

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 Patel¹, Nira Hadar¹, Joungee Lee¹, Barry A. Siegel², Bruce E. Hillner³, and Joseph Lau¹

¹*Institute of Clinical Research and Health Policy Studies, Tufts Medical Center, Boston, Massachusetts;* ²*Division of Nuclear Medicine, Mallinckrodt Institute of Radiology and the Siteman Cancer Center, Washington University School of Medicine, St. Louis, Missouri;* and ³*Department of Internal Medicine and the Massey Cancer Center, Virginia Commonwealth University, Richmond, Virginia*

PET and PET/CT are widely used for surveillance of patients after 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 cancers. **Methods:** We searched MEDLINE and Cochrane Library databases from 1996 to March 2012 for English-language studies of PET or PET/CT used for surveillance of 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: 6 lymphoma ($n = 767$ patients), 2 colorectal cancer ($n = 96$), and 4 head and neck cancer ($n = 194$). All studies lacked a uniform definition of surveillance and scan protocols. Half the studies were retrospective, and a third were rated as low quality. The majority reported sensitivities and specificities in the range of 90%–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 preclude making an informed judgment on the value of PET for this potential indication.

Key Words: surveillance; PET; PET/CT; lymphoma; colorectal

J Nucl Med 2013; 54:1518–1527

DOI: 10.2967/jnumed.112.119362

PET using the glucose analog ¹⁸F-FDG has become an important modality for cancer imaging because of the characteristically increased use of glucose by malignant cells. Since its introduction in 2000, PET/CT has progressively replaced conventional PET, and nearly all scanners now in use worldwide are PET/CT scanners (1). Compared with conventional PET, PET/CT provides greater accuracy in localizing ¹⁸F-FDG uptake, with resultant improvement in observer performance (2,3). Hereafter in this article,

the term *PET* will be used to refer to PET or PET/CT; distinctions will be made where needed.

PET is used for many cancers for diagnosis, initial staging, assessment of treatment response (4,5), restaging, detection of clinically suspected recurrence, and surveillance (6–9). Using advanced imaging, including PET, for posttreatment surveillance of patients is controversial and generally not recommended for most cancers (10,11). The widely held yet anecdotal impression is 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). Although systematic reviews have been conducted for a range of PET uses, none have focused on the use of PET for surveillance (14,15).

A common conceptual framework for evaluating diagnostic test technologies categorizes studies into 6 assessment levels (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, the largest numbers of patients undergoing posttreatment surveillance. We also gathered data from studies that did not meet the inclusion criteria to inform future research recommendations.

MATERIALS AND METHODS

In carrying out this systematic review, we adhered to the PRISMA statement (Preferred Reporting Items for Systematic reviews and Meta-Analyses) (17).

Literature Search Strategy

We searched the MEDLINE and Cochrane Central Register of Controlled Trials from 1996 to March 2012 for English-language studies examining the use of PET in lymphoma, colorectal cancer, and head and neck cancer. We looked for additional studies by reviewing the reference lists of studies that met our inclusion criteria and relevant Cochrane systematic reviews. A variety of keywords and Medical Subject Heading (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 1 of 4 authors, with questionable studies being adjudicated by all authors. Surveillance imaging was defined as imaging performed at least 6 mo after completion of treatment with curative intent among patients who were

Received Jan. 9, 2013; revision accepted Mar. 18, 2013.
For correspondence or reprints contact: Joseph Lau, Brown University, Box G-S121-8, Providence, RI 02912.
E-mail: joseph_lau@brown.edu
Published online Jun. 17, 2013.
COPYRIGHT © 2013 by the Society of Nuclear Medicine and Molecular Imaging, Inc.

considered to be disease-free by clinical examination or other imaging 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 1-time basis or a 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 used either individual patients or individual scans as the unit of analysis and either reported test accuracy (e.g., sensitivity, specificity, positive predictive value, negative predictive value, positive likelihood ratio, and negative likelihood ratio) or presented data in 2×2 tables that allowed calculation of test 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 and confirmed by another. Discrepancies were reconciled by three of us. 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. Although some studies performed surveillance scans at more than one time point, 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 time point at which surveillance scans occurred, at a minimum of 6 mo after treatment completion. Where possible, we computed the “yield” of screening, defined as the percentage of positive studies (true-positive plus false-positive) in the scanned population. When not provided by the study, test accuracy measures (sensitivity, specificity, positive and negative predictive values, and likelihood ratios) and confidence intervals were calculated using Stata, version 11.0 (StataCorp LP).

Study Quality Assessment

We extracted information on study design, conduct, and reporting and used the Quality Assessment of Diagnostic Accuracy Studies tool (18) to evaluate the quality of the studies assessing test accuracy. For comparative studies reporting on clinical impact outcomes, we combined this tool with selected items from the Cochrane Risk of Bias tool (19) that were applicable to diagnostic testing studies. 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 or reporting deficiencies.

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

RESULTS

The literature search yielded 1,813 citations, from which 146 full-text articles were evaluated (Fig. 1). Twelve studies (7 PET, 5 PET/CT) met our inclusion criteria and provided test accuracy

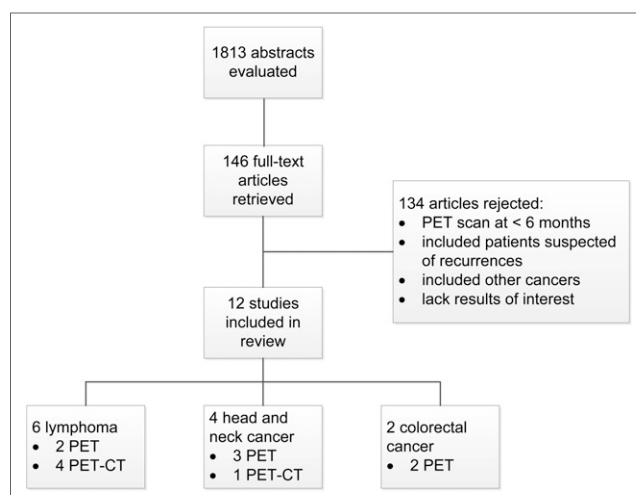


FIGURE 1. Literature flow of evaluated abstracts, retrieved studies, and included studies.

data. One randomized controlled trial 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 6 mo after completion of treatment or scanning in order to assess treatment response or for restaging.

Table 1 shows the characteristics of the included studies: 6 lymphoma studies (2 PET, 4 PET/CT), 2 PET studies in colorectal cancer, and 4 in head and neck cancer (3 PET, 1 PET/CT). All 5 PET/CT studies used 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 schedule for repeated scans. The duration between the final surveillance PET and the last clinical follow-up examination ranged from 2.3 to 31 mo. In 7 studies, patients were scanned serially, in 4 studies patients were scanned once, and the frequency was indeterminate in another. The reference standard used to verify PET results varied among studies and included CT alone, as well as a combination of laboratory and imaging findings and an absence of symptoms.

Although patients were deemed to be disease-free after treatment completion in all studies, 10 studies did not indicate the means by which disease status was confirmed. Patients in 2 studies were deemed disease-free by negative results on 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). Information was unclear as to how sensitivity and specificity were calculated when a patient had conflicting scan results at 2 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 and specific grading criteria. There were 1 quality A study, 8 B studies, and 3 C studies.

Lymphoma

For lymphoma, there were 4 retrospective PET/CT studies and 2 prospective PET studies. Four were rated as quality B (22–26) and 2 as quality C (21,27). The 2 quality C studies listed in the tables

TABLE 1
Characteristics of Studies Evaluating Surveillance PET or PET/CT

Study	Type of cancer	N	Age (y)	Male (%)	Tumor stage	Study design and duration*	Type of scanner	Time (mo) from treatment end to first PET/CT scan	Scanning frequency	Reference standard
PET/CT: lymphoma										
Crocchiolo (22), 2009, Italy	Hodgkin lymphoma	27	Median, 35	ND	I, 4%; II, 44%; III, 33%; IV, 19%	Retrospective cohort (1999–2006)	Discovery LS or Discovery ST; GE Healthcare (CT used for attenuation correction)	ND	Every 4 mo during first year, every 6 mo in second and third years, yearly thereafter	Contrast-enhanced CT, MR imaging, bone marrow biopsy, clinical examination
El-Galaly (23), 2011, Denmark	Non-Hodgkin lymphoma	52	Mean, 61	60	Ann Arbor I, 25%; II, 15%; III, 27%; IV, 33%	Retrospective cohort (2006–2009)	Discovery VCT; GE Healthcare	ND	Mean, 2.6 routine PET/CT scans per patient	Biopsy or additional imaging/follow-up
Lee (21), 2010, United States	Hodgkin lymphoma	192	Median, 33	50	Early (IA–IIA), 48%; advanced, IIB–IV, 52%	Retrospective cohort (2003–2006)	ND	ND	Variable	CT, biopsy
Rhodes (24), 2006, United States	Hodgkin and non-Hodgkin lymphoma	41	Median, 13	56	I or II, 66%; III or IV, 34%	Retrospective cohort (1999–2004)	Discovery LS (CT used for attenuation correction)	ND	ND	Clinical examination, lab results, biopsy
PET/CT: head and neck cancer										
Abgral (25), 2009, France	Squamous cell carcinoma	91	Mean, 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 GXL; 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 examination

TABLE 1 (Continued)

Study	Type of cancer	N	Age (y)	Male (%)	Tumor stage	Study design and duration*	Type of scanner	Time (mo) from treatment end to first PET/CT scan	Scanning frequency	Reference standard
PET:										
lymphoma										
Hosein (27), 2011, Italy	Mantle cell lymphoma	34	Mean, 69.1	73	Unknown, 6%; I, 20.5%; II A+B, 3.8%; III A+B+C, 69.7%	Prospective cohort (2004–2011)	Axis; Marconi, and Gemini TF; Philips	ND	Three times, at 6, 12, and 24 mo	Biopsy or curative surgery
Zinzani (26), 2009, Denmark	Lymphoma	421	ND	ND	ND	Prospective cohort (2002–2007)	Discovery; GE Healthcare	6	Every 6 mo for first 2 y, and then annually	Imaging or biopsy or clinical examination
PET: head and neck cancer										
Lowe (29), 2000, France	Head and neck	30	ND	ND	III or IV, 100%	Prospective cohort (ND)	ECAT 951/31; Siemens	10 (also PET at 2 mo)	Twice (also PET at 2 mo)	Biopsy
Périer (30), 2007, France	Squamous cell carcinoma	43	Mean, 56.8 (range, 41–78)	86	III or IV, 100%	Prospective cohort (2000–2003)	C-PET; ADAC	12–14	Once	Imaging, biopsy, or cytology
Salaun (31), 2007, France	Squamous cell carcinoma	30	Mean, 59.3 (SD, 13.2)	77	I, 7%; II, 7%; III, 27%; IV, 40%	Retrospective cohort (2002–2005)	Allegro; Philips	21 (SD, 13.7)	Once	Imaging, histology
PET: colorectal										
Selvaggi (28), 2003, United States	Colorectal	31	Mean, 61 (range, 43–79)	61	B2, 26%; C1, 42%; C2, 32%	Retrospective cohort (1993–1998)	EXACT 47; Siemens	24	Once	CT, biopsy, histology, surgery

TABLE 1 (Continued)

Study	Type of cancer	N	Age (y)	Male (%)	Tumor stage	Study design and duration*	Type of scanner	Time (mo) from treatment end to first PET/CT scan	Scanning frequency	Reference standard
Sobhani (20), 2008, United States	Colorectal	65	PET: mean, 58.1 (SD, 11.2); control: mean, 62.0 (SD, 12.1)	ND	PET: IV, 12.1%; Control: IV, 13.8%	Randomized controlled (2001–2004)	CPET; Philips	9	Twice, at 9 and 15 mo	Histology from biopsy or curative surgery, or clinical examination, tumor markers, and imaging procedures

*Period during which patients were treated and seen for posttreatment consultation.
ND = no data.

are not included in the synthesis. Sample sizes of quality B studies ranged from 27 to 421 patients, with a total of 541 patients. Only 1 study analyzed children (24).

Table 3 shows diagnostic accuracy by cancer and imaging modality. The 3 quality B PET/CT studies included 120 patients, had a per-patient level sensitivity of 100%, and had specificities ranging from 43% to 92% (22–24). One PET study with 421 patients was rated quality B and reported a sensitivity of 89% and a specificity of 100% (26). Among the 4 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 (28). The randomized trial compared a surveillance strategy that included CT at 9 and 15 mo after surgery ($n = 65$) with a strategy that included both PET and CT scans ($n = 65$) at the same time points. This trial assessed impact on therapeutic decision making, impact on mortality, and test accuracy. The retrospective study was graded quality C because of likely selection bias.

The randomized trial ended recruitment early because of ethical and methodologic concerns when PET/CT scanning became available in 2004 at the institution. For clinical impact outcomes, the study was rated quality A. Using a per-protocol analysis with 60 patients in the PET group (5 fewer than in the intention-to-treat analysis because of 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 mo) than in patients in the control group (15.4 ± 6 mo; $P = 0.01$). Therapy was started sooner, but not significantly so, in the PET group (14.8 ± 4.1 vs. 17.5 ± 6.0 mo, $P = 0.09$). Surgery for recurrent disease was performed more frequently in the PET group (15 of 23 [65%] vs. 2 of 21 [9.5%], $P < 0.0001$). Moreover, the frequency of curative resection of recurrences was higher in the PET group (43.8% vs. 9.5%, $P < 0.01$). Intention-to-treat analyses gave similar results. Using a per-protocol analysis, the study also found that a nonsignificantly greater number of patients with recurrences died during the study period (with a maximum follow-up of 24 mo) in the control group than in the PET group (28.5% vs. 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 1 prospective study (25) and by PET in 2 prospective (29,30) and 1 retrospective (31) studies. The PET/CT study was rated quality A, and the PET studies were rated quality B. The PET studies had unclear reporting and 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 3 PET studies comprised 103 patients and included 2 studies examining squamous cell carcinoma (30,31) and 1 including all cell types (29). Sensitivities ranged from 75% to 100%, and specificities ranged from 92% to 95%. The 4 head and neck cancer studies had positive predictive values between 0.5 and 0.9, a negative predictive value of 1.0, and a yield of positive PET scans ranging from 14% to 57%.

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?	Masked 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
PET/CT: 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
PET: head and neck cancer							
Lowe (29)	Yes	Yes	Yes	Yes	Yes	No	B
Périé (30)	Yes	Yes	Yes	Yes	Yes	Yes	B
Salaun (31)	No	Yes	No	Yes	Yes	Yes	B
PET: colorectal cancer							
Selvaggi (28)	No	Yes	Yes	Yes	Yes	Yes	C
Sobhani (20)	Yes	Yes	No	Yes	Yes	No	B

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 on these cancers described the scans as being for surveillance. In none of the colorectal cancer studies were patients verified to be disease-free, and in only one of these were the scans described as being for surveillance.

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

DISCUSSION

This systematic review of PET for posttreatment surveillance of patients with lymphoma, colorectal cancer, or head and neck cancer found only a single 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 on therapeutic decision making and may improve patient outcomes when used for surveillance of colorectal cancer—one of the few cancers for which evidence exists supporting intensive posttreatment surveillance (34). Most trials reported sensitivity and specificity in the range of 90% and 100%, but some reported much lower values.

Because of the inconsistent definition of surveillance, the variations in imaging protocols, and the few studies using a particular imaging modality in a given cancer type, we did not conduct a metaanalysis. In addition, the literature was of inferior quality—7 of 12 studies used a retrospective design and half lacked masked outcome assessments. The retrospective studies had an inconsistent or no predefined scanning frequency and interval. Prospective studies used widely ranged scanning schedules—from multiple scans every 6 mo to a single scan at roughly 2 y after treatment completion.

Our finding of a lack of evidence supporting PET/CT in posttreatment surveillance is reflected in practice guidelines (10,11). Current National Comprehensive Cancer Network 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, nor is PET recommended in the non-Hodgkin lymphoma guideline. Nevertheless, PET/CT is commonly used for surveillance (35). Possible risks of using PET/CT for surveillance include overtreatment based on false-positives and unnecessary radiation exposure (36).

Our review highlights the 2 failures that dominate problems with surveillance. The first is the lack of a common definition for surveillance (the minimal time since last treatment and the absence of clinical or other diagnostic suspicion of recurrence), and the second is the lack of a 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, which has been shown in each of these cancers to have its own declining pattern with time.

Few studies met the 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

TABLE 3
Summary of Test Performance for Surveillance PET/CT and PET

Study	No. of patients or scans	TP	FN	TN	FP	Sensitivity, with 95% CI	Specificity, with 95% CI	PPV and NPV, with 95% CIs	LR+ and LR-, with 95% CIs	Time of scan
PET/CT: lymphoma										
Crocchiolo (22)	27 patients	6	0	15	6	100, 54–100	71, 48–88	0.50, 0.25–0.75; 1.00, 0.80–1.00	3.50, 1.78–6.88; incalculable	First year, every 4 mo; second and third year, every 6 mo; annually thereafter
El-Galaly (23)	52 patients	1	0	47	4	100, 3–100*	92, 80–97*	0.20, 0.04–0.62; 1.00, 0.92–1.00	12.8, 4.98–32.7; incalculable	Half-yearly scan for first 2 y after treatment; mean 2.6 scans per patient during first 2 y
Lee (21)	192 patients	9	3	162	18	75, 43–93	90, 84–94	0.33, 0.19–0.52; 0.98, 0.95–0.99	7.50, 4.34–13.0; 0.28, 0.10–0.74	Variable timing and frequency
Rhodes (24)										Variable, within 30 mo after negative posttreatment PET/CT
	41 patients	6	0	15	20	100, 54–100	43, 27–60	0.23, 0.11–0.42; 1.00, 0.80–1.00	1.75, 1.31–2.33; incalculable	(Per-patient calculation, equivocal scans considered positive)
		6	0	28	7	100, 54–100	80, 63–91	0.46, 0.23–0.71; 1.00, 0.88–1.00	5.0, 2.58–9.70; incalculable	(Per-patient calculation, equivocal scans considered negative)
	247 scans	18	1	154	74	95, 72–100	68, 61–73	0.20, 0.13–0.29; 0.99, 0.96–1.00	2.92, 2.35–3.62; 0.08, 0.01–0.53	(Per-scan calculation, equivocal scans considered positive)

TABLE 3 (Continued)

Study	No. of patients or scans	TP	FN	TN	FP	Sensitivity, with 95% CI	Specificity, with 95% CI	PPV and NPV, with 95% CIs	LR+ and LR-, with 95% CIs	Time of scan
		18	1	212	16	95, 72–100	93, 89–96	0.53, 0.37–0.69; 1.00, 0.97–1.00	13.5, 8.3–21.9; 0.06, 0.01–0.38	(Per-scan calculation, equivocal scans considered negative)
PET/CT: head and neck cancer										
Abgral (25)	91 patients	30	0	52	9	100, 88–100	85, 73–93	0.77, 0.62–0.87; 1.00, 0.93–1.00	6.78; 3.71–12.4	One scan at mean of 11.6 mo; follow-up in 6 mo
PET: lymphoma										
Hosein (27)	34 patients	ND	ND	ND	ND	ND	ND	ND	ND	Every 6 mo for 3 y
Zinzani (26)	162 scans 421 patients	41	5	375	0	83, 36–97 89, 76–96	64, 56–72 100, 99–100†	1.00, 0.91–1.00; 0.99, 0.97–0.99	Incalculable; 0.11, 0.05–0.25	Every 6 mo first 2 y, then annually
		46	0	358	17	100, 92–100	95, 93–97†	0.73, 0.61–0.82; 1.00, 0.99–1.00	22.1, 13.9–35.1; incalculable	
PET: head and neck cancer										
Lowe (29)	30 patients	16	0	13	1	100, 79–100	93, 64–100	0.94, 0.73–0.99; 1.00, 0.77–1.00	14, 2.12–92.6; incalculable	Once at 10 mo
Périé (30)	43 patients	3	1	36	3	75, 22–99	92, 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 y
Salaun (31)	30 patients	8	0	21	1	100, 63–100	95, 75–100	0.89, 0.57–0.98; 1.00, 0.85–1.00	22.0, 3.24–149; incalculable	Once at mean of 21 (±13.7) mo
PET: colorectal										
Selvaggi (28)	31 patients	4	0	26	1	100, 40–100	96, 79–100	0.80, 0.38–0.96; 1.00, 0.87–1.00	27.0, 3.95–184; incalculable	At 2 y, after negative body CT and MR imaging at 1 and 2 y
Sobhani (20)	65 patients (ITT)	ND	ND	ND	ND	91	92	ND	ND	At 9 and 15 mo

*Result of first scan 6 mo after treatment.

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

TP = true-positive; FN = false-negative; TN = true-negative; FP = false-positive; CI = confidence interval; PPV = positive predictive value; NPV = negative predictive value; LR+ = positive likelihood ratio; LR- = negative likelihood ratio; ND = No data; ITT = intention to treat.

TABLE 4
Characteristics of Studies Not Included After Full-Text Screening

Characteristic	Lymphoma, n = 41 (40%)	Head and neck cancer, n = 38 (37%)	Colorectal cancer, n = 24 (23%)
Prospective	6 (15%)	17 (45%)	5 (21%)
Masked scan interpretation	9 (22%)	10 (26%)	8 (33%)
PET only	17 (41%)	16 (42%)	18 (75%)
PET or PET/CT	0 (0%)	2 (5%)	3 (12%)
PET/CT only	24 (58%)	20 (53%)	3 (12%)
Accuracy results reported	25 (61%)	29 (76%)	21 (88%)
PET schedule clearly reported	11 (27%)	13 (34%)	2 (8%)
Patients disease-free at time of scan	3 (7%)	5 (13%)	0 (0%)
Scans described as being for "surveillance"	10 (24%)	10 (26%)	1 (4%)

amount of evidence on surveillance scanning, we collected data from rejected studies to better understand characteristics of studies that may still fall under the definition of surveillance scanning. We found that most 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 had several limitations. Results were difficult to synthesize because of the lack of a standard surveillance definition. Studies were generally of poor quality, with more than half being retrospective. In studies conducting multiple scans as part of a surveillance protocol, we were unable to use all available data because test accuracy was not consistently defined and reporting was incomplete. One study on head and neck cancer 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 2 generations of PET technologies—PET alone and PET/CT. There is substantial evidence across many cancers 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 results with PET alone, results from PET-only studies set a baseline of performance that is likely only to be improved by PET/CT. Finally, we did not assess publication bias. Although there is always a concern in systematic reviews that unpublished negative studies may negate the positive results, the paucity of evidence in favor of using surveillance PET lessens this concern. There is also a lack of reliable methods to assess publication bias (37).

Future research should provide detailed descriptions of the surveillance protocols and patient populations; the need for such details has been suggested in previous systematic reviews of colorectal cancer surveillance (34,38). Better-conducted studies will help to answer the questions of which patients would be helped most by surveillance (e.g., patients having different disease severities) and which surveillance protocols are most effective for different cancers. Because of the few randomized trials conducted in this area, it is even more important for prospective trials to clearly report protocols and patient populations in order for the efficacy of PET/CT surveillance strategies to be understood. Retrospective studies can help answer the question of PET/CT surveillance test accuracy, but prospective studies are needed to address aspects of clinical impact. As suggested by Baca et al. (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 cancers 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 cancers.

DISCLOSURE

The costs of publication of this article were defrayed in part by the payment of page charges. Therefore, and solely to indicate this fact, this article is hereby marked "advertisement" in accordance with 18 USC section 1734. This work was supported in part by a National Cancer Institute Grand Opportunity Grant, RC2CA148259. Barry A. Siegel is on the advisory boards of Radiology Corporation of America, Siemens Molecular Imaging, and GE Healthcare and is a stockholder of Radiology Corporation of America. No other potential conflict of interest relevant to this article was reported.

REFERENCES

1. Mujoondar M, Moulton K, Nkansah E. *Positron Emission Tomography (PET) in Oncology: A Review of the Clinical Effectiveness and Indications for Use*. Ottawa, Canada: Canadian Agency for Drugs and Technologies in Health—Health Technology Inquiry Service; 2009.
2. von Schultheiss GK, Steinert HC, Hany TF. Integrated PET/CT: current applications and future directions. *Radiology*. 2006;238:405–422.
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–S26.
4. Meyer RM, Ambinder RF, Stroobants S. Hodgkin's lymphoma: evolving concepts with implications for practice. *Hematology Am Soc Hematol Educ Program*. 2004;1:184–202.
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;42:586–590.
6. la Fougère C, 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;33:1417–1425.
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;45:639–644.

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;44:1797–1803.
9. Adams S, Baum RP, Stuckensen T, Bitter K, Hör G. Prospective comparison of ^{18}F -FDG PET with conventional imaging modalities (CT, MR imaging, US) in lymph node staging of head and neck cancer. *Eur J Nucl Med*. 1998;25:1255–1260.
10. Special report: positron emission tomography for the indication of post-treatment surveillance of cancer. BlueCross BlueShield Association Web site. http://www.bcbs.com/blueresources/tec/vols/25/25_03.pdf. Published November 2010. Accessed May 29, 2013.
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–S22.
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;53:1113–1116.
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;49:1928–1935.
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;38:2083–2095.
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;33:210–222.
16. Tatsioni A, Zarin DA, Aronson N, et al. Challenges in systematic reviews of diagnostic technologies. *Ann Intern Med*. 2005;142:1048–1055.
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:W65–W94.
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;155:529–536.
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.
20. Sobhani I, Tiret E, Lebtahi R, et al. Early detection of recurrence by ^{18}F -FDG-PET in the follow-up of patients with colorectal cancer. *Br J Cancer*. 2008;98:875–880.
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;116:3835–3842.
22. Crocchiolo R, Fallanca F, Giovacchini G, et al. Role of ^{18}F -FDG-PET/CT in detecting relapse during follow-up of patients with Hodgkin's lymphoma. *Ann Hematol*. 2009;88:1229–1236.
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;52:597–603.
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. *J Pediatr Hematol Oncol*. 2006;28:300–306.
25. Abgral R, Querellou S, Potard G, et al. Does ^{18}F -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? *J Nucl Med*. 2009;50:24–29.
26. Zinzani PL, Stefoni V, Tani M, et al. Role of [^{18}F]fluorodeoxyglucose positron emission tomography scan in the follow-up of lymphoma. *J Clin Oncol*. 2009;27:1781–1787.
27. Hosein PJ, Pastorini VH, Paes FM, et al. Utility of positron emission tomography scans in mantle cell lymphoma. *Am J Hematol*. 2011;86:841–845.
28. Selvaggi F, Cuocolo A, Sciaudone G, Maurea S, Giuliani A, Mainolfi C. FDG-PET in the follow-up of recurrent colorectal cancer. *Colorectal Dis*. 2003;5:496–500.
29. Lowe VJ, Boyd JH, Dunphy FR, et al. Surveillance for recurrent head and neck cancer using positron emission tomography. *J Clin Oncol*. 2000;18:651–658.
30. Périé 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;137:647–653.
31. Salaun PY, Abgral R, Querellou S, et al. Does ^{18}F -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;29:1115–1120.
32. Rabalais AG, Walvekar R, Nuss D, et al. Positron emission tomography-computed tomography surveillance for the node-positive neck after chemoradiotherapy. *Laryngoscope*. 2009;119:1120–1124.
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;29:986–995.
34. Jeffery M, Hickey BE, Hider PN. Follow-up strategies for patients treated for non-metastatic colorectal cancer. *Cochrane Database Syst Rev*. January 24, 2007: CD002200.
35. Wagner-Johnston ND, Bartlett NL. Role of routine imaging in lymphoma. *J Natl Compr Canc Netw*. 2011;9:575–584.
36. Huang B, Law MW, Khong PL. Whole-body PET/CT scanning: estimation of radiation dose and cancer risk. *Radiology*. 2009;251:166–174.
37. Lau J, Ioannidis JP, Terrin N, Schmid CH, Olkin I. The case of the misleading funnel plot. *BMJ*. 2006;333:597–600.
38. Baca B, Beart RW Jr, Etzioni DA. Surveillance after colorectal cancer resection: a systematic review. *Dis Colon Rectum*. 2011;54:1036–1048.