Accuracy of Detecting Residual Disease After Neoadjuvant Chemoradiotherapy for Esophageal Cancer

A Systematic Review and Meta-analysis

Ben M. Eyck, MD, Barbera D. Onstenk, BSc,* Bo J. Noordman, MD, PhD,* Daan Nieboer, MSc,† Manon C. W. Spaander, MD, PhD,‡ Roelf Valkema, MD, PhD,§ Sjoerd M. Lagarde, MD, PhD,* Bas P. L. Wijnhoven, MD, PhD,* and J. Jan B. van Lanschot, MD, PhD**

Objective: The aim of this study was to perform a meta-analysis on the accuracy of endoscopic biopsies, EUS, and 18F-FDG PET(-CT) for detecting residual disease after neoadjuvant chemoradiotherapy (nCRT) for esophageal cancer.

Summary of Background Data: After nCRT, one-third of patients have a pathologically complete response in the resection specimen. Before an active surveillance strategy could be offered to these patients, clinically complete responders should be accurately identified.

Methods: Embase, Medline, Cochrane, and Web-of-Science were searched until February 2018 for studies on accuracy of endoscopic biopsies, EUS, or PET(-CT) for detecting locoregional residual disease after nCRT for squamous cell- or adenocarcinoma. Pooled sensitivities and specificities were calculated using random-effect meta-analyses.

Results: Forty-four studies were included for meta-analyses. For detecting residual disease at the primary tumor site, 12 studies evaluated endoscopic biopsies, 11 qualitative EUS, 14 qualitative PET, 8 quantitative PET using maximum standardized uptake value (SUVmax), and 7 quantitative PET using percentage reduction of SUVmax (%ΔSUVmax). Pooled sensitivities and specificities were 33% and 95% for endoscopic biopsies, 96% and 8% for qualitative EUS, 74% and 52% for qualitative PET, 69% and 72% for PET-SUVmax, and 73% and 63% for PET-%ΔSUVmax. For detecting residual nodal disease, 11 studies evaluated qualitative EUS with a pooled sensitivity and specificity of 68% and 57%, respectively. In subgroup analyses, sensitivity of PET-%ΔSUVmax and EUS for nodal disease was higher in squamous cell carcinoma than adenocarcinoma.

Conclusions: Current literature suggests insufficient accuracy of endoscopic biopsies, EUS, and 18F-FDG PET(-CT) as single modalities for detecting residual disease after nCRT for esophageal cancer.

Keywords: accuracy, endoscopic biopsies, esophageal cancer, EUS, neoadjuvant chemoradiotherapy, PET, residual disease, restaging

(*Ann Surg* 2019;xx:xxx–xxx)

Esophageal cancer is an aggressive disease. Less than half of the patients can be offered curative treatment at first presentation. To improve survival and prognosis after surgery, the value of neoadjuvant treatment has been investigated extensively. After potentially curative neoadjuvant chemoradiotherapy (nCRT) followed by surgery, 5-year overall survival varies from 47% to 60%.[1,2] After this treatment, one-third of patients have a pathologically complete response (pCR) in the resection specimen, defined as the absence of viable tumor cells at the resected primary tumor site and in the regional lymph nodes as determined by conventional histopathological examination.[1,2] These patients are perhaps unnecessarily exposed to the risk of surgery, which includes perioperative mortality rates of 1% to 5% in high volume centers, severe postoperative morbidity, and a large impact on health-related quality of life.[3,4] Therefore, the question arises whether a standard esophagectomy after nCRT is necessary in all patients, or if patients can be identified who might benefit from a postponed or even omitted resection.

Active surveillance after nCRT, in which patients undergo frequent clinical examinations instead of standard esophagectomy, has been proposed as a novel treatment option.[5] In this organ-sparing approach, surgical resection is offered only to patients with evidence or high suspicion of locoregional recurrence after nCRT without distant metastases. Only patients without signs of locoregional residual disease and distant metastases after nCRT are eligible for active surveillance. To identify patients with locoregional residual disease after nCRT, the disease should be restaged during clinical response evaluations (CREs). These CREs should distinguish patients with locoregional residual and/or disseminated disease from patients with a (near) complete response after nCRT. In current clinical practice, endoscopic biopsies, endoscopic ultrasonography (EUS), and 18F-fluoro-2-deoxy-D-glucose (18F-FDG) positron emission tomography with or without computed tomography (PET(-CT)) are used for pretreatment staging and for restaging during CREs.

The aim of the present study was to provide a systematic review and meta-analysis of the literature regarding the accuracy of endoscopic biopsies, EUS and 18F-FDG PET(-CT) for detecting residual disease after nCRT in potentially curable esophageal cancer patients.

From the *Department of Surgery, Erasmus MC - University Medical Center, Rotterdam, The Netherlands; †Department of Public Health, Erasmus MC - University Medical Center, Rotterdam, The Netherlands; ‡Department of Gastroenterology, Erasmus MC - University Medical Center, Rotterdam, The Netherlands; and §Department of Radiology and Nuclear Medicine, Erasmus MC - University Medical Center, Rotterdam, The Netherlands.

Authors' contributions: BME and BDO contributed to the study design, conducted the systematic review and meta-analysis, and drafted the manuscript. BJN contributed to the conception, study design, and drafting of the manuscript and supervised the systematic review. DN advised in methodology, assisted in statistical analysis, and critically revised the manuscript. MCWS, RV, SML, and BPL had roles in study conception and design and critically revised the manuscript. JJBvL advised in data interpretation, supervised drafting of the manuscript, and critically revised the manuscript. All authors approved the manuscript.

No means of funding were received for this contribution.

The authors declare no conflicts of interest.

BME and BDO equally contributed to this work and should be acknowledged with first authorship.

Supplemental digital content is available for this article. Direct URL citations appear in the printed text and are provided in the HTML and PDF versions of this article on the journal's Web site (www.annalsofsurgery.com).

Reprints: Ben M. Eyck, MD, Department of Surgery, Erasmus MC - University-Medical Center, Dr. Molewaterplein 40, 3015 GD Rotterdam, The Netherlands. E-mail: b.eyck@erasmusmc.nl.

ISSN: 0003-4932/16/XXXX-0001

DOI: 10.1097/SLA.0000000000003397

METHODS

Literature Search

The protocol of this study was registered in the PROSPERO database (CRD42018116649) and the study was performed according to the PRISMA-guidelines for systematic reviews and meta-analyses.[6,7] A systematic literature search was performed by a Health Sciences librarian with expertise in systematic review searching to identify studies that reported on the accuracy of endoscopic biopsy, EUS, and/or PET(-CT) for detection of residual disease after nCRT for esophageal or esophagogastric junctional cancer. The literature search was limited to English language and human studies. Embase, Medline, Cochrane Central libraries, and Web-of-Science were searched until February 2018. The full search strategy is presented in Supplementary Table 1, http://links.lww.com/SLA/B663. References of included studies and reviews of similar subjects were screened for relevance.

Study Selection

Studies were considered eligible if (1) the study population consisted of patients with adenocarcinoma or squamous cell carcinoma of the esophagus or esophago-gastric junction; (2) endoscopic biopsy, EUS, and/or 18F FDG PET(-CT) were investigated; (3) the index tests evaluated detection of residual disease after nCRT at the primary tumor site or in regional lymph nodes; (4) histopathological examination of the surgical resection specimen was used as reference standard; and (5) the study contained sufficient data for construction of a 2 × 2 contingency table. If studies had insufficient data to construct 2 × 2 contingency tables, corresponding authors of each study were contacted by email ≥3 times to provide missing or incomplete data. Studies written in other languages than English, conference abstracts, letters to the editor, editorials, reviews, and studies including <10 patients were excluded. Also, studies reporting on cervical esophageal cancer only were excluded because the current standard of care with curative intent for these tumors is definitive chemoradiotherapy.

The results of the literature search were collected and managed in EndNote reference management software version X7.5 (Thomas Reuters, New York, NY). Records were deduplicated. If duplicates were found during the formal screening process, the record that was published earliest was included. Titles and abstracts were independently evaluated by 2 authors (BE and BO). Potentially relevant reports were screened independently on full text by the same authors. Discrepancies were resolved by consensus discussion. In case of disagreement, a third author (BN) gave a binding verdict.

Data Extraction

Data were extracted by 2 authors (BE and BO) and recorded in predefined data-extraction forms. Study, patient, and test characteristics for each diagnostic modality were extracted from the selected studies. Values of true-positives (TP), false-positives (FP), true-negatives (TN), and false-negatives (FN) were extracted from each study or from additional data provided by the authors to construct 2 × 2 contingency tables. If studies investigated multiple threshold for one index test modality, TP, FP, TN, and FN values produced by the optimal cutoff were chosen for data extraction. For data comparison, pathological response criteria were equated. Studies that used pathological response criteria similar to pathologically complete versus incomplete response were redefined, that is, percentage viable tumor cells (0% vs >0%), American Joint Committee on Cancer TNM stage (T0 vs T+, N0 vs N+, and T0N0 vs T+N+), Mandard and modified Mandard classifications of tumor regression grade (TRG1 vsTRG2–4), Japanese Esophageal Society response evaluation criteria (grade 3 vs grade 0–2), histomorphologic regression grading according to Schneider (grade IV vs grade I–III), WHO and RECIST criteria (complete response vs noncomplete response).[8–14] Likewise, studies that could be categorized as <10% versus >10% and <33.3% versus >33.3% residual disease were redefined as such.

Quality Assessment

The quality of the included studies was independently appraised by 2 authors (BE and BO) according to the revised Quality Assessment of Diagnostic Accuracy Studies (QUADAS-2) tool.[15] Disagreements were resolved by consensus discussion. The QUADAS-2 tool assesses risk of bias regarding 4 key domains: patient selection, index test, reference standard, and flow and timing. Some adjustments were made. Regarding patient selection, the use of induction chemotherapy before concurrent chemoradiotherapy was considered as low risk of bias. In the domain of the index test, the use of non-prespecified thresholds was considered as high risk of bias if these thresholds had not been validated previously. The use of dichotomous outcome measures or validated thresholds was not considered as a risk of bias if not prespecified. Knowledge of the outcome of the reference standard was only considered as a risk of bias when the index test results were reviewed after notification of pathology of the resection specimen. Regarding the reference standard, pathological examination without blinding for the index test was not considered as a potential risk of bias because pathological examination is mainly an independent procedure. In the domain of flow and timing, a time interval between index test and surgery of >4 weeks was considered a high risk of bias because a longer interval increases the probability of a varying index test and pathology outcome.

Statistical Analysis

For individual studies, sensitivity and specificity along with 95% confidence intervals for distinguishing patients with residual disease from patients with pathologically complete response were calculated from TP, FP, TN, and FN and were displayed in forest plots generated with RevMan version 5.3 (The Cochrane Collaboration). Sensitivity was defined as the percentage of patients with residual disease after nCRT who are correctly identified as such.

Random-effect meta-analyses were performed for index test modalities that were evaluated by a minimum of 4 studies and used pathologically complete response (pCR) as pathological response criterion. Summary data were presented in summary receiver-operating characteristic (SROC) plots. The hierarchical summary receiver-operating characteristic (HSROC) model was used to produce SROC curves. The bivariate model was used to generate summary operating points, along with 95% confidence regions and 95% prediction regions. The summary operating point reflected pooled sensitivity and specificity of an index test. Precision of the summary operating point was visualized by a 95% confidence region, which showed the variability for the pooled sensitivity and specificity. Lower variability represented a higher reliability for the index test in identifying residual disease. Between-study heterogeneity was visualized by the 95% prediction region.[16,17] The existence of between-study heterogeneity was primarily assessed through visually inspecting forest plots for the degree of overlapping confidence intervals. The extent of heterogeneity was assessed by visual inspection of 95% prediction regions in SROC plots, where high heterogeneity was depicted by larger 95% prediction regions than 95% confidence regions. Positive predictive value (PPV) and negative predictive value (NPV) were computed from pooled sensitivities and specificities and a representative prevalence of the target condition. After nCRT approximately one-third of patients have pCR.[1,2]

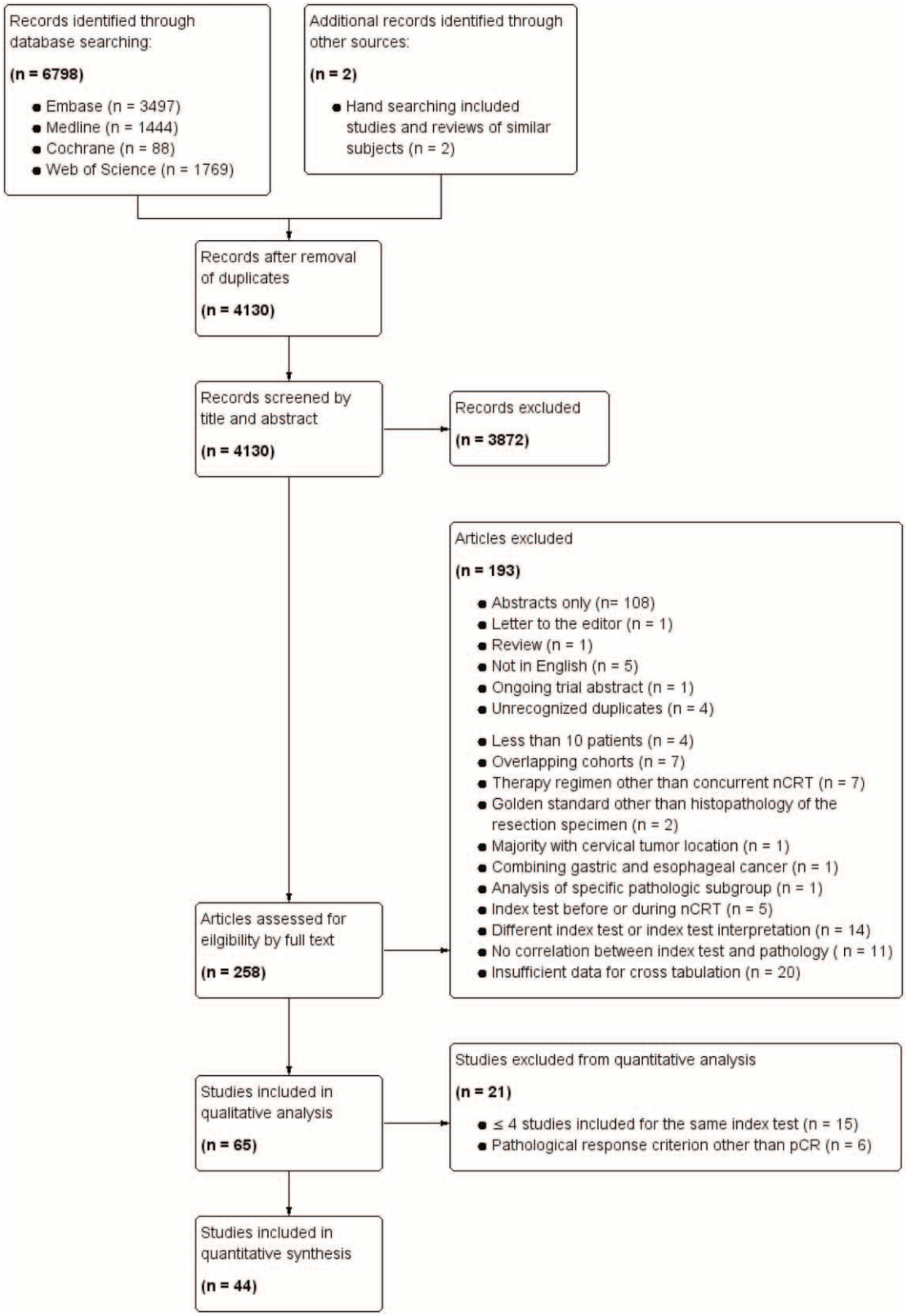

FIGURE 1. Flowchart summarizing search results and study selection.

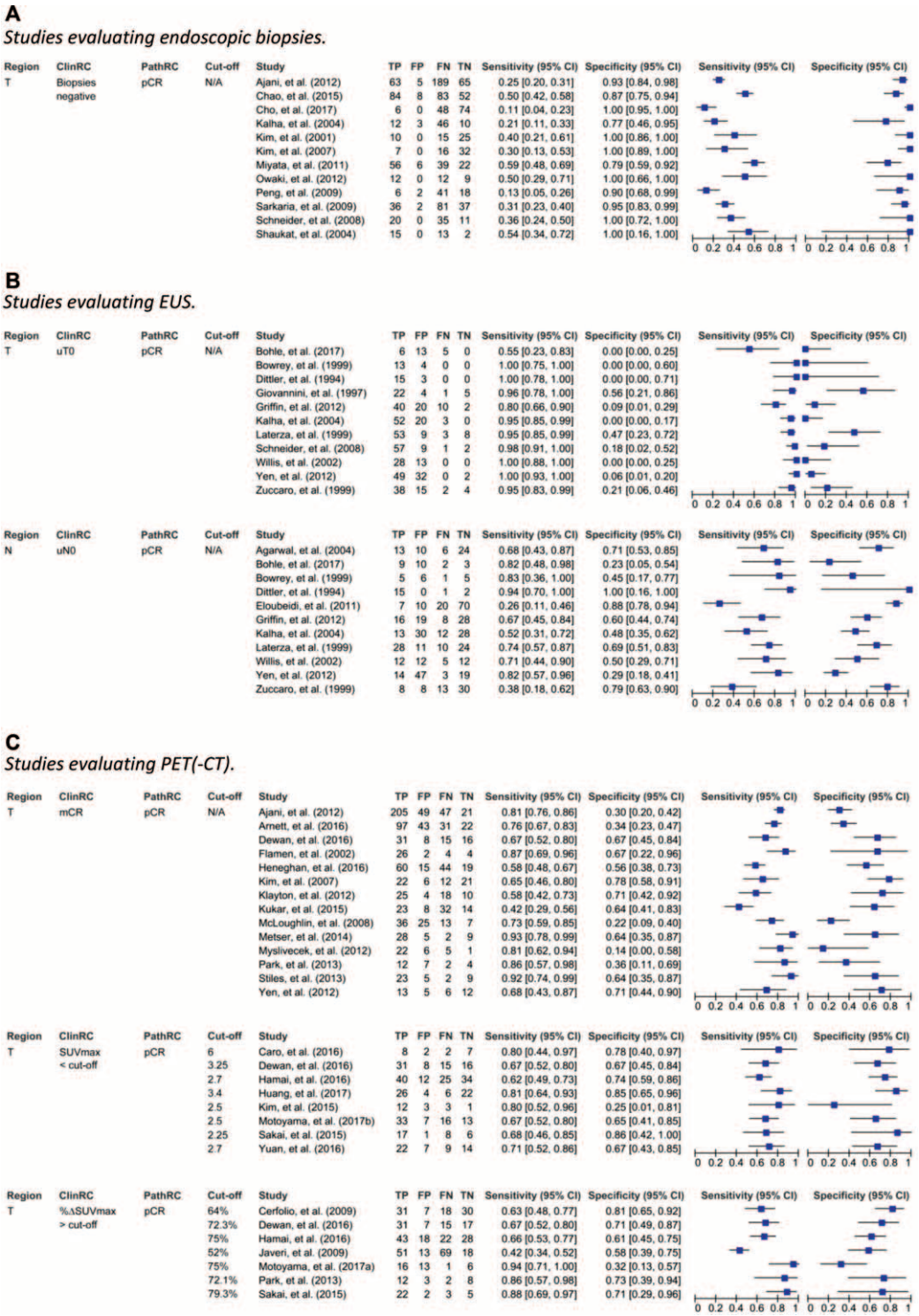

FIGURE 2. Forest plots demonstrating sensitivities and specificities from individual studies included in quantitative synthesis. %ΔSUVmax indicates percentage reduction of SUVmax; ClinRC, clinical response criterion; EUS, endoscopic ultrasonography; mCR, metabolically complete response; N, lymph nodes; N/A, not applicable; PathRC, pathological response criterion; pCR, pathologically complete response; PET(-CT), positron emission tomography with or without computed tomography; SUVmax, maximum standardized uptake value; T, primary tumor; uN0, ultrasonographic nodal stage 0; uT0, ultrasonographic tumor stage 0. (A) Studies evaluating endoscopic biopsies. (B) Studies evaluating EUS. (C) Studies evaluating PET(-CT).

Therefore, PPV and NPV were calculated with a prevalence of 66% for having residual esophageal cancer after nCRT.

Subgroup analyses were performed to investigate sources of heterogeneity (histology, definition of pCR, PET imaging technique, and cutoff of quantitative PET parameters) by extending bivariate models with covariates for both logit sensitivity and logit specificity. Two-sided statistical significance level was set at $P < 0.050$. Statistical analyses were performed in STATA version 15.1 (StataCorp LLC, College Station, TX) and were performed in accordance with the Cochrane Handbook for Diagnostic Test Accuracy Meta-Analysis.[18] Further explanation of statistical analysis is provided in Supplementary information 1, http://links.lww.com/SLA/B663.

RESULTS

Eligible Studies

The systematic literature search identified 4130 records after deduplication. The inclusion criteria were met by title and abstract in 258 records. After full-text review, 65 articles comprising one or more index tests of interest were included for qualitative analysis (Fig. 1).[19–83] Endoscopic biopsies were evaluated by 13 articles, EUS was evaluated by 16 articles, and PET(-CT) by 40 articles. Twenty-one studies were excluded from quantitative synthesis because a pathological response criterion other than pCR was used or because <4 studies were included that evaluated the same index test or the same combination of index tests. Data of studies excluded from quantitative synthesis are shown in Supplementary Information 2 and 3, Supplementary Tables 2–5 and Supplementary Figures 1 and 2, http://links.lww.com/SLA/B663. Forty-four studies were included for quantitative synthesis, comprising 6 index test modalities.[19–21,23,24,27,28,30,32–38,40–43,45–51,55–63,67,68,71,72,76,80–83]

Endoscopy With Biopsies

Twelve studies comprising a total of 1328 patients evaluated endoscopic biopsies for detecting any residual disease at the primary tumor site as positive biopsy versus negative biopsies (Supplementary Table 2, http://links.lww.com/SLA/B663).[20,30,32,45,46,48,57,61,63,68,71,72] Three out of 12 studies were prospective studies.[46,48,71] Patients' median age ranged from 50 to 63 years, the majority was male (89.3%), and more than half had squamous cell carcinoma (55.7%). Patients received concurrent nCRT with a total radiation dose ranging from 30 to 50.4 Gy. Chemotherapy regimens were based on a fluoropyrimidine with a platinum compound in 11 of 12 studies. Intervals between the end of nCRT and endoscopy with biopsies ranged from 7 to 42 days and intervals from endoscopy with biopsies to surgery ranged from <10 days to 42 days. Only 2 of 12 studies reported on the number of biopsies that were taken, which had a median of 4 biopsies.[71,72] Locations of biopsies were not reported. High risk of bias was present in 6 studies that evaluated endoscopy with biopsies.[45,57,61,63,68,71] One of these studies had a high risk of bias in 2 domains (Supplementary Information 2, Supplementary Figure 1A, http://links.lww.com/SLA/B663).[63]

Sensitivity and specificity of positive versus negative biopsies to detect any residual disease at the primary tumor site ranged from 0.11 to 0.59 and from 0.77 to 1.00, respectively (Fig. 2A). The summary operating point consisted of a pooled sensitivity of 0.33 (95% CI, 0.24–0.43) and a pooled specificity of 0.95 (95% CI, 0.88–0.98) (Fig. 3A, Table 1). There was a higher variability for sensitivity than for specificity. The forest plot (Fig. 2A) and the 95% prediction region in the SROC plot (Fig. 3A) demonstrated substantial heterogeneity between studies. Computed PPV was 0.92 (95% CI, 0.83–0.97) and NPV was 0.42 (95% CI, 0.39–0.45) (Table 1). For studies evaluating endoscopic biopsies for detecting any residual disease at the primary tumor site, histology (>80% adenocarcinoma vs >80% squamous cell carcinoma) and definition of pCR (ypT0 vs ypT0N0) had no significant impact on diagnostic performance (Table 2A).

EUS

For EUS, 13 studies were included in quantitative synthesis (Supplementary Table 3, http://links.lww.com/SLA/B663).[19,23,24,34,35,37,38,45,51,71,80,81,83] Of these, 11 studies comprising a total of 563 patients evaluated qualitative EUS for residual disease at the primary tumor site.[23,24,34,37,38,45,51,71,80,81,83] Another 11 studies comprising a total of 629 patients evaluated qualitative EUS for regional lymph nodes.[19,23,24,34,35,38,45,51,80,81,83] Three out of 13 studies were prospective studies.[51,71,80] Patients' median age ranged from 55 to 62 years, the majority was male (86.0%), and less than half had squamous cell carcinoma (42.8%). Patients received concurrent nCRT with a total radiation dose ranging from 30 to 50.4 Gy. Chemotherapy regimens were based on a fluoropyrimidine with a platinum compound in 10 of 13 studies. Intervals between the end of nCRT and EUS ranged from <14 days to 42 days and intervals from EUS to surgery ranged from 6 to 22 days. Six of 11 studies that evaluated EUS for regional lymph nodes reported on whether or not fine-needle aspiration (FNA) was used.[23,35,45,80,81,83] Of these, only one used FNA for cytological confirmation of positive lymph nodes.[35] High risk of bias was present in 7 studies that evaluated EUS for the primary tumor site,[24,37,38,45,51,71,83] and in 7 studies that evaluated EUS for regional lymph nodes.[23,24,35,38,45,51,83] None of these had a high risk of bias in more than one domain (Supplementary Information 2, Supplementary Figure 1B, http://links.lww.com/SLA/B663).

Sensitivity and specificity of qualitative EUS uT+ versus uT0 for detecting any residual disease at the primary tumor site ranged from 0.55 to 1.00 and from 0.00 to 0.56, respectively (Fig. 2B). The summary operating point consisted of a pooled sensitivity of 0.96 (95% CI, 0.89–0.99) and a pooled specificity of 0.08 (95% CI, 0.03–0.24) (Fig. 3B, Table 1). There was a slightly higher variability for specificity than for sensitivity. The forest plot (Fig. 2B) and the 95% prediction region in the SROC plot (Fig. 3B) demonstrated substantial heterogeneity between studies. Computed PPV was 0.67 (95% CI, 0.65–0.70) and NPV was 0.51 (95% CI, 0.19–0.85) (Table 1). Sensitivity and specificity of qualitative EUS uN+ versus uN0 for detecting any residual nodal disease ranged from 0.26 to 0.94 and from 0.23 to 1.00, respectively (Fig. 2C). The summary operating point consisted of a pooled sensitivity of 0.68 (95% CI, 0.54–0.80) and a pooled specificity of 0.57 (95% CI, 0.43–0.70) (Fig. 3C, Table 1). Variability for sensitivity and specificity was comparable. The forest plot (Fig. 2C) and the 95% prediction region in the SROC plot (Fig. 3C) demonstrated substantial heterogeneity between studies. Computed PPV was 0.75 (95% CI, 0.71–0.79) and NPV was 0.48 (95% CI, 0.41–0.55) (Table 1). For studies evaluating qualitative EUS for residual nodal disease, histology had a significant impact on sensitivity ($P = 0.0138$) (Table 2B). Sensitivity was 0.52 (95% CI, 0.35–0.69) for studies including >80% adenocarcinoma versus 0.81 (95% CI, 0.67–0.90) for studies including >80% squamous cell carcinoma. Corresponding specificities were 0.68 (95% CI, 0.50–0.82) for studies including >80% adenocarcinoma versus 0.52 (95% CI, 0.23–0.79) for studies including >80% squamous cell carcinoma, but did not significantly differ ($P = 0.2301$). For studies evaluating qualitative EUS for the primary tumor site, histology had no significant impact on diagnostic performance.

PET(-CT)

For PET(-CT), 24 studies were included in quantitative synthesis (Supplementary Table 4, http://links.lww.com/SLA/B663).[20,21,27,28,33,36,40–43,47–50,55,56,58–60,62,67,76,81,82] Of these, 14 studies comprising a total of 1213 patients evaluated qualitative

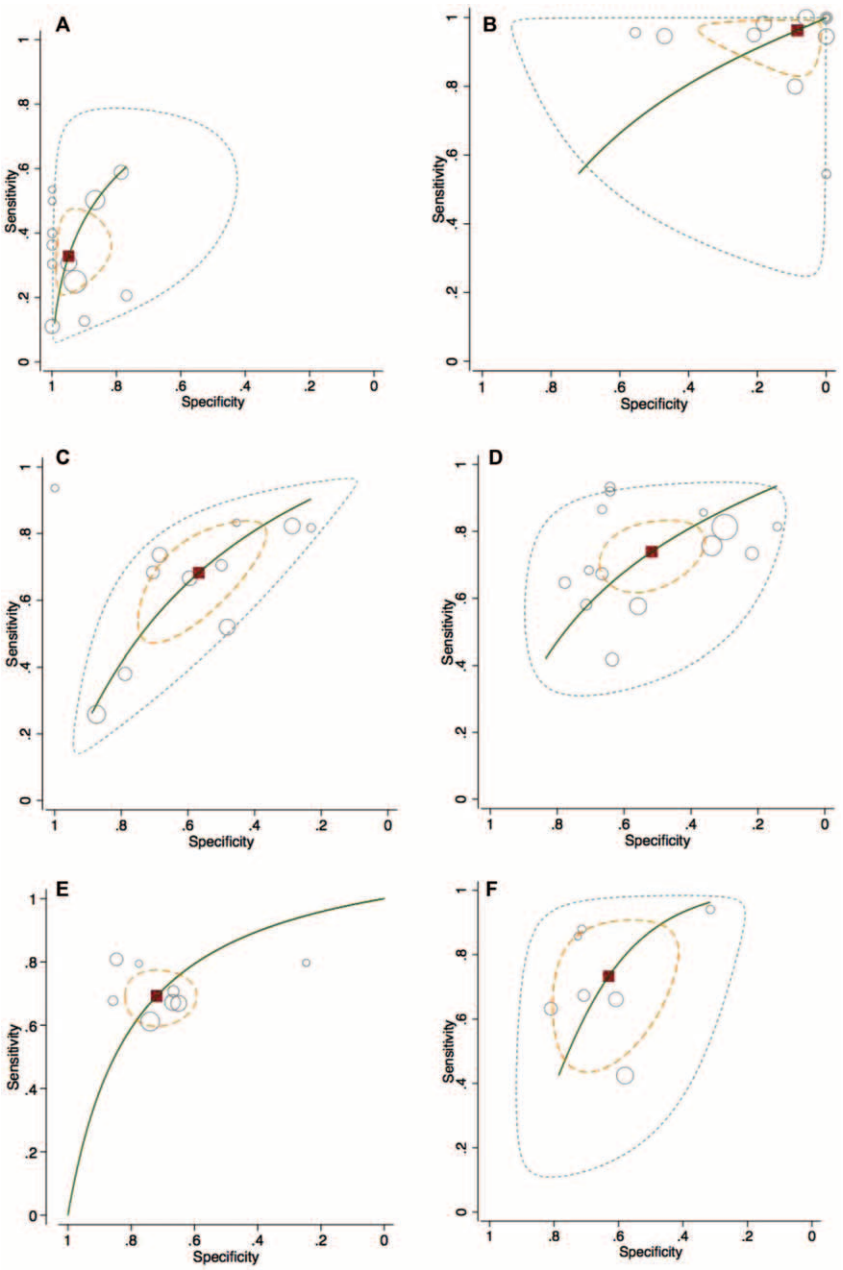

FIGURE 3. Summary receiver operating characteristic (SROC) plots demonstrating the diagnostic performance of index tests. The grey circles represent the individual studies and sizes. Summary operating points (red block) along with 95% confidence regions (orange dotted lines) are added to the SROC plots to reflect average observed accuracy and 95% prediction regions (grey dotted lines) are added to demonstrate between-study heterogeneity. The continuous green line presents the SROC curve. %ΔSUVmax indicates percentage reduction of SUVmax; EUS, endoscopic ultrasonography; mCR, metabolically complete response; pCR, pathologically complete response; PET(-CT), positron emission tomography with or without computed tomography; SUVmax, maximum standardized uptake value; uN0, ultrasonographic nodal stage 0; uT0, ultrasonographic tumor stage 0. (A) SROC of endoscopic biopsies evaluating the primary tumor site using negative biopsies as clinical response criterion and pCR as pathological response criterion. (B) SROC of EUS evaluating the primary tumor site using uT0 as clinical response criterion and pCR as pathological response criterion. (C) SROC of EUS evaluating lymph nodes using uN0 as clinical response criterion and pCR as pathological response criterion. (D) SROC of PET(-CT) evaluating the primary tumor site using mCR as clinical response criterion and pCR as pathological response criterion. (E) SROC of PET(-CT) evaluating the primary tumor site using SUVmax $<$ cutoff as clinical response criterion and pCR as pathological response criterion. (F) SROC of PET(-CT) evaluating the primary tumor site using %ΔSUVmax $>$ cutoff as clinical response criterion and pCR as pathological response criterion.

TABLE 1. Summary Diagnostic Performance of Studies Included in Quantitative Synthesis

Index Test	Region	ClinRC	PathRC	Median Cutoff (range)	No. Studies	Summary Sensitivity (95% CI)	Summary Specificity (95% CI)	Computed PPV*	Computed NPV*
Endoscopic biopsies	T	Biopsies negative	pCR	N/A	12	0.33 (0.24–0.43)	0.95 (0.88–0.98)	0.92 (0.83–0.97)	0.42 (0.39–0.45)
EUS	T	uT0	pCR	N/A	11	0.96 (0.89–0.99)	0.08 (0.03–0.24)	0.67 (0.65–0.70)	0.51 (0.19–0.85)
EUS	N	uN0	pCR	N/A	11	0.68 (0.54–0.80)	0.57 (0.43–0.70)	0.75 (0.71–0.79)	0.48 (0.41–0.55)
PET(-CT)	T	mCR	pCR	N/A	14	0.74 (0.68–0.79)	0.52 (0.44–0.60)	0.75 (0.70–0.79)	0.51 (0.43–0.58)
PET(-CT)	T	SUVmax < cutoff	pCR	2.7 (2.25–6)	8	0.69 (0.64–0.74)	0.72 (0.64–0.78)	0.83 (0.79–0.86)	0.55 (0.50–0.60)
PET(-CT)	T	%ΔSUVmax > cutoff	pCR	72.3% (52%–79.3%)	7	0.73 (0.57–0.85)	0.63 (0.51–0.74)	0.79 (0.74–0.84)	0.55 (0.42–0.67)

%ΔSUVmax indicates percentage reduction of SUVmax; ClinRC, clinical response criterion; EUS, endoscopic ultrasound; mCR, metabolically complete response; N, lymph nodes; N/A, not applicable; NPV, negative predictive value; PathRC, pathological response criterion; pCR, pathologically complete response; PET(-CT), positron emission tomography with or without computed tomography; PPV, positive predictive value; SUVmax, maximum standardized uptake value; T, primary tumor; uN0, ultrasonographic nodal stage 0; uT0, ultrasonographic tumor stage 0.

*PPV and NPV were computed with the summary sensitivity, summary specificity and a prevalence of 66% for having residual disease after nCRT.

PET as metabolically non-complete response versus metabolically complete response (mCR) for residual disease at the primary tumor site.[20,21,33,36,41,48–50,55,56,60,62,76,81] Another 8 studies comprising a total of 430 patients evaluated quantitative PET for the primary tumor site by using maximum standardized uptake value (SUVmax),[27,33,40,42,47,59,67,82] and 7 studies comprising a total of 511 patients by using percentage reduction of SUVmax (%ΔSUVmax).[28,33,40,43,58,62,67] Four out of 24 studies were prospective.[27,33,36,48] Patients' median age ranged from 55 to 67 years, the majority was male (83.7%) and less than half had squamous cell carcinoma (48.8%). Patients received concurrent nCRT with a total radiation dose ranging from 36 to 50.4 Gy. Chemotherapy regimens were based on a fluoropyrimidine with a platinum compound in 21 of 24 studies. Intervals between the end of nCRT and PET ranged from 14 to 52 days and intervals from PET to surgery ranged from 2 to 42 days. For studies evaluating SUVmax, the median cutoff was 2.7 (range 2.25–6.0). For %ΔSUVmax, median cutoff was 72.3% (range 52%–79.3%). High risk of bias was present in 7 studies that evaluated qualitative PET.[20,21,33,50,55,56,60] Two of these studies had a risk of bias in 2 domains.[33,50] High risk of bias was present in 7 studies that evaluated PET-SUVmax.[27,33,40,42,47,67,82] Four of these studies had a risk of bias in ≥2 domains.[27,33,67,82] High risk of bias was present in 4 studies that evaluated PET-%ΔSUVmax.[28,33,40,67] Three of these studies had a risk of bias in 2 domains (Supplementary Information 2, Supplementary Figure 1C, http://links.lww.com/SLA/B663).[28,33,67]

Sensitivity and specificity of qualitative PET metabolically noncomplete response versus mCR for detecting any residual disease at the primary tumor site ranged from 0.42 to 0.93 and from 0.14 to 0.78, respectively (Fig. 2D). The summary operating point consisted of a pooled sensitivity of 0.74 (95% CI, 0.68–0.79) and a pooled specificity of 0.52 (95% CI, 0.44–0.60) (Fig. 3D, Table 1). There was a higher variability for specificity than for sensitivity. The forest plot (Fig. 2D) and the 95% prediction region in the SROC plot (Fig. 3D) demonstrated substantial heterogeneity between studies. Computed PPV was 0.75 (95% CI, 0.70–0.79) and NPV was 0.51 (95% CI, 0.43–0.58) (Table 1). Sensitivity and specificity of quantitative PET-SUVmax for detecting any residual disease at the primary tumor site ranged from 0.62 to 0.80 and from 0.25 to 0.86, respectively (Fig. 2D). Cutoffs ranged from 2.5 to 6.0. The summary operating point consisted of a pooled sensitivity of 0.69 (95% CI, 0.64–0.74) and a pooled specificity of 0.72 (95% CI, 0.64–0.78) (Fig. 3E, Table 1). Variability for sensitivity and specificity was comparably low. Besides the outlying specificity of Kim et al, the forest plot demonstrated low heterogeneity between studies (Fig. 2D).[47] The low heterogeneity leads to identical 95% confidence and 95% prediction regions. Computed PPV was 0.83 (95% CI, 0.79–0.86) and NPV was 0.55 (95% CI, 0.50–0.60) (Table 1). Sensitivity and specificity of quantitative PET-%ΔSUVmax for detecting residual disease at the primary tumor site ranged from 0.42 to 0.94 and from 0.32 to 0.81, respectively (Fig. 2D). Cutoffs ranged from 52% to 79.3%. The summary operating point consisted of a pooled sensitivity of 0.73 (95% CI, 0.57–0.85) and a pooled specificity of 0.63 (95% CI, 0.51–0.74) (Fig. 3F, Table 1). Variability was higher for sensitivity than for specificity. The forest plot (Fig. 2D) and the 95% prediction region in the SROC plot (Fig. 3F) demonstrated substantial heterogeneity between studies. Computed PPV was 0.79 (95% CI, 0.74–0.84) and NPV was 0.55 (95% CI, 0.42–0.67) (Table 1). For studies evaluating quantitative PET-%ΔSUVmax for detecting any residual disease at the primary tumor site, histology had a significant impact on sensitivity ($P = 0.0403$) (Table 2C). Sensitivity was 0.43 (95% CI, 0.34–0.51) for studies including >80% adenocarcinoma versus 0.80 (95% CI, 0.64–0.90) for studies including >80% squamous cell carcinoma. Corresponding specificities were 0.58 (95% CI, 0.40–0.74) for studies including >80% adenocarcinoma versus 0.57 (95% CI, 0.44–0.70) for studies including >80% squamous cell carcinoma, but did not significantly differ ($P = 0.9662$). For the other PET modalities, subgroups had no significant impact on diagnostic performance.

DISCUSSION

This systematic review and meta-analysis suggest that endoscopic biopsies, qualitative EUS, qualitative PET(-CT), and quantitative PET(-CT) with SUVmax or %ΔSUVmax as single modalities can correctly identify residual esophageal cancer at the primary tumor site after nCRT with summary sensitivities of 33%, 96%, 74%, 69%, and 73%, respectively. Corresponding summary specificities for correctly identifying a complete response were 95%, 8%, 52%, 72%, and 63%, respectively. Qualitative EUS can correctly identify residual nodal disease after nCRT with a sensitivity of 68% and can identify complete response with a specificity of 57%.

TABLE 2. Results From Study-level Subgroup Analyses

A. Studies Evaluating Endoscopic Biopsies

Region	ClinRC	PathRC	Parameter	Category	N	*P* Value for Global Effect	Sensitivity (95% CI)	*P*	Specificity (95% CI)	*P*
T	Biopsies negative	pCR	Histology	>80% AC	4	0.0129	0.25 (0.15–0.40)	0.3315	0.91 (0.81–0.96)	0.1847
				>80% SCC	6		0.36 (0.24–0.52)		0.99 (0.72–1.00)	
			Definition of pCR	ypT0	6	0.5590	0.32 (0.18–0.50)	0.2841	0.94 (0.84–0.98)	0.7766
				ypT0N0	2		0.20 (0.07–0.44)		0.97 (0.69–1.00)	

B. Studies Evaluating EUS

Region	ClinRC	PathRC	Parameter	Category	N	*P* Value for Global Effect	Sensitivity (95% CI)	*P*	Specificity (95% CI)	*P*
T	uT0	pCR	Histology	>80% AC	2	0.0098	0.88 (0.71–0.95)	0.0708	0.04 (0.01–0.20)	0.3082
				>80% SCC	3		0.98 (0.90–1.00)		0.18 (0.05–0.49)	
			Definition of pCR	ypT0	11	N/A	0.96 (0.89–0.99)	N/A	0.08 (0.03–0.24)	N/A
				ypT0N0	0		N/A		N/A	
N	uN0	pCR	Histology	>80% AC	4	0.0456	0.52 (0.35–0.69)	0.0138	0.68 (0.50–0.82)	0.2301
				>80% SCC	3		0.81 (0.67–0.90)		0.52 (0.23–0.79)	
			Definition of pCR	ypN0	11	N/A	0.68 (0.54–0.80)	N/A	0.57 (0.43–0.70)	N/A
				ypT0N0	0		N/A		N/A	

C. Studies Evaluating PET (-CT)

Region	ClinRC	PathRC	Parameter	Category	N	*P* Value for Global Effect	Sensitivity (95% CI)	*P*	Specificity (95% CI)	*P*
T	mCR	pCR	Histology	>80% AC	4	0.0540	0.61 (0.47–0.74)	0.5848	0.56 (0.42–0.69)	0.1954
				>80% SCC	4		0.76 (0.62–0.86)		0.38 (0.25–0.54)	
			Definition of pCR	ypT0	3	0.3917	0.76 (0.51–0.90)	0.5042	0.63 (0.49–0.75)	0.2423
				ypT0N0	6		0.77 (0.72–0.82)		0.46 (0.26–0.67)	
			Imaging technique	PET	3	0.1839	0.86 (0.77–0.92)	0.0662	0.49 (0.26–0.72)	0.5434
				PET-CT	9		0.70 (0.58–0.79)		0.53 (0.42–0.65)	
T	SUVmax < cutoff	pCR	Histology	>90% AC	0	N/A	N/A	N/A	N/A	N/A
				>90% SCC	6		0.69 (0.64–0.74)		0.72 (0.64–0.78)	
			Definition of pCR	ypT0	5	0.4660	0.71 (0.61–0.80)	0.8145	0.77 (0.66–0.85)	0.2224
				ypT0N0	2		0.69 (0.58–0.78)		0.66 (0.50–0.79)	
			Imaging technique	PET	0	N/A	N/A	N/A	N/A	N/A
				PET-CT	7		0.69 (0.63–0.75)		0.71 (0.64–0.78)	
T	%ΔSUVmax > cutoff	pCR	Histology	>80% AC	1	0.0460	0.43 (0.34–0.51)	0.0403	0.58 (0.40–0.74)	0.9662
				>80% SCC	5		0.80 (0.64–0.90)		0.57 (0.44–0.70)	
			Definition of pCR	ypT0	2	0.7072	0.74 (0.58–0.85)	0.6514	0.64 (0.53–0.73)	0.9131
				ypT0N0	2		0.76 (0.60–0.87)		0.66 (0.43–0.83)	

TABLE 2. (*Continued*)

C. Studies Evaluating PET (-CT)

Region	ClinRC	PathRC	Parameter	Category	N	*P* Value for Global Effect	Sensitivity (95% CI)	*P*	Specificity (95% CI)	*P*
			Imaging technique	PET	0	N/A	N/A	N/A	N/A	N/A
				PET-CT	5		0.70 (0.52–0.83)		0.63 (0.50–0.74)	

%ΔSUVmax indicates percentage reduction of SUVmax; ClinRC, clinical response criterion; EUS, endoscopic ultrasound; mCR, metabolically complete response; N, lymph nodes; N/A, not applicable; PathRC, pathological response criterion; pCR, pathologically complete response; PET, positron emission tomography with or without computed tomography; SUVmax, maximum standardized uptake value; T, primary tumor; uN0, ultrasonographic nodal stage 0; uT0, ultrasonographic tumor stage 0.

In the light of an active surveillance strategy, sensitivity is an important diagnostic parameter because FN results cause delay in detecting residual disease. This delay allows for tumor growth and potential distant dissemination, jeopardizing oncological safety. However, corresponding specificity has its importance as well. As the number of FP increases and therewith specificity decreases, more patients will be incorrectly classified as having residual disease. Consequently, patients in an active surveillance program might be unnecessarily exposed to operative risks. Considering this, endoscopic biopsies, EUS, qualitative PET(-CT), and quantitative PET(-CT) with SUVmax or %ΔSUVmax seem insufficiently accurate for individually detecting residual disease at the primary tumor site after nCRT. EUS with uN0 as clinical response criterion seems also insufficiently accurate for detecting residual nodal disease.

For the quantitative synthesis in this study, only studies using pathologically complete response as pathological response criterion were included, which reflected the actual accuracy of index tests. However, it is debatable how accurate index tests should be to safely perform an active surveillance strategy. Although ideally the smallest amount of residual disease should be detected, microscopic residue is often missed during preoperative clinical response evaluations in current clinical practice. Yet, existing studies show no decline in oncological outcome for patients who underwent active surveillance with similar diagnostic tests (ie, endoscopic biopsies and PET(-CT) instead of standard esophagectomy after nCRT).[84,85] This might be explained by regrowth of microscopic residual disease to a detectable and still resectable amount of tumor during active surveillance, resulting in oncological outcomes similar to immediate resection. Consequently, patients might undergo postponed radical resection with comparable oncological outcomes. Moreover, the known decrease in health-related quality of life after esophagectomy as well as high postoperative morbidity and mortality rates should be taken into account when considering an active surveillance strategy. A discrete choice experiment showed that in a hypothetical situation patients with esophageal cancer are willing to trade off 16% 5-year overall survival to reduce the risk of an esophagectomy from 100% (standard surgical treatment after nCRT) to 35% (active surveillance in case of cCR after nCRT). Moreover, it should be taken into consideration whether patients have squamous cell carcinoma or adenocarcinoma. After primary surgery, patients with squamous cell carcinoma have a higher risk of locoregional recurrence than patients with adenocarcinoma. This might suggest that the level of error for detecting residual disease in patients with squamous cell carcinoma should be lower. On the contrary, squamous cell carcinoma tends to respond better to nCRT than adenocarcinoma, resulting in comparable locoregional recurrence rates if patients are treated with nCRT followed by surgery.[86] For example, pCR rate for patients receiving nCRT according to the CROSS regimen is 23% for adenocarcinoma compared with 49% for squamous cell carcinoma.[87] These patients with squamous cell carcinoma might benefit more from an organ-preserving strategy than patients with adenocarcinoma. Moreover, after nCRT for squamous cell carcinoma less patients have TRG3–4 residual tumor. It could, therefore, be argued that the level of error for missing TRG3–4 residual tumor can be higher for patients with squamous cell carcinoma. Providing an exact level of error for detecting residual disease after nCRT has disadvantages because there are many factors that have to be taken into consideration. Only well-designed prospective trials that compare immediate surgery with active surveillance for patients with cCR after nCRT that take all these issues into account can provide more evidence on the acceptable level of error for this complex concept.

For endoscopic biopsies, EUS and PET(-CT), respectively, 1, 6, and 23 clinical response criteria were excluded from quantitative syntheses. Interestingly, some studies excluded from quantitative synthesis show promising results. One study quantified EUS measurements as maximum tumor thickness after nCRT (yMTT).[44] Although feasibility has yet to be confirmed, this method showed a favorable sensitivity (0.86) and specificity (0.64). The ratio of maximum tumor thickness after and before nCRT (yMTT/MTT) also showed favorable sensitivity (0.79) and specificity (0.82).[44] Also, several quantitative PET measurements showed promising results. Percentage decrease in tumor length (sensitivity of 0.92 and specificity of 0.90), percentage reduction of standardized uptake value of tumor volume (sensitivity of 0.70 and specificity of 0.95), percentage reduction of PET area (sensitivity of 0.93 and specificity of 0.68), percentage reduction of standardized uptake value of tumor area (sensitivity of 1.00 and specificity of 0.68), percentage reduction of tumor diameter (sensitivity of 0.89 and specificity of 0.91), and percentage reduction of diameter multiplied by standardized uptake value of tumor area with cutoff 56% (sensitivity of 0.93 and 0.91) all showed good accuracy. However, results should be confirmed because these studies were performed in one hospital with overlapping cohorts.[64–66] Moreover, combining index test modalities to obtain an optimal set for response evaluation can improve diagnostic accuracy.[39,48,53]

Several limitations were present in the included studies. According to the QUADAS-2 tool, most studies were of low quality. The majority was retrospectively designed and had insufficient statistical power. Furthermore, most studies did not determine the optimal combination of tests for response evaluation, but investigated index tests separately. Also, the clinical investigations had not been aimed to detect residual disease for distinguishing between patients who might benefit from active surveillance and patients who might not. Because of this lack of clinical focus of the clinical operator, accuracy may not have been optimal for response evaluations for a future active surveillance strategy.

The present study also has several limitations. First, substantial between-study heterogeneity existed. Studies included in the meta-analyses comprised many variables, which have led to a wide range of sensitivity and specificity values. Interpreting the pooled

data with wide confidence intervals should therefore be done with caution. Second, the majority of the studies included in the qualitative analysis and all of the studies included in the quantitative synthesis investigated a single modality. To identify patients with residual disease after nCRT, a combination of modalities would probably have better diagnostic performance. Moreover, because most studies investigated a single modality, between-study heterogeneity is potentially greater which could complicate comparison of modalities. Third, studies of low quality based on the QUADAS-2 were not excluded because well-designed, sufficiently powered, prospective studies on this topic are scarce. Fourth, different histological subtypes and a wide range of tumor and nodal stages were initially analyzed together. It has been shown that patients with squamous cell carcinoma tend to respond better to nCRT than patients with adenocarcinoma, making residual disease less likely.[87] Moreover, patients with lower T and N stages have a higher chance of having pCR. In contrast to positive and negative predictive values, however, sensitivity and specificity are not influenced by prevalence of disease. Subgroup analyses were performed on histological subtypes (>80% adenocarcinoma vs >80% squamous cell carcinoma) to investigate heterogeneity on study level. Sensitivity values of EUS for detecting residual nodal disease and of PET-%ΔSUVmax for detecting any residual disease at the primary tumor were better for squamous cell carcinoma than for adenocarcinoma. This suggests that histology might have had a significant impact on diagnostic performance for studies evaluating these diagnostic tests. Fifth, studies with different definitions for discriminating pathological responders from nonresponders were included. Most studies used pathologically complete versus incomplete response to discriminate between both groups. Studies that used similar pathological response criteria were redefined to pathologically complete versus incomplete response. Some studies defined pCR as ypT0 and some as ypT0N0. To investigate the influence of these different interpretations of pCR, subgroup analyses were performed. No sources of heterogeneity were found in the definition of pCR subgroup. However, subgroups consisted of <10 studies and therefore results should be interpreted with caution.

Future studies should focus on further improving diagnostic accuracy of clinical response evaluation, thereby decreasing the number of patients potentially endangered by unresectable regrowth after FN response evaluation and decreasing the number of patients with unnecessary surgical resections after FP response evaluation. The recently published prospective diagnostic preSANO trial showed that combining diagnostic tests can improve detection of residual disease.[88] In this study, a combination of PET-CT, endoscopy with bite on-bite biopsies and EUS with fine-needle aspiration of suspected lymph nodes detected TRG3–4 tumors with a sensitivity of 90%. In addition, PET-CT detected interval metastases in 10% of patients, which prevented unnecessary surgical resection. This diagnostic strategy is currently tested in the Dutch randomized phase III SANO trial, which compares active surveillance with standard resection in patients with a clinically complete response after nCRT.[5] In the preSANO trial, sensitivity of detecting TRG2–4 residual tumor improved from 54% to 77% by using bite-on-bite biopsies. This is most likely because residual disease is often found in the submucosal layer, whereas the mucosal layer is free of residual tumor.[89] Alternative esophageal sampling techniques such as wide-area transepithelial sampling for wider and deeper sampling and the Cytosponge for entire esophageal sampling might lower false-negative rates as well.[90,91] Prolonging the interval between nCRT and response evaluation could potentially increase accuracy as well. Twelve weeks after completing nCRT, radiation-induced esophagitis still causes noise on PET-CT. Because inflammation of the esophageal wall and therewith noise decreases over time, serial quantitative PET-CT will most likely be of considerably additional value during follow-up as the 18F-FDG signal is expected to increase during tumor regrowth. Moreover, additional imaging techniques such as dynamic contrast-enhanced MRI, diffusion-weighted MRI, and more advanced quantitative PET analyses could further increase accuracy of response evaluations.[92,93] Also, other novel techniques such as multianalyte blood tests and circulating cell-free tumor DNA (liquid biopsies) might in future prove of additional value in evaluating response to nCRT and detecting disease recurrence early during active surveillance.[94,95]

In conclusion, current literature suggests that endoscopy with biopsies, endoscopic ultrasonography, or 18F-FDG PET(-CT) as single modalities are moderately accurate for detecting locoregional residual esophageal cancer after neoadjuvant chemoradiotherapy. These accuracies are regarded insufficient to direct therapeutic management in individual patients.

ACKNOWLEDGMENTS

The authors thank Mrs. Gerdien B. de Jonge (Medical Library, Erasmus MC - University Medical Center, Rotterdam, Netherlands) for her assistance with the literature search.

REFERENCES

1. Shapiro J, van Lanschot JJB, Hulshof M, et al. Neoadjuvant chemoradiotherapy plus surgery versus surgery alone for oesophageal or junctional cancer (CROSS): long-term results of a randomised controlled trial. *Lancet Oncol*. 2015;16:1090–1098.
2. Yang H, Liu H, Chen Y, et al. Neoadjuvant chemoradiotherapy followed by surgery versus surgery alone for locally advanced squamous cell carcinoma of the esophagus (NEOCRTEC5010): a phase III multicenter, randomized, open-label clinical trial. *J Clin Oncol*. 2018;36:2796–2803.
3. Djarv T, Lagergren J, Blazeby JM, et al. Long-term health-related quality of life following surgery for oesophageal cancer. *Br J Surg*. 2008;95:1121–1126.
4. Noordman BJ, Verdam MGE, Lagarde SM, et al. Effect of neoadjuvant chemoradiotherapy on health-related quality of life in esophageal or junctional cancer: results from the randomized CROSS trial. *J Clin Oncol*. 2018;36:268–275.
5. Noordman BJ, Wijnhoven BPL, Lagarde SM, et al. Neoadjuvant chemoradiotherapy plus surgery versus active surveillance for oesophageal cancer: a stepped-wedge cluster randomised trial. *BMC Cancer*. 2018;18:142.
6. Moher D, Liberati A, Tetzlaff J, et al. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *BMJ*. 2009;339:b2535.
7. Eyck BM, Onstenk BD, Noordman BJ, et al. Accuracy of detecting residual disease after neoadjuvant chemoradiotherapy for esophageal cancer: a systematic review and meta-analysis. PROSPERO 2018 CRD42018116649 Available at: http://www.crd.york.ac.uk/PROSPERO/display_record.php?ID=CRD42018116649. Accessed December 4, 2018.
8. Tong DK, Law S, Kwong DL, et al. Histological regression of squamous esophageal carcinoma assessed by percentage of residual viable cells after neoadjuvant chemoradiation is an important prognostic factor. *Ann Surg Oncol*. 2010;17:2184–2192.
9. Society JE. Guidelines for clinical and pathologic studies on carcinoma of the esophagus, ninth edition: Preface, general principles, part I. *Esophagus*. 2004;1:61–88.
10. Schneider PM, Baldus SE, Metzger R, et al. Histomorphologic tumor regression and lymph node metastases determine prognosis following neoadjuvant radiochemotherapy for esophageal cancer: implications for response classification. *Ann Surg*. 2005;242:684–692.
11. Miller AB, Hoogstraten B, Staquet M, et al. Reporting results of cancer treatment. *Cancer*. 1981;47:207–214.
12. Mandard AM, Dalibard F, Mandard JC, et al. Pathologic assessment of tumor regression after preoperative chemoradiotherapy of esophageal carcinoma. Clinicopathologic correlations. *Cancer*. 1994;73:2680–2686.
13. Eisenhauer EA, Therasse P, Bogaerts J, et al. New response evaluation criteria in solid tumours: Revised RECIST guideline (version 1.1). *Eur J Cancer*. 2009;45:228–247.
14. Edge SB, Compton CC. The American Joint Committee on Cancer: the 7th edition of the AJCC cancer staging manual and the future of TNM. *Ann Surg Oncol*. 2010;17:1471–1474.

15. 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.
16. Rutter CM, Gatsonis CA. A hierarchical regression approach to meta-analysis of diagnostic test accuracy evaluations. *Stat Med*. 2001;20:2865–2884.
17. Reitsma JB, Glas AS, Rutjes AW, et al. Bivariate analysis of sensitivity and specificity produces informative summary measures in diagnostic reviews. *J Clin Epidemiol*. 2005;58:982–990.
18. Macaskill P, Gatsonis C, Deeks JJ, et al. Chapter 10: Analysing and Presenting Results. In: Deeks JJ, Bossuyt PM, Gatsonis C (editors), Cochrane Handbook for Systematic Reviews of Diagnostic Test Accuracy Version 1.0. The Cochrane Collaboration, 2010. Available at: http://srdta.cochrane.org/. Accessed March 28, 2018.
19. Agarwal B, Swisher S, Ajani J, et al. Endoscopic ultrasound after preoperative chemoradiation can help identify patients who benefit maximally after surgical esophageal resection. *Am J Gastroenterol*. 2004;99:1258–1266.
20. Ajani JA, Correa AM, Hofstetter WL, et al. Clinical parameters model for predicting pathologic complete response following preoperative chemoradiation in patients with esophageal cancer. *Ann Oncol*. 2012;23:2638–2642.
21. Arnett ALH, Merrell KW, Macintosh EM, et al. *Utility of (18)F-FDG PET for predicting histopathologic response in esophageal carcinoma following chemoradiation*. 2016.
22. Arslan N, Miller TR, Dehdashti F, et al. Evaluation of response to neoadjuvant therapy by quantitative 2-deoxy-2-[18F]fluoro-D-glucose with positron emission tomography in patients with esophageal cancer. *Mol Imaging Biol*. 2002;4:301–310.
23. Bohle W, Kasper M, Zoller WG. Prognostic relevance of serial endoscopic ultrasound after chemoradiation in esophageal cancer. *Dis Esophagus*. 2017;30:1–8.
24. Bowrey DJ, Clark GW, Roberts SA, et al. Serial endoscopic ultrasound in the assessment of response to chemoradiotherapy for carcinoma of the esophagus. *J Gastrointest Surg*. 1999;3:462–467.
25. Brucher BLDM, Weber W, Bauer M, et al. Neoadjuvant therapy of esophageal squamous cell carcinoma: response evaluation by positron emission tomography. *Ann Surg*. 2001;233:300–309.
26. Bruzzi JF, Swisher SG, Truong MT, et al. Detection of interval distant metastases: clinical utility of integrated CT-PET imaging in patients with esophageal carcinoma after neoadjuvant therapy. *Cancer*. 2007;109:125–134.
27. Caro M, Font A, Comas S, et al. Preoperative low-dose weekly cisplatin and continuous infusion fluorouracil plus hyperfractionated radiotherapy in stage II–III esophageal carcinoma. *Clin Transl Oncol*. 2016;18:1106–1113.
28. Cerfolio RJ, Bryant AS, Talati AA, et al. Change in maximum standardized uptake value on repeat positron emission tomography after chemoradiotherapy in patients with esophageal cancer identifies complete responders. *J Thorac Cardiovasc Surg*. 2009;137:605–609.
29. Chao YK, Chuang WY, Yeh CJ, et al. Anatomical distribution of residual cancer in patients with oesophageal squamous cell carcinoma who achieved clinically complete response after neoadjuvant chemoradiotherapy. *Eur J Cardiothorac Surg*. 2018;53:201–208.
30. Chao YK, Yeh CJ, Lee MH, et al. Factors associated with false-negative endoscopic biopsy results after neoadjuvant chemoradiotherapy in patients with esophageal squamous cell carcinoma. *Medicine*. 2015;94:e588.
31. Cheedella NKS, Suzuki A, Xiao L, et al. Association between clinical complete response and pathological complete response after preoperative chemoradiation in patients with gastroesophageal cancer: analysis in a large cohort. *Ann Oncol*. 2013;24:1262–1266.
32. Cho CJ, Kang HJ, Park SJ, et al. *A novel endoscopic categorization for prediction of chemoradiotherapy response in locally advanced esophageal cancer*. 2017.
33. Dewan A, Sharma S, Dewan A, et al. Impact on radiological and pathological response with neoadjuvant chemoradiation and its effect on survival in squamous cell carcinoma of thoracic esophagus. *J Gastrointest Cancer*. 2016;1–8.
34. Dittler HJ, Fink U, Siewert GR. Response to chemotherapy in esophageal cancer. *Endoscopy*. 1994;26:769–771.
35. Eloubeidi MA, Cerfolio RJ, Bryant AS, et al. Efficacy of endoscopic ultrasound in patients with esophageal cancer predicted to have N0 disease. *Eur J Cardiothorac Surg*. 2011;40:636–641.
36. Flamen P, Van Cutsem E, Lerut A, et al. Positron emission tomography for assessment of the response to induction radiochemotherapy in locally advanced oesophageal cancer. *Ann Oncol*. 2002;13:361–368.
37. Giovannini M, Seitz JF, Thomas P, et al. Endoscopic ultrasonography for assessment of the response to combined radiation therapy and chemotherapy in patients with esophageal cancer. *Endoscopy*. 1997;29:4–9.
38. Griffin JM, Reed CE, Denlinger CE. Utility of restaging endoscopic ultrasound after neoadjuvant therapy for esophageal cancer. *Ann Thorac Surg*. 2012;93:1855–1860.
39. Guillem P, Fabre S, Mariette C, et al. Surgery after induction chemoradiotherapy for oesophageal cancer. *Eur J Surg Oncol*. 2003;29:158–165.
40. Hamai Y, Hihara J, Emi M, et al. Ability of fluorine-18 fluorodeoxyglucose positron emission tomography to predict outcomes of neoadjuvant chemoradiotherapy followed by surgical treatment for esophageal squamous cell carcinoma. *Ann Thorac Surg*. 2016;102:1132–1139.
41. Heneghan HM, Donohoe C, Elliot J, et al. Can CT-PET and endoscopic assessment post-neoadjuvant chemoradiotherapy predict residual disease in esophageal cancer? *Ann Surg*. 2016;264:831–838.
42. Huang YC, Lu HI, Huang SC, et al. FDG PET using SUVmax for preoperative T-staging of esophageal squamous cell carcinoma with and without neoadjuvant chemoradiotherapy. *BMC Med Imaging*. 2017;17:1.
43. Javeri H, Xiao L, Rohren E, et al. The higher the decrease in the standardized uptake value of positron emission tomography after chemoradiation, the better the survival of patients with gastroesophageal adenocarcinoma. *Cancer*. 2009;115:5184–5192.
44. Jost C, Binek J, Schuller JC, et al. Endosonographic radial tumor thickness after neoadjuvant chemoradiation therapy to predict response and survival in patients with locally advanced esophageal cancer: a prospective multicenter phase ll study by the Swiss Group for Clinical Cancer Research (SAKK 75/02). *Gastrointest Endosc*. 2010;71:1114–1121.
45. Kalha I, Kaw M, Fukami N, et al. The accuracy of endoscopic ultrasound for restaging esophageal carcinoma after chemoradiation therapy. *Cancer*. 2004;101:940–947.
46. Kim JH, Choi EK, Kim SB, et al. Preoperative hyperfractionated radiotherapy with concurrent chemotherapy in resectable esophageal cancer. *Int J Radiat Oncol Biol Phys*. 2001;50:1–12.
47. Kim JJ, Park JK, Moon SW. Usefulness of positron emission tomography-computed tomography in pre-operative evaluation of intra-thoracic esophageal cancer. *Thorac Cancer*. 2015;6:687–694.
48. Kim MK, Ryu JS, Kim SB, et al. Value of complete metabolic response by 18F-fluorodeoxyglucose-positron emission tomography in oesophageal cancer for prediction of pathologic response and survival after preoperative chemoradiotherapy. *Eur J Cancer*. 2007;43:1385–1391.
49. Klayton T, Li T, Yu JQ, et al. The role of qualitative and quantitative analysis of F18-FDG positron emission tomography in predicting pathologic response following chemoradiotherapy in patients with esophageal carcinoma. *J Gastrointest Cancer*. 2012;43:612–618.
50. Kukar M, Alnaji RM, Jabi F, et al. Role of repeat ^{18}F-fluorodeoxyglucose positron emission tomography examination in predicting pathologic response following neoadjuvant chemoradiotherapy for esophageal adenocarcinoma. *JAMA Surg*. 2015;150:555–562.
51. Laterza E, De Manzoni G, Guglielmi A, et al. Endoscopic ultrasonography in the staging of esophageal carcinoma after preoperative radiotherapy and chemotherapy. *Ann Thorac Surg*. 1999;67:1466–1469.
52. Levine EA, Farmer MR, Clark P, et al. Predictive value of 18-fluoro-deoxy-glucose-positron emission tomography (18F-FDG-PET) in the identification of responders to chemoradiation therapy for the treatment of locally advanced esophageal cancer. *Ann Surg*. 2006;243:472–478.
53. Liu SL, Xi M, Yang H, et al. Is there a correlation between clinical complete response and pathological complete response after neoadjuvant chemoradiotherapy for esophageal squamous cell cancer? *Ann Surg Oncol*. 2016;23:273–281.
54. Mamede M, Abreu-E-Lima P, Oliva MR, et al. FDG-PET/CT tumor segmentation-derived indices of metabolic activity to assess response to neoadjuvant therapy and progression-free survival in esophageal cancer: crrelation with histopathology results. *Am J Clin Oncol Cancer Clin Trials*. 2007;30:377–388.
55. McLoughlin JM, Melis M, Siegel EM, et al. Are patients with esophageal cancer who become PET negative after neoadjuvant chemoradiation free of cancer? *J Am Coll Surg*. 2008;206:879–886.
56. Metser U, Rashidi F, Moshonov H, et al. 18F-FDG-PET/CT in assessing response to neoadjuvant chemoradiotherapy for potentially resectable locally advanced esophageal cancer. *Ann Nucl Med*. 2014;28:295–303.
57. Miyata H, Yamasaki M, Takiguchi S, et al. Prognostic value of endoscopic biopsy findings after induction chemoradiotherapy with and without surgery for esophageal cancer. *Ann Surg*. 2011;253:279–284.
58. Motoyama S, Sato Y, Maruyama K, et al. Metabolic rather than pathological response to preoperative chemoradiotherapy is a stronger predictor of survival in cStage IIB-IV esophageal cancer. *Anticancer Res*. 2017;37:4189–4194.

59. Motoyama S, Sato Y, Sasaki T, et al. Efficacy and safety of neoadjuvant chemoradiotherapy following esophagectomy with japanese-style extended 3-field lymphadenectomy for thoracic esophageal cancer. *Anticancer Res*. 2017;37:5837–5843.
60. Myslivecek M, Neoral C, Vrba R, et al. The value of 18F-FDG PET/CT in assessment of metabolic response in esophageal cancer for prediction of histopathological response and survival after preoperative chemoradiotherapy. *Biomed Pap Med Fac Univ Palacky Olomouc Czech Repub*. 2012;156:171–179.
61. Owaki T, Matsumoto M, Okumura H, et al. Endoscopic ultrasonography is useful for monitoring the tumor response of neoadjuvant chemoradiation therapy in esophageal squamous cell carcinoma. *Am J Surg*. 2012;203: 191–197.
62. Park JS, Choi JY, Moon SH, et al. Response evaluation after neoadjuvant chemoradiation by positron emission tomography-computed tomography for esophageal squamous cell carcinoma. *Cancer Res Treat*. 2013;45: 22–30.
63. Peng HQ, Halsey K, Sun CC, et al. Clinical utility of postchemoradiation endoscopic brush cytology and biopsy in predicting residual esophageal adenocarcinoma. *Cancer Cytopathol*. 2009;117:463–472.
64. Roedl JB, Colen RR, Holalkere NS, et al. Adenocarcinomas of the esophagus: Response to chemoradiotherapy is associated with decrease of metabolic tumor volume as measured on PET-CT. Comparison to histopathologic and clinical response evaluation. *Radiother Oncol*. 2008;89:278–286.
65. Roedl JB, Halpern EF, Colen RR, et al. Metabolic tumor width parameters as determined on PET/CT predict disease-free survival and treatment response in squamous cell carcinoma of the esophagus. *Mol Imaging Biol*. 2009;11:54–60.
66. Roedl JB, Harisinghani MG, Colen RR, et al. Assessment of treatment response and recurrence in esophageal carcinoma based on tumor length and standardized uptake value on positron emission tomography-computed tomography. *Ann Thorac Surg*. 2008;86:1131–1138.
67. Sakai M, Sohda M, Miyazaki T, et al. Usefulness of 18f-fluorodeoxyglucose positron emission tomography for predicting the pathological response of neoadjuvant chemoradiotherapy for T4 esophageal squamous cell carcinoma. *Hepatogastroenterology*. 2015;62:898–901.
68. Sarkaria IS, Rizk NP, Bains MS, et al. Post-treatment endoscopic biopsy is a poor-predictor of pathologic response in patients undergoing chemoradiation therapy for esophageal cancer. *Ann Surg*. 2009;249:764–767.
69. Sasaki K, Uchikado Y, Okumura H, et al. Role of 18F-FDG-PET/CT in esophageal squamous cell carcinoma after neoadjuvant chemoradiotherapy. *Anticancer Res*. 2017;37:859–864.
70. Schmidt M, Bollschweiler E, Dietlein M, et al. Mean and maximum standardized uptake values in [18F]FDG-PET for assessment of histopathological response in oesophageal squamous cell carcinoma or adenocarcinoma after radiochemotherapy. *Eur J Nucl Med Mol Imaging*. 2009;36:735–744.
71. Schneider PM, Metzger R, Schaefer H, et al. Response evaluation by endoscopy, rebiopsy, and endoscopic ultrasound does not accurately predict histopathologic regression after neoadjuvant chemoradiation for esophageal cancer. *Ann Surg*. 2008;248:902–908.
72. Shaukat A, Mortazavi A, Demmy T, et al. Should preoperative, post-chemoradiotherapy endoscopy be routine for esophageal cancer patients? *Dis Esophagus*. 2004;17:129–135.
73. Smithers BM, Couper GC, Thomas JM, et al. Positron emission tomography and pathological evidence of response to neoadjuvant therapy in adenocarcinoma of the esophagus. *Dis Esophagus*. 2008;21:151–158.
74. Song YS, Kim JH, Jin SR, et al. FDG-PET in the prediction of pathologic response after neoadjuvant chemoradiotherapy in locally advanced, resectable esophageal cancer. *Int J Radiat Oncol Biol Phys*. 2005;63:1053–1059.
75. Stiekema J, Vermeulen D, Vegt E, et al. Detecting interval metastases and response assessment using ^{18}F-FDG PET/CT after neoadjuvant chemoradiotherapy for esophageal cancer. *Clin Nucl Med*. 2014;39:862–867.
76. Stiles BM, Salzler G, Jorgensen A, et al. Complete metabolic response is not uniformly predictive of complete pathologic response after induction therapy for esophageal cancer. *Ann Thorac Surg*. 2013;96:1820–1825.
77. Swisher SG, Erasmus J, Maish M, et al. 2-Fluoro-2-deoxy-d-glucose positron emission tomography imaging is predictive of pathologic response and survival after preoperative chemoradiation in patients with esophageal carcinoma. *Cancer*. 2004;101:1776–1785.
78. Swisher SG, Maish M, Erasmus JJ, et al. Utility of PET, CT, and EUS to identify pathologic responders in esophageal cancer. *Ann Thorac Surg*. 2004;78:1152–1160.
79. Wieder HA, Brücher BLDM, Zimmermann F, et al. Time course of tumor metabolic activity during chemoradiotherapy of esophageal squamous cell carcinoma and response to treatment. *J Clin Oncol*. 2004;22:900–908.
80. Willis J, Cooper GS, Isenberg G, et al. Correlation of EUS measurement with pathologic assessment of neoadjuvant therapy response in esophageal carcinoma. *Gastrointest Endosc*. 2002;55:655–661.
81. Yen TJ, Chung CS, Wu YW, et al. Comparative study between endoscopic ultrasonography and positron emission tomography-computed tomography in staging patients with esophageal squamous cell carcinoma. *Dis Esophagus*. 2012;25:40–47.
82. Yuan H, Tong DKH, Vardhanabhuti V, et al. PET/CT in the evaluation of treatment response to neoadjuvant chemoradiotherapy and prognostication in patients with locally advanced esophageal squamous cell carcinoma. *Nucl Med Commun*. 2016;37:947–955.
83. Zuccaro G Jr, Rice TW, Goldblum J, et al. Endoscopic ultrasound cannot determine suitability for esophagectomy after aggressive chemoradiotherapy for esophageal cancer. *Am J Gastroenterol*. 1999;94:906–912.
84. Taketa T, Xiao L, Sudo K, et al. Propensity-based matching between esophagogastric cancer patients who had surgery and who declined surgery after preoperative chemoradiation. *Oncology*. 2013;85:95–99.
85. Castoro C, Scarpa M, Cagol M, et al. Complete clinical response after neoadjuvant chemoradiotherapy for squamous cell cancer of the thoracic oesophagus: is surgery always necessary? *J Gastrointest Surg*. 2013;17:1375–1381.
86. Oppedijk V, van der Gaast A, van Lanschot JJ, et al. Patterns of recurrence after surgery alone versus preoperative chemoradiotherapy and surgery in the CROSS trials. *J Clin Oncol*. 2014;32:385–391.
87. van Hagen P, Hulshof MC, van Lanschot JJ, et al. Preoperative chemoradiotherapy for esophageal or junctional cancer. *N Engl J Med*. 2012; 366:2074–2084.
88. Noordman BJ, Spaander MCW, Valkema R, et al. Detection of residual disease after neoadjuvant chemoradiotherapy for oesophageal cancer (preSANO): a prospective multicentre, diagnostic cohort study. *Lancet Oncol*. 2018;19: 965–974.
89. Shapiro J, ten Kate FJ, van Hagen P, et al. Residual esophageal cancer after neoadjuvant chemoradiotherapy frequently involves the mucosa and submucosa. *Ann Surg*. 2013;258:678–688.
90. Vennalaganti PR, Naag Kanakadandi V, Gross SA, et al. Inter-observer agreement among pathologists using wide-area transepithelial sampling with computer-assisted analysis in patients with Barrett's esophagus. *Am J Gastroenterol*. 2015;110:1257–1260.
91. Ross-Innes CS, Debiram-Beecham I, O'Donovan M, et al. Evaluation of a minimally invasive cell sampling device coupled with assessment of trefoil factor 3 expression for diagnosing Barrett's esophagus: a multi-center case-control study. *PLoS Med*. 2015;12:e1001780.
92. Yip SSF, Coroller TP, Sanford NN, et al. Relationship between the temporal changes in positron-emission tomography-imaging-based textural features and pathologic response and survival in esophageal cancer patients. *Front Oncol*. 2016;29:72.
93. Heethuis SE, van Rossum PS, Lips IM, et al. Dynamic contrast-enhanced MRI for treatment response assessment in patients with oesophageal cancer receiving neoadjuvant chemoradiotherapy. *Radiother Oncol*. 2016;120:128–135.
94. Cohen JD, Li L, Wang Y, et al. Detection and localization of surgically resectable cancers with a multi-analyte blood test. *Science*. 2018;359:926–930.
95. Creemers A, Krausz S, Strijker M, et al. Clinical value of ctDNA in upper-GI cancers: a systematic review and meta-analysis. *Biochim Biophys Acta*. 2017;1868:394–403.