# Annals of Surgery

## Accuracy of detecting residual disease after neoadjuvant chemoradiotherapy for esophageal cancer: a systematic review and meta-analysis

---Manuscript Draft---

<table>
<thead>
<tr>
<th>Manuscript Number:</th>
<th>Accuracy of detecting residual disease after neoadjuvant chemoradiotherapy for esophageal cancer: a systematic review and meta-analysis</th>
</tr>
</thead>
<tbody>
<tr>
<td>Full Title:</td>
<td>Accuracy of detecting residual disease after neoadjuvant chemoradiotherapy for esophageal cancer: a systematic review and meta-analysis</td>
</tr>
<tr>
<td>Article Type:</td>
<td>Meta-Analysis</td>
</tr>
<tr>
<td>Keywords:</td>
<td>esophageal cancer; restaging; residual disease; neoadjuvant chemoradiotherapy; endoscopic biopsies; EUS; PET; accuracy</td>
</tr>
<tr>
<td>Corresponding Author:</td>
<td>Ben M. Eyck, M.D. Erasmus MC Rotterdam, Zuid-Holland NETHERLANDS</td>
</tr>
<tr>
<td>Corresponding Author Secondary Information:</td>
<td></td>
</tr>
<tr>
<td>Corresponding Author's Institution:</td>
<td>Erasmus MC</td>
</tr>
<tr>
<td>Corresponding Author's Secondary Institution:</td>
<td></td>
</tr>
<tr>
<td>First Author:</td>
<td>Ben M. Eyck, M.D.</td>
</tr>
<tr>
<td>First Author Secondary Information:</td>
<td></td>
</tr>
</tbody>
</table>
| Order of Authors:  | Ben M. Eyck, M.D.  
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  
J. Jan B. van Lanschot, MD, PhD |
| Order of Authors Secondary Information: | |
| Manuscript Region of Origin: | NETHERLANDS |
| Additional Information: | |

<table>
<thead>
<tr>
<th>Question</th>
<th>Response</th>
</tr>
</thead>
<tbody>
<tr>
<td>Does your manuscript describe a Randomized Controlled Trial?</td>
<td>No</td>
</tr>
<tr>
<td>Briefly indicate the type of study your paper describes. as follow-up to &quot;Does your manuscript describe a Randomized Controlled Trial?&quot;</td>
<td>meta-analysis</td>
</tr>
<tr>
<td>Have all coauthors viewed and approved of this version manuscript?</td>
<td>Yes - All coauthors have seen &amp; approved</td>
</tr>
<tr>
<td>Do you or any of your coauthors have any potential conflicts of interest you need to</td>
<td>The authors declare there are no conflicts of interest.</td>
</tr>
<tr>
<td>Question</td>
<td>Answer</td>
</tr>
<tr>
<td>------------------------------------------------------------------------</td>
<td>------------------------------------------------------------------------</td>
</tr>
<tr>
<td>Was this submission funded by an outside agency (e.g., business, government agency, etc.)</td>
<td>No funding was received in support of this work.</td>
</tr>
<tr>
<td>If your submission contains figures, are any of them in color?</td>
<td>Yes, my submission has colored figures.</td>
</tr>
<tr>
<td>Are you and your coauthors willing to pay for color printing?</td>
<td>No, please print the figures in black and white.</td>
</tr>
<tr>
<td>Will reprints be available from the authors?</td>
<td>Yes</td>
</tr>
<tr>
<td>Will the corresponding author also be the contact for reprints?</td>
<td>Yes</td>
</tr>
<tr>
<td>RETAINED RIGHTS: Except for copyright, other proprietary rights related to the Work (e.g., patent or other rights to any process or procedure) shall be retained by the author. To reproduce any text, figures, tables, or illustrations from this Work in future works of their own, the author must obtain written permission from Wolters Kluwer Health, Inc. (&quot;WKH&quot;).</td>
<td>I agree</td>
</tr>
<tr>
<td>ORIGINALITY: Each author warrants that his or her submission to the Work is original, does not infringe upon, violate, or misappropriate any copyright or other intellectual property rights, or any other proprietary right, contract or other right or interest of any third party, and that he or she has full power to enter into this agreement. Neither this Work nor a similar work has been published nor shall be submitted for publication elsewhere while under consideration by this Publication.</td>
<td></td>
</tr>
<tr>
<td>AUTHORSHIP RESPONSIBILITY: Each author warrants that he or she has</td>
<td></td>
</tr>
</tbody>
</table>
participated sufficiently in the intellectual content, the analysis of data, if applicable, and the writing of the Work to take public responsibility for it. Each has reviewed the final version of the Work, believes it represents valid work, and approves it for publication. Moreover, should the editors of the Publication request the data upon which the work is based, they shall produce it.

PREPRINTS: Upon acceptance of the article for publication, each author warrants that he/she will promptly remove any prior versions of this Work (normally a preprint) that may have been posted to an electronic server.

DISCLAIMER: Each author warrants that this Work contains no libelous or unlawful statements and does not infringe or violate the publicity or privacy rights of any third party, libel or slander any third party, contain any scandalous, obscene, or negligently prepared information, or infringe or violate any other personal or proprietary right of others. Each author warrants that the Work does not contain any fraudulent, plagiarized or incorrectly attributed material. Each author warrants that all statements contained in the Work purporting to be facts are true, and any formula or instruction contained in the Work will not, if followed accurately, cause any injury, illness, or damage to the user. If excerpts (e.g., text, figures, tables, illustrations, or audio/video files) from copyrighted works are included, a written release will be secured by the author prior to submission, and credit to the original publication will be properly acknowledged. Each author further warrants that he or she has obtained, prior to submission, written releases from patients whose names or likenesses are submitted as part of the Work. Should the Editor or WKH request copies of such written
releases, the author shall provide them in a timely manner.

DISCLOSURES/CONFLICT OF INTEREST
Each author must identify any financial interests or affiliations with institutions, organizations, or companies relevant to the manuscript by completing the form below. Additionally, any financial associations involving a spouse, partner or children must be disclosed as well.

Note: Some sections below come from the ICMJE Uniform Disclosure Form for Potential Conflicts of Interest at http://www.icmje.org/downloads/coi_disclosure.pdf (dated July 2010).

<table>
<thead>
<tr>
<th>Question</th>
<th>Answer</th>
</tr>
</thead>
<tbody>
<tr>
<td>Did you or your institution at any time receive payment or support in kind for any aspect of the submitted work (including but not limited to grants, consulting fee or honorarium, support for travel to meetings for the study or other purposes, fees for participation in review activities such as data monitoring boards, statistical analysis, end point committees, and the like, payment for writing or reviewing the manuscript, provision of writing assistance, medicines, equipment, or administrative support, etc...)?</td>
<td>No</td>
</tr>
<tr>
<td>Other: Did you or your institution at any time receive additional payments or support in kind for any aspect of the submitted work?</td>
<td>No</td>
</tr>
<tr>
<td>Please indicate whether you have financial relationships (regardless of amount of compensation) with entities. You should report relationships that were present during the 36 months prior to submission including board membership, consultancy, employment, expert testimony, grants/grants pending, payment for lectures including service on speakers bureaus, payment for manuscript preparation, patents (planned,</td>
<td>No</td>
</tr>
</tbody>
</table>
pending or issued), royalties, payment for development of educational presentations, stock/stock options, travel/accommodations/meeting expenses unrelated to activities listed (for example, if you report a consultancy above there is no need to report travel related to that consultancy), etc.

<table>
<thead>
<tr>
<th>Other (err on the side of full disclosure): Please indicate whether you have any additional financial relationships (regardless of amount of compensation) with entities. You should report relationships that were present during the 36 months prior to submission.</th>
</tr>
</thead>
<tbody>
<tr>
<td>No</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Other Relationships</th>
</tr>
</thead>
<tbody>
<tr>
<td>No other relationships/conditions/circumstances that present potential conflict of interest</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>AUTHOR'S OWN WORK: In consideration of WKH's publication of the Work, the author hereby transfers, assigns, and otherwise conveys all his/her copyright ownership worldwide, in all languages, and in all forms of media now or hereafter known, including electronic media such as CD-ROM, Internet, and Intranet, to WKH. If WKH should decide for any reason not to publish the Work, WKH shall give prompt notice of its decision to the corresponding author, this agreement shall terminate, and neither the author nor WKH shall be under any further liability or obligation. Each author grants WKH the rights to use his or her name and biographical data (including professional affiliation) in the Work and in its or the journal's promotion. Notwithstanding the foregoing, this paragraph shall not apply, and any transfer made pursuant to this paragraph shall be null and void if (i) the Work has been accepted by WKH for publication, and (ii) the author chooses to have the Work published by WKH as an open access publication.</th>
</tr>
</thead>
<tbody>
<tr>
<td>I agree</td>
</tr>
</tbody>
</table>

Powered by Editorial Manager® and ProduXion Manager® from Aries Systems Corporation
WORK MADE FOR HIRE: If this Work or any element thereof has been commissioned by another person or organization, or if it has been written as part of the duties of an employee, an authorized representative of the commissioning organization or employer must also sign this form stating his or her title in the organization.

GOVERNMENT EMPLOYEES: If the Work or a portion of it has been created in the course of any author's employment by the United States Government, check the "Government" box at the end of this form. A work prepared by a government employee as part of his or her official duties is called a "work of the U.S. Government" and is not subject to copyright. If it is not prepared as part of the employee's official duties, it may be subject to copyright.

INSTITUTIONAL REVIEW BOARD/ANIMAL CARE COMMITTEE APPROVAL: Each author warrants that his or her institution has approved the protocol for any investigation involving humans or animals and that all experimentation was conducted in conformity with ethical and humane principles of research.

WARRANTIES: Each author warranty made in this form is for the benefit of WKH and the Editor; each author agrees to defend, indemnify, and hold harmless those parties for any breach of such warranties.

The journal will permit the author(s) to deposit for display a "final peer-reviewed manuscript" (the final manuscript after peer-review and acceptance for publication but prior to the publisher's copyediting, design, formatting, and other services) 12 months after publication of the final article on the author's personal web site, university's institutional repository or employer's
* You may only deposit the final peer-reviewed manuscript.

* You may not update the final peer-reviewed manuscript text or replace it with a proof or with the final published version.

* You may not include the final peer-reviewed manuscript or any other version of the article on any commercial site or in any repository owned or operated by any third party. For authors of articles based on research funded by the National Institutes of Health ("NIH"), Wellcome Trust, Howard Hughes Medical Institute ("HHMI"), or other funding agency, see below for the services that WKH will provide on your behalf to comply with "Public Access Policy" guidelines.

* You may not display the final peer-reviewed manuscript until twelve months after publication of the final article.

* You must attach the following notice to the final peer-reviewed manuscript: "This is a non-final version of an article published in final form in (provide complete journal citation)".

* You shall provide a link in the final peer-reviewed manuscript to the journal website.

**"Public Access Policy" Funding Disclosure**

Please disclose below if you have received funding for research on which your article is based from any of the following organizations:

<table>
<thead>
<tr>
<th>Please select:</th>
<th>Author's Own Work</th>
</tr>
</thead>
<tbody>
<tr>
<td>Any additional comments?</td>
<td>No</td>
</tr>
</tbody>
</table>
## Compliance with RCUK and Wellcome Trust Open Access Policies

Both the Research Councils UK (RCUK) and the Wellcome Trust have adopted policies regarding Open Access to articles that have been funded by grants from the RCUK or the Wellcome Trust. If either "Wellcome Trust" or "Research Councils UK (RCUK)" has been selected above, and the authors of the applicable article choose to have the article published as an open access publication, the following policies will apply:

* If the article is to be published pursuant to the "Gold" route of Open Access, both the RCUK and the Wellcome Trust require that WKH make the article freely available immediately pursuant to the Attribution 4.0 Creative Commons License, currently found at [http://creativecommons.org/licenses/by/4.0/legalcode](http://creativecommons.org/licenses/by/4.0/legalcode) (the “CC BY License”). The CC BY License is the most accommodating of the Creative Commons licenses and allows others to distribute, remix, tweak, and build upon the article, even commercially, as long as they credit the authors for the original creation.

* If the article is to be published pursuant to the "Green" route of Open Access, both the RCUK and the Wellcome Trust require that WKH make the article freely available within six months pursuant to the Attribution-NonCommerical 4.0 Creative Commons License, currently found at [http://creativecommons.org/licenses/by-nc/4.0/legalcode](http://creativecommons.org/licenses/by-nc/4.0/legalcode) (the “CC BY-NC License”). The CC BY-NC License allows others to remix, tweak, and build upon the article non-commercially, and although their new works must also acknowledge the authors for the original creation and
be non-commercial, they don't have to license their derivative works on the same terms.

As a service to our authors, WKH will identify the National Library of Medicine (NLM) articles that require deposit pursuant to the RCUK and Wellcome Trust policies described in this section. This Copyright Transfer Agreement provides the mechanism for identifying such articles.

WKH will transmit the final peer-reviewed manuscript of an article based on research funded in whole or in part by either RCUK or the Wellcome Trust to Pub Med Central.

Upon NIH request, it remains the legal responsibility of the author to confirm with NIH the provenance of his/her manuscript for purposes of deposit. Author will not deposit articles him/herself. Author will not alter the final peer-reviewed manuscript already transmitted to NIH.

With respect to the “Green” route of Open Access, author will not authorize the display of the final peer-reviewed manuscript prior to 6 months following publication of the final article.

Authors of articles that have been funded from grants from the RCUK or the Wellcome Trust are required to sign the WKH Open Access License Agreement prior to publication of the applicable article. Please contact the Editorial Office of the applicable journal to receive the Open Access License Agreement that is to be signed in connection with the publication of the article.

I am the person in question for this I agree
<table>
<thead>
<tr>
<th><strong>CME/CE Disclosure</strong></th>
<th>I agree</th>
</tr>
</thead>
<tbody>
<tr>
<td>Each author must identify and disclose any financial associations involving a spouse, partner or children by completing the Family Disclosure question below, and whether any off-label uses or unapproved drugs or devices are discussed in his/her manuscript by completing the Off-Label Use/Unapproved Drugs or Products question below. In the event that the Work is published as a continuing education or continuing medical education article, this information will be provided to the accrediting body and may be included in the published article. When applicable, articles accepted for publication may need to comply with additional standards related to CME or CE accreditation. Please refer to guidelines for authors for details. WKH and its affiliates reserve the right to publish the manuscript as a continuing education article.</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th><strong>Family Disclosure</strong></th>
<th>No other relationships/conditions/circumstances that present potential conflict of interest</th>
</tr>
</thead>
<tbody>
<tr>
<td>Do your children or your spouse or partner have financial relationships with entities that have an interest in the content of the submitted work?</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th><strong>Off-Label Use/Unapproved Drugs or Products</strong></th>
<th>I will not discuss unlabeled/investigational uses of any commercial product or device</th>
</tr>
</thead>
<tbody>
<tr>
<td>If your manuscript discusses an unlabeled use of a commercial product or device or an investigational use of a product or device not yet approved by the FDA for any purpose, you must specifically disclose in the manuscript that the product is not labeled for the use under discussion or that the product is still investigational. Please check the item below that applies to you</td>
<td></td>
</tr>
</tbody>
</table>
B.M. Eyck  
Department of Surgery, Erasmus MC - University Medical Center  
Dr. Molewaterplein 40  
3015 GD Rotterdam  
The Netherlands  
b.eyck@erasmusmc.nl

Dr. Keith D. Lillemoe  
Editor-in-Chief  
Annals of Surgery

Dear Dr. Lillemoe,

Rotterdam, December 12, 2018

On behalf of the co-authors, I wish to submit our systematic review and meta-analysis entitled ‘Accuracy of detecting residual disease after neoadjuvant chemoradiotherapy (nCRT) for esophageal cancer: a systematic review and meta-analysis’ for consideration to publish in the Annals of Surgery.

Previously, we have shown that nCRT consisting of carboplatin and paclitaxel with 41.4 Gy radiotherapy significantly improves survival, compared to surgery alone (van Hagen et al, N Engl J Med 2012; Shapiro et al, Lancet Oncol 2015). After this nCRT regimen plus surgery, nearly a third (29%) of patients has a pathologically complete response in the resection specimen. This provides a rationale for an active surveillance approach, in which patients are subjected to frequent clinical investigations after nCRT, and esophagectomy is offered only to those with a proven locoregional recurrence without distant metastases. It is unknown which diagnostic tests are adequate for detecting residual disease after nCRT. The current study evaluates current literature on the accuracy of endoscopic biopsies, endoscopic ultrasound (EUS) and 18F-fluoro-2-deoxy-D-glucose positron emission tomography with or without computed tomography (18F-FDG PET(-CT)) for detecting residual disease, which are the major diagnostic techniques used for response evaluation in current clinical practice.

Since three diagnostic modalities are evaluated, the primary search term yielded 65 eligible articles for systematic review. Including essential references, we exceeded your reference limitation of 75 with 19 additional references. We believe that this manuscript is of interest to the readers of Annals of Surgery because it discusses a timely and relevant topic in the surgical field of esophageal cancer.

The manuscript has not been published and is not under consideration for publication elsewhere. No funding was received for this contribution and we do not have any conflicts of interest. Please do not hesitate to contact us, should you have any questions or comments.

Yours sincerely,

B.M. Eyck, MD, corresponding author
MINI-ABSTRACT

An active surveillance strategy has been proposed for clinically complete responders after neoadjuvant chemoradiotherapy for esophageal cancer. This systematic review and meta-analysis suggest that endoscopic biopsies, EUS and 18F-FDG PET(-CT) as single modalities are insufficiently accurate to detect residual disease after neoadjuvant chemoradiotherapy.
STRUCTURED ABSTRACT

Objective: 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, eight quantitative PET using maximum standardized uptake value (SUVmax) and seven 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.
Accuracy of detecting residual disease after neoadjuvant chemoradiotherapy for esophageal cancer: a systematic review and meta-analysis

Ben M. Eyck, MD, *1 Barbera D. Onstenk, BSc, *2 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, * J. Jan B. van Lanschot, MD, PhD*

*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
§Department of Radiology and Nuclear Medicine, Erasmus MC - University Medical Center, Rotterdam, the Netherlands

*Both authors contributed equally and should be acknowledged with first authorship.

Corresponding author and reprints: Ben M. Eyck, MD, Department of Surgery, Erasmus MC - University Medical Center, Dr. Molewaterplein 40, 3015 GD Rotterdam, the Netherlands, Telephone: 0031(0)10-7040704, Email: b.eyck@erasusmc.nl.

No means of funding were received for this contribution.

The authors declare no conflict of interests.

Words: text 4.974 (max. 5.000), abstract 250 (max. 250); References: 94 (max. 75); Tables & figures: 5 (max. 7)

Short running title: Detecting residual esophageal cancer after nCRT
MINI-ABSTRACT

An active surveillance strategy has been proposed for clinically complete responders after neoadjuvant chemoradiotherapy for esophageal cancer. This systematic review and meta-analysis suggest that endoscopic biopsies, EUS and 18F-FDG PET(-CT) as single modalities are insufficiently accurate to detect residual disease after neoadjuvant chemoradiotherapy.
STRUCTURED ABSTRACT

**Objective:** 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, eight quantitative PET using maximum standardized uptake value (SUVmax) and seven 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.
INTRODUCTION

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, five-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 1-5% in high volume centers, severe post-operative 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 re-staged 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.
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. 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. 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 2x2 contingency table. If studies had insufficient data to construct 2x2 contingency tables, corresponding authors of each study were contacted by email at least 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 less than 10 patients were excluded. Also, studies reporting on cervical esophageal cancer only were excluded since 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, USA). 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 two authors (B.E. and B.O.). 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 (B.N.) gave a binding verdict.

Data Extraction

Data were extracted by two authors (B.E. and B.O.) 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 2x2 contingency tables. If studies investigated multiple threshold for one index test modality, TP, FP, TN, and FN values produced by the optimal cut-off were chosen for data extraction. For data comparison, pathological response criteria were equated. Studies that used pathological response criteria similar to pathologically complete vs. incomplete response were redefined, i.e.: 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 vs.TRG2-4), 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. non-complete response). Likewise, studies that could be categorized as <10% vs. > 10% and <33.3% vs. >33.3% residual disease were redefined as such.

Quality Assessment

The quality of the included studies was independently appraised by two authors (B.E. and B.O.) according to the revised Quality Assessment of Diagnostic Accuracy Studies (QUADAS-2) tool. Disagreements were resolved by consensus discussion. The QUADAS-2 tool assesses risk of bias regarding four key domains; patient selection, index test, reference standard and flow and timing. Some adjustments were made. Regarding patient selection, the use of induction chemotherapy prior to 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 since pathological examination is mainly an independent procedure. In the domain of flow and timing, a time interval between index test and surgery of more than 4 weeks was considered a high risk of bias, since 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 true-positives, false-positives, true-negatives, and false-negatives 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 four 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 receiving 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.\textsuperscript{16,17} The existence of between-study heterogeneity was primarily assessed through visually inspecting forest plots for the degree of overlapping confidence intervals. Subsequently, 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.
Subgroup analyses were performed to investigate sources of heterogeneity (histology, definition of pCR, PET imaging technique, and cut-off 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, USA) and were performed in accordance with the Cochrane Handbook for Diagnostic Test Accuracy Meta-Analysis. Further explanation of statistical analysis is provided in Supplementary information 1.
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 (Figure 1).\textsuperscript{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 less than four studies were included that evaluated the same index test or because a pathological response criterion other than pCR was used. Forty-four studies were included for quantitative synthesis, comprising six index test modalities.\textsuperscript{19-21, 23, 24, 27, 30, 32-38, 40-43, 45-51, 55-63, 67, 68, 71, 72, 76, 80-83}

Study Characteristics

Studies included in quantitative synthesis

Endoscopy with biopsies (Supplementary Table 2)

Twelve studies comprising a total of 1328 patients evaluated endoscopic biopsies for detecting any residual disease at the primary tumor site as positive biopsy vs. negative biopsies (Supplementary Table 2).\textsuperscript{20, 30, 32, 45, 46, 48, 57, 61, 63, 68, 71, 72} Three out of 12 studies were prospective studies.\textsuperscript{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 two of 12 studies reported on the number of biopsies that were taken, which had a median of 4 biopsies.\textsuperscript{71, 72} Locations of biopsies were not reported.
EUS (Supplementary Table 3)

For EUS, 13 studies were included in quantitative synthesis (Supplementary Table 3). Of these, 11 studies comprising a total of 563 patients evaluated qualitative EUS for residual disease at the primary tumor site. Another 11 studies comprising a total of 629 patients evaluated qualitative EUS for regional lymph nodes. Three out of 13 studies were prospective studies. 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 days 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. Of these, only one used FNA for cytological confirmation of positive lymph nodes.

PET(-CT) (Supplementary Table 4)

For PET(-CT), 24 studies were included in quantitative synthesis (Supplementary Table 4). Of these, 14 studies comprising a total of 1213 patients evaluated qualitative PET as metabolically non-complete response vs. metabolically complete response (mCR) for residual disease at the primary tumor site. Another eight studies comprising a total of 430 patients evaluated quantitative PET for the primary tumor site by using maximum standardized uptake value (SUVmax), and seven studies comprising a total of 511 patients by using percentage reduction of SUVmax (%∆SUVmax). Four out of 24 studies were prospective. 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 days to 52 days and intervals from PET to surgery...
ranged from 2 days to 42 days. For studies evaluating SUVmax, the median cut-off was 2.7 (range 2.25 – 6.0).
For %ΔSUVmax, median cut-off was 72.3% (range 52% – 79.3%).

Studies excluded from quantitative synthesis

Characteristics of 21 studies excluded from quantitative synthesis are described in Supplementary Table 2-4.

Quality Assessment

Studies included in quantitative synthesis

Endoscopy with biopsies (Supplementary Figure 1A)

Of 12 studies evaluating endoscopic biopsies that were included in quantitative analysis, risk of bias was present in six studies.45, 57, 61, 63, 68, 71 Risk of bias regarding patient selection was high in five studies evaluating endoscopic biopsies because patients were not consecutively included.45, 57, 61, 63, 68 Risk of bias concerning the index test was high in one study because the index test result was not interpreted without the knowledge of the result of the reference standard.63 With regard to the reference standard, risk of bias was low in all studies. In the domain of flow and timing, risk of bias was unclear in two studies because the interval between index test and reference standard was not given,30, 68 and high in one study since the number of initially included patients differed from the number of patients included in the analysis without clarification.71

EUS (Supplementary Figure 1B)

Of 11 studies evaluating qualitative EUS for the primary tumor site, risk of bias was present in eight studies.24, 37, 38, 45, 51, 71, 83 Of 11 studies evaluating qualitative EUS for regional lymph nodes, risk of bias was also present in eight studies.23, 24, 35, 38, 45, 51, 83 Risk of bias regarding patient selection was high in six studies evaluating EUS because patients were not consecutively included.24, 35, 37, 45, 51, 83 Risk of bias concerning the index test and reference standard was low in all studies. In the domain of flow and timing, risk of bias was considered unclear in three studies because the interval between index test and reference standard was not given,34, 35, 51 and
considered high in two studies because the number of initially included patients differed from the number of patients included in the analysis without clarification.\(^{38,71}\)

**PET(-CT) (Supplementary Figure 1C)**

Of 14 studies evaluating qualitative PET, risk of bias was present in seven studies.\(^{20,21,33,50,55,56,60}\) Of eight studies evaluating PET-SUV\(_{\text{max}}\), risk of bias was also present in seven studies.\(^{27,33,40,42,47,67,82}\) Of seven studies evaluating PET-%ΔSUV\(_{\text{max}}\), risk of bias was present in four studies.\(^{28,33,40,67}\) Risk of bias regarding patient selection was high in seven studies evaluating PET because patients were not consecutively included.\(^{21,27,33,50,55,56,67}\) Risk of bias concerning the index test was high in nine studies since the thresholds for quantitative PET analyses were not prespecified.\(^{27,28,33,40,42,50,62,67,82}\) Risk of bias concerning the reference standard was low in all studies. In the domain of flow and timing, risk of bias was considered unclear in three studies because the interval between index test and reference standard was not given,\(^{43,49,50}\) and considered high in five studies because the interval between index test and reference standard was more than four weeks,\(^{28,47,82}\) or because the number of initially included patients differed from the number of patients included in the analysis without clarification.\(^{27,60}\)

**Studies excluded from quantitative synthesis**

Quality assessment of studies excluded from quantitative synthesis is addressed in Supplementary Information 2.

**Diagnostic Accuracy**

**Studies included in quantitative synthesis (Table 1)**

**Endoscopy with biopsies**

Sensitivity and specificity of positive vs. 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 (Figure 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)
(Figure 3A). There was a higher variability for sensitivity than for specificity. The forest plot (Figure 2A) and the 95% prediction region in the SROC plot (Figure 3A) demonstrated substantial heterogeneity between studies.

**EUS**

Sensitivity and specificity of qualitative EUS uT+ vs. 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 (Figure 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) (Figure 3B). There was a slightly higher variability for specificity than for sensitivity. The forest plot (Figure 2B) and the 95% prediction region in the SROC plot (Figure 3B) demonstrated substantial heterogeneity between studies.

Sensitivity and specificity of qualitative EUS uN+ vs. uN0 for detecting any residual nodal disease ranged from 0.26 to 0.94 and from 0.23 to 1.00, respectively (Figure 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) (Figure 3C). Variability for sensitivity and specificity was comparable. The forest plot (Figure 2C) and the 95% prediction region in the SROC plot (Figure 3C) demonstrated substantial heterogeneity between studies.

**PET(-CT)**

Sensitivity and specificity of qualitative PET metabolically non-complete response vs. 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 (Figure 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) (Figure 3D). There was a higher variability for specificity than for sensitivity. The forest plot (Figure 2D) and the 95% prediction region in the SROC plot (Figure 3D) demonstrated substantial heterogeneity between studies.

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 (Figure 2D). Cut-offs 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) (Figure 3E). Variability for sensitivity and specificity was comparably low. Besides the outlying specificity of Kim et al. the forest plot demonstrated low heterogeneity between studies (Figure 2D).47 The low heterogeneity lead to identical 95% confidence and 95% prediction regions.

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 (Figure 2D). Cut-offs 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) (Figure 3F). Variability was higher for sensitivity than for specificity. The forest plot (Figure 2D) and the 95% prediction region in the SROC plot (Figure 3F) demonstrated substantial heterogeneity between studies.

**Studies excluded from quantitative synthesis**

Diagnostic accuracy of studies excluded from quantitative synthesis is explicated in Supplementary Information 4.

**Subgroup Analyses**

**Endoscopy with biopsies**

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 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 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 $\%\Delta$SUVmax 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%.

In the light of an active surveillance strategy, sensitivity is an important diagnostic parameter, since false-negative 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 false-positives 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 $\%\Delta$SUVmax 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. While ideally the smallest amount of residual disease should be detected, microscopic residue is often missed during preoperative clinical response evaluations in current clinical practice. However, existing studies show no decline in oncological outcome for patients who underwent active surveillance with similar diagnostic tests (i.e. endoscopic biopsies and PET(-CT) instead of standard esophagectomy after nCRT. 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.

For endoscopic biopsies, EUS and PET(-CT) respectively one, six 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). 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/MMT) also showed favorable sensitivity (0.79) and specificity (0.82). 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 cut-off 56% (sensitivity of 0.93 and 0.91) all showed good accuracy. However, results should be confirmed since these studies were performed in one hospital with overlapping cohorts. Moreover, combining index test modalities to obtain an optimal set for response evaluation can improve diagnostic accuracy.

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, studies of low quality based on the QUADAS-2 were not excluded because well-designed, sufficiently powered, prospective studies on this topic are scarce. Second, different histological subtypes 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. However, in contrast to positive- and negative predictive value, sensitivity and specificity are not influenced by prevalence of esophageal cancer. Therefore, 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. Third, studies with different definitions for discriminating pathological responders from non-responders were included. Most studies used pathologically complete vs. incomplete response to discriminate between both groups. Studies that used similar pathological response criteria were redefined to pathologically complete vs. 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 less than 10 studies and therefore results should be interpreted with caution.

The recently published prospective preSANO trial aimed to accurately determine the optimal combination of tests for clinical response evaluations to identify patients who might benefit from active surveillance. The study showed that the combination of endoscopic biopsies and EUS with fine-needle aspiration of suspected lymph nodes detected TRG2-4 tumor with a sensitivity of 54% and a specificity of 69%. Sensitivity improved to 77% with a specificity of 72% using bite-on-bite biopsies. This is most likely because residual disease is often found in the submucosal layer while the mucosal layer is free of residual tumor. Importantly, the investigators of the preSANO trial hypothesized that microscopic amounts of residual tumor (i.e. TRG2 according to Mandard) can be missed during initial response evaluations in an active surveillance strategy as malignancy will likely be detected in a resectable stage during follow-up. Therefore, the primary analysis considered TRG1 and TRG2 (<10 % residual disease) as negative condition to determine the accuracy of clinical response evaluations.

The combination of endoscopy with bite on-bite biopsies at the primary tumor site and EUS with fine-needle aspiration of all suspected lymph nodes detected TRG3-4 tumors correctly in 90%. In addition, PET-CT detected interval metastases in 10% of patients, which prevented unnecessary surgical resection. Of all patients treated with neoadjuvant chemoradiotherapy according to CROSS, 40% will develop interval metastases. Besides patients with a pCR after nCRT, these patients might benefit from an active surveillance strategy as well. 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. 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 false-negative
response evaluation and decreasing the number of patients with unnecessary surgical resections after false-positive response evaluation. 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.\textsuperscript{89, 90} 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. Since 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.\textsuperscript{91, 92} Also, other novel techniques such as multi-analyte 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.\textsuperscript{93, 94}

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.
ACKNOWLEDGEMENTS

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


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(22):5184-5192.


List of Tables

- Table 1. Summary sensitivity and specificity values of studies included in quantitative synthesis.

- Table 2. Results from study-level subgroup analyses.
  - A. Studies evaluating endoscopic biopsies.
  - B. Studies evaluating EUS.
  - C. Studies evaluating PET(CT).
List of Figures and Figure Legend

- **Figure 1.** Flowchart summarizing search results and study selection.
- **Figure 2.** Forest plots demonstrating sensitivities and specificities from individual studies included in quantitative analyses.

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

  - **A. Studies evaluating endoscopic biopsies.**
  - **B. Studies evaluating EUS.**
  - **C. Studies evaluating PET(-CT).**

- **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.

EUS: endoscopic ultrasonography, PET(-CT): positron emission tomography with or without computed tomography, uT0: ultrasonographic tumor stage 0, uN0: ultrasonographic nodal stage 0, mCR: metabolically complete response, SUVmax: maximum standardized uptake value, %ΔSUVmax: percentage reduction of SUVmax, pCR: pathologically complete response.
o A. SROC of endoscopic biopsies evaluating the primary tumor site using negative biopsies as clinical response criterion and pCR as pathological response criterion.

o B. SROC of EUS evaluating the primary tumor site using uT0 as clinical response criterion and pCR as pathological response criterion.

o C. SROC of EUS evaluating lymph nodes using uN0 as clinical response criterion and pCR as pathological response criterion.

o D. SROC of PET(-CT) evaluating the primary tumor site using mCR as clinical response criterion and pCR as pathological response criterion.

o E. SROC of PET(-CT) evaluating the primary tumor site using SUVmax < cut-off as clinical response criterion and pCR as pathological response criterion.

o F. SROC of PET(-CT) evaluating the primary tumor site using %ΔSUVmax > cut-off as clinical response criterion and pCR as pathological response criterion.
List of Supplementary Digital Content

Supplementary information

- Supplementary Information 1. Explanation of the statistical analysis.

- Supplementary Information 2. Quality assessment of studies included in the qualitative analysis.
  - Endoscopy with biopsies (Supplementary Figure 1A)
  - EUS (Supplementary Figure 1B)
  - PET(-CT) (Supplementary Figure 1C)
  - Combinations of index tests (Supplementary Figure 1D)

- Supplementary Information 3. Diagnostic accuracy of studies included in the qualitative analysis.
  - Endoscopy with biopsies (Supplementary Figure 2A)
  - EUS (Supplementary Figure 2B)
  - PET(-CT) (Supplementary Figure 2C)
  - Combinations of index tests (Supplementary Figure 2D)

Supplementary tables

- Supplementary Table 1. Search strategy and results as of February, 2018.
- Supplementary Table 2. Characteristics of the studies evaluating endoscopic biopsies.
- Supplementary Table 3. Characteristics of the studies evaluating EUS.
- Supplementary Table 4. Characteristics of the studies evaluating PET(-CT)
- Supplementary Table 5. Characteristics of the studies evaluating combined index tests.

Supplementary figures

- Supplementary Figure 1. Quality assessment according to the QUADAS-2 tool of all included studies.
  - A. Studies evaluating endoscopic biopsies.
  - B. Studies evaluating EUS.
  - C. Studies evaluating PET(-CT).
  - D. Studies evaluating combined index tests.
- Supplementary Figure 2. Forest plots demonstrating sensitivities and specificities from individual studies excluded from quantitative synthesis.
A. Studies evaluating endoscopic biopsies.
B. Studies evaluating EUS.
C. Studies evaluating PET(-CT).
D. Studies evaluating combined index tests.
Table 1. Summary sensitivity and specificity values of studies included in quantitative synthesis.

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

<table>
<thead>
<tr>
<th>Index test</th>
<th>Region</th>
<th>ClinRC</th>
<th>PathRC</th>
<th>Median cut-off (range)</th>
<th>Number of studies</th>
<th>Summary sensitivity (95% CI)</th>
<th>Summary specificity (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Endoscopic biopsies</td>
<td>T</td>
<td>Biopsies negative</td>
<td>pCR</td>
<td>N/A</td>
<td>12</td>
<td>0.33 [0.24-0.43]</td>
<td>0.95 [0.88-0.98]</td>
</tr>
<tr>
<td>EUS</td>
<td>T</td>
<td>uTO</td>
<td>pCR</td>
<td>N/A</td>
<td>11</td>
<td>0.96 [0.89-0.99]</td>
<td>0.08 [0.03-0.24]</td>
</tr>
<tr>
<td>EUS</td>
<td>N</td>
<td>uNO</td>
<td>pCR</td>
<td>N/A</td>
<td>11</td>
<td>0.68 [0.54-0.80]</td>
<td>0.57 [0.43-0.70]</td>
</tr>
<tr>
<td>PET(CT)</td>
<td>T</td>
<td>mCR</td>
<td>pCR</td>
<td>N/A</td>
<td>14</td>
<td>0.74 [0.68-0.79]</td>
<td>0.52 [0.44-0.60]</td>
</tr>
<tr>
<td>PET(CT)</td>
<td>T</td>
<td>SUVmax &lt; cut-off</td>
<td>pCR</td>
<td>2.7 (2.25-6)</td>
<td>8</td>
<td>0.69 [0.64-0.74]</td>
<td>0.72 [0.64-0.78]</td>
</tr>
<tr>
<td>PET(CT)</td>
<td>T</td>
<td>%ΔSUVmax &gt; cut-off</td>
<td>pCR</td>
<td>72.3% (52%-79.3%)</td>
<td>7</td>
<td>0.73 [0.57-0.85]</td>
<td>0.63 [0.51-0.74]</td>
</tr>
</tbody>
</table>
Table 2. Results from study-level subgroup analyses.

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

A. Studies evaluating endoscopic biopsies.

<table>
<thead>
<tr>
<th>Region</th>
<th>ClinRC</th>
<th>PathRC</th>
<th>Parameter</th>
<th>Category</th>
<th>N</th>
<th>P value for global effect</th>
<th>Sensitivity (95% CI)</th>
<th>P value</th>
<th>Specificity (95% CI)</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td>T</td>
<td>Biopsies negative</td>
<td>pCR</td>
<td>Histology</td>
<td>&gt;80% AC</td>
<td>4</td>
<td>0.0129</td>
<td>0.25 (0.15-0.40)</td>
<td>0.3315</td>
<td>0.91 (0.81-0.96)</td>
<td>0.1847</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>&gt;80% SCC</td>
<td>6</td>
<td></td>
<td>0.36 (0.24-0.52)</td>
<td></td>
<td>0.99 (0.72-1.00)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Definition of pCR</td>
<td>ypT0</td>
<td>6</td>
<td>0.5590</td>
<td>0.32 (0.18-0.50)</td>
<td>0.2841</td>
<td>0.94 (0.84-0.98)</td>
<td>0.7766</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>ypT0N0</td>
<td>2</td>
<td></td>
<td>0.20 (0.07-0.44)</td>
<td>0.97</td>
<td>0.69 (0.69-1.00)</td>
<td></td>
</tr>
</tbody>
</table>

B. Studies evaluating EUS.

<table>
<thead>
<tr>
<th>Region</th>
<th>ClinRC</th>
<th>PathRC</th>
<th>Parameter</th>
<th>Category</th>
<th>N</th>
<th>P value for global effect</th>
<th>Sensitivity (95% CI)</th>
<th>P value</th>
<th>Specificity (95% CI)</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td>T</td>
<td>uT0</td>
<td>pCR</td>
<td>Histology</td>
<td>&gt;80% AC</td>
<td>2</td>
<td>0.0098</td>
<td>0.88 (0.71-0.95)</td>
<td>0.0708</td>
<td>0.04 (0.01-0.20)</td>
<td>0.3082</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>&gt;80% SCC</td>
<td>3</td>
<td></td>
<td>0.98 (0.90-1.00)</td>
<td></td>
<td>0.18 (0.05-0.49)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Definition of pCR</td>
<td>ypT0</td>
<td>11</td>
<td>N/A</td>
<td>0.96 (0.89-0.99)</td>
<td>N/A</td>
<td>0.08 (0.03-0.24)</td>
<td>N/A</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>ypT0N0</td>
<td>0</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>N</td>
<td>uN0</td>
<td>pCR</td>
<td>Histology</td>
<td>&gt;80% AC</td>
<td>4</td>
<td>0.0456</td>
<td>0.52 (0.35-0.69)</td>
<td>0.0138</td>
<td>0.68 (0.50-0.82)</td>
<td>0.2301</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>&gt;80% SCC</td>
<td>3</td>
<td></td>
<td>0.81 (0.67-0.90)</td>
<td></td>
<td>0.52 (0.23-0.79)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Definition of pCR</td>
<td>ypN0</td>
<td>11</td>
<td>N/A</td>
<td>0.68 (0.54-0.80)</td>
<td>N/A</td>
<td>0.57 (0.43-0.70)</td>
<td>N/A</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>ypT0N0</td>
<td>0</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
</tr>
</tbody>
</table>
### C. Studies evaluating PET(-CT).

<table>
<thead>
<tr>
<th>Region</th>
<th>ClinRC</th>
<th>PathRC</th>
<th>Parameter</th>
<th>Category</th>
<th>N</th>
<th>P value for global effect</th>
<th>Sensitivity (95% CI)</th>
<th>P value</th>
<th>Specificity (95% CI)</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>T</td>
<td>mCR</td>
<td>pCR</td>
<td>Histology</td>
<td>&gt;80% AC</td>
<td>4</td>
<td>0.0540</td>
<td>0.61 (0.47-0.74)</td>
<td>0.5848</td>
<td>0.56 (0.42-0.69)</td>
<td>0.1954</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>&gt;80% SCC</td>
<td>4</td>
<td></td>
<td>0.76 (0.62-0.86)</td>
<td>0.38 (0.25-0.54)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Definition of pCR</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>ypTO</td>
<td>3</td>
<td>0.3917</td>
<td>0.76 (0.51-0.90)</td>
<td>0.5042</td>
<td>0.63 (0.49-0.75)</td>
<td>0.2423</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>ypT0N0</td>
<td>6</td>
<td></td>
<td>0.77 (0.72-0.82)</td>
<td>0.46 (0.26-0.67)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Imaging technique</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>PET</td>
<td>3</td>
<td>0.1839</td>
<td>0.86 (0.77-0.92)</td>
<td>0.0662</td>
<td>0.49 (0.26-0.72)</td>
<td>0.5434</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>PET-CT</td>
<td>9</td>
<td></td>
<td>0.70 (0.58-0.79)</td>
<td>0.53 (0.42-0.65)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>T</td>
<td></td>
<td></td>
<td>SUVmax &lt; cut-off</td>
<td>pCR</td>
<td>&gt;90% AC</td>
<td>0</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>&gt;90% SCC</td>
<td>6</td>
<td></td>
<td>0.69 (0.64-0.74)</td>
<td>0.72 (0.64-0.78)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Definition of pCR</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>ypTO</td>
<td>5</td>
<td>0.4660</td>
<td>0.71 (0.61-0.80)</td>
<td>0.8145</td>
<td>0.77 (0.66-0.85)</td>
<td>0.2224</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>ypT0N0</td>
<td>2</td>
<td></td>
<td>0.69 (0.58-0.78)</td>
<td>0.66 (0.50-0.79)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Imaging technique</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>PET</td>
<td>0</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>PET-CT</td>
<td>7</td>
<td></td>
<td>0.69 (0.63-0.75)</td>
<td>0.71 (0.64-0.78)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>T</td>
<td></td>
<td></td>
<td>%ΔSUVmax &gt; cut-off</td>
<td>pCR</td>
<td>&gt;80% AC</td>
<td>1</td>
<td>0.0460</td>
<td>0.43 (0.34-0.51)</td>
<td>0.0403</td>
<td>0.58 (0.40-0.74)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>&gt;80% SCC</td>
<td>5</td>
<td></td>
<td>0.80 (0.64-0.90)</td>
<td>0.57 (0.44-0.70)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Definition of pCR</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>ypTO</td>
<td>2</td>
<td>0.7072</td>
<td>0.74 (0.58-0.85)</td>
<td>0.6514</td>
<td>0.64 (0.53-0.73)</td>
<td>0.9131</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>ypT0N0</td>
<td>2</td>
<td></td>
<td>0.76 (0.60-0.87)</td>
<td>0.66 (0.43-0.83)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Imaging technique</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>PET</td>
<td>0</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>PET-CT</td>
<td>5</td>
<td></td>
<td>0.70 (0.52-0.83)</td>
<td>0.63 (0.50-0.74)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Figure 1. Flowchart summarizing search results and study selection.

- Records identified through database searching: (n = 6798)
  - Embase (n = 3437)
  - Medline (n = 1444)
  - Cochrane (n = 88)
  - Web of Science (n = 1786)
- Additional records identified through other sources: (n = 2)
  - Hand searching included studies and reviews of similar subjects (n = 2)

Records after removal of duplicates (n = 4130)

Records screened by title and abstract (n = 4130)

Records excluded (n = 3872)

Articles excluded (n = 193)
- Abstracts only (n = 168)
- Letter to the editor (n = 1)
- Review (n = 1)
- Not in English (n = 5)
- Ongoing trial abstract (n = 1)
- Unrecognized duplicates (n = 4)
- Less than 10 patients (n = 4)
- Overlapping cohorts (n = 7)
- Therapy regimen other than concurrent nCRT (n = 7)
- Golden standard other than histopathology of the resection specimen (n = 2)
- Majority with cervical tumor location (n = 1)
- Combining gastric and esophageal cancer (n = 1)
- Analysis of specific pathologic subgroup (n = 1)
- Index test before or during nCRT (n = 5)
- Different index test or index test interpretation (n = 14)
- No correlation between index test and pathology (n = 11)
- Insufficient data for cross tabulation (n = 20)

Articles assessed for eligibility by full text (n = 258)

Studies included in qualitative analysis (n = 65)

Studies excluded from quantitative analysis (n = 21)
- ≤ 4 studies included for the same index test (n = 15)
- Pathological response criterion other than pCR (n = 6)

Studies included in quantitative synthesis (n = 44)
Figure 2. Forest plots demonstrating sensitivities and specificities from individual studies included in quantitative analyses.

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

A. Studies evaluating endoscopic biopsies.

<table>
<thead>
<tr>
<th>Region</th>
<th>ClinRC</th>
<th>PathRC</th>
<th>Cut-off</th>
<th>Study</th>
<th>TP</th>
<th>FP</th>
<th>TN</th>
<th>Sensitivity (95% CI)</th>
<th>Specificity (95% CI)</th>
<th>Sensitivity (95% CI)</th>
<th>Specificity (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>T</td>
<td>EUS</td>
<td>pCR</td>
<td>N/A</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A.</td>
<td>Biopsy</td>
<td>negative</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A.</td>
<td>Biopsy</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

B. Studies evaluating EUS.

<table>
<thead>
<tr>
<th>Region</th>
<th>ClinRC</th>
<th>PathRC</th>
<th>Cut-off</th>
<th>Study</th>
<th>TP</th>
<th>FP</th>
<th>TN</th>
<th>Sensitivity (95% CI)</th>
<th>Specificity (95% CI)</th>
<th>Sensitivity (95% CI)</th>
<th>Specificity (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>N</td>
<td>uT0</td>
<td>pCR</td>
<td>N/A</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>B.</td>
<td>EUS</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
C. Studies evaluating PET-CT.

<table>
<thead>
<tr>
<th>Region</th>
<th>ClinIC</th>
<th>PathIC</th>
<th>Cut-off</th>
<th>Study</th>
<th>TP</th>
<th>FP</th>
<th>FN</th>
<th>TN</th>
<th>Sensitivity (95% CI)</th>
<th>Specificity (95% CI)</th>
<th>Sensitivity (95% CI)</th>
<th>Specificity (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>T</td>
<td>mCR</td>
<td>pCR</td>
<td>N/A</td>
<td></td>
<td>205</td>
<td>49</td>
<td>47</td>
<td>21</td>
<td>0.81 [0.76, 0.86]</td>
<td>0.36 [0.20, 0.42]</td>
<td>0.36 [0.23, 0.47]</td>
<td>0.33 [0.24, 0.45]</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>97</td>
<td>43</td>
<td>31</td>
<td>22</td>
<td>0.76 [0.67, 0.83]</td>
<td>0.34 [0.23, 0.47]</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>31</td>
<td>8</td>
<td>15</td>
<td>16</td>
<td>0.67 [0.52, 0.80]</td>
<td>0.67 [0.45, 0.84]</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>26</td>
<td>2</td>
<td>4</td>
<td>4</td>
<td>0.67 [0.59, 0.83]</td>
<td>0.67 [0.22, 0.96]</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>60</td>
<td>15</td>
<td>44</td>
<td>19</td>
<td>0.58 [0.48, 0.67]</td>
<td>0.58 [0.28, 0.73]</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>23</td>
<td>6</td>
<td>12</td>
<td>21</td>
<td>0.58 [0.46, 0.68]</td>
<td>0.58 [0.58, 0.61]</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>28</td>
<td>5</td>
<td>2</td>
<td>9</td>
<td>0.83 [0.78, 0.89]</td>
<td>0.64 [0.35, 0.87]</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>22</td>
<td>5</td>
<td>2</td>
<td>9</td>
<td>0.83 [0.78, 0.89]</td>
<td>0.64 [0.35, 0.87]</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>23</td>
<td>5</td>
<td>2</td>
<td>9</td>
<td>0.83 [0.78, 0.89]</td>
<td>0.64 [0.35, 0.87]</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>23</td>
<td>5</td>
<td>2</td>
<td>9</td>
<td>0.83 [0.78, 0.89]</td>
<td>0.64 [0.35, 0.87]</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>13</td>
<td>5</td>
<td>6</td>
<td>12</td>
<td>0.68 [0.43, 0.87]</td>
<td>0.71 [0.44, 0.98]</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Region | ClinIC | PathIC | Cut-off | Study | TP | FP | FN | TN | Sensitivity (95% CI) | Specificity (95% CI) | Sensitivity (95% CI) | Specificity (95% CI) |
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>T</td>
<td>6.0%&lt;max</td>
<td>pCR</td>
<td></td>
<td></td>
<td>8</td>
<td>2</td>
<td>2</td>
<td>7</td>
<td>0.80 [0.79, 0.81]</td>
<td>0.70 [0.40, 0.91]</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>3.25</td>
<td></td>
<td></td>
<td></td>
<td>31</td>
<td>8</td>
<td>15</td>
<td>16</td>
<td>0.67 [0.52, 0.80]</td>
<td>0.67 [0.45, 0.84]</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>2.7</td>
<td></td>
<td></td>
<td></td>
<td>40</td>
<td>12</td>
<td>25</td>
<td>24</td>
<td>0.62 [0.49, 0.73]</td>
<td>0.74 [0.56, 0.86]</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>3.4</td>
<td></td>
<td></td>
<td></td>
<td>26</td>
<td>4</td>
<td>6</td>
<td>22</td>
<td>0.81 [0.64, 0.92]</td>
<td>0.85 [0.65, 0.95]</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>2.5</td>
<td></td>
<td></td>
<td></td>
<td>12</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>0.80 [0.52, 0.86]</td>
<td>0.75 [0.51, 0.91]</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>2.5</td>
<td></td>
<td></td>
<td></td>
<td>33</td>
<td>7</td>
<td>16</td>
<td>13</td>
<td>0.67 [0.52, 0.80]</td>
<td>0.65 [0.41, 0.88]</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>2.5</td>
<td></td>
<td></td>
<td></td>
<td>17</td>
<td>1</td>
<td>8</td>
<td>6</td>
<td>0.68 [0.46, 0.82]</td>
<td>0.86 [0.42, 1.00]</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>2.7</td>
<td></td>
<td></td>
<td></td>
<td>22</td>
<td>7</td>
<td>9</td>
<td>14</td>
<td>0.71 [0.52, 0.86]</td>
<td>0.67 [0.43, 0.85]</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Region | ClinIC | PathIC | Cut-off | Study | TP | FP | FN | TN | Sensitivity (95% CI) | Specificity (95% CI) | Sensitivity (95% CI) | Specificity (95% CI) |
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>T</td>
<td>&gt;6.0%&lt;max</td>
<td>pCR</td>
<td></td>
<td></td>
<td>64</td>
<td>1</td>
<td>6</td>
<td>30</td>
<td>0.63 [0.48, 0.77]</td>
<td>0.81 [0.65, 0.92]</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>75%</td>
<td></td>
<td></td>
<td></td>
<td>31</td>
<td>7</td>
<td>15</td>
<td>17</td>
<td>0.67 [0.52, 0.80]</td>
<td>0.71 [0.46, 0.87]</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>75%</td>
<td></td>
<td></td>
<td></td>
<td>43</td>
<td>18</td>
<td>22</td>
<td>28</td>
<td>0.66 [0.53, 0.77]</td>
<td>0.61 [0.45, 0.75]</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>75%</td>
<td></td>
<td></td>
<td></td>
<td>25</td>
<td>15</td>
<td>47</td>
<td>21</td>
<td>0.42 [0.34, 0.52]</td>
<td>0.56 [0.30, 0.70]</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>75%</td>
<td></td>
<td></td>
<td></td>
<td>16</td>
<td>13</td>
<td>1</td>
<td>6</td>
<td>0.94 [0.71, 1.00]</td>
<td>0.32 [0.13, 0.57]</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>72.1%</td>
<td></td>
<td></td>
<td></td>
<td>12</td>
<td>3</td>
<td>2</td>
<td>8</td>
<td>0.86 [0.57, 0.98]</td>
<td>0.73 [0.36, 0.94]</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>70.3%</td>
<td></td>
<td></td>
<td></td>
<td>22</td>
<td>3</td>
<td>2</td>
<td>5</td>
<td>0.88 [0.65, 0.97]</td>
<td>0.71 [0.49, 0.98]</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
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.

EUS: endoscopic ultrasonography, PET(-CT): positron emission tomography with or without computed tomography, uT0: ultrasonographic tumor stage 0, uN0: ultrasonographic nodal stage 0, mCR: metabolically complete response, SUVmax: maximum standardized uptake value, %ΔSUVmax: percentage reduction of SUVmax, pCR: pathologically complete response.
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 < cut-off as clinical response criterion and pCR as pathological response criterion.

F. SROC of PET(-CT) evaluating the primary tumor site using %ΔSUVmax > cut-off as clinical response criterion and pCR as pathological response criterion.
**Author 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.
Click here to access/download
Supplemental Data File
181211_Meta-Analysis_Supplementary_AOS_v2.docx
Click here to access/download
Supplemental Data File
181211_Meta-Analysis_PRISMAchecklist.docx