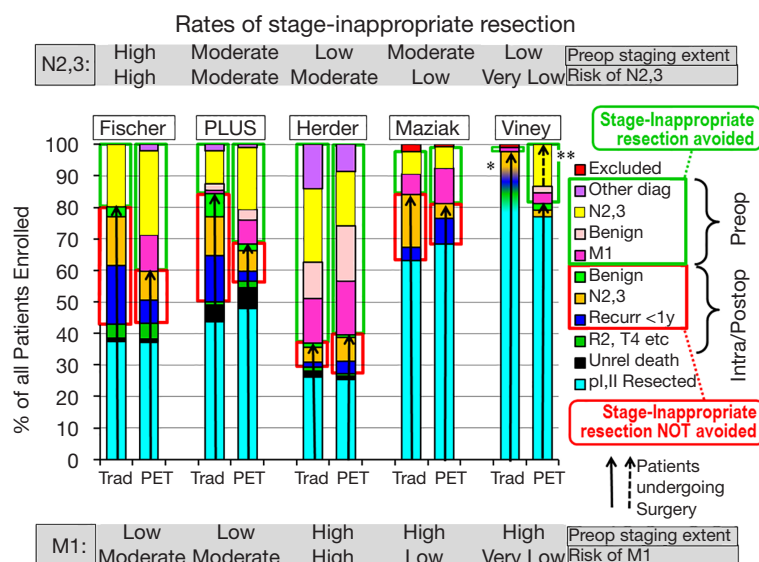


Figure 3 Rate of inappropriate resection. Rate of inappropriate resection avoided or not avoided for all evaluable patients. Black arrows (solid and dashed) indicate the percent of patients operated. *, this study reported several categories of patients together; **, in this study 12% underwent stage-inappropriate resection (for N2,3 positive disease), but if mediastinoscopy had been done for suspicious nodes (assuming 80% sensitivity) then 3% would have undergone stage-inappropriate resection. The dashed black arrow represents these patients, who were operated in this study but would not have been if a more standard policy of investigation of suspicious N2,3 nodes and avoidance of surgery in N2,3 positive patients had been followed. Diag, diagnoses; intraop, intraoperatively; preop, preoperatively; postop, postoperatively (i.e., within 1 year); Recurr, recurrence; Trad, traditional; Unrel, unrelated; yr, year.

Analyzing the reported data relative to consistent specific endpoints allows the RCTs to be compared (Table 4). This analysis suggests several conclusions. First, the benefit of PET to identify benign disease is moderate after evaluation by a generalist and limited investigation but low after a specialist's evaluation and more extensive pre-enrollment testing.

Second, the benefit of PET in detecting N2,3 or M1 disease is low if the clinical evaluation and chest CT suggest a low risk of metastasis. PET is of little value for identification of N2,3 disease if the rate of invasive staging is low—and it may be detrimental if it is used to lower the rate of invasive mediastinal staging. However, in cohorts with at least moderate a risk of N2,3 involvement, PET appears to be of benefit even if invasive staging is done in most patients (perhaps by directing attention to suspicious nodes).

Finally, PET is of value in identifying M1 disease if the risk is at least moderate and little traditional imaging is done. If the patient undergoes extensive traditional imaging, there is little additional impact of PET in identifying M1 disease.

Taking everything together, PET is useful if there is at least a moderate risk of metastases (N2,3 or M1), the extent

of traditional imaging is low, and the rate of invasive staging is high (Figure 4). In these settings PET appears to reduce the rate of stage-inappropriate resection. However, with initial evaluation by a specialist and extensive traditional imaging the impact of adding PET is low. PET does not appear to obviate the need for invasive nodal staging. Furthermore, the rate of potentially misleading false-positive PET results is substantial, suggesting a potential detriment if confirmation of positive PET findings (i.e., mediastinal or distant metastases) is not undertaken diligently.

We avoided the endpoint “futile thoracotomy” because we perceive this to be biased and not conducive to an objective scientific assessment. “Futile” has a strong negative connotation, and is one-sided, ignoring the converse risk of a missed stage-appropriate, potentially curative resection (Considering only the stage-inappropriate resection mandates that all surgery be avoided because then no stage-inappropriate resection will occur). Furthermore, counting unrelated (random, unpredictable) deaths as “futile” resections seems inappropriate as an outcome to assess pre-treatment evaluation. Finally, the majority of lung cancer resections are performed by thoracoscopy, not thoracotomy in cutting-edge thoracic surgical units.

Table 4 Key conclusions			
Outcome	Benefit of PET	Little impact of PET	Detriment of PET
Preoperative identification of benign disease	Suspicion by generalist with limited preoperative evaluation	Clinical diagnosis by specialist or biopsy proven	
Preoperative identification of N2,3 disease	High risk of N2,3 disease and high rate of invasive staging	Low risk of N2,3 disease Limited invasive mediastinal staging	Using PET to lower the rate of invasive biopsy of N2,3 nodes
Preoperative identification of M1 disease	Moderate to high risk of M1 and limited preoperative investigation	Low risk of M1 disease Extensive traditional imaging for M1	
Avoidance of stage-inappropriate surgery	High risk of N2,3 or M1 and limited preoperative investigation	Low risk of N2,3 or M1 disease Extensive traditional investigation	Possible if there is limited confirmation of PET results ^a

^a, detriment due to a risk of missed stage-appropriate surgery.

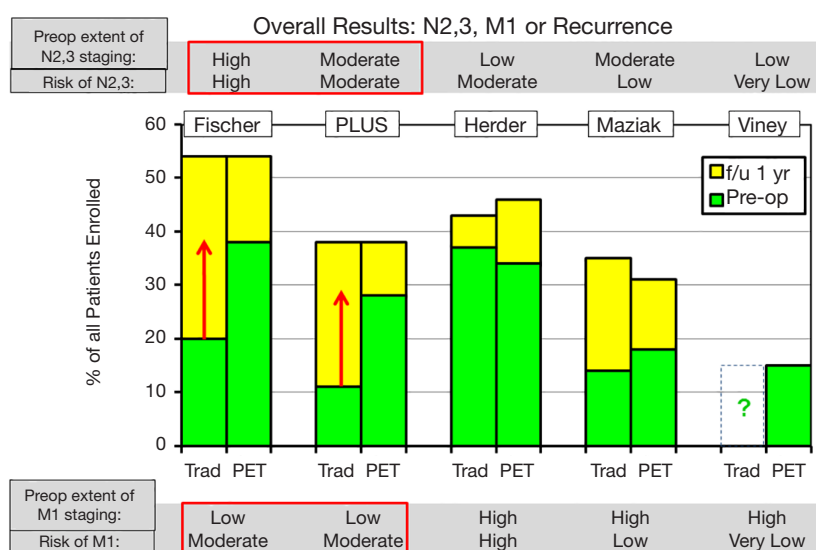


Figure 4 Overall results—rate of N2,3/M1 disease or recurrence. Overall results—rate of either N2,3 or M1 disease diagnosed preoperatively *vs.* the rate of N2,3 involvement diagnosed intra-operatively or early recurrence (within 1 year). Upwards solid red arrows indicate a benefit to PET. f/u, follow-up; pre-op, preoperatively; Trad, traditional; yr, year.

The quality of the PET imaging and interpretation in the RCTs appears to have been good. We did not find that differences in technical details had an impact (Table 2). The use of PET/CT *vs.* stand-alone PET does not alter the findings, suggesting that the setting in which PET is implemented may have a greater impact than the technology itself. Furthermore, PET/CT is not universally available: in the US about half of PET imaging involves stand-alone PET. In the RCTs PET imaging was relatively centralized, whereas in the US PET is performed at many smaller institutions and using mobile scanners. What effect this has on the accuracy of interpretation of the PET scans is unknown.

A difficulty in this analysis are discrepancies in the Maziak study between the randomized arms in the overall rate of N2,3 and M1 disease. There is no discernible explanation for this. These discrepancies drive some of the face-value conclusions of the study, namely that PET identifies more patients with metastases; without this imbalance it is unclear whether this result would hold up.

Outcomes studies suggest PET has an impact in lung cancer, primarily by identifying distant metastases in cIII patients (15–17). Other studies suggest little impact in stage cI patients (15,18). PET detected N2,3 or M1 involvement in 7% of the subset of cI patients in the ACoSOG study, but at a price of falsely suggesting N2,3/M1 disease in 14% (18).

This study also found that while PET could reduce the rate of biopsy for benign lesions from 21% to 11%, this would cause missed (or delayed) resection in 13% of cancers (18). Finally, many studies have consistently reported that ~25% of patients with central or cII or cIII tumors by CT harbor N2,3 involvement despite a PET that is negative in the mediastinum (10,19-24). Thus, other (non-RCT) studies corroborate a significant benefit with PET in some clinical settings but also a potential detriment when the false negative and false positive rates of PET in particular clinical settings are not recognized and the PET results are not appropriately confirmed.

Conclusions

This analysis suggests that while PET can be useful, it depends on many factors. PET appears to be of benefit when the chance of N2,3 or M1 involvement is moderate or high, when the extent of traditional imaging (abdominal/pelvis CT, bone scan) is low, and when the rate of invasive mediastinal staging is high. The impact of PET is low in other settings. If PET is used to avoid invasive mediastinal staging in clinical settings in which the risk of N2,3 involvement is moderately high, PET can lead to lower preoperative and higher intraoperative detection rates of N2 disease. Finally, the data suggests a significant risk of missed curative-intent treatment if positive PET findings are not interpreted carefully. Accurate evaluation is a complex interplay of various clinical aspects (symptoms), risk of metastases, extent of imaging, confirmation of suspicious findings, the thoroughness of intra-operative assessment and of follow-up.

The overriding conclusion of this analysis of RCTs comparing PET to traditional evaluation of lung cancer patients is that the results are dependent on the clinical setting. A blanket recommendation for PET may be too simplistic—various clinical aspects affect the value of PET. A more judicious use of PET may lower costs without negatively impacting the accuracy of evaluation of patients with lung cancer.

Acknowledgements

Disclosure: The authors declare no conflict of interest.

References

1. Herder GJ, Kramer H, Hoekstra OS, et al. Traditional versus up-front [18F] fluorodeoxyglucose-positron emission tomography staging of non-small-cell lung cancer: a Dutch cooperative randomized study. *J Clin Oncol* 2006;24:1800-6.
2. Maziak DE, Darling GE, Inculet RI, et al. Positron Emission Tomography in Staging Early Lung Cancer. *Ann Intern Med* 2009;151:221-8.
3. Viney RC, Boyer MJ, King MT, et al. Randomized controlled trial of the role of positron emission tomography in the management of stage I and II non-small-cell lung cancer. *J Clin Oncol* 2004;22:2357-62.
4. van Tinteren H, Hoekstra OS, Smit EF, et al. Effectiveness of positron emission tomography in the preoperative assessment of patients with suspected non-small-cell lung cancer: the PLUS multicentre randomised trial. *Lancet* 2002;359:1388-93.
5. Fischer B, Lassen U, Mortensen J, et al. Preoperative Staging of Lung Cancer with Combined PET-CT. *N Engl J Med* 2009;361:32-9.
6. Wong G, Greenhalgh T, Westhorp G, et al. Realist methods in medical education research: what are they and what can they contribute? *Medical Education* 2012;46:89-96.
7. Wong G, Greenhalgh T, Westhorp G, et al. RAMESES publication standards: realist syntheses. *BMC Med* 2013;11:21.
8. Pawson R, Greenhalgh T, Gill H, et al. Realist review—a new method of systematic review designed for complex policy interventions. *J Health Serv Res Policy* 2005;10:21-34.
9. Silvestri GA, AV Gonzalez, Jantz M, et al. Methods for staging non-small cell lung cancer: Diagnosis and management of lung cancer, 3rd ed: American College of Chest Physicians evidence-based clinical practice guidelines. *Chest* 2013;143:e211S-50S.
10. Darling GE, Maziak D, Inculet R, et al. Positron Emission Tomography-Computed Tomography Compared with Invasive Mediastinal Staging in Non-small Cell Lung Cancer—Results of mediastinal staging in the early lung positron emission tomography trial. *J Thorac Oncol* 2011;6:1367-72.
11. Detterbeck F, Puchalski J, Rubinowitz A, et al. Classification of the thoroughness of mediastinal staging of lung cancer. *Chest* 2010;137:436-42.
12. Silvestri GA, Gould MK, Margolis ML, et al. Non-invasive staging of non-small cell lung cancer: ACCP evidenced-based clinical practice guidelines (2nd Edition). *Chest* 2007;132:178S-201S.

13. Crinò L, Weder W, van Meerbeeck J, et al. Early stage and locally advanced (non-metastatic) non-small-cell lung cancer: ESMO Clinical Practice Guidelines for diagnosis, treatment and follow-up. *Ann Oncol* 2010;21:v103-v115.
14. Ettinger DS, Akerley W, Borghaei H, et al. Non-small cell lung cancer. *J Natl Compr Canc Netw* 2012;10:1236-71.
15. Morgensztern D, Goodgame B, Baggstrom M, et al. The effect of FDG-PET on the stage distribution of non-small cell lung cancer. *J Thorac Oncol* 2008;3:135-9.
16. Morgensztern D, Waqar S, Subramanian J, et al. Improving survival for stage IV non-small cell lung cancer: a surveillance, epidemiology, and end results survey from 1990 to 2005. *J Thorac Oncol* 2009;4:1524-9.
17. Morgensztern D, Ng SH, Gao F, et al. Trends in stage distribution for patients with non-small cell lung cancer: a National Cancer Database survey. *J Thorac Oncol* 2010;5:29-33.
18. Kozower BD, Meyers BF, Reed CE, et al. Does Positron Emission Tomography Prevent Nontherapeutic Pulmonary Resections for Clinical Stage IA Lung Cancer? *Ann Thorac Surg* 2008;85:1166-9.
19. Gould MK, Kuschner WG, Rydzak CE, et al. Test performance of positron emission tomography and computed tomography for mediastinal staging in patients with non-small-cell lung cancer: a meta-analysis. *Ann Intern Med* 2003;139:879-92.
20. Pozo-Rodríguez F, Martín de Nicolás JL, Sánchez-Nistal MA, et al. Accuracy of Helical Computed Tomography and [18F] Fluorodeoxyglucose Positron Emission Tomography for Identifying Lymph Node Mediastinal Metastases in Potentially Resectable Non-Small-Cell Lung Cancer. *J Clin Oncol* 2005;23:8348-56.
21. Serra M, Cirera L, Rami-Porta R, et al. Routine positron tomography (PET) and selective mediastinoscopy is as good as routine mediastinoscopy to rule out N2 disease in non-small cell lung cancer (NSCLC). *J Clin Oncol* 2006;2006:24.
22. Verhagen AFT, Bootsma GP, Tjan-Heijnen VCG, et al. FDG-PET in staging lung cancer: How does it change the algorithm? *Lung Cancer* 2004;44:175-81.
23. Cerfolio RJ, Bryant AS, Ojha B, et al. Improving the inaccuracies of clinical staging of patients with NSCLC: a prospective trial. *Ann Thorac Surg* 2005;80:1207-13; discussion 1213-4.
24. Dietlein M, Weber K, Gandjour A, et al. Cost-effectiveness of FDG-PET for the management of potentially operable non-small cell lung cancer; priority for a PET-based strategy after nodal-negative CT results. *Eur J Nucl Med* 2000;27:1598-609.

Cite this article as: Detterbeck FC, Figueroa Almanzar S. Lung cancer staging: the value of PET depends on the clinical setting. *J Thorac Dis* 2014;6(12):1714-1723. doi: 10.3978/j.issn.2072-1439.2014.11.16

Title			
√	1	Mention realist review or synthesis	
Abstract			
√	2	Brief description of background, objectives, search, methods of selection, appraisal, analysis and synthesis or sources, main results and implications for practice	
Introduction			
√	3	Rationale for Review	Why is this needed, likely to be useful
√	4	Objectives and focus	State objectives or question, justify focus
Methods			
NA	5	Changes in Review Process	Describe and justify changes that were made
√	6	Why realist synthesis?	Why is realist synthesis appropriate for the study?
√	7	Scoping the literature	Describe, justify initial process of scoping literature
√	8	Searching Processes	Provide rationale for iterative searching, sources accessed, dates
√	9	Selection, Appraisal	Describe inclusion, exclusion judgments
√	10	Data extraction	What was extracted and why?
√	11	Analysis and Synthesis	What constructs were analyzed, how?
Results			
*	12	Flow diagram	Numbers of papers, reasons for exclusion etc.
√	13	Characteristics	Describe characteristics of papers selected
√	14	Main Findings	Key findings with specific focus on theory building and testing
Discussion			
√	15	Summary of findings	Summarize relative to focus, objective
√	16	Strengths/limitations	Include S/L of process followed, need for future studies
√	17	Compare to literature	Compare to existing literature, other reviews
√	18	Conclusion/recommendations	List implications and recommendations for policy or practice
√	19	Funding	Describe any funding sources, role of organizations
*, Available on request; NA, not applicable.			

Figure S1 RAMSES checklist.

Table S1 Setting of study

Study	N	Region	Accrual years	Primary question	Primary outcome measure	PET technique	No. of PET scanners	No. of centers participating
Fisher	189	Denmark	2002-2007	Preop eval accuracy	“Futile” thoracotomy ^a	PET/CT	1	3
Plus	188	Netherlands	1998-1999	Preop eval accuracy	“Futile” thoracotomy ^a	PET + CT	1	9
Herder	465	Netherlands	1999-2001	Role of upfront PET	Number of tests needed	PET ^b	2	22
Maziak	337	Canada	2004-2007	Preop eval accuracy	Detection of IIIb, IV	PET/CT	5	8
Viney	183	Australia	1999-2000	Preop eval accuracy for stage I,II	Detection of IIIb, IV ^c	PET ^d	1	6

CT, computed tomography; PET, positron emission tomography; PET/CT, integrated PET/CT scanner; PET + CT, PET scan with visual correlation with a separate CT; Preop eval, preoperative evaluation. ^a, futile thoracotomy defined as thoracotomy despite unrecognized N2,3 or M1 involvement, thoracotomy for benign disease, recurrence within 1 year, or any death within 1 year regardless of the cause; ^b, PET read without CT scan available for comparison; ^c, the explicit endpoint was the rate of thoracotomy, but effectively the endpoint is identification of distant metastases because patients with suspected mediastinal involvement underwent thoracotomy nevertheless; ^d, CT scan available but was not used during PET interpretation; PET reader knew that the mediastinum was deemed normal on all CT scans.

Table S2 Study endpoints-N2,3 involvement

Study	% N2,3 anytime		% N2-3 preop		% N2-3 intraop		% other ^a intraop		Impact of PET on N2,3	Prevalence of cN2,3 disease	Extent of preop N2,3 staging
	Trad	PET	Trad	PET	Trad	PET	Trad	PET			
Fischer	35	36	20	27	15	9	4	5	↑↑	High	High
Plus	22	27	10	20	12	7	1	2	↑↑	Moderate	Moderate
Herder ^b	27	25	23	17	4	8	1	1	↓↓	Moderate	Low ^b
Maziak	24	12	7	7	17	5	–	–	(↑) ^c	Low	Moderate
Viney	–	14	–	12 ^d	–	2 ^d	–	–	–	Minimal	Low

Note: all results calculated on intent to treat basis for entire cohort (arm) in each study. Intraop, intraoperatively identified; PET, positron emission tomography; preop, preoperatively identified; Trad, traditional evaluation arm. ^a, e.g., unresectability, T4, SCLC; metastatic extrathoracic cancer; ^b, mediastinoscopy done significantly less often in PET arm (13% vs. 34%); ^c, a higher rate of intra-operative identification of N2,3 disease in the conventional w/u arm appears to be due to an discrepancy in the overall prevalence of N2,3 disease despite randomization; ^d, in the study 2% had preoperatively identified nodes and 12% intraoperatively identified nodes, because the local policy was to ignore suspicious N2,3 nodes and proceed with resection; if mediastinoscopy had been done for suspicious nodes (assuming 80% sensitivity) then 12% would have been identified preoperatively and 2% would have been identified intraoperatively.

Table S3 Study endpoints-distant disease or early recurrence

Study	% M1 anytime		Death (not cancer) within 1 yr		% M1 diagnosed preop		% M1/recur diagnosed within 1 yr		Impact of PET on M1	Risk of advanced stage ^a	Extent of preop M1 staging
	Trad	PET	Trad	PET	Trad	PET	Trad	PET			
Fischer	19	18	1	1	0	11	19	7	↑↑↑↑	+++	Low
Plus	16	12	5	6	1	8	15	3	↑↑↑	+++	Low
Herder	16	21	4	2	14	17	2	4	↑	++++	High
Maziak	11	20	–	–	7	11	4	8	(=) ^b	++	High
Viney	–	–	–	–	1	3	–	–	(=) ^c	+	High

Note: all results calculated on intent to treat basis for entire cohort (arm) in each study. PET, positron emission tomography; preop, preoperatively identified; recur, recurrence; Trad, traditional evaluation arm; yr, year. ^a, risk based on presence of clinical markers of advanced disease (>5% weight loss, performance status ≥2); ^b, differences in pre- vs. intra/post-operative identification of M1 disease or recurrence are conflicting and appear to be due to a discrepancy in the overall prevalence of M1 disease/recurrence despite randomization; ^c, impact is not fully able to be assessed because data on early recurrence was not reported.

Table S4 Study endpoints-appropriateness of resection

Study	% undergoing surgery		% stage-inappropriate resection ^a		At risk for missed surgery ^b		Impact of PET on stage-appropriate resection	Impact of PET on N2,3	Impact of PET on M1	Any N2,3/M1 preop		Any N2,3/M1 intraop or in 1 yr	
	Trad	PET	Trad	PET	Trad	PET				Trad	PET	Trad	PET
Fischer	80	61	42	22	–	–	↑↑↑	↑↑	↑↑↑↑	20	38	34	16
Plus	81	65	33	15	–	42	↑↑↑	↑↑↑	↑↑↑	11	28	27	10
Herder	38	41	6	10	1	1	↓	↓↓	↑	37	34	6	12
Maziak	78	81	27	16	1	5	(↑↑) ^c	(↑) ^c	(=) ^c	14	19	21	13
Viney	98	96	–	3 ^d	–	10	–	–	=	1	15	–	2

Note: all results calculated on intent to treat basis for entire cohort (arm) in each study. It is assumed that preoperatively identified N2,3 or M1 disease is inappropriate to resect (although there are exceptions to this treatment policy). Intraop, intraoperatively identified; PET, positron emission tomography; preop, preoperatively identified; Trad, traditional evaluation arm. ^a, resection but N2,3, M1 disease or recurrence in <1 year; ^b, imaging that falsely suggested benign disease, N2,3 or M1 disease (which could have led to a stage inappropriate lack of resection if further work-up had not been pursued); ^c, due to discrepancies in the overall number of N2,3 patients per arm and M1 patients per arm despite randomization; ^d, in the study 12% underwent stage-inappropriate resection (for N2,3 positive disease), but if mediastinoscopy had been done for suspicious nodes (assuming 80% sensitivity) then 3% would have undergone stage-inappropriate resection.