Published in final edited form as:
*J Nucl Med*. 2013 September ; 54(9): 1518–1527. doi:10.2967/jnumed.112.119362.

# The Lack of Evidence for PET or PET-CT Surveillance of Patients with Treated Lymphoma, Colorectal Cancer, and Head and Neck Cancer: A Systematic Review

**Kamal Patel, MPH, MBA**[1], **Nira Hadar, MS**[1], **Jounghee Lee, PhD**[1], **Barry A. Siegel, MD**[2], **Bruce E. Hillner, MD**[3], and **Joseph Lau, MD**[1]

[1]Institute of Clinical Research and Health Policy Studies, Tufts Medical Center, Boston, MA

[2]Division of Nuclear Medicine, Mallinckrodt Institute of Radiology and the Siteman Cancer Center, Washington University School of Medicine, St. Louis, MO

[3]Department of Internal Medicine and the Massey Cancer Center, Virginia Commonwealth University, Richmond, VA

## Abstract

**Rationale—**Positron emission tomography (PET) and positron emission tomography–computed tomography (PET-CT) are widely used for surveillance purposes in patients following cancer treatments. We conducted a systematic review to assess the diagnostic accuracy and clinical impact of PET and PET-CT used for surveillance in several cancer types.

**Methods—**We searched MEDLINE and Cochrane Library databases from 1996 to March 2012 for relevant English-language studies of PET or PET-CT used for surveillance in patients with lymphoma, colorectal cancer, or head and neck cancer. We included prospective or retrospective studies that reported test accuracy and comparative studies that assessed clinical impact.

**Results—**Twelve studies met our inclusion criteria: six in lymphoma (n=767 patients), two in colorectal cancer (n=96), and four in head and neck cancer (n=194). All studies lacked a uniform definition of surveillance and scan protocols. One-half were retrospective studies and one-third were rated as low quality. The majority reported sensitivities and specificities in the range of 90% to 100%, although several studies reported lower results. The only randomized controlled trial, a colorectal cancer study with 65 patients in the surveillance arm, reported earlier detection of recurrences with PET and suggested improved clinical outcomes.

**Conclusion—**There is insufficient evidence to draw conclusions on the clinical impact of PET or PET-CT surveillance for these cancers. The lack of standard definitions for surveillance, heterogeneous scanning protocols, and inconsistencies in reporting test accuracy precludes making an informed judgment of the value of PET for this potential indication.

### Keywords

surveillance; PET; PET-CT; lymphoma; colorectal

Corresponding Author: Joseph Lau, MD, jlau1@tuftsmedicalcenter.org, ph 617-6364750 fax 617-636-8628, Tufts Medical Center, 800 Washington Street, Box 63, Boston, MA 02111.

**Conflict of Interest**: BAS: Advisory Board and stockholder, Radiology Corporation of America; Advisory Board, Siemens Molecular Imaging; Advisory Board, GE Healthcare.

## Background

Positron emission tomography (PET) using the glucose analog 2-[$^{18}$F]fluoro-2-deoxy-D-glucose (FDG) has become an important modality for cancer imaging because of the characteristically increased utilization of glucose by malignant cells. Since its introduction in 2000, PET integrated with computed tomography (PET-CT) has progressively replaced conventional PET and nearly all scanners now used worldwide are PET-CT scanners.(1) Compared with conventional PET, PET-CT provides greater accuracy in localizing FDG uptake, with resultant improvement in observer performance.(2), (3) Hereafter, PET will refer to PET or PET-CT; distinctions will be made where indicated.

PET is used for a wide variety of cancer types and clinical purposes, including diagnosis, initial staging, assessment of treatment response(4), (5), restaging, detection of clinically suspected recurrence and surveillance.(6), (7), (8), (9) Using advanced imaging, including PET, for post-treatment surveillance of patients after treatment is controversial and generally is not recommended for most cancer types.(10), (11) It is a widely held yet anecdotal impression that surveillance PET imaging is common, yet there are few published estimates of utilization rates for this indication.(12) The National Oncologic PET Registry does not specifically gather data on the use of PET for surveillance purposes.(13) While systematic reviews have been conducted for a range of PET uses, none have focused on the use of PET specifically for surveillance.(14), (15)

A common conceptual framework for evaluating diagnostic test technologies categorizes studies into six levels of assessments.(16) In this systematic review, we searched for evidence to assess the diagnostic accuracy and clinical impact of surveillance PET (i.e., impact of scans on use of other diagnostic tests, impact on therapeutic decisions, and effect on patient outcomes). We focused *a priori* on lymphoma, colorectal cancer, and head and neck cancer, as these have the most studies and, in our experience, have the largest number of patients undergoing post-treatment surveillance. We also gathered data from studies that did not meet the inclusion criteria to inform future research recommendations.

## Methods

In carrying out this systematic review, we adhered to the PRISMA statement for reporting systematic reviews and meta-analyses.(17)

### Literature Search Strategy

We searched the MEDLINE and Cochrane Central Trials Registry databases from 1996 to March 2012 for English-language studies examining the use of PET in lymphoma, colorectal cancer, and head and neck cancer. In addition, we searched the Cochrane Database of Systematic Reviews to identify relevant reviews and manually reviewed the reference lists of studies that met our inclusion criteria. A variety of keywords and MESH terms were used, including terms used to describe PET devices and terms related to surveillance (e.g., “monitoring” and “follow-up”).

### Study Selection

The abstracts were reviewed for eligibility by one of four authors (KP, JLau, JLee, and NH) with questionable studies being adjudicated by all the authors. Surveillance imaging was defined as imaging performed at least six months after completion of treatment with curative intent among patients who were considered to be disease free by clinical examination or other imaging prior at the time of PET. We included reports evaluating patients with lymphoma, colorectal cancer, or head and neck cancer at any cancer stage before treatment.

Studies were excluded if results were not separately reported for patients considered to be disease free or if patients were suspected by any clinical signs or symptoms of having recurrent disease. Scans could be performed on a one-time basis or periodic schedule. Acceptable reference standards for recurrence included histology, other imaging modalities, laboratory tests, clinical examination, or some combination as defined by the study authors.

For studies of test accuracy, we included prospective or retrospective studies. We accepted studies that (1) used either individual patients or individual scans as the unit of analysis and (2) either reported test accuracy (e.g., sensitivity, specificity, positive predictive value (PPV), negative predictive value (NPV), positive likelihood ratio (LR+), and negative likelihood ratio (LR-)) or presented data in 2×2 tables allowing calculating accuracy. For studies assessing clinical impact, we considered only comparative studies.

## Data Extraction and Calculation of Test Accuracy

Data from each study were extracted by one of us (KP, NH) and confirmed by another. Discrepancies were reconciled by three of us (JLau, KP, and NH). Information was collected on cancer type, patient characteristics, details of the surveillance protocol, the reference standard used, and relevant measures for diagnostic accuracy and clinical impact outcomes. While some studies performed surveillance scans at more than one timepoint, test accuracy metrics were typically not reported for all time points, and surveillance protocols often were unclear as to which patients were included in later scans. Thus for each study, we extracted data for the first timepoint at which surveillance scans occurred, at a minimum of six months post treatment completion. Where possible, we also computed the "yield" of screening, defined as the percentage of positive studies (true positive plus false positive) in the scanned population. When they were not provided by the study, test accuracy measures (sensitivity, specificity, positive and negative predictive values, and likelihood ratios) as well as confidence intervals were calculated using STATA version 11.0.

## Study Quality Assessment

We extracted information on the design, conduct, and reporting and used the Quality Assessment of Diagnostic Accuracy Studies (QUADAS) tool to evaluate the quality of the studies assessing test accuracy.(18) For comparative studies reporting on clinical impact outcomes, we combined QUADAS together with selected items from the Cochrane Risk of Bias tool that were applicable to diagnostic testing studies.(19) The primary data extractor assessed the study quality and another reviewer confirmed the quality grade.

We rated each study using an "A", "B", or "C" letter grade according to predefined criteria. Quality A studies adhered to recognized standards of conduct for diagnostic test studies and provided clear descriptions of the design, population, test, reference standard, and outcomes. These studies also had no major reporting omissions or errors and no obvious source of bias. Quality B studies had some deficiencies in these criteria, but these deficiencies were considered unlikely to result in a major bias (retrospective studies start with a lower grade of "B"). Quality C studies had serious design and/or reporting deficiencies.

Results are summarized by cancer type, and separately for PET and PET-CT. While we reported information on quality "C" studies, we drew test accuracy conclusions only from quality "A" and "B" studies.

## Role of Funding Source

This study was funded by the National Cancer Institute. The funder had no role in informing selection of the topic or in protocol development, and did not review the manuscript.

## Results

The literature search yielded 1,813 citations, from which 146 full-text articles were evaluated (Figure 1). Twelve studies (seven PET, five PET-CT) met our inclusion criteria and provided test accuracy data. One study, a randomized controlled trial, also provided data on impact on therapeutic decision-making and clinical outcomes.(20) Studies were most often excluded for failing to meet our definition of surveillance; most commonly for scanning less than six months after completion of treatment or scanning performed for assessing treatment response or restaging. Additional studies were excluded for a variety of other reasons, such as including patients suspected of having recurrent disease, lacking test accuracy results, and including cancer types outside the scope of this review.

Table 1 shows the characteristics of the included studies: six lymphoma studies (two PET, four PET-CT), two PET studies in colorectal cancer, and four in head and neck cancer (three PET, one PET-CT). All five PET-CT studies utilized CT for attenuation correction and localization of PET findings. One study used contrast-enhanced CT for diagnostic purposes. (21)

There was no standard definition of surveillance across all studies or within cancer types, nor was there a consistent time schedule used for repeated scans. The duration between the final surveillance PET and the last clinical follow-up examination ranged from 2.3 months to 31 months. In seven studies, patients were scanned serially, in four studies patients were scanned once, and in one the frequency could not be determined. The reference standard used to verify PET results varied between studies, and included CT alone, as well as a combination of laboratory and imaging findings and an absence of symptoms.

While patients were deemed to be disease-free after treatment completion in all studies, the specific means of confirmation of disease status was not provided in ten. Patients in two studies were deemed disease free by negative PET-CT done for restaging after treatment. (22), (23)

In colorectal cancer and head and neck cancer, all studies reported diagnostic accuracy using patients as the unit of analysis. Two lymphoma reports used scans as the unit of analysis. (24), (25) No clear information was provided as to how sensitivity and specificity were calculated in the cases where a patient had conflicting scan results at two different time points (i.e., if a patient had a negative scan followed by a positive scan).(26)

Table 2 lists for each study our overall quality ratings as well as specific grading criteria. There was one quality A study, eight quality B studies, and three quality C studies.

### Lymphoma

There were four retrospective PET-CT studies and two prospective PET studies. Four were rated as quality B,(22), (23), (24), (26) and two quality C.(21), (27) The two quality C studies are listed in the results tables but are not included in the synthesis of the body of literature because of their low quality. Sample sizes of B quality studies ranged from 27 to 421 patients with a total of 541 patients. Only one study analyzed children.(24)

Table 3 shows diagnostic accuracy by cancer type and imaging modality. The three quality B PET-CT studies included 120 patients, had a per patient level sensitivity of 100%, and specificities ranging from 43% to 92%.(22), (23), (24) One PET study with 421 patients was rated quality B, and reported a sensitivity of 89% and a specificity of 100%.(26) Among the four lymphoma studies with sufficient data to calculate predictive values, positive predictive

values ranged from 0.2 to 1.0 and negative predictive values ranged from 0.98 to 1.0; the yield of positive PET scans in these studies ranged from 9.6% to 63%.

### Colorectal Cancer

Two PET studies evaluated patients with colorectal cancer. One was a randomized controlled trial of 130 patients(20) and the other a retrospective study with 31 patients.[24,](28) The randomized trial compared a conventional surveillance strategy that included CT at 9 and 15 months after surgery (n=65) with a strategy that included both PET and CT scans (n=65) at the same timepoints. This trial assessed impact on therapeutic decision-making and mortality, as well as test accuracy. The retrospective study performed one PET scan at 2 years after treatment and was graded C quality because of likely selection bias.

The randomized trial ended recruitment early due to ethical and methodological concerns when PET-CT scanning became available in 2004 at their institution. For clinical impact outcomes, the study was rated quality A, with groups balanced in baseline characteristics and adequate reporting of data. Using a per-protocol analysis with 60 patients in the PET group (5 fewer than in the intention-to-treat analysis due to missing data) and 65 in the control group, the study found that recurrences were detected sooner after baseline in patients in the PET group (12.1±4.1 months) compared to patients in those in the control group (15.4±6 months; p=0.01). Therapy was started sooner, but not significantly so, in the PET group (14.8±4.1 versus 17.5 ±6.0 months, p=0.09). Surgery for recurrent disease was performed more frequently in those in the PET group (15 of 23 [65%] versus 2 of 21 [9.5 %], p<0.0001). Moreover, the frequency of curative resection of recurrences was higher in the PET group (43.8% versus 9.5%, p<0.01). Intention-to-treat analyses gave similar results. Using a per-protocol analysis, the study also found that a non-significantly greater number of patients with recurrences died during the study period (with a maximum follow-up of 24 months) in the control group than in the PET group (28.5% versus 13%, p=0.33). This study was rated quality B for assessment of test accuracy with sensitivity of 100% and specificity of 96%. Yield could not be calculated.

### Head and Neck Cancer

Patients with head and neck cancer were evaluated by PET-CT in one prospective study(25) and by PET in three (two prospective(29), (30), one retrospective(31)) PET studies. The PET-CT study was rated quality A and the PET studies were rated quality B. The PET studies had unclear reporting as well as possible selection bias.

The prospective PET-CT study enrolled 91 squamous cell carcinoma patients, and reported a sensitivity of 100% and a specificity of 85%. The three PET studies comprised 103 patients, and included two studies examining squamous cell carcinoma(30), (31) and one including all cell types.(29) Sensitivities ranged from 75% to 100% and specificities ranged from 92% to 95%. The four head and neck cancer studies had positive predictive values between 0.5 and 0.9, negative predictive values of 1.0, and a yield of positive PET scans ranging from 14% to 57%.

### Additional Analysis of Studies Not Included in the Review

Less than 10% of retrieved full-text articles met our inclusion criteria. Table 4 summarizes selected characteristics leading to exclusion. Less than a quarter of lymphoma and colorectal cancer and roughly half of head and neck cancer studies had prospective designs. Less than 15% of lymphoma and head and neck cancer studies included patients that were considered to be disease free at the time of imaging, and approximately a quarter of studies in these cancers described the scans as surveillance. In none of the colorectal cancer studies were

patients verified to be disease-free, and only in one of these were the scans described as surveillance.

Several studies met the majority of inclusion criteria but failed to either adequately report the surveillance protocol or clearly describe their patient population. For example, one study described scans as being for the purpose of surveillance, but these were performed at a median of 12 weeks after treatment completion (and thus would be more properly classified as restaging).(32) Another study performed scans at a median time post-treatment completion of 6.6 months but the range was 1.6 to 166 months and 28 of 35 scans were for suspected recurrence.(33)

## Discussion

This systematic review of PET used for post-treatment surveillance of patients with lymphoma, colorectal cancer, or head and neck cancer found only one comparative study examining its impact on patient management and few studies that assessed test accuracy. The sole randomized trial suggests that PET may have an important clinical impact in therapeutic decision making and improved patient outcomes when used for surveillance of colorectal cancer; one of the few cancers for which evidence exists supporting intensive post-treatment surveillance.(34) The majority of trials reported sensitivity and specificity in the range of 90% and 100%, but there were others which were much lower.

Due to the inconsistent definition of surveillance, variations in imaging protocols, and few studies employing a particular imaging modality in a given cancer type, we did not conduct a meta-analysis of this heterogeneous data. In addition, the literature was of inferior quality —seven out of 12 studies used a retrospective design and half lacked blinded outcome assessments. The retrospective studies had an inconsistent or an absence of a pre-defined scanning frequency and interval. Prospective studies used widely ranged scanning schedules – ranging from multiple scans every 6 months to performing one scan at roughly 2 years after treatment completion.

Our finding of the absence of evidence in support of PET-CT in post-treatment surveillance is reflected in practice guidelines.(10), (11) Current NCCN guidelines do not recommend surveillance. For head and neck cancer, PET is recommended for restaging in patients with higher stage disease (III and IV), but not thereafter. Similarly, PET is now the standard of care for end of treatment response assessment in Hodgkin lymphoma and aggressive Non-Hodgkin lymphoma, but not for surveillance. The Hodgkin lymphoma guideline explicitly states that surveillance PET should not be done because of the risk of false-positives, and PET is also not recommended in the Non-Hodgkin lymphoma guideline. Despite this, PET-CT is commonly used for the purpose of surveillance.(35) Possible risks of using PET-CT for surveillance include overtreatment based on false-positives and unnecessary radiation exposure.(36)

Our review highlights that the problems of surveillance are dominated by two failures. First, there is a lack of common definition or taxonomy of surveillance (the minimal time since last treatment and the absence of clinical or other diagnostic suspicion of recurrence). Second, there is no well-thought out, prospective protocol based on cancer type and stage at last treatment. Testing intervals should be tailored to the relative risk of recurrence that has been shown in each of these cancers to have its own declining pattern with time.

Few studies met the inclusion criteria of our review. However, some studies that were excluded from the review may have included patients who had surveillance scans. Because of the limited amount of poor quality evidence on surveillance scanning, we collected data from rejected studies to better understand characteristics of studies that didn't meet our

inclusion criteria but still may fall under the definition of surveillance scanning. We found that the large majority of these studies did not include patients who were disease-free at the time of the scan, and most did not clearly report the details of the scanning protocol. Our review has a number of limitations. Results were difficult to synthesize due to lack of a standard surveillance definition. Studies were generally of poor quality with more than half being of retrospective design. In studies conducting multiple scans as part of a surveillance protocol, we were unable to utilize all available results data because of a lack of consistent definitions of test accuracy and incomplete reporting. One head and neck cancer study included in this review examined the hypothetical therapeutic impact of PET surveillance, but this outcome was not included in our results because of the lack of a comparison group. (30) In addition, our review included two generations of PET technologies—PET alone and PET-CT. There is substantial evidence across many cancer types and indications that PET alone is more sensitive and specific than conventional imaging methods. Thus, even though PET-CT results are usually more accurate than those with PET alone, results from PET only studies set a baseline of performance, which are likely only to be improved upon by PET-CT. Finally, we did not assess publication bias. While this is always a concern in systematic review that unpublished negative studies may negate the positive results, the paucity of evidence in favor of using surveillance PET scans made this less of a concern. There is also a the lack of reliable methods to assess publication bias.(37)

Future research should provide detailed descriptions of the surveillance protocols and patient populations, which has been suggested in previous systematic reviews of colorectal cancer surveillance.(38), (34) More well-conducted studies will help to answer the questions of which patients would be helped most by surveillance (e.g., patients of different disease severities) as well as which surveillance protocols are most effective for different cancer types. Due to the small number of RCTs conducted in this area, it is even more important for prospective trials to be specific in describing protocols and patient populations in order to understand the efficacy of PET-CT surveillance strategies. Retrospective studies can inform the question of PET-CT surveillance test accuracy, but prospective studies are needed to address aspects of clinical impact, such as impact on therapeutic decision making, mortality and morbidity, and effect on usage of other imaging tests. As suggested by Baca,(38) adapting the parameters of surveillance protocols (such as frequency and duration of surveillance) to patient risk levels is an intriguing study design that would allow a more targeted approach to surveillance. Different cancer types with different natural histories may dictate variable surveillance durations, as the benefits and risks of follow-up vary over time. (38)

The results of this review point to the need to establish common definitions of surveillance and surveillance protocols. Broadly, surveillance can be defined as evaluation of an asymptomatic patient with no clinical evidence of disease to assess for otherwise occult disease. In addition to a need for improved reporting, there is a need for comparative studies of surveillance that are powered to look at clinically relevant outcomes. Future high-quality prospective studies, including randomized trials, are necessary to answer the question of what role surveillance scanning should play and for what duration in different cancer types.

## Acknowledgments

This work was supported in part by a National Cancer Institute Grand Opportunity Grant RC2CA148259.

## Reference List


1. Mujoomdar, M. MKNE. Positron Emission Tomography (PET) in Oncology: A Systematic Review of Clinical Effectiveness and Indications for Use. Ottawa: Canadian Agency for Drugs and Technologies in Health; Jan 4. 2010 Ref Type: Generic
2. von Schulthess GK, Steinert HC, Hany TF. Integrated PET/CT: current applications and future directions. Radiology. 2006; 238(2):405–422. [PubMed: 16436809]
3. Podoloff DA, Ball DW, Ben-Josef E, et al. NCCN task force: clinical utility of PET in a variety of tumor types. J Natl Compr Canc Netw. 2009; 7(Suppl 2):S1–26. [PubMed: 19555588]
4. Meyer RM, Ambinder RF, Stroobants S. Hodgkin's lymphoma: evolving concepts with implications for practice. Hematology Am Soc Hematol Educ Program. 2004:184–202. [PubMed: 15561683]
5. Meta J, Seltzer M, Schiepers C, et al. Impact of 18F-FDG PET on managing patients with colorectal cancer: the referring physician's perspective. J Nucl Med. 2001; 42(4):586–590. [PubMed: 11337546]
6. la FC, Hundt W, Brockel N, et al. Value of PET/CT versus PET and CT performed as separate investigations in patients with Hodgkin's disease and non-Hodgkin's lymphoma. Eur J Nucl Med Mol Imaging. 2006; 33(12):1417–1425. [PubMed: 16858568]
7. Freudenberg LS, Rosenbaum SJ, Beyer T, Bockisch A, Antoch G. PET versus PET/CT dual-modality imaging in evaluation of lung cancer. Radiol Clin North Am. 2007; 45(4):639–44. v. [PubMed: 17706528]
8. Cohade C, Osman M, Leal J, Wahl RL. Direct comparison of (18)F-FDG PET and PET/CT in patients with colorectal carcinoma. J Nucl Med. 2003; 44(11):1797–1803. [PubMed: 14602862]
9. Adams S, Baum RP, Stuckensen T, Bitter K, Hor G. Prospective comparison of 18F-FDG PET with conventional imaging modalities (CT, MRI, US) in lymph node staging of head and neck cancer. Eur J Nucl Med. 1998; 25(9):1255–1260. [PubMed: 9724374]
10. Special report: positron emission tomography for the indication of post-treatment surveillance of cancer. Technol Eval Cent Assess Program Exec Summ. 2010; 25(3):1–2.
11. Podoloff DA, Advani RH, Allred C, et al. NCCN task force report: positron emission tomography (PET)/computed tomography (CT) scanning in cancer. J Natl Compr Canc Netw. 2007; 5(Suppl 1):S1–22. [PubMed: 17509259]
12. Abel GA, Vanderplas A, Rodriguez MA, et al. High rates of surveillance imaging for treated diffuse large B-cell lymphoma: findings from a large national database. Leuk Lymphoma. 2012; 53(6):1113–1116. [PubMed: 22098406]
13. Hillner BE, Siegel BA, Shields AF, et al. Relationship between cancer type and impact of PET and PET/CT on intended management: findings of the national oncologic PET registry. J Nucl Med. 2008; 49(12):1928–1935. [PubMed: 18997054]
14. Gupta T, Master Z, Kannan S, et al. Diagnostic performance of post-treatment FDG PET or FDG PET/CT imaging in head and neck cancer: a systematic review and meta-analysis. Eur J Nucl Med Mol Imaging. 2011; 38(11):2083–2095. [PubMed: 21853309]
15. Isles MG, McConkey C, Mehanna HM. A systematic review and meta-analysis of the role of positron emission tomography in the follow up of head and neck squamous cell carcinoma following radiotherapy or chemoradiotherapy. Clin Otolaryngol. 2008; 33(3):210–222. [PubMed: 18559026]
16. Tatsioni A, Zarin DA, Aronson N, et al. Challenges in systematic reviews of diagnostic technologies. Ann Intern Med. 2005; 142(12 Pt 2):1048–1055. [PubMed: 15968029]
17. Liberati A, Altman DG, Tetzlaff J, et al. The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate health care interventions: explanation and elaboration. Ann Intern Med. 2009; 151(4):W65–W94. [PubMed: 19622512]
18. Whiting PF, Rutjes AW, Westwood ME, et al. QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. Ann Intern Med. 2011; 155(8):529–536. [PubMed: 22007046]
19. Higgins JP, Altman DG, Gotzsche PC, et al. The Cochrane Collaboration's tool for assessing risk of bias in randomised trials. BMJ. 2011; 343:d5928. [PubMed: 22008217]

20. Sobhani I, Tiret E, Lebtahi R, et al. Early detection of recurrence by 18FDG-PET in the follow-up of patients with colorectal cancer. British Journal of Cancer. 2008; 98(5):875–880. [PubMed: 18301402]
21. Lee AI, Zuckerman DS, Van den Abbeele AD, et al. Surveillance imaging of Hodgkin lymphoma patients in first remission: a clinical and economic analysis. Cancer. 2010; 116(16):3835–3842. [PubMed: 20564135]
22. Crocchiolo R, Fallanca F, Giovacchini G, et al. Role of 18FDG-PET/CT in detecting relapse during follow-up of patients with Hodgkin's lymphoma. Annals of Hematology. 2009; 88(12): 1229–1236. [PubMed: 19468730]
23. El-Galaly T, Prakash V, Christiansen I, et al. Efficacy of routine surveillance with positron emission tomography/computed tomography in aggressive non-Hodgkin lymphoma in complete remission: status in a single center. Leuk Lymphoma. 2011; 52(4):597–603. [PubMed: 21314488]
24. Rhodes MM, Delbeke D, Whitlock JA, et al. Utility of FDG-PET/CT in follow-up of children treated for Hodgkin and non-Hodgkin lymphoma. Journal of Pediatric Hematology/Oncology. 2006; 28(5):300–306. [PubMed: 16772881]
25. Abgral R, Querellou S, Potard G, et al. Does 18F-FDG PET/CT improve the detection of posttreatment recurrence of head and neck squamous cell carcinoma in patients negative for disease on clinical follow-up? Journal of Nuclear Medicine. 2009; 50(1):24–29. [PubMed: 19091901]
26. Zinzani PL, Stefoni V, Tani M, et al. Role of [18F]fluorodeoxyglucose positron emission tomography scan in the follow-up of lymphoma. Journal of Clinical Oncology. 2009; 27(11): 1781–1787. [PubMed: 19273712]
27. Hosein PJ, Pastorini VH, Paes FM, et al. Utility of positron emission tomography scans in mantle cell lymphoma. Am J Hematol. 2011; 86(10):841–845. [PubMed: 21922524]
28. Selvaggi F, Cuocolo A, Sciaudone G, Maurea S, Giuliani A, Mainolfi C. FGD-PET in the follow-up of recurrent colorectal cancer. Colorectal Disease. 2003; 5(5):496–500. [PubMed: 12925088]
29. Lowe VJ, Boyd JH, Dunphy FR, et al. Surveillance for recurrent head and neck cancer using positron emission tomography. Journal of Clinical Oncology. 2000; 18(3):651–658. [PubMed: 10653881]
30. Perie S, Hugentobler A, Susini B, et al. Impact of FDG-PET to detect recurrence of head and neck squamous cell carcinoma. Otolaryngol Head Neck Surg. 2007; 137(4):647–653. [PubMed: 17903585]
31. Salaun PY, Abgral R, Querellou S, et al. Does 18fluoro-fluorodeoxyglucose positron emission tomography improve recurrence detection in patients treated for head and neck squamous cell carcinoma with negative clinical follow-up? Head & Neck. 2007; 29(12):1115–1120. [PubMed: 17636537]
32. Rabalais AG, Walvekar R, Nuss D, et al. Positron emission tomography-computed tomography surveillance for the node-positive neck after chemoradiotherapy. Laryngoscope. 2009; 119(6): 1120–1124. [PubMed: 19358193]
33. Connell CA, Corry J, Milner AD, et al. Clinical impact of, and prognostic stratification by, F-18 FDG PET/CT in head and neck mucosal squamous cell carcinoma. Head Neck. 2007; 29(11):986–995. [PubMed: 17563906]
34. Jeffery M, Hickey BE, Hider PN. Follow-up strategies for patients treated for non-metastatic colorectal cancer. Cochrane Database Syst Rev. 2007; (1):CD002200. [PubMed: 17253476]
35. Wagner-Johnston ND, Bartlett NL. Role of routine imaging in lymphoma. J Natl Compr Canc Netw. 2011; 9(5):575–584. [PubMed: 21550970]
36. Huang B, Law MW, Khong PL. Whole-body PET/CT scanning: estimation of radiation dose and cancer risk. Radiology. 2009; 251(1):166–174. [PubMed: 19251940]
37. Lau J, Ioannidis JP, Terrin N, Schmid CH, Olkin I. The case of the misleading funnel plot. BMJ. 2006; 333(7568):597–600. [PubMed: 16974018]
38. Baca B, Beart RW Jr, Etzioni DA. Surveillance after colorectal cancer resection: a systematic review. Dis Colon Rectum. 2011; 54(8):1036–1048. [PubMed: 21730795]

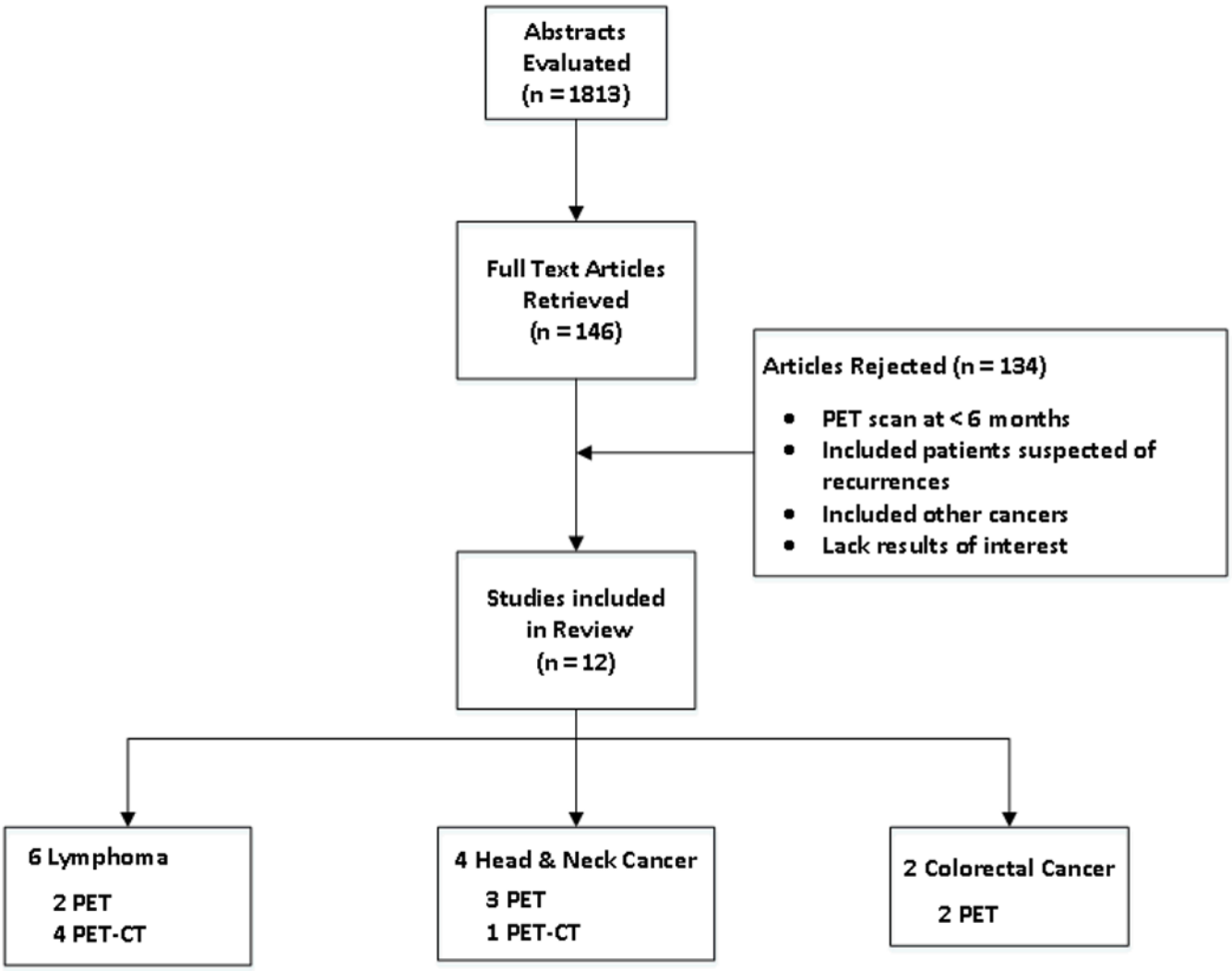


**Figure 1.**
Literature flow – This figure enumerates abstracts as well as retrieved and included studies for the review.

**Table 1**

**Characteristics of Studies Evaluating Surveillance PET or PET-CT**

| Author Year Country | Cancer | N | Age, y (Mean or Median, range) | Male (%) | Tumor stage | Study design (study duration*) | PET or PET-CT manufacturer, description | Time from treatment end to first PET-CT scan | PET or PET-CT frequency | Reference standard |
|---|---|---|---|---|---|---|---|---|---|---|
| ***PET-CT*** | | | | | | | | | | |
| ***Lymphoma*** | | | | | | | | | | |
| Crocchiolo 2009(22) Italy | Hodgkin lymphoma | 27 | 35 (median) | nd | Stage I 4% Stage II 44% Stage III 33% Stage IV 19% | Retrospective cohort (1999-2006) | -Discovery LS or Discovery ST, GE Medical Systems -CT used for attenuation correction | nd | Every 4 months during first year, every 6 months in second and third year, yearly afterwards | Contrast enhanced CT scans, MRI, bone marrow biopsy, clinical exam |
| El-Galaly 2011(23) Denmark | Non-Hodgkin lymphoma | 52 | 61 | 60 | Ann Arbor stage I: 25% Ann Arbor stage II: 15% Ann Arbor stage III: 27% Ann Arbor stage IV: 33% | Retrospective cohort (2006-2009) | -Discovery VCT, GE Healthcare | nd | Mean 2.6 routine PET-CTs per patient | Biopsy and/or additional imaging/follow-up |
| Lee 2010(21) US | Hodgkin lymphoma | 192 | 33 (median) | 50 | Early (stage IA-IIA) 48% Advanced (stage IIB-IV) 52% | Retrospective cohort (2003-2006) | -nd | nd | Variable | CT, biopsy |
| Rhodes 2006(24) US | Hodgkin and Non-Hodgkin lymphoma | 41 | 13 (median) | 56 | I or II- 66% III or IV- 34% | Retrospective cohort (1999-2004) | -Discovery LS -CT used for attenuation correction | nd | nd | Clinical exam, lab results, biopsy |
| ***Head and Neck Cancer*** | | | | | | | | | | |
| Abgral 2009(25) France | Squamous cell carcinoma | 91 | 57.4 | 86 | Negative scans: I- 7.7% II- 21.2% III- 17.3% IV- 53.8% Positive scans: I- 2.6% II- 17.9% III- 23.1% IV- 56.4% | Prospective cohort (2005-2008) | -Gemini GXLi, Philips -CT used for attenuation correction | 12.3 ± 4.1 for negative scans, 10.7 ± 4.7 for positive scans | once | Physician interpretation of clinical exam |
| ***PET*** | | | | | | | | | | |
| ***Lymphoma*** | | | | | | | | | | |

| Author Year Country | Cancer | N | Age, y (Mean or Median, range) | Male (%) | Tumor stage | Study design (study duration*) | PET or PET-CT manufacturer, description | Time from treatment end to first PET-CT scan | PET or PET-CT frequency | Reference standard |
|---|---|---|---|---|---|---|---|---|---|---|
| Hosein(27)<br>2011<br>Italy | Mantle cell lymphoma | 34 | 69.1 | 73 | Unknown- 6%<br>I- 20.5%<br>II A+B - 3.8%<br>III A+B+C – 69.7% | Prospective cohort (2004-2011) | Axis- Marconi Medical Systems, and Gemini TF-Philips Medical Systems | nd | Three times, at 6, 12, and 24 mo | Biopsy or curative surgery |
| Zinzani(26)<br>2009<br>Denmark | Lymphoma | 421 | nd | nd | nd | Prospective cohort (2002-2007) | GE Discovery tomograph, GE Medical Systems | 6 mo | Every 6 mo for first 2 yr, then annually | Imaging and/or biopsy and/ or clinical exam |
| ***Colorectal*** | | | | | | | | | | |
| Selvaggi(28)<br>2002<br>US | Colorectal | 31 | 61<br>(43-79) | 61 | B2- 26%;<br>C1- 42%,<br>C2- 32% | Retrospective cohort (1993-1998) | PET EXACT 47, Siemens | 24 mo | Once | CT, biopsy, histology, surgery |
| Sobhani(20)<br>2008<br>US | Colorectal | 65 | PET: 58.1 (11.2); Control: 62.0 (12.1) | nd | PET: 12.1% stage IV<br>Control: 13.8% stage IV | Randomized controlled trial (2001-2004) | C-PET tomograph, Philips | 9 mo | Twice, at 9 and 15 mo | Histology from biopsy or curative surgery, or clinical exam, tumor markers, and imaging procedures |
| ***Head and Neck Cancer*** | | | | | | | | | | |
| Lowe(29)<br>2000<br>France | Head and neck | 30 | nd | nd | 100% stage III or IV | Prospective cohort (nd) | ECAT 951/31, Siemens | 10 mo (also PET at 2 mo) | Twice (also PET at 2 mo) | Biopsy |
| Perie(30)<br>2007<br>France | Squamous cell carcinoma | 43 | 56.8 (41-78) | 86 | 100% stage III or IV | Prospective cohort (2000-2003) | C-PET, Adac laboratories | 12-14 mo | Once | Imaging, biopsy, or cytology |
| Salaun(31)<br>2007<br>France | Squamous cell carcinoma | 30 | 59.3 (13.2) | 77 | Stage 1 7%, stage 27%, stage 3 27%, stage 4 40% | Retrospective cohort (2002-2005) | Allegro, Philips | 21 (13.7) mo | Once | Imaging, histology |

nd = No data

*Period during which patients were treated and seen for post-treatment consultation.

## Table 2

### Quality of Surveillance PET and PET-CT studies

| Study | Prospective study design? | Clear eligibility criteria? | Selection bias likely? | Index and reference tests adequately described? | Blinded outcome assessment? | Clear reporting with no discrepancies? | Overall grade |
|---|---|---|---|---|---|---|---|
| **PET-CT** | | | | | | | |
| **Lymphoma** | | | | | | | |
| Crocchiolo(22) | No | Yes | No | Yes | No | Yes | **B** |
| El-Galaly(23) | No | Yes | No | Yes | No | Yes | **B** |
| Lee(21) | No | Yes | Yes | No | No | Yes | **C** |
| Rhodes(24) | No | Yes | No | Yes | No | Yes | **B** |
| **Head and Neck Cancer** | | | | | | | |
| Abgral(25) | Yes | Yes | No | Yes | Yes | Yes | **A** |
| **PET** | | | | | | | |
| **Lymphoma** | | | | | | | |
| Hosein(27) | No | Yes | Yes | Yes | No | Yes | **C** |
| Zinzani(26) | Yes | Yes | No | Yes | No | No | **B** |
| **Head and Neck Cancer** | | | | | | | |
| Lowe(29) | Yes | Yes | Yes | Yes | Yes | No | **B** |
| Perie(30) | Yes | Yes | Yes | Yes | Yes | Yes | **B** |
| Salaun(31) | No | Yes | No | Yes | Yes | Yes | **B** |
| **Colorectal Cancer** | | | | | | | |
| Selvaggi(28) | No | Yes | Yes | Yes | Yes | Yes | **C** |
| Sobhani(20) | Yes | Yes | No | Yes | Yes | No | **B** |

**Table 3**

**Summary of Test Performance for Surveillance PET-CT and PET**

| Author, Year | Participants *n* | TP | FN | TN | FP | Sensitivity 95% CI | Specificity 95% CI | PPV, 95% CI<br>NPV, 95% CI | LR+, 95% CI<br>LR-, 95% CI | Time of PET / PET-CT Scan |
|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | | ***PET-CT*** | | | | |
| ***Lymphoma*** | | | | | | | | | | |
| Crocchiolo, 2009(22) | 27 | 6 | 0 | 15 | 6 | 100<br>54-100 | 71<br>48-88 | 0.50, 0.25-0.75;<br>1, 0.80-1.00 | 3.50, 1.78-6.88;<br>not calculable; | -1st year every 4 mo<br>-2nd/3rd year every 6 mo<br>-Yearly thereafter |
| El-Galaly, 2011(23) | 52 | 1 | 0 | 47 | 4 | 100<br>3-100* | 92<br>80-97* | 0.20, 0.04-0.62; 1.00,<br>0.92-1.00 | 12.8, 4.98-32.7;<br>not calculable | Half-yearly scan for the first 2 yr after treatment; mean 2.6 scans per patient during the first 2 yr. |
| Lee, 2010(21) | 192 | 9 | 3 | 162 | 18 | 75<br>43-93 | 90<br>84-94 | 0.33, 0.19-0.52; 0.98,<br>0.95-0.99 | 7.50, 4.34-13.0;<br>0.28, 0.10-0.74 | Variable timing and frequency |
| Rhodes, 2006(24) | 41 | 6 | 0 | 15 | 20 | 100<br>54-100 | 43<br>27-60 | 0.23, 0.11-0.42; 1,<br>0.80-1.00 | 1.75, 1.31-2.33;<br>not calculable | Variable, within 30 mos after negative post-treatment PET-CT *(Per-patient calculation, equivocal scans considered positive)* |
| | | 6 | 0 | 28 | 7 | 100<br>54-100 | 80<br>63-91 | 0.46, 0.23-0.71; 1.00,<br>0.88-1.00 | 5.0, 2.58-9.70;<br>not calculable | *(Per-patient calculation, equivocal scans considered negative)* |
| | 247 scans | 18 | 1 | 154 | 74 | 95<br>72-100 | 68<br>61-73 | 0.20, 0.13-0.29; 0.99,<br>0.96-1.00 | 2.92, 2.35-3.62; 0.08,<br>0.01-0.53 | *(Per-scan calculation, equivocal scans considered positive)* |
| | | 18 | 1 | 212 | 16 | 95<br>72-100 | 93<br>89-96 | 0.53, 0.37-0.69; 1,<br>0.97-1.00 | 13.5, 8.3-21.9; 0.06,<br>0.01-0.38 | *(Per-scan calculation, equivocal scans considered negative)* |
| ***Head and Neck Cancer*** | | | | | | | | | | |
| Abgral, 2009(25) | 91 | 30 | 0 | 52 | 9 | 100<br>88-100 | 85<br>73-93 | 0.77, 0.62-0.87; 1,<br>0.93-1.00 | 6.78; 3.71-12.4 | One scan at mean 11.6 mo<br>Follow-up in 6 mo |
| | | | | | | ***PET*** | | | | |
| ***Lymphoma*** | | | | | | | | | | |
| Hosein 2011(27) | 34 | nd | nd | nd | nd | nd | Nd | nd | nd | Every 6 mo for 3 yr |
| | 162 scans | | | | | 83<br>36-97 | 64<br>56-72 | | | |

| Author, Year | Participants *n* | TP | FN | TN | FP | Sensitivity 95% CI | Specificity 95% CI | PPV, 95% CI<br>NPV, 95% CI | LR+, 95% CI<br>LR-, 95% CI | Time of PET / PET-CT Scan |
|---|---|---|---|---|---|---|---|---|---|---|
| Zinzani 2009(26) | 421 | 41 | 5 | 375 | 0 | 89<br>76-96 | 100<br>99-100** | 1.00, 0.91-1.00; 0.99, 0.97-0.99 | not calculable; 0.11, 0.05-0.25 | Every 6 mo first 2 years, then annually |
| | | 46 | 0 | 358 | 17 | 100<br>92-100 | 95<br>93-97** | 0.73, 0.61-0.82; 1.00, 0.99-1.00 | 22.1, 13.9-35.1; not calculable | |
| ***Head and Neck Cancer*** | | | | | | | | | | |
| Lowe, 2000(29) | 30 | 16 | 0 | 13 | 1 | 100<br>79-100 | 93<br>64-100 | 0.94, 0.73-0.99; 1, 0.77-1.00 | 14, 2.12-92.6; not calculable | Once at 10 mo |
| Perie, 2007(30) | 43 | 3 | 1 | 36 | 3 | 75<br>22-99 | 92<br>78-98 | 0.50, 0.19-0.81; 0.97, 0.86-1.00 | 9.75, 2.86-33.2; 0.27, 0.05-1.48 | Once at 1 yr |
| Salaun, 2007(31) | 30 | 8 | 0 | 21 | 1 | 100<br>63-100 | 95<br>75-100 | 0.89, 0.57-0.98; 1.00, 0.85-1.00 | 22.0, 3.24-149; not calculable | Once at mean of 21 (+/-13.7) mo |
| ***Colorectal*** | | | | | | | | | | |
| Selvaggi, 2002(28) | 31 | 4 | 0 | 26 | 1 | 100<br>40-100 | 96<br>79-100 | 0.80, 0.38, 0.96; 1, 0.87-1.00 | 27.0, 3.95-184; not calculable | At 2 yr, after negative body CT and MRI at 1, 2 yr |
| Sobhani, 2008(20) | 65 (ITT) | nd | nd | nd | nd | 91 | 92 | nd | nd | At 9 mo and 15 mo |

nd = No data; N/A = Not applicable

ITT = intention to treat

PPV = positive predicted value, NPV = negative predicted value

LR+ = positive likelihood ratio, LR- = negative likelihood ratio

* result of the first scan 6 months after treatment;

** 2×2 table not provided in study. To estimate test accuracy, we performed two analyses--treating inconclusive tests as negative in the first row and as positive in the second row.

## Table 4
**Characteristics of studies not included after full-text screening**

| | Lymphoma 41 (40%) | Head & Neck Cancer 38 (37%) | Colorectal Cancer 24 (23%) |
|---|---|---|---|
| **Prospective** | 6 (15%) | 17 (45%) | 5 (21%) |
| **Blinded scan interpretation** | 9 (22%) | 10 (26%) | 8 (33%) |
| **PET only** | 17 (41%) | 16 (42%) | 18(75%) |
| **PET / PET-CT** | 0 (0%) | 2 (5%) | 3 (12%) |
| **PET-CT only** | 24 (58%) | 20 (53%) | 3 (12%) |
| **Reported accuracy results** | 25 (61%) | 29 (76%) | 21 (88%) |
| **Reported PET schedule clearly** | 11 (27%) | 13 (34%) | 2 (8%) |
| **Patients disease free at time of scan** | 3 (7%) | 5 (13%) | 0 (0%) |
| **Describes scans as ‘surveillance’** | 10 (24%) | 10 (26%) | 1 (4%) |